Qualitative Inquiry in Geoscience Education Research

THE GEOLOGICAL SOCIETY OF AMERICA®

Special Paper 474

Edited by Anthony D. Feig and Alison Stokes

Qualitative Inquiry in Geoscience Education Research

edited by

Anthony D. Feig
Department of Geology & Meteorology
Central Michigan University
Mount Pleasant, Michigan 48859
USA

Alison Stokes
Experiential Learning Centre for Excellence in Teaching and Learning
University of Plymouth
PL4 8AA, UK

THE
GEOLOGICAL
SOCIETY
OF AMERICA®

Special Paper 474

3300 Penrose Place, P.O. Box 9140 ▪ Boulder, Colorado 80301-9140, USA

2011

Copyright © 2011, The Geological Society of America (GSA), Inc. All rights reserved. GSA grants permission to individual scientists to make unlimited photocopies of one or more items from this volume for noncommercial purposes advancing science or education, including classroom use. For permission to make photocopies of any item in this volume for other noncommercial, nonprofit purposes, contact The Geological Society of America. Written permission is required from GSA for all other forms of capture or reproduction of any item in the volume including, but not limited to, all types of electronic or digital scanning or other digital or manual transformation of articles or any portion thereof, such as abstracts, into computer-readable and/or transmittable form for personal or corporate use, either noncommercial or commercial, for-profit or otherwise. Send permission requests to GSA Copyright Permissions, 3300 Penrose Place, P.O. Box 9140, Boulder, Colorado 80301-9140, USA. GSA provides this and other forums for the presentation of diverse opinions and positions by scientists worldwide, regardless of their race, citizenship, gender, religion, or political viewpoint. Opinions presented in this publication do not reflect official positions of the Society.

Copyright is not claimed on any material prepared wholly by government employees within the scope of their employment.

Published by The Geological Society of America, Inc.
3300 Penrose Place, P.O. Box 9140, Boulder, Colorado 80301-9140, USA
www.geosociety.org

Printed in U.S.A.

GSA Books Science Editors: Marion E. Bickford and Donald I. Siegel

Library of Congress Cataloging-in-Publication Data

Qualitative inquiry in geoscience education research / edited by Anthony D. Feig, Alison Stokes.
p. cm. — (Special paper ; 474)
Includes bibliographical references.
ISBN 978-0-8137-2474-4 (pbk.)
1. Geology—Study and teaching—United States. 2. Qualitative research. I. Stokes, Alison.
QE40.Q35 2011
550.7—dc22

2010043051

Cover: An opportunity to contemplate the Chippewa River, Isabella County, Michigan. View to the southeast. Photograph by Cathy Willermet; front-cover image concept by Anthony D. Feig.

10 9 8 7 6 5 4 3 2 1

Contents

Theoretical Foundations

Methodology and Models

Complete Studies

Introduction

Anthony D. Feig
Department of Geology & Meteorology, Central Michigan University,
Mount Pleasant, Michigan 48859, USA

Alison Stokes
Experiential Learning CETL (Centre for Excellence in Teaching and Learning),
University of Plymouth, PL4 8AA, UK

The enthusiastic response to our initial suggestion for this volume, and willingness of colleagues to contribute manuscripts, are clear indicators that qualitative methods are gaining in acceptance and popularity within the geoscience education and geocognition community. What we currently lack, however, is an extensive, formal literature base for qualitative inquiry into how novice and expert geoscientists think and learn. This volume provides the first reference of its kind for geoscience education researchers to explore in-depth qualitative theory and methods, and to examine case studies documenting both the application of these methods, and the contribution made by qualitative inquiry to geoscience education and geocognition. It is our intention to provide a reference framework aimed at, but by no means restricted to, two key audiences: geoscience educators who are (or who aspire to be) qualitative practitioners; and geoscientists who are new to, and may even be openly critical of, qualitative inquiry. In this volume the "rules" of qualitative inquiry are parsed, the theories are dissected, and examples are provided to allow skeptics and practitioners alike to see examples of how we do what we do.

Geoscience education and geocognition researchers are an interesting group. As geoscientists, we work in the world of natural processes, and we speak a language that quantifies and categorizes our observations in an orderly fashion. As education researchers, however, we enter a different world. Here, we often find ourselves confronted with problems and data that are difficult to measure, that resist experimentation, and that are quite often impossible to quantify. "Reality" may become fuzzy, multiplying from our expected single, objective version to something iterative and subjective. In these situations, we realize that our trusted tools of observation, experiment, and objectivity fail us, so we turn to the tools of qualitative inquiry to provide the insight that we seek. But here we hit some interesting, and often frustrating, hurdles. First of all, it is an unfortunate fact that many of us have little or no formal training in qualitative research methods. Usually working in isolation, we enter an entirely new literature base; we engage with unfamiliar and, at times, uncomfortable ways of thinking and practicing. Each application of a new method or approach is, in a sense, a private re-invention of the wheel. The inevitable outcome of this private labor is that we tend to work in isolation—we are an archipelago, laboriously discussing in our publications the theory behind qualitative convention and justifying standard processes ("…is well established in the social and behavioral sciences…"). Having negotiated this challenging (but eminently rewarding!) process, we then find that our geoscientist peers are often highly skeptical of our methods, results, and interpretations. Sometimes skepticism becomes criticism without critique. The following comments, or variants of them, will be familiar to many geoscience education researchers:

- It's all subjective!
- That's not an interpretation! That's just what you wanted to say!
- Those aren't data! Those are anecdotes!
- You need to run some statistics on this before it's considered valid!

- There is no way you could replicate that!
- Show me the numbers!

And perhaps the most bothersome of all:

- This is just a bunch of edu-speak!

This is not surprising. As scientists we naturally want to quantify and assign statistical significance to our research findings, and to test which of our multiple hypotheses provide the most viable explanation for what we observe. We perceive learning as a complex yet understandable system, which we seek to understand in much the same way as Earth systems. What is lacking from this approach, however, is insight into the causal factors influencing the patterns or relationships that we deduce from our quantitative data. Qualitative data are what provide this insight…they provide a window into the *how* and *why*, rather than simply the *what* or *how much*, of learning. Qualitative inquiry is finding particular traction in terms of revealing how novices and experts think and reason about geoscientific concepts. Understanding of, for example, how novices move from observation, to inference, to interpretation cannot be gained using quantitative methods. Likewise, experts cannot assume to know the different ideas that novices will hold about a particular concept—this information can only be gained through qualitative exploration of novice conceptions. Nor can instructors claim to "know" the field experience of their students, since each student constructs and inhabits his or her own reality.

We tend to categorize research approaches as either "qualitative" or "quantitative," but the two are not mutually exclusive; this is a continuum and not a dichotomy. Within the research methods literature we find the term "qualitative inquiry" applied to a wide range of approaches from ethnography and narrative analysis, which are typically characterized by unstructured and naturalistic methods, to mixed-method and semi-quantitative approaches, which blend numerical with non-numerical or non-metric data. It is the latter mixed-method approach that commonly, but not always, acts as the entry point to qualitative inquiry for many geoscience education researchers. Typical scenarios might include combining qualitative methods, e.g., student interviews, with quantitative methods, e.g., attitudinal surveys, to explore a particular aspect of student learning, or quantifying qualitative data to explore the frequency with which particular conceptions occur within a population. Again, this is not surprising given the research paradigm within which geoscientists typically operate. The geoscientific worldview lends itself well to a "mixed-methods" approach. As an example, consider the process of grain analysis. The size of a grain is quantitative if it is expressed in terms of millimeters, but it may also be non-metrically expressed by use of the term "sand-sized." Despite the absence of a specific number, this term is immediate in its meaning; the mental image of sand comes unbidden to the mind's eye, as does the quantity ranging between 1/16 to 2 millimeters. This familiarity with non-metric data eases our transition into the world of qualitative inquiry; we begin a limited exploration of these tools.

Our transition, however, comes with one caveat: Sometimes geoscientists tend to categorize all non-numerical and non-metric data as "qualitative." However, non-metric and qualitative data are not the same, as explored within this volume. This act of "lumping" is natural for researchers taking their first steps into unfamiliar territory. Beyond that, however, it is woolly thinking, leading to a dichotomous metaphor that we (and others) might call "little *q* vs. Big *Q*" (*q*/*Q* standing, of course, for "qualitative"). Here *q* would apply to methods and approaches which incorporate elements of qualitative inquiry, but which are underpinned by post-positivist philosophies, and a tendency to force the application of "scientific" methods in qualitative problems. We do not intend to criticize empiricism, rather those who would wield it inflexibly when it is not the best approach. By comparison *Q* rejects the notion of a detached, objective observer, relying instead on purely interpretive methods in which the researcher is the instrument. Phenomenology and ethnography would be examples of *Q* approaches. In the wider sense, *q* represents where we are now: Our shortcomings, our misapplications of technique, our unfamiliarity with this theoretical and methodological world; the defensive postures we (need to) take in our praxis; the paucity and diffusion of relevant literature. On the other hand, *Q* represents where we need to go, and where we can go: Building a canonical literature base; establishing qualitative inquiry as a "valid" and useful mode of inquiry; building our skill and praxis in qualitative inquiry. We see the contributions in this volume as a first step toward *Q*.

The chapters in this volume fit into naturally emerging categories, which made our editing job somewhat akin to a qualitative data-coding process. The volume is divided into three sections: Theoretical

Foundations, Methodology and Models, and Complete Studies. While there is undoubtedly some overlap between these sections, each has a specific theme and purpose.

Section I focuses on the theoretical foundations of qualitative inquiry, exploring some of the philosophies and theories that are rooted in the social sciences and educational research. This section presents four papers which vary from purely theoretical, to those providing data and case studies to support and exemplify theoretical discussions. The section opens with Anthony D. Feig's discussion of methodology and location in the context of qualitative inquiry. This paper lays out some important foundations for the remainder of the volume by exploring the role of the researcher in the inquiry process, and by making explicit the distinction between theoretical framework, methodology, and method. These concepts are then contextualized by Christopher L. Atchison and Anthony D. Feig, who offer some theoretical perspectives on constructing a qualitative research investigation into the experience of mobility-impaired students engaging with alternative field-based learning environments. In the third contribution to this section, Alison Stokes provides a detailed overview of phenomenography, a research tool for exploring variations in individual lived experiences, applied here to investigating how undergraduates conceptualize geoscience as an academic discipline. Finally in this section, Julie C. Libarkin and Emily Geraghty Ward revisit the early history of the Geoscience Concept Inventory (GCI), sharing analyses and reflections on the qualitatively oriented development of this particular research instrument.

Part II of the volume explores the "nuts-and-bolts" processes of qualitative inquiry by focusing on methodology, and on the process and outcomes of generating models in qualitative research. In the first of four papers in this section, anthropologist Deborah Williams and geoscientist Steven Semken present a rigorous and detailed, ethnographically informed treatment of a place-based pedagogical experience. This is followed by Matthew Alles and Eric Riggs' application of a grounded theory approach to investigating how novice geology students acquire and express spatial skills. Data generated using this approach form the basis for a conceptual model describing the acquisition of visual penetrative ability within students. Scott K. Clark and Julie C. Libarkin then provide a detailed account of the procedures followed in designing and creating a mixed-format survey and scoring rubric, used in this case to investigate expert and novice conceptions of plate tectonics. The final contribution to part II comes from Leilani Arthurs and Thomas Marchitto, who discuss the theoretical and methodological background of concept inventory design, documenting the step-by-step process involved in designing a concept inventory for the ocean sciences.

The third and final section presents five complete investigations that provide excellent examples of the contribution to be made by qualitative inquiry to geoscience education and geocognition research. In the first of these, Renee M. Clary and James H. Wandersee explore learning opportunities in U.S. fossil parks, synthesizing data generated over many years of immersed study. Leilani Arthurs then presents a detailed investigation into students' alternate conceptions and cognitive models, in which she evaluates the effectiveness of different qualitative approaches in revealing student thinking. The final three papers in this section all provide examples of mixed-methods research, whereby qualitative inquiry is blended with quantitative methods to explore learners' understandings of complex geoscientific concepts. Toru Ishikawa and colleagues present some intriguing insights into environmental policy students' ability to understand and evaluate climate forecast data, and explore how this understanding is applied in authentic decision-making scenarios. Following this, Karen M. Kortz and colleagues report on findings from a collaborative study that applies the mixed-format survey instrument designed by Clark and Libarkin to investigating students' conceptions of tectonic plates and boundaries. This section concludes with an investigation by Sandra Swenson and Kim Kastens into the ways in which school students perceive and interpret visual representations of complex data, in this case a global elevation map.

Ultimately, we hope that this volume will serve as a useful reference for geoscience education and geocognition researchers embarking on qualitative study, as well as practicing geoscientists looking to understand qualitative data and methodologies. Its relevance and exploration of new intellectual territory will enhance the growing subdisciplines of geoscience education and geocognition in terms of their application, rigor, and literature base. Compiling the volume has been a thoroughly rewarding experience, reinforcing the notion that geoscience educators are nimble and flexible thinkers, willing to enter and engage with new (and sometimes frightening!) territory. We are deeply grateful to Pat Bickford and Joanne Ranz at GSA Books for their assistance with and commitment to this volume. Finally, we would like to thank all of our contributors wholeheartedly for their rapid response to our requests, for their collegiality, and for their high-quality contributions, and we look forward to continued discussions and future collaborations.

The Geological Society of America
Special Paper 474
2011

Methodology and location in the context of qualitative data and theoretical frameworks in geoscience education research

Anthony D. Feig
Department of Geology and Meteorology, Central Michigan University, Mt. Pleasant, Michigan 48859, USA

ABSTRACT

Successful and rigorous qualitative research requires careful planning of purpose, methods, and theoretical frameworks. The qualitative researcher must locate the study in time, space, and culture, and must also locate herself or himself in the study. This is in order to thoroughly and publicly explore his or her purpose, role, and potential biases. Through this process, the researcher defines the ways in which these and other factors inform the research. Finally, the qualitative researcher must take thought of methodology—as opposed to method—and must understand the difference between the two.

In this paper, I review basic principles of qualitative inquiry with regard to the nature of qualitative data and theoretical frameworks. I then explore the issues of location and methodology as applied to qualitative inquiry in geoscience education research through examples relevant to the discipline. I describe the process of locating the study, and the ways in which the researcher defines his or her place therein. I then discuss the differences between method and methodology. Finally, I review four specific methodologies, including hermeneutics, phenomenology, ethnography, and policy analysis.

INTRODUCTION

Geoscience education can be defined as the scholarship of teaching and learning applied to the geosciences. The scholarship of geoscience teaching includes the study of teaching innovations, evolving classroom practices, field skills acquisition, and educational policy, to name a few items. Some examples of learner-focused scholarship include their alternative conceptions of Earth processes, their cognitive and meta-cognitive processes, and the affective factors that shape their learning. The examples I have listed here are quantitatively observable or measureable either directly, or as a function of their impact upon teachers, students, and/or an educational setting. These measurable parameters can be student outcomes (how much/how well did they learn?), analyses of variance between outcomes of two teaching techniques, or identification of systematic patterns in student misconceptions. Even policy analysis lends itself to quantitative inquiry. For example, how does a high-stakes test in a public school system affect geoscience teaching in terms of student outcomes, numbers of misconceptions, or enrollments in postsecondary geoscience courses? These kinds of quantitative inquiries are fundamentally empirical, and firmly rooted in a hypothesis-driven, positivistic tradition. The instrumentation associated with these measurements are usually in the form of questionnaires, surveys, concept inventories, and pre- and post-tests.

Survey instruments and concept inventories can and do yield extensive and useful data, but they have inherent limitations. Most of the time, quantitatively oriented researchers can safely ignore these limitations, because numerical methods are appropriate for their research questions. The first limitation of instruments and

Feig, A.D., 2011, Methodology and location in the context of qualitative data and theoretical frameworks in geoscience education research, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 1–10, doi:10.1130/2011.2474(01).

inventories is that responses are generally constrained to prewritten options. In choosing between A, B, C, and D, the respondent may wish for an E to select, or may feel that a mix of A and C is most appropriate. The second limitation is the inherent difficulty for the researcher in quantifying open-ended questions, especially where a "correct" answer does not exist. A third limitation is that instrument data compose a snapshot of an end condition. While surveys and inventories yield an understanding of a particular perceived truth, it is much more difficult to figure out how the person *constructed* that truth by using a survey or inventory. Furthermore, the instrument does not allow you to see this construction happen in real time. This is because personal truths are constructed through lived experiences, and these both resist quantitative study. Quantitative inquiry can tell a researcher *what* and *how much* of something happens, but the question of *why* is problematic. In pursuit of this line of inquiry, a researcher might find herself awash in information about her students' attitudes, perceptions, lived experiences, values, and memories. She recognizes these as qualitative data in the form of communicated truths: that is, words, stories, and descriptions, and sometimes nonverbal expressions. These data are not subject to validity tests or manipulation by experiment. In order to extract meaning from these data, the researcher must turn to qualitative inquiry.

Many geoscientists are skeptical of non-numerical modes of inquiry. These concerns are fueled in part by a lack of familiarity with qualitative methods. Another issue is the fact that qualitative inquiry is generally not conducted within a framework of "scientific" empiricism. Experimental and control groups are not established; norm- or criterion-referenced metrics are not collected; dependent and independent variables are not identified. Furthermore, reality itself takes on a more nuanced meaning. Geoscientists are accustomed to making our findings and observations correspond to reality. For example, an observed formation has a measurable thickness; its rocks have certain characteristics observable by others; those rocks in turn represent a particular environment. We might argue over details, but we are likely to assume that only one realistic interpretation exists, which is best articulated by the most thorough observations, the most logical arguments, and the most replicable experiments. By contrast, the qualitative researcher embraces the notion of multiple realities. This is largely because of the nature of the data we work with. As scientists, we understand our world through the data we amass. An examination of the contrasts of "reality" in the context of "data" is a good conceptual starting place for scientists unfamiliar with qualitative inquiry.

A PRIMER ON QUALITATIVE DATA

In scientific research, data are *collected*, because they exist independent of the scientist. For example, we say strikes and dips are collected because the rocks had an orientation before the geologist arrived, and continue to be oriented after she or he leaves the field. The data exist, and the geologist goes and gets them. Qualitative data are *generated* because they do not exist until the researcher goes after them. For example, asking a geology student to describe her experiences mapping with and without a global positioning system (GPS) unit is an example of generating data because (1) it is possible she did not consciously compare those experiences before she was asked; (2) her responses will be different depending on how she is asked and who is asking; and (3) her responses will vary depending upon the mapping situation she is in. Other reasons likely exist. In short, qualitative data have a tendency to be iteratively variable.

I must make an important aside here: not all non-numerical data are qualitative. For example, when documenting the luster of, say, the mineral galena, we describe it as "metallic." It is true that this is not a number. However, it is possible to generate consensus, even among large numbers of geologists, that the luster of galena is metallic. This physical property is not iterative, and it exists independent of the observer. Luster does not depend upon the observer as much as it does on the mineral's composition and other consistent, physical properties. So in this case, rather than being "qualitative," that is, a communicated truth from a social actor based on lived experience, this datum is "nonmetric." Other examples of nonmetric, geological data types include grain sorting and angularity, and relative bedding thickness, i.e., not expressed in units but as either "thinly bedded" or "thickly bedded."

So what are qualitative data, and what makes them resistant to empirical manipulation? Qualitative data can be thought of as "ontological objects." "Ontology" as I use it is synonymous with "metaphysics"; that is, the study of reality and being, and the things that constitute the world (Schwandt, 2001). This definition is not limited to material objects but also includes items from the "mental life" of those being studied. In the science education literature, ontology refers to how people ascribe meaning to phenomena (Chi et al., 1994; Libarkin and Kurdziel, 2006). In terms of mental life, ontological objects include observed behaviors, responses to verbal questions, nonverbal cues (e.g., body language), individual choices, student preconceptions or alternate conceptions (Libarkin, 2005), the ways in which students cope with and process novel field spaces (Orion, 1993), their previously lived experiences, and how they respond to stimuli. Ontological objects like these are in the form of communicated truths (Gadamer, 1975), from participant to researcher, not objective truths, such as the dip of a package of rocks, or the composition of those rocks. Rocks are physical objects, subject to third-party verification and validity analysis. Ontological objects in qualitative inquiry do not correspond with a single reality, but exist within multiple realities. The researcher uses them to assemble mental models (e.g., Brodaric et al., 2004). One can argue that these models represent objective reality, but in truth they should be considered a snapshot of one single reality among many. Models of teaching and learning processes may have wide applicability and general application, but they could just as easily fail to accommodate a given situation or set of learners, where reality may be different.

Ontological objects are real to those who hold and live them, but they are not subject to verification. For example, a student in my class might arrive with a concept of plate tectonics wherein solid plates float on top of a homogeneous, completely liquid mantle. The mantle has been established to be much more complex than this simple representation, so his conception is "wrong" ("alternate"). His conceptualization may be faulty, but nevertheless it is what he presents to me, so therefore it exists. I can add it to my data pool, along with other ontologies from other learners, if these objects are the focus of my research.

Another issue that may confront the geoscientist unfamiliar with qualitative inquiry consists of distinguishing between qualitative data and anecdotes. In quantitative inquiry, we occasionally obtain information that we do not intend to quantify. Depending on the researcher's intent, this information may be anecdotal, or it may in fact be a pool of generated qualitative data. For example, teaching evaluations often have quantitative and qualitative components. Students respond to a questionnaire containing ordinal items such as, "Rate the instructor's ability to motivate you (on a scale of 1 to 5)." Institutions compile descriptive statistics, and a score is produced for each question. Students are also given the opportunity to write open-ended comments, such as, "What did you like best/least about this course?" These comments are frequently not scored against any rubric, but are simply aggregated and sent back to the instructor. If the institution bases merit and promotion solely on the numerical results, then the student comments are not important to that process, and, in that context, they are anecdotes. Furthermore, the comments may have been influenced by external factors, such as the difficulty of the upcoming final exam, or the donuts the instructor brought in for his class.

On the other hand, the student comments may yield patterns upon close inspection. This pattern-identification is systematic, but it is not necessarily repeatable; different workers might produce different interpretations. This makes the process iterative in a manner dependent upon the situation and the investigator, and therefore is not "scientific" in the way we practice traditional geoscientific investigations. For example, the students may be struggling with the course management software; the instructor's approach to collaborative learning may need adjustment; the instructor may have displayed cultural insensitivity in his interactions with students. If the instructor systematically analyzes the comments he receives and acts on them, they are no longer anecdotes: they are data. This is especially true if a large percentage of students chose, on the quantitative portion, to score him arbitrarily (picking "all threes," or "all fives," to finish the survey faster). He could have high quantitative scores, but still see a need for improvement yielded by the comments. Systematic analysis and action are what mark the difference between anecdotes and qualitative data. While it could be argued that these data exist independent of the instructor, the bottom line is that they did not exist until the students were asked to contribute them. This is why qualitative data such as these are different in nature from quantitative items like strike and dip.

Yet another point of contention for geoscientists who are new to working with qualitative data is the issue of how much to collect. A striking contrast between quantitative and qualitative research is that in the former, a large population is sampled, while the latter extracts meaning from a much smaller pool of participants. The meanings parsed via qualitative inquiry concerning the "how" and "why" of teaching and learning are often much deeper and more fully developed than in quantitative approaches. The purposes of sampling in numerical analysis are to mathematically and statistically extrapolate results from the sampled population to an entire population. The rigor of the extrapolation is directly proportional to the amount of sampling. However, the qualitative researcher is not interested in representing a population (Mason, 2002), but rather illustrating a process, documenting events, or understanding specific ontological realities. These require not a representative sampling, but rather a purposive or theoretical strategy (Schwandt, 2001). Purposive sampling is guided by the question of relevance to the phenomenon being studied. For example, a qualitative researcher studying how field students use GPS technology would sample students in a single field course that uses such technology. The researcher would not sample a larger pool that includes students who are not mapping with GPS, even if the available population (n) is small. This small group may or may not be demographically or cognitively representative of all geology students; however, they are the group in the field interacting with the technology. They have the highest relevance to the question of how students use, depend on, and conceptualize GPS technology. Documenting and understanding what happens with these students is critical to understanding the phenomenon.

Another factor that reduces sample size is that of data saturation (Mason, 2002). Saturation takes place when enough data have been generated so that the researcher has a picture of what is going on, and any further generation would result in the data repeating themselves. For example, a researcher examining barriers to understanding the concept of geologic time could interview students. Through the course of the interviews, ten students express a dissonance between geologic time scales and their religious beliefs. The researcher interviews three more, and all three express the same dissonance. The researcher could interview 87 more students, or eight more, and expect the theme of dissonance to recur. However, it is safe for the researcher to conclude that she or he has a reasonable picture that "religious dissonance" impacts student conception of geologic time. This dissonance is an ontological object, an emergent theme expressed by the participants as a group. What the researcher does with this theme depends on his or her chosen theoretical framework.

THEORETICAL FRAMEWORKS: A SELECTIVE REVIEW

Scientific inquiry in the geosciences is conducted wholly within the realm of mechanistic, positivistic logical empiricism (Nagel, 1961). The "laws" of physics and chemistry underlie every Earth process, and those processes are investigated through hypothesis, experiment, observation, and subsequent hypothesis revision/rejection. Quantitative inquiry in geoscience education research also lies entirely within the realm of logical empiricism. Relationships are parsed, correlations discovered, and inferences made according to the rules and assumptions contained within mathematical, statistical, and psychological approaches. As such, geoscientific and quantitative educational research is informed by only one theoretical foundation, that of logical empiricism, i.e., the scientific method.

An inquiry guided by empiricism is feasible in qualitative inquiry, if the data are somehow quantifiable (so-called "mixed methods"), or if they lend themselves to experimental manipulation. Ontological objects (e.g., lived experiences, communicated truths, attitudes) do not readily lend themselves to empirical study. The scientific method is not a suitable theoretical foundation for working with most qualitative data types; however, qualitative inquiry can be *informed* by logical empiricism. In my own research, I have alternately operated within two different frameworks, those of grounded theory and critical theory. The former is, in part, informed by the scientific method. The latter has a long history in the social sciences, but has only recently emerged in science education research.

Grounded Theory

Grounded theory is a data-driven approach to understanding a central phenomenon (Creswell, 1998). The outcome of a grounded theory study is a model, or some other theoretical construct, applicable to multiple settings. Ontological objects are constantly compared and analyzed for concept indicators (Glaser and Strauss, 1965, 1967; Strauss, 1987). Concept indicators can be thought of as data categories or "flags." The researcher may establish the concept indicators in advance of analysis, depending on the problem being considered and his or her theoretical perspectives. In this situation, data are sorted into categories, for example, high-visual penetrative ability (VPA) or low-VPA (Alles and Riggs, this volume). Alternatively, concept indicators may arise inductively during observation/analysis. The framework of grounded theory is familiar to scientists: pattern recognition, consistency across different settings, visual models of processes, and unifying explanations of phenomena. In establishing models, researchers look for discounting or disconfirming evidence (e.g., Morrow and Smith, 1995). Grounded theory studies typically have a more scientific and objective language and feel (Creswell, 1998). Grounded theory is an appropriate framework for modeling teaching and learning processes. Kusnick (2002) applied a grounded theory framework in her work to outline a specific process (model) for how students come to understand sedimentary rock-forming processes. She articulated a common set of conceptual blocks that students navigate during their learning process. Alles and Riggs (this volume) work in a grounded theory framework in their model for the development of three-dimensional visualization skills among students.

Some workers (e.g., Creswell, 1998; Willis, 2007) classify grounded theory not as a guiding framework, but rather as a methodology. They consider grounded theory to be any set of procedures for constructing models that are based on data. However, I cast it here as a theoretical framework in the context of geoscience education research. This is because one purpose of this paper is to acquaint geoscientists with specific elements of qualitative inquiry. It is my assumption that many readers are unfamiliar with the territory. Because of the more "scientific" feel of grounded theory, it serves as a useful conceptual bridge between the more familiar theoretical framework of logical empiricism and other, less familiar frameworks such as critical theory.

Grounded theory is suitable for analysis of policy, especially in historical context. Those workers exploring patterns and consistencies in alternative conceptions are served by a grounded theory approach as well. To a point: Workers seeking to document situated power relationships, or document a lived experience will find grounded theory a limiting framework. Additionally, qualitative data that are site-specific and/or iteratively complex resist grounded theory analysis. In these cases, critical theory may be an appropriate framework.

Critical Theory

Critical theorists are researchers whose work is intended to be transformational (affecting change), liberationist (breaking down barriers and promoting freedom from literal and figurative oppression), and deconstructive (identifying and breaking down power relationships). These workers address social and educational problems such as systemic oppression and racism (e.g., Freire, 2000; Gould, 1993; Haymes, 1995), and sex and gender inequities (e.g., Christ, 1979; Barton, 1998). Critical theorists view science as a major tool in the construction of social realities (Kvale, 1995). Physical scientists seek to understand the workings of natural systems, and critical theorists seek to ensure that "Nature" is not pitted against "Man" in an adversarial relationship. Critical theory is counter to the use of physical science for dominion or oppression. An example of this is found in the work of Stephen Jay Gould (1993), who analyzed the nineteenth-century practice of craniometry, and the perversion of Darwin's theory of evolution by Victorian society. Gould (1993) deconstructed the Victorian application of evolution to justify British conquest and oppression of African and Asian peoples. As Gould points out, the Victorians felt that evolution resulted in an ultimately superior human phenotype, i.e., Caucasians, which gave the Europeans license for conquest. Craniometry was selectively applied to assert that non-Europeans had smaller cranial capacity, and therefore lower intelligence, and could be conquered (or domesticated or civilized) much like any nonhuman species (Gould, 1993). Gould was a noted paleontologist, and a critical theorist as well.

Feminist inquiry in science is rooted in critical theory. Barton (1998) presented a comprehensive history of the development of three waves of feminist science, beginning in the 1960s. The three waves evolved from liberationists attacking equity issues within the patriarchal scientific institution, to exploring multiple perspectives on the nature of science and ways of knowing, to challenging how science is positioned as a school subject (Barton, 1998).

Ostensibly, feminism works for the liberation and advancement of women, but true feminism and feminist science visualizes a (scientific) world that is inclusive (e.g., Mayberry and Rees, 1996) and does not speak a hegemonic (i.e., aggressive, colonialistic) language (e.g., Summa, 1995). This is a world in which social and gender realities are not merely relevant, but must play a central role in how scientists construct their worldviews (Harding, 1991; Nairn, 1996). As such, feminist theory has given rise to other critical-theoretical frameworks such as queer theory and ecofeminism. Queer theory challenges notions of social and scientific categorization of people (Sullivan, 2003), especially with regard to sexuality and sexual identification (e.g., Nairn, 2003). Ecofeminism synthesizes critical theory, feminist theory, and multiculturalism (Schwartz, 1999). It is concerned with the treatment of women, as well as nature, at the hand of patriarchal western science (Christ, 1979; Gadon, 1989; Ruether, 1992; Schwartz, 1999). Ecofeminists view Earth and nature not as passive entities that exist for the benefit of [Man], but rather Nature is cast as benevolent maternal figure and caretaker, and holds the ultimate power over the planet, in contrast to human dominion and exploitation (Feig, 2004). The Gaia hypothesis (Lovelock, 1972) is a central theme of ecofeminism. Critical theory is largely informed by postmodernism, which itself can be labeled a framework (e.g., Creswell, 1998). However, I have treated critical theory in such a manner as to "bundle" it with postmodernism.

Grounded theory and critical theory are two of many guiding frameworks for the qualitative researcher, informing his or her approach. Another approach is that of symbolic interactionism. This framework is based on the notion that human behavior is predicated on the identification of people and objects as symbols with attendant meanings (Blumer, 1986). Social interaction is interpreted as entirely symbolic, and researchers interpret the deeper meanings of the symbols. This approach is not common in geoscience education research; much of our work is grounded in behavioral-psychology approaches (e.g., Arthurs, this volume; Libarkin, 2005; Petcovic and Libarkin, 2007). Indeed, behaviorism is an empirical outlook that assumes rules and metrics for human behavior (Skinner, 1953). Because the behaviorist approach is dominant in the geoscience education literature, I do not provide a comprehensive discussion of it here.

It is important to understand that theoretical frameworks themselves do not define the purpose of a qualitative study. Understanding the purpose of a qualitative study, either as its producer or consumer, lies in understanding the study's elements of location. Locating the study and the researcher, and exploring the roles of the researcher reveal a great deal about the researcher's purpose and potential biases as well as what informs, or drives, the research.

LOCATING THE STUDY AND THE RESEARCHER

Geoscientists typically think about "location" as a physical parameter. Geologic investigations are located in a particular terrain or part of a state or province. Physical location is also important to qualitative inquiry, but this parameter takes on additional meaning. The study itself has a temporal location in multiple spaces, and the researcher has a location within the study.

Locating the Study in Time, Space, and Culture

Geoscience education research can focus on the present, it can be a historical investigation of some process or policy, or it can be predictive, leading to a model. The researcher must locate the study temporally in order to define the context of his or her work. Is the research intended to support a cognitive model of student learning, focused on future events and phenomena? Is the study a snapshot in time of a phenomenon? Does the study examine a phenomenon from a longitudinal perspective? Spatially, geoscience education research can take place in a classroom, in the field, "on the street," and in cyberspace. The less obvious factor is the notion of "cultural space."

We tend to equate culture with ethnicity, with a social "otherness" that emerges when contrasted with our own cultural identity as, for example, Americans, Britons, Anglos, Hispanics, or scientists. The anthropologist Harry Wolcott (1990) defined the subjects of qualitative study as "culture sharing groups." This definition does not focus on shared, inherent traits like ethnicity, but rather on patterns of observed behavior. For example, students in an introductory class express a culture; they share a set of similar experiences by virtue of their interaction with the instructor and with the course content; they take exams, participate in laboratory exercises, and attend field trips. Their culture-sharing group is independent of their ages, ethnicities, or other demographics. However, these demographics, together with their value systems, and political and religious positions inform and impinge upon the overall classroom culture.

This impingement compares well with what Vélez-Ibáñez (1997) defined as cultural "bumping." This is the notion that human populations are never isolated enough to not interact with each other on some level and thus remain unchanged. In this context, Vélez-Ibáñez was concerned with the bumping and intercultural interactions between indigenous peoples, Hispanics, and Europeans in what eventually became the southwestern United States. However, analogs exist between these social actors and the subjects of qualitative inquiry in geoscience education. For example, a researcher might be interested in how science majors and nonscience majors learn in the collaborative setting of his class. These two groups have their own cultural spaces, which bump each other in the overall cultural space of the class. Because the researcher is conducting qualitative, intergroup comparisons, he must work to understand the two culture-sharing groups, and describe this understanding in his research. His study has a location in cultural space. Through documenting the cultural location of his study, the researcher outlines who is being studied and how they interact. This additional context benefits the consumers of his research.

Sometimes culture-sharing groups and their bumping are more ephemeral, and more subtly defined. For example, in my own ethnographic research of field camp (Feig, 2010), I encountered the phenomenon of technology (GPS) dependence among the students. I interpreted these students in terms of cultural space as digital natives (Sheffield, 2007); they were comfortable with technology and never questioned it. However, the instructional staff of the field camp was, culturally (in the words of one student) "old-school" (Feig, 2010). The instructors valued nontechnological approaches to solving field problems. "Bumping" happened as the two groups juxtaposed their unaligned values, decisions, and viewpoints in the field. In this situation, cultural bumping yielded significant insight to my observation of field learning. I also discovered different cultural spaces within the student group. For example, some students were risk-takers in the physical environment, while others were risk-averse. This latter group spent considerable time planning traverses to minimize, as much as they could, exceptionally rough terrain, steep ridgelines, and sheer drops. By contrast, members of the former group moved across the landscape with minimal thought to topography or even, in some cases, to personal safety. Cultural bumping took place when risk-avoiders were paired with risk-takers to map the area, and when avoiders encountered instructors who insisted they negotiate a particular topographic feature. Risk behavior is a cultural classification in this context, because responses to risk are classifiable as a set of common observed behaviors. In both of these examples, my "big-picture" insights and the themes that I identified were heavily dependent upon my understanding and descriptions of my study's location in cultural space.

Qualitative inquiry requires the researcher to publicly address the question, "Who is being studied?" Another requisite question is, "Who is the researcher?" The answer to this question comes through the process of the researcher locating himself or herself in the study.

Locating the Researcher

In geological research, we are accustomed to assuming the role of a detached observer, as something of a disembodied eye, observing a single reality (Varela et al., 1991). We objectively collect data and make inferences about processes that operate independent of our thought or presence. It is possible to assume this role in educational research as well. Public policy, historical records, and aggregate test scores are examples of processes and data that exist independently of the researcher and can be studied objectively. This is not feasible when studying the ontological objects that are the subjects of qualitative inquiry. The role, purpose, potential bias, and background of the researcher each inform the generation of data and its subsequent interpretation. I provide three categorical examples of a researcher's role and potential purposes: (1) the researcher-observer; (2) the researcher-participant; and (3) the action-researcher.

The researcher-observer generates data by both passive and active means. Passive means include detached observation of a classroom or field setting. Participant behavior is recorded by audio, video, or in the form of field notes. In this sense, the researcher is "looking over the shoulder" (Wolcott, 2001, p. 117) of the participants. This role is useful for documenting student choices, how they cope with novel situations, or how they respond to a teaching innovation. Active observation encompasses passive methods, but is interactive in nature. An active observer engages participants either conversationally, or via interviews and focus groups. In this volume (Feig and Stokes, 2011), examples of studies in which the authors were located as researcher-observers include Clary and Wandersee, Ishikawa et al., and Stokes.

The researcher-participant is most commonly found generating data on his or her own classroom. This is the role assumed by those studying and documenting "best practices" in their own classroom and field-learning settings. For example, an instructor wants to understand the efficacy of a technological or pedagogical innovation beyond outcomes (quiz scores). A major purpose of this inquiry is to craft the innovation further, and to consider its application to future classes that she or he teaches. Additionally, the researcher may wish to share the innovation with the wider community. The *Journal of Geoscience Education* contains many examples of researcher-participant roles (e.g., Boundy and Condit, 2004; Basu and Middendorf, 2004; Earle, 2004), although the majority are quantitative inquiries. The researcher-participant observes his or her students engaged in or with the innovation, and may conduct interviews and focus groups. The instructor is not only a researcher, but is also a participant in the research, by virtue of (1) studying his or her own students, (2) having designed the innovation in question, and (3) using the results to improve and apply the innovation in future classes. Researcher-participant studies in this volume (Feig and Stokes, 2011) include Alles and Riggs, Arthurs, Arthurs and Marchitto, Atchison and Feig, Kortz et al., and Swenson and Kastens.

The action-researcher tackles educational questions in the context of social problems. An example of action-research is the work of Riggs et al. (2007) and Riggs (2005) in their efforts to address, respectively, increasing the participation of Native American students in the geosciences, and integrating geosciences and indigenous knowledge. The purposes of their research went beyond understanding phenomena of teaching and learning. Rather, they sought to address ostensible problems, such as broadening participation in Earth sciences, and incorporating multiple ways of knowing. Research based on service-learning and community-outreach efforts (e.g., Feig and Girón, 2001; Prakash and Richardson, 1999, respectively) is action-research. The action-researchers in the present volume (Feig and Stokes, 2011) are Williams and Semken.

A comprehensive statement of location made by the researcher is a public, transparent exploration of why the study was conducted, how the researcher fits into the study, and provides context in which to address potential bias. For example, by declaring a participant-observer role, the researcher has the opportunity to address the appropriateness of his or her methods, and how those methods may have influenced the data generation. The researcher-participant has the opportunity to address how she or he impacted student perceptions of and performance on a teaching innovation. An action-researcher has the opportunity to provide the framework of the social problem in question. In all of these cases, an understanding of the researcher's location gives the consumers of the research insight into emergent themes and how the researcher interpreted those themes.

A qualitative study is enriched by an articulation of the temporal and cultural location of the study, as well as the researcher's location within it. Doing so answers questions about why the study was conducted, who was studied, and who was investigating, but what about the actual practice of conducting qualitative research? This is a question of methodology.

METHODOLOGY AS SEPARATE AND DISTINCT FROM METHOD

A researcher will make choices about methodology and method, in part, based on his or her ontological and epistemological frameworks. In empirical science, "method" describes how research is conducted. Electron microscopy, statistical analysis, and disaggregation of sediment are all methods of geoscientific inquiry. As geoscientists, we tend not to make a distinction between our methods and our methodologies. In fact, we occasionally use those terms interchangeably. However, it is possible to distinguish them in geoscientific research. Consider the how these concepts might differ from one another:

(1) field methodology;
(2) laboratory methodology;
(3) mathematical modeling methodology.

The procedures (methods) used in the field differ from those used in the laboratory. They differ still from a mathematical modeling approach. For example, using a Jacob staff to measure section is a field method, but not something done in a laboratory. "Jaking" fits neither into laboratory nor mathematical methodologies. X-ray diffraction is a method used in the laboratory, but not in the field. This fits into the laboratory methodology.

One is hard pressed to find more than occasional instances in the literature where geologists have made the method–methodology distinction. It may be that many geoscientists would find it either artificial or useless, or both. However, this is not the case in qualitative inquiry. Observation, interviewing and conducting focus groups, comparing policy outcomes, and examining historical records are all examples of methods. Each one of these, however, can be applied across multiple methodologies. I discuss four examples of methodology with which I have experience as a qualitative researcher: hermeneutics, phenomenology, ethnography and policy analysis. Other valid methodologies exist, such as case study, phenomenography, narrative analysis, and biography. However, they are outside the scope of this paper.

Hermeneutics

A researcher-participant working in his or her own classroom moves multiple times between the roles of researcher and teacher. She or he is working with the intention of using the research to improve his or her teaching. The data move from being generated in a research environment to being put into day-to-day practice. These movements are defined as hermeneutic (Balfour and Mesaros, 1994). A hermeneutic approach seeks to understand a larger process through the understanding of smaller parts of that process, which in turn requires an understanding of that greater process itself (Schwandt, 2001). This is not circular reasoning; it is a shifting of perspectives. To understand why a student thinks or says something about, say, plate tectonics, it is important to perceive that thought or statement from multiple perspectives. We want to know what the thought/statement says about the larger phenomenon of learning in general (e.g., alternative conceptions studies; Libarkin and Anderson, 2005). We try to "get in the student's head" to improve our understanding of learning, but we need to have an understanding of the larger process of learning in the geosciences in order to get into the student's head. This is an example of a hermeneutic process, and it is applied in Kortz (this volume). If this student is in our class, then we are both the practitioner (teacher) and the researcher. In the purest sense, hermeneutics is not really a self-contained methodology. Rather, it is best thought of as a modifier for other methodologies, such as phenomenology or ethnography. Phenomenology, for example, can be hermeneutic or not.

Phenomenology

A phenomenologist seeks to understand the "essence" of things such as the everyday lived experiences of people engaged in a particular activity or process, and the values that drive them (Feig, 2004). Phenomenology, therefore, is a highly descriptive process (Schwandt, 2001). The kinds of data common in phenomenological studies include personal accounts and narratives, nonverbal behaviors, interpersonal interactions, individual choices, strategies, and attitudes. The ontological objects in phenomenology include basic realities, people, people as social actors, emotion, memory, consciousness, understandings and interpretations, ideas and perceptions, attitudes, beliefs, and belief systems. These data and ontological properties form and reside in communicated truths. A phenomenological methodology is appropriate for those workers who are seeking an intimate understanding of how reality is constructed (e.g., alternative conceptions), how preconceptions are acted on, or how students cope with new situations, i.e., novel spaces (Orion, 1993). Often, phenomenology is combined with ethnography for a blended methodology (e.g., Feig, 2004, 2010).

Ethnography

Ethnography is the careful and thorough documentation and description of a culture-sharing group with the goal of understanding that group (Wolcott, 1990). This is accomplished through immersive observation. Ethnographic observation is the act of living and working among one's subjects for an extended period of time. How long that time should be has been the subject of some debate. Anthropologist Harry Wolcott (2001) suggests a twelve-month minimum, while Margaret Mead (1970) argued for a far shorter period of time. Both workers ultimately agreed that the amount of time spent in the field should be enough to gain an intimate insight into the culture, persons, or processes being observed.

Ethnographers extract meaning by coding themes from interviews, conversations, and their own observations. These latter data are in the form of field notes. Ethnographic studies have a very different look and feel from other kinds of qualitative and quantitative studies. Data are typically excerpted rather than presented in full. The common format is that recommended by Wolcott (1994), which provides for a description, an analysis, and an interpretation. The description is essentially a narrated story, describing the events that contribute to a thematic understanding. The setting of those events is also described (e.g., Feig, 2010).

An ethnographic analysis constructs an argument out of raw data, such as student responses to the researcher's questions. Items that occur multiple times or are otherwise significant (meaningful) are "flagged" in the coding process. For further reference, Libarkin (2005) discussed qualitative analysis relevant to the geosciences, and the anthropologists Ryan and Bernard (2000, 2003) provided in-depth discussions on thematic analysis in ethnography. The interpretation portion of an ethnography asks the question, "What is to be made of the group being studied?" (Wolcott, 1994). This is where the ethnographer discusses the implications of emergent themes and places them in the larger world context of the phenomenon being studied. An example of the applications of ethnography to geoscience education problems is described in the place-based education research conducted by Semken (2005) and Semken and Butler Freeman (2008).

Policy Analysis

Policy analysis is the systematic investigation of the function of a set of rules, requirements, or norms. The major players are identified: the authors of the policy, those who it applies to, and those meant to enforce it (Anderson, 1996). Policy analysis examines the implied assumptions and values made by the authors of the policy. Winners and losers are identified, as well as unexpected outcomes (Anderson, 1996). The longitudinal effects of, and compliance with, a given policy are identified. Finally, the fate of the policy is explored, depending on whether it is continued or terminated. Topical examples of policy analysis studies include high-stakes testing in Texas public schools (McNeil, 2000), renewals of the Elementary and Secondary Education Act by Congress (e.g., Applegate, 2001), and ongoing educational reform (e.g., Geary and Groat, 1994).

SUMMARY

Qualitative inquiry is a powerful means for gaining deep insight into a process, event, or culture-sharing group. Qualitative study transcends the limitations of empiricism, the constraints of codified metric analysis, and the notion of a single, objective reality. In geoscience education, qualitative inquiry allows for the analysis of such data as communicated truths about student conceptions; feelings and perceptions about teaching and learning; lived experiences in the classroom and field; and attitudes and beliefs. These data are ontological objects that resist empirical and numerical analysis.

Qualitative researchers may choose from multiple theoretical frameworks. Those who wish to model a teaching or learning process by generating qualitative data work from a grounded theory perspective. Those who wish to directly and publicly confront social or educational problems through their research work in a critical theory framework, and those who conduct mixed-methods studies blend empiricism into their theoretical approach.

The qualitative researcher claims a location in his or her research, which varies depending on his or her purpose. The location can be that of researcher-observer, researcher-participant, or action-researcher. The grounded theorist constructing a model is usually a researcher-observer or a researcher-participant. The critical theorist seeking to affect change is usually an action-researcher. These options have limited utility in hypothesis-driven, empirical study. I do not suggest that qualitative approaches are better or more valuable. Rather, they allow for the parsing of educational problems where multiple realities exist, the data cannot be manipulated, and/or a call for change is needed. Qualitative study is appropriate for those workers who move back and forth hermeneutically between roles, such as a teacher studying his or her own classroom.

Qualitative inquiry requires the researcher to distinguish between methodology and method. This distinction is less important in empirical research, but the qualitative worker must select a methodology with his or her purpose, location, and group to be studied in mind. The qualitative researcher who is seeking to document the essence of a phenomenon chooses phenomenology; for a detailed understanding of a culture-sharing group, ethnography is an appropriate methodology. Those workers who seek to understand the purpose, intent, and detailed impact of rules or procedures select policy analysis as their methodology. Each of these options allows the use of multiple methods, including document review, direct observation, and interviews.

Qualitative inquiry is unfamiliar territory to many geoscientists. The notion of multiple realities is daunting; the need to consider location is novel; and the nature and iterative variability of qualitative data are potentially intimidating. However, the richness of meaning that can be extracted from these data, together with the potential for real change and impact as a result are worth every effort.

ACKNOWLEDGMENTS

My thanks go to A. Stokes, E. Riggs, C. Willermet, and an anonymous reviewer for helpful discussion. Ongoing inspiration was provided by I. Withnail and the S.P. Group. I especially thank M.P. Bickford and the Geological Society of America Books staff for cheerful aid and assistance.

REFERENCES CITED

Alles, M., and Riggs, E.M., 2011, this volume, Developing a process model for visual penetrative ability, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(06).

Anderson, J., 1996, Public Policy Making (3rd ed.): New York, Houghton Mifflin College Division, 352 p.

Applegate, D., 2001, Monkey business: The sense of the Senate: Geotimes, v. 46, p. 19–46.

Arthurs, L., 2011, this volume, What college-level students think: Student alternate conceptions and their cognitive models of geoscience concepts, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(10).

Arthurs, L., and Marchitto, T., 2011, this volume, Qualitative methods applied in the development of an introductory oceanography concept inventory survey, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(08).

Atchison, C., and Feig, A.D., 2011, this volume, Theoretical perspectives on an investigation of constructing experience through alternative field-based learning environments for students with mobility impairments, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(02).

Balfour, D., and Mesaros, W., 1994, Connecting the local narratives: Public administration as a hermeneutic science: Public Administration Review, v. 54, p. 559–564, doi:10.2307/976676.

Barton, A., 1998, Feminist Science Education: New York, Teachers College Press, 155 p.

Basu, A., and Middendorf, J., 2004, Demanding and enforcing high expectations in freshman courses: Journal of Geoscience Education, v. 52, p. 320–323.

Blumer, H., 1986, Symbolic Interactionism: Perspective and Method: Berkeley, University of California Press, 208 p.

Boundy, T., and Condit, C., 2004, Bringing the field into the classroom by using dynamic digital maps to engage undergraduate students in petrology research: Journal of Geoscience Education, v. 52, p. 313–319.

Brodaric, B., Gahegan, M., and Harrap, R., 2004, The art and science of mapping: Computing geological categories from field data: Computers & Geosciences, v. 30, p. 719–740, doi:10.1016/j.cageo.2004.05.001.

Chi, M., Slotta, J., and de Leeuw, N., 1994, From things to processes: A theory of conceptual change for learning science concepts: Learning and Instruction, v. 4, p. 27–43, doi:10.1016/0959-4752(94)90017-5.

Christ, C., 1979, Why women need the Goddess: Phenomenological, psychological, and political reflections, *in* Plaskow, J., and Christ, C., eds., Womanspirit Rising: San Francisco, Harper & Row, p. 71–86.

Clary, R.M., and Wandersee, J.H., 2011, this volume, Geobiological opportunities to learn at U.S. fossil parks, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(09).

Creswell, J., 1998, Qualitative Inquiry and Research Design: Choosing among Five Traditions: Thousand Oaks, California, Sage Publications, 403 p.

Earle, S., 2004, A simple paper model of a transform fault at a spreading ridge: Journal of Geoscience Education, v. 52, p. 391–392.

Feig, A.D., 2004, Identifying Ecoliteracy in the El Paso Borderplex (Ed.D. thesis): El Paso, University of Texas, 144 p.

Feig, A.D., 2010, Technology, accuracy and scientific thought in field camp: An ethnographic study: Journal of Geoscience Education, v. 58, p. 213–220.

Feig, A.D., and Girón, H., 2001, Building community connections in the Earth science laboratory, *in* International Sun Conference on Teaching and Learning, 1st Annual Meeting, Abstracts: El Paso, Texas, p. 2.

Feig, A.D., and Stokes, A., eds., 2011, Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, 212 p.

Freire, P., 2000, Pedagogy of the Oppressed, 30th Anniversary Edition: New York, Continuum, 183 p.

Gadamer, H.G., 1975, Truth and Method (G. Bardon and J. Cumming, trans.): New York, Seabury Press, 601 p.

Gadon, E., 1989, The Once and Future Goddess: San Francisco, Harper and Rowe, 432 p.

Geary, E., and Groat, C., 1994, Implementing science education reform: Geotimes, v. 39, p. 13–15.

Glaser, B., and Strauss, A., 1965, Awareness of Dying: Chicago, Aldine, 305 p.

Glaser, B., and Strauss, A., 1967, The Discovery of Grounded Theory: Chicago, Aldine, 271 p.

Gould, S.J., 1993, American polygeny and craniometry before Darwin, *in* Harding, S., ed., The "Racial" Economy of Science: Bloomington, Indiana University Press, p. 77–91.

Harding, S., 1991, Whose Science? Whose Knowledge? Thinking from Women's Lives: Ithaca, New York, Cornell University Press, 319 p.

Haymes, S., 1995, Race, Culture, and the City: Albany, New York, State University of New York Press, 167 p.

Ishikawa, T., Barnston, A.G., Kastens, K.A., and Louchouarn, P., 2011, this volume, Understanding, evaluation, and use of climate forecast data by environmental policy students, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(11).

Kortz, K.M., Clark, S.K., Gray, K., Smay, J.J., Viveiros, B., and Steer, D., 2011, this volume, Counting tectonic plates: A mixed-methods study of college student conceptions of plates and boundaries, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(12).

Kusnick, J., 2002, Growing pebbles and conceptual prisms: Understanding the source of student misconceptions about rock formation: Journal of Geoscience Education, v. 50, p. 31–39.

Kvale, S., 1995, Themes of postmodernity, *in* Anderson, W., ed., The Truth about the Truth: New York, Penguin Putnam, p. 18–25.

Libarkin, J., 2005, Conceptions, cognition, and change: Student thinking about the Earth: Journal of Geoscience Education, v. 54, p. 342.

Libarkin, J., and Anderson, S., 2005, Assessment of learning in entry-level geoscience courses: Results from the geoscience concept inventory: Journal of Geoscience Education, v. 53, p. 394–401.

Libarkin, J., and Kurdziel, J., 2006, Ontology and the teaching of earth system science: Journal of Geoscience Education, v. 54, p. 408–413.

Lovelock, J., 1972, Gaia: A New Look at Life on Earth: Oxford, UK, Oxford University Press, 176 p.

Mason, J., 2002, Qualitative Researching: London, Sage Publications, 223 p.

Mayberry, M., and Rees, M., 1996, Feminist pedagogy and interdisciplinary science education; working for social and scientific transformation: Geological Society of America Abstracts with Programs, v. 28, no. 7, p. A-330.

McNeil, L., 2000, Contradictions of Reform: The Educational Costs of Standardization: New York, Routledge, 256 p.

Mead, M., 1970, The art and technology of fieldwork, *in* Naroll, R., and Cohen, R., eds., Handbook of Method in Cultural Anthropology: Garden City, New Jersey, Natural History Press, p. 426–265.

Morrow, S., and Smith, M., 1995, Constructions of survival and coping by women who have survived childhood sexual abuse: Journal of Counseling Psychology, v. 42, p. 24–33, doi:10.1037/0022-0167.42.1.24.

Nagel, E., 1961, The Structure of Science: Problems in the Logic of Scientific Explanation: London, Routledge Kegan Paul, 618 p.

Nairn, K., 1996, Parties on geography field trips: Embodied fieldwork?: New Zealand Women's Studies Journal, v. 12, p. 86–97.

Nairn, K., 2003, What has the geography of sleeping arrangements got to do with the geography of our teaching spaces?: Gender, Place and Culture, v. 10, p. 67–81, doi:10.1080/0966369032000052667.

Orion, N., 1993, A model for the development and implementation of field trips as an integral part of the science curriculum: School Science and Mathematics, v. 93, p. 325–331.

Petcovic, H.L., and Libarkin, J.C., 2007, Research in science education: The expert–novice continuum: Journal of Geoscience Education, v. 55, p. 333–339.

Prakash, M., and Richardson, H., 1999, From human waste to the gift of soil, *in* Smith, G., and Williams, D., eds., Ecological Education in Action: On Weaving Education, Culture and the Environment: New York, State University of New York Press, p. 65–78.

Riggs, E.M., 2005, Field-based education and indigenous knowledge: Essential components of geoscience education for Native American communities: Science Education, v. 89, p. 296–313, doi:10.1002/sce.20032.

Riggs, E., Robbins, E., and Darner, R., 2007, Sharing the land: Attracting Native American students to the geosciences—Special Edition on Broadening Participation in the Earth Sciences: Journal of Geoscience Education, v. 55, p. 478–485.

Ruether, R., 1992, Gaia and God: An Ecofeminist Theology of Earth-Healing: San Francisco, HarperCollins, 320 p.

Ryan, G., and Bernard, H.R., 2000, Data management and analysis methods, *in* Denzin, N., and Lincoln, Y., eds., Handbook of Qualitative Research (2nd ed.): Thousand Oaks, California, Sage Publications, 1065 p.

Ryan, G., and Bernard, H.R., 2003, Techniques to identify themes: Field Methods, v. 15, p. 85–109, doi:10.1177/1525822X02239569.

Schwandt, T.A., 2001, Dictionary of Qualitative Inquiry: Thousand Oaks, California, Sage Publications, 281 p.

Schwartz, E., 1999, Exploring children's picture books through ecofeminist literacy, *in* Smith, G., and Williams, D., eds., Ecological Education in Action: On Weaving Education, Culture and the Environment: New York, State University of New York Press, 244 p.

Semken, S., 2005, Sense of place and place-based introductory geoscience teaching for American Indian and Alaska Native undergraduates: Journal of Geoscience Education, v. 53, p. 149–157.

Semken, S., and Butler Freeman, C., 2008, Sense of place in the practice and assessment of place-based science teaching: Science Education, v. 92, p. 1042–1057, doi:10.1002/sce.20279.

Sheffield, C., 2007, Technology and the gifted adolescent: Higher order thinking, 21st century literacy, and the digital native: Meridian: A Middle School Computer Technologies Journal, v. 10, http://www.ncsu.edu/meridian/sum2007/gifted/index.htm (accessed 9 September 2009).

Skinner, B.F., 1953, Science and Human Behavior: New York, The Free Press, 461 p.

Stokes, A., 2011, this volume, A phenomenographic approach to investigating students' conceptions of geoscience as an academic discipline, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(03).

Strauss, A., 1987, Qualitative Analysis for Social Scientists: Cambridge, UK, Cambridge University Press, 336 p.

Sullivan, N., 2003, A Critical Introduction to Queer Theory: New York, New York University Press, 240 p.

Summa, C., 1995, The language of geology: Can it change and should it change?: Geological Society of America Abstracts with Programs, v. 27, no. 6, p. A-148.

Swenson, S., and Kastens, K., 2011, this volume, Student interpretation of a global elevation map: What it is, how it was made, and what it is useful for, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(13).

Varela, F.J., Thompson, E., and Rosch, E., 1991, The Embodied Mind: Cognitive Science and Human Experience: Cambridge, Massachusetts Institute of Technology Press, 308 p.

Vélez-Ibáñez, C., 1997, Border Visions: Mexican Cultures of the Southwest United States: Tucson, University of Arizona Press, 361 p.

Williams, D., and Semken, S., 2011, this volume, Ethnographic methods in analysis of place-based geoscience curriculum and pedagogy, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(05).

Willis, J., 2007, Foundations of Qualitative Research: Thousand Oaks, California, Sage Publications, 392 p.

Wolcott, H., 1990, Writing Up Qualitative Research: Newbury Park, California, Sage Publishing, 96 p.

Wolcott, H., 1994, Transforming Qualitative Data: Description, Analysis and Interpretation: Thousand Oaks, California, Sage Publishing, 440 p.

Wolcott, H., 2001, The Art of Fieldwork: Walnut Creek, California, AltaMira Press, 285 p.

MANUSCRIPT ACCEPTED BY THE SOCIETY 23 JUNE 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

Theoretical perspectives on constructing experience through alternative field-based learning environments for students with mobility impairments

Christopher L. Atchison
School of Teaching and Learning, Ohio State University, Columbus, Ohio 43210, USA

Anthony D. Feig
Department of Geology and Meteorology, Central Michigan University, Mt. Pleasant, Michigan 48859, USA

ABSTRACT

Students with physical disabilities encounter challenges in any scientific discipline, yet the geosciences have extremely low participation levels for persons with disabilities. Because of the emphasis placed on field research at the undergraduate level, persons with mobility impairments face limited opportunities for progressing in the geosciences. One strategy to address this is the application of adaptive technologies, such as virtual field trips (VFTs), as a supplement to traditional field instruction. A common goal of VFTs and other adaptive technologies is to promote equal access to undergraduate geoscience curricula for physically impaired students. If the scientific talents of these students are embraced and accommodated, regardless of their physical ability, the overall welfare of the geosciences as a discipline is enhanced.

This paper describes ongoing research into the development of one specific VFT: an electronic re-creation of Mammoth Cave National Park for the *Introduction to Cave and Karst Systems* field course at a Midwestern research university. This paper focuses on the theoretical processes necessary to conduct qualitative inquiry for the purpose of developing an accessible, alternative field-based learning environment. Grounded theory and critical theory are contrasted as two possible guiding frameworks. Three roles for the researcher are compared: researcher-as-observer, participant-researcher, and action-researcher. Phenomenology is discussed as the preferred methodological choice for this research, and attendant methods are described. Finally, a discussion of validity and reliability issues is provided. This paper is intended to serve as a guide for future researchers embarking on qualitative studies similar to this one.

INTRODUCTION

Qualitative researchers must have a good working knowledge of theoretical frameworks and methodologies, but how are they actually applied? Consider the compelling issue of field education and the barriers it presents to students with mobility impairments (Cooke et al., 1997, Hall et al., 2004; Locke, 2005; Stokes and Boyle, 2009). Research in this area represents the intersections among student learning, classroom innovations, challenges to geoscience curricula, and general policy applied to a subset of students. The design and subsequent study of adaptive technologies, for example, is one nexus in which these topics all

Atchison, C.L., and Feig, A.D., 2011, Theoretical perspectives on constructing experience through alternative field-based learning environments for students with mobility impairments, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 11–21, doi:10.1130/2011.2474(02). For permission to copy, contact editing@geosociety.org.

reside. Adaptive technology may be first implemented as a classroom innovation. The impact of the implemented technology on student learning may then be (quantitatively) assessed. Its application could then be compared to extant pedagogical practice, which itself may be anecdotally considered an idealized baseline condition. For example, a web-based field trip may be designed by an instructor to provide his or her mobility-impaired students with a field-based learning opportunity. Subsequent research into this innovation could address the following questions/objectives: (1) How was it designed and executed? (2) What contribution did the innovation make to student learning? (3) How does it compare as a supplement to—or replacement of—a traditional field trip?

These research objectives would ideally be approached in a mixed-methods manner. Quantitative analyses of outcomes between experimental and control groups of students are appropriate measures of efficacy. Qualitative inquiry provides a deep picture of students' lived experiences with the innovation, guiding its future applications, and providing an additional illustration of its efficacy. Qualitative inquiry allows the researcher's role to be presented and understood. This in turn allows the research to be deliberately applied to address a social problem, such as providing equal access to the sciences for disabled students.

PURPOSE OF THIS PAPER AND GENERAL PROBLEM STATEMENT

This paper is intended to serve as one example of the conceptualization of a qualitative study. It does not present a completed study, with a description of methods, data, results, and implications. Instead, it is a theoretical examination of the earliest stages of a research and development project. Guiding theoretical frameworks and roles of the researchers are discussed in detail. Methodological choices are described and justified. Strategies for ensuring reliability and trustworthiness are then outlined.

The basic problem to be addressed by this research is that of improving accessibility to the geosciences for mobility-impaired students, which is defined more deliberately in the population characteristics section of this paper. One avenue to accessibility is the application of adaptive technologies, specifically virtual field trips (VFTs). A desire to understand the utility of VFTs in the context of accessibility generates three specific research objectives to be addressed by the present qualitative inquiry. Prior to a discussion of these three objectives, however, the necessary background on geoscience curricula and mobility-impaired students is provided.

BACKGROUND

In the most recent and thorough study documenting the enrollment of geoscience students with disabilities, it was determined that by the year 2000, total college enrollment in the United States approached 15.3 million persons (NCES, 2008). Of that total, 10.6%, or ~1.4 million students, reported having a disability of some kind (NCES, 1999; Locke, 2005). According to the American Geological Institute (AGI), during the 1995–1996 school year, the total U.S. undergraduate enrollment in the geosciences was estimated at 32,932; of this group, only 59 students (0.17%) were identified as disabled (AGI, 2006; Locke, 2005). In 2006, AGI also reported that of all science, technology, engineering, and mathematics (STEM) bachelor degree recipients, only 6.45% were individuals documenting some kind of a physical disability (AGI, 2006). Furthermore, as of 2004, only 7% of the science and engineering workforce was composed of persons with disabilities (NSF, 2004). The downside of these figures is that they combine all physical disabilities into one category, and do not categorize a disability that may be a hindrance to one's mobility. Despite the fact that science education reform is widely accepted (McCarthy, 2005), little research has been conducted to determine how the practice of science education addresses the needs of individuals with disabilities. This determination would begin to satisfy the nondiscriminatory provisions of the Americans with Disabilities Act of 1990 (P.L. 101-336) and of Section 504 of the Rehabilitation Act of 1973 (P.L. 93-112) (Cooke et al., 1997).

A comparison of the proportions of students with disabilities in the geosciences to the overall college population suggests that students with disabilities are poorly represented in geosciences. This discrepancy represents opportunities to (1) diversify the geoscience student population, (2) improve the inclusiveness of geoscience curricula and (3) explore instructional innovations, particularly with respect to field-based education, which will potentially improve field-based education for all student populations (Cooke et al., 1997; Norman, 2002).

Challenges and Opportunities of Field-Based Education

Students with disabilities encounter unique challenges in any scientific discipline, yet geoscience remains one of the sciences with the lowest participation levels for persons with disabilities (Locke, 2005). With the emphasis placed on field research at the undergraduate level, persons with various physical ability impairments face profound barriers to obtaining a higher education in the geosciences. Geologic field study is considered a key component of a well-rounded understanding in geology and earth sciences (e.g., Elkins and Elkins, 2007; Maskall and Stokes, 2008; Riggs et al., 2009). The basic method of classroom instruction is not enough; no analog exists for traversing a landscape (Riggs et al., 2009). First-hand observation and construction of field knowledge, which are associated with the aspect of "embodied field work" (Nairn, 1999), are especially important for a novice geology student with limited field knowledge and experience, regardless of their physical ability (Elkins and Elkins, 2007). The embodiment of the fieldwork experience is often represented by the effect that the field experience has on the student, both cognitively and physically. It is depicted by the ways in which the student begins to understand the content and operate as a field practitioner rather than just a student. This requires a physical interaction with the field environment as well as the learning experience. Professional geoscientists maintain that "field competence is an essential skill" for undergraduate geoscience students and should be emphasized as a primary part of the geoscience curriculum (Whitmeyer and Mogk, 2009). Additionally, an increase in geoscience field-based coursework offerings in most geology and earth science departments is reappearing nationwide (AGI, 2009; Whitmeyer and Mogk, 2009). This suggests a potential revitalization of the importance of field instruction in the geoscience curriculum that had been on a slight decline in recent years, and, therefore, an increased need for accessibility.

A student's identity within the geosciences may also be profoundly shaped by field study. Locke (2005) identified a traditional conceptualization of fieldwork that ultimately suggests that "geoscience careers are only for the strong and able-bodied" (Locke, 2005, p. 2). Those students who do not fit this able-bodied profile are therefore marginalized, and excluded from traditional geoscience fieldwork (Nairn, 1999; Hall et al., 2004; Hall and Healey, 2005). Based on the findings of Hall et al., an assertion can be made that students with physical disabilities should not be treated as *fragile* students, but as collaborative learning partners who can offer an alternative view of reality based on their own perspective. All students have needs, and in order to improve access and opportunities, as well as to increase diversity in the geosciences, all needs must be taken into consideration when preparing inclusive, educational field excursions (Cooke et al., 1997; Healey et al., 2002). Increasing the awareness of barriers to complete participation by all students will limit the difficulties to overcoming these barriers through the course design process (Healey et al., 2002). It is therefore imperative to develop innovative learning environments that will transform and improve learning experiences for all students. So, in order to accommodate students with mobility-impairments who primarily face the physical barriers of field-based coursework, adaptive technologies must be designed and developed.

Adaptive Technologies in Field-Based Education

Fieldwork is primarily designed as a physical experience; students go to the field, traverse the landscape, take measurements, collect rock samples, and make observations of the geology by navigating through it. This experience engages multiple senses, yet the overall sensory experience is narrow, relying heavily on the physical abilities of the student (Hall et al., 2002). The traditional field-based learning experience is predicated on a given level of mobility. Therefore, mobility impairments negatively impact the field learning experience. This is a barrier to students with disabilities, especially if no modified methods of field instruction exist.

Visualization technologies, such as virtual field trips (VFTs), are not only a practical supplement to traditional field methods, but they are also a potential alternative instructional method for all students requiring improved accessibility to a field experience. VFTs range from simple and static narratives: hot-linked web pages or animations, to dynamic and complex interactive experiences (Spicer and Stratford, 2001; Qiu and Hubble, 2002). Interactive VFTs are capable of analyzing students' conceptualizations of the environment through the development of mental models of environmental scenarios (Shepardson et al., 2007), three-dimensional (3-D) computer-based simulations, visualizations, and computational modeling (McKinney, 1997; Schwert et al., 1999; Barab et al., 2000; Bakas and Mikropoulos, 2003; Saat, 2004). Such environments have been developed to inform students on reducing potential hazards in a natural field environment, serve as a supplemental resource to enhance instruction and interpretation skills both before and after a field trip, or completely replace trips to a location that may be inaccessible or no longer in existence. In addition, the realism and effectiveness of VFTs continue to improve with the advancements of technology and visualization. The 3-D displays of stereographic representation, surround display of Cave Automatic Virtual Environment (CAVE) systems, supercomputer-based computational modeling and immersive virtual reality (VR) including the potential for real-time tele-presence visualization, a real-time synchronous interaction between two live users within two different locations, provide effective alternatives to field courses without the need to enter the field (Jackson and Winn, 1999; Lascara et al., 1999; Tan and Subramaniam, 2003; Stredney et al., 2008).

Alternative field environments, however, should not be considered superior to traditional field courses. It is not the intent for technology to replace traditional field studies, but rather enhance them by increasing accessibility for persons with mobility impairments who would otherwise be potentially excluded from the field excursion. However, opportunities for alternative field-based education would offer utilities that are an advantage to all students (Cooke et al., 1997; Norman, 2002). It is also reasonable to assume that virtual field studies will provide alternative ways to enhance existing geoscience field courses by supplementing pretrip content knowledge acquisition or post-trip review of concepts (Holt, 1996; Spicer and Stratford, 2001; Qiu and Hubble, 2002). These virtual programs are finding respect within several geoscience programs as a supplement of virtual and hypermedia materials and resources for pre- and post–field-trip activities (Spicer and Stratford, 2001). Through strong instructional design, these supplementary materials could reduce the duration of the field study as well as increase the informational transfer while in the field (Qiu and Hubble, 2002). Also, supplementary information to the field content would potentially boost the confidence of novice geoscience students, allowing them to refresh the material during a post-trip review. The greatest potential impacts of VFTs, however, are the opportunities presented to students who do not have access to the field because of mobility impairments.

DESIGN AND DEVELOPMENT OF THE VFT

As stated already, virtual field environments can take on a variety of formats, from low-technology text and image-based, self-paced field reference guides to a high-technology, complete virtual presence with an immersive 360°, 3-D visual display of the environment in reference. Additionally, task-specific activities contained in a large number of VFTs found on the internet have a wide spectrum of content and quality variance (Qiu and Hubble, 2002). However, recent studies have determined that higher levels of immersion in virtual environments have increased benefits of spatial understanding and a correlation to natural reality (Schuchardt and Bowman, 2007; Stredney et al., 2008). Through 3-D display, geographic distances and sizes of comparable objects can be realistically rendered. The added scenery promotes a more accurate content acquisition than current desktop systems (Schuchardt and Bowman, 2007).

We investigated a process aimed at developing a complete immersive, multidimensional, abstract re-creation of a natural cave environment. An immersive technology environment experience is being conducted in order to test the multimodal aspects of different VR environments, and to determine the best environment to address the learning objectives and create a realistic field experience. Deliberately broad, the goal is to narrow down the type of visualization technology based on effectiveness of re-creation and participant usability. The aspects of modality include, but are not limited to, seamless and realistic image rendering, visual movement within the virtual environment, and interactivity of the user-simulation interface design. This investigation looks at three separate, virtual environments, two of which have been developed during former studies at Ohio State University, and are available to be utilized for this specific purpose. The first of the two virtual environments was developed as a re-creation of a hazardous scenario using farming equipment in which the virtual re-creation allows for an enhanced level of safety and control for the user (Stredney et al., 2008). The other is a complete re-creation of archaeological ruins that are no longer accessible to public visitors: an interactive model of an ancient sun-dial calendar site in Chaco Canyon, New Mexico (Nicoli et al., 2008). The rationale for using these predeveloped environments is that they are readily available on site with the required equipment and software needed, and will provide a diverse representation of the VR options needed for the initial design of the pilot VR model.

The first environment is a more visually immersive agricultural simulation; users wear a full, head-mounted display interface and have the ability to visualize their own hands within the environment through wrist-tracking devices. This simulation provides a significant amount of user-environment interface interaction that controls their experience within the fully rendered, interactive virtual environment. This simulation also boasts elements of motion from the simulated objects within the environment that are independent of the user.

The second environment is a full re-creation of a real archaeological site using a high-resolution 10′ × 6′ stereoscopic three-dimensional (3-D) display; the user is capable of interacting with other users, while controlling 360° of visual mobility around the virtual site. The medium is a rear-screen 3-D projection system where the user wears polarized glasses.

The third simulation will be a real-time tele-presence experience also using technology available at OSU. This two-dimensional (2-D) user-to-user video display will allow a laboratory/field student exploration team (one student user in the interface laboratory while the other user is remotely exploring in the field environment) to work in tandem in a field exploration experience. The laboratory student will receive visual imagery through a monitor or projection display from the perspective of the explorer and be able to have full vocal interaction with the explorer. The laboratory student will have control over the environment to the extent that the explorer responds and obliges. The intention will be to then begin translating this into an interactive virtual explorer, controlled by the laboratory student.

The three VR environments differ significantly from each other in scope of technology, display, interactivity, and perception. Utilizing the personal experiences of the students, we will be able to determine which features of the simulation are most effective for them and develop a completely innovative and interactive design for the final VFT. The students involved in this investigation will assist in assessing the capability of the virtual field environment to represent geologic content, and they will provide an interactive presence and an accurate visual perspective.

Context of the Field Site

Mammoth Cave National Park can be described as one of the world's most fascinating places for detailing Earth's history as well as recent human history. This cave system, which is considered the longest in the world, contains well-preserved evidence from past civilizations as well as the early history of the United States of America. This is a location that is utilized by many research and educational institutions for the vast learning opportunities it possesses. In fact, Mammoth Cave National Park, and the surrounding region, is the primary focus for one of the field courses at The Ohio State University in cave and karst processes within the School of Earth Sciences. However, like many field-intensive courses, there is an accessibility issue with this excursion, which excludes many mobility-impaired students from taking part in the course. The extremely limited access to the park's resources and challenges of descending stairs and very narrow passages further reduce the likelihood that these students would consider taking part in this field course.

The VFT will be a virtual re-creation of portions of the cave's interior—an innovative solution to improve accessibility to the park's extensive resources. This project is facilitated through collaboration between the Ohio Supercomputer Center, the Advanced Computing Center for Arts and Design, Mammoth Cave National Park, and Ohio's STEM Ability Alliance, and it is supported by National Science Foundation funding (NSF GEO-0939645).

SETTING THE STAGE FOR THE PRESENT STUDY: THEORY AND LOCATION IN QUALITATIVE INQUIRY

Quantitative inquiry and scientific empiricism are both predicated on the notion of a measurable, objective reality and the ability to replicate experiments and observations. Geoscientific investigators have a unifying purpose: to discover the "truth" about a phenomenon and how it works. By contrast, qualitative inquiry considers problems that are much more "messy" in nature. Attitudes, perceptions, and individual communicated truths are the data generated by qualitative inquiry (Mason, 2002; Feig, this volume). These data do not lend themselves well to experimental manipulation, numerical analysis, or external judgments of validity.

Researchers addressing geological problems, such as the structural history of a region, the age of a pluton, or the evolution of a landscape, are working wholly within the theoretical framework of scientific empiricism. They do not need to consciously reflect on their purpose and their role as the researcher, and then choose a theoretical framework in which to conduct their work. To understand a natural process, to verify or reject a hypothesis about that process, and to apply a model of that process across settings are empirical and logical procedures, based on experimentation and replication.

Researchers addressing educational problems through qualitative inquiry, however, have a choice of theoretical frameworks. Their choices are based on their purpose in conducting the research, as well as their place within it. The options for theoretical frameworks and the researchers' place in qualitative study are described next, and the choices made for the present study are outlined.

Choosing a Theoretical Framework

Qualitative researchers may choose from a variety of theoretical frameworks, including quasi-empiricism, behaviorism, symbolic interactionism, grounded theory, or critical theory, depending upon their motives and goals. A common approach in geoscience education research is a blend of behaviorism and grounded theory. Behaviorism is predicated on the notion that human behavior is a set of responses to stimuli that follow basic rules and patterns (Skinner, 1953). Grounded theory is a data-driven approach to building a systematic model, or other type of theoretical construct (Creswell, 1998). This model should have some applicability beyond the study that generated it. The model is constructed by means familiar to geoscientists: recognition of patterns (concept indicators), unifying explanations of a phenomenon, and the outlining of visual, stepwise sequences in a process. One example of grounded theory in geoscience education research is that of Riggs et al. (2009). These workers equipped students in a field mapping class with global positioning system (GPS) transmitters that recorded their movements in the field. These movements were subsequently overlain onto a geographic information system (GIS) grid. Riggs, Leader, and Balliet then identified navigation paths and coded them to produce consistent models of student performance as a function of land navigation. The outcome of their study was a new tool (model) for gauging student performance in geologic field problem solving.

By contrast, critical theory is an approach suitable for those researchers seeking to affect social or educational change (Mayo, 2007). In this context, "change" means addressing the social and educational problems of equal access (e.g., Freire, 2000), power relationships (e.g., Gould, 1993), or race- or sex-discrimination (e.g., Haymes, 1995; Barton, 1998). In the geosciences, an example of research informed by critical theory is that of Williams and Semken (this volume) and Semken and Brandt (2010). The former workers seek to affect change in educational praxis through challenging the status quo of geoscience teaching. They seek to transform education from static, exclusionary dispensation of facts into a dynamic, inclusive process that connects learners to a specific landscape. The latter workers take on broader societal issues of cultural sustainability and ecojustice. They approach these problems by calling for the integration of Euro-American science with indigenous knowledge. Semken and Brandt (2010), through applications of place-based geoscience education, seek to give Native American peoples and their knowledge a voice.

In this study, we ultimately seek improved accessibility within the geosciences for a previously marginalized population; this is a challenge to the status-quo preconception that some students can "do" geology, while others cannot. An eventual outcome of this research will be the development of a generalizable model of effective student-VFT interfaces. However, the main thrust of the present study is to challenge current geoscience field-based accessibility. Therefore, this present study is to be conducted in the framework of critical theory. We, as authors, are critical theorists.

CHARACTERISTICS OF THE STUDY POPULATION

As with most qualitative studies, the population size of the present investigation will be small ($n \approx 20$). Half of the participants ($n = 10$) within this study population are mobility impaired; the rest ($n = 10$) are trained personal assistants for the mobility-impaired students, who will not be compared to the students, but rather be observed working and interacting with them throughout the learning experience. For the purpose of this study, the term mobility impairment represents any physical condition that prevents an individual from performing a key life activity through movement; conditions that limit their ability to use stairs or traverse rough terrain, not due to sensory or psychological impairment. The rationale for this study is to determine how experience in a geologic field environment assists in the overall construction of knowledge. Given this idea, most traditional field environments are inaccessible to mobility-impaired students. The data gathered from these individuals will be necessary for the development of an alternative field environment that will permit students to experience the field site virtually, within a controlled facility, learning within the field environment without the physical or emotional stress of negotiating the natural field site.

Participants were identified through the assistance of the Office of Disability Services (ODS) at Wright State University (WSU), as well as the Project Coordinator for Ohio's STEM Ability Alliance (OSAA). All participants in this study are to be enrolled in an introductory cave geology course within the Earth and Environmental Sciences department at WSU entitled *Introduction to Cave and Karst Systems*, which is only being offered to students registered with ODS and their personal assistants. The principal investigator will gain access to these students as a result of being a co-instructor for the geology course.

Factors to determine participant eligibility include the following:

(1) mobility impairment, registered with the University Office of Disability Services (ODS) *or* a personal assistant to a student registered with ODS;

(2) must be in good academic standing; and

(3) enrollment in EES 199, *Introduction to Cave and Karst Systems*.

LOCATING THE RESEARCHERS

The location in time of this study is the present; it is not a historical or longitudinal study. The spatial locations of this study are the face-to-face (F2F) classroom and field site at Mammoth Cave National Park, as well as the interface laboratory at the Ohio Supercomputer Center for the study of the virtual environments associated with this investigation. In a qualitative study, the researcher must also describe his or her role in the research. Ultimately, the questions that must be answered are, "Who is the researcher, and why is she or he conducting this research?" Such questions are highly relevant in qualitative inquiry (Feig, this volume; Mason, 2002; Wolcott, 1999) because they reveal the purpose of the study. Three possibilities exist for a researcher's location: the researcher-observer, the researcher-participant, and the action-researcher. These are discussed in detail in Feig (this volume), but a brief overview is included here.

The researcher-observer documents, among other things, how students cope with novel situations, how they respond to a teaching innovation, and how they navigate tangible or intangible barriers. This researcher generates data either by passive observation, or active processes such as interviewing participants. Clary and Wandersee (this volume) are researcher-observers. They engaged in direct observation of verbal and nonverbal behaviors to document student reaction to, and experience in, an informal learning environment.

Researcher-participants study their own classrooms. These are the instructors who have created a classroom innovation, and wish to document its efficacy and impact. They often seek to refine the innovation, or expand its application beyond their setting. The instructor is not only a researcher, but is also a participant in the research, because she or he is studying his or her own students, and is using the results to improve and apply the innovation in his or her future classes. The researcher-participant moves back and forth between the roles of "detached" observer and active participant in the research.

Action-researchers seek to address social problems through their research. This can be a reflexive process, as when an instructor seeks to modify his/her own praxis (Hunter, 2007); alternatively, the researcher may be working on an external social/educational problem. For example, research that is meant to broaden participation in geoscience by underrepresented groups (e.g., Riggs, 2005), or promote environmental activism and stewardship (e.g., Smith and Williams, 1999) is action research. Action-researchers are typically critical theorists, because they are issuing a call to action or a challenge to the status quo, beyond understanding or modeling a process. The present study is best categorized as action-research, because the overarching goal is to increase geoscience access to a commonly marginalized population at the study institution. This is a direct challenge to current educational practice, compelling educators to evaluate and, as necessary, rethink their preconceptions of the educational ability of mobility-impaired students.

In the present study Atchison, as the principal investigator, is a critical theorist, and has a location as an action-researcher. He is seeking to solve the educational problem of mobility-impaired students being excluded from traditional field-based education opportunities. The "call to action" here is for instructors to see the application of a VFT as a viable educational opportunity for this population of students in particular, and ultimately all geoscience students.

RESEARCH OBJECTIVES

The general research problem of improving access to fieldwork with VFTs generates three specific research objectives for qualitative inquiry. First, successful application of VFTs as an adaptive technology for mobility-impaired students requires an understanding of how students construct geological knowledge in the face of field-related barriers. Second, evaluating VFTs as a supplement to field instruction requires an understanding of how students interface with their environment, i.e., understanding their lived experience with (and within) a traditional field site. Third, the overall effectiveness of a VFT must be documented in terms of learning outcomes and other measures of student achievement. This documentation will begin to determine how the virtual re-creation adequately promotes an authentic interaction with the natural environment.

Critical theory is the consistent guiding framework in which each of these objectives will be addressed. Another consistent factor is the location of the authors as action-researchers. The variable, however, is the methodology and methods used for each objective, as well as the processes of ensuring reliability and trustworthiness. For the sake of convenience, these variables are grouped as the "mechanics" of each research question, and are discussed accordingly. Figure 1 provides a process map listing each research objective, applied methods and methodologies, and expected outcomes.

Research Objective 1: The Construction of Geological Knowledge in the Face of Field-Related Barriers among Mobility-Impaired Students

Substantial work has been conducted on student learning in the field (Thrift, 1975; McKenzie et al., 1986; Orion, 1993; Garrison and Endsley, 2005; Elkins and Elkins, 2007; Potter et al.,

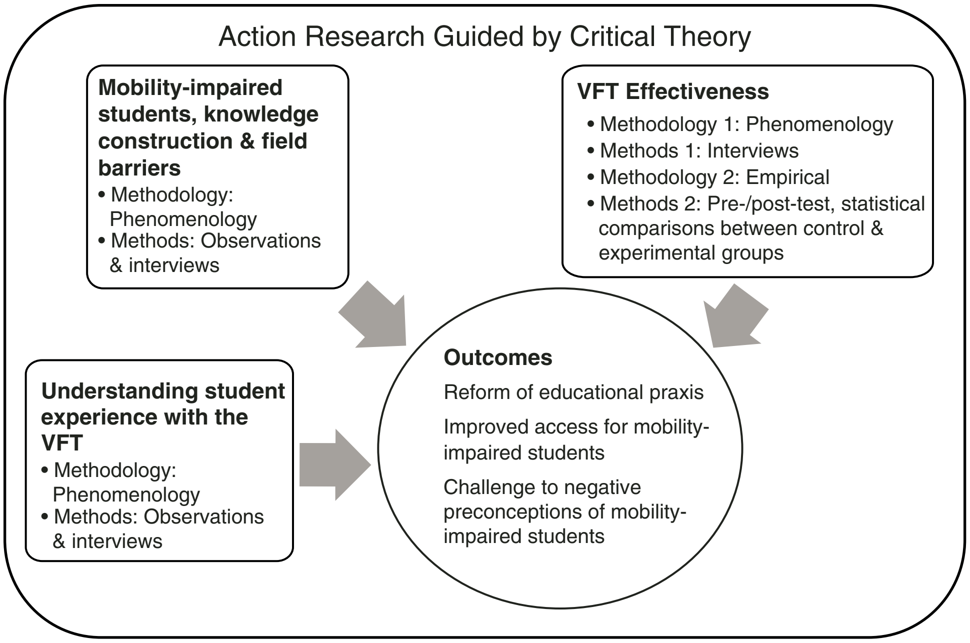


Figure 1. Research problems (in bold), methodologies and methods (bulleted), and their contributions to potential outcomes of this study.

2009; Thomas and Roberts, 2009). Much less work has been done on students with disabilities attempting to learn in the field (Cooke et al., 1997; Hall et al., 2004; Healey et al., 2002; Locke, 2005; Hall and Healey, 2005), and even fewer on field-based education with students with mobility impairments (Norman, 2002; Stokes and Boyle, 2009). In terms of specific challenges of the field environment, Orion and Hofstein (1994) articulated the concept of "novelty space": geographic, cognitive, and psychological barriers that field students negotiate in the field-based learning process. All students face aspects of novelty space with respect to field-based learning environments. However, many studies on field-based education assume a situation in which the learners are physically able-bodied. As a result, the documentation of knowledge construction did not take into account the physical and psychological barriers that are significant to learners with mobility impairments that are evident in Orion's concept of novelty space.

This first research objective addresses this issue by seeking to discover the exact nature of the potential barriers related to the Mammoth Cave National Park field site, and how mobility-impaired students negotiate them in the learning process. In terms of qualitative inquiry, this is an objective of the "essence" (Creswell, 1998; Mason, 2002) of a particular experience. An appropriate methodology here is phenomenology, which is a meticulous documentation and understanding of the essence of a lived experience (Schwandt, 2001; Feig, 2008). Understanding comes through the process of extracting meaning from data, and relevant data include behaviors, feelings, choices, attitudes, personal accounts, and interpersonal interactions (Feig, this volume). As such, this methodology is appropriate for the objective of documenting how students construct knowledge through experience.

The methods used to generate these data include individual and focus group interviews, attitudinal and motivational surveys, student journaling, and observations. The interviews may be unstructured, semistructured, or highly structured, depending on the researcher's intentions (Wolcott, 1999). For example, in their work to refine the "Geoscience Concept Inventory," Clark and Libarkin (this volume) asked novices and experts specific questions about tectonic processes in order to document their conceptions of how the process worked. Their questioning was directed, but it allowed participants to guide the discussion wherever it might go. Clark and Libarkin did not know what the specific conceptions would turn out to be, and so they allowed the conceptions to emerge naturally through conversation. This is an example of semistructured interviewing. In the present research, a semistructured interview process will be applied to generate discussion about potential barriers to learning and the participants' negotiation of those barriers, while allowing thick descriptive data to emerge naturally through conversation.

Research Objective 2: Understanding Student Interactions with Their Environment

This research objective considers mobility-impaired students in the context of both the natural and virtual field-trip experiences. Phenomenology is again an appropriate methodology, since the objective is to carefully document the essence of individual experiences with the VFT. The methods used would also include observation and structured interviews both before and after using the VFT. These methods allow participants to be guided through the process of reporting and reflecting on their experiences. The data of interest here include accounts of using the VFT, students' perceptions of the VFT, and changes in conceptions as a result of interaction with the VFT, all while relating it to the natural field location. Participants will be interviewed about their understanding of cave formation, cave features that are used to assist the interpretation of that formation, and the physical accessibility issues associated with the natural environment. One potential outcome of this procedure is the construction of a model that describes student interaction with the natural environment, either holistically or as progressive stages, that can be mimicked in a virtual environment. This model would be useful to other instructors and applicable to groups of student-users beyond those studied here. The chosen strategy lays a foundation for a descriptive portrait of an individual's perspective and lived experience. Without having a documented, first-person account of the experience, researchers are left to detail an event through hearsay or second-hand evidence, i.e., describing one's perspective without having the experience to do so.

Research Objective 3: Effectiveness of an Actual VFT in Engaging Mobility-Impaired Students

The effectiveness of the VFT within the given student population may be assessed both qualitatively and quantitatively. In a qualitative process, a phenomenological methodology would be applied to document effectiveness in the form of individual "successes," including evolving conceptions and so-called "Aha! moments." The appropriate methods here are individual and focus group interviews, student journaling, a survey of motivation in science learning, and direct observation in order to build a narrative of the VFT as an effective strategy to remove barriers to learning. Applying these methods will allow participants to narrate their experience with the VFT, identifying threshold concepts of the content, and the ways in which, in detail, using the VFT altered their conceptions. The generated data will be in the form of oral and written narratives.

This research objective is also well suited for quantitative study. For example, student outcomes in the study population can be measured against criterion-referenced learning outcomes. In the Mammoth Cave field course, these learning outcomes include the ability to conduct scientific measurements and observations, formulate interpretations, and present findings based on gathered evidence and newly constructed knowledge. This is an empirical methodology, and two strategies for methods may be applied. In the first, a pretest of field skills or content knowledge is applied to the study population before they interact with the VFT, and a subsequent post-test is administered. The two tests may then be compared statistically for significant differences. In the second method, the study population may be compared against a control group (mobility-impaired or otherwise) who did not use the VFT. Alternately, the study population could be compared with a non-mobility-impaired group who also used the VFT. If such empirical measurements were to be conducted simultaneously with the qualitative inquiry, this would be a true mixed-methods study, with the same question being addressed by different methods.

ENSURING RELIABILITY AND TRUSTWORTHINESS

Raw qualitative data will include field notes written by an observer, transcripts of recorded interviews, and journal entries provided by participants. In phenomenological methodology, meaning is extracted by applying coding processes to these data. The coding process breaks data into manageable segments (Schwandt, 2001) and groups them by category. The categories may emerge naturally as the researcher looks for concept indicators, or the researcher may specify categories beforehand. An example of the latter would be sorting data into the two categories of "Dependent on student mobility" and "Independent of mobility." One common coding process is the constant comparison method, developed by Glaser and Strauss (1965). In this process, data from one interview/observation are grouped as themes, and then compared to the themes from other interviews/observations. It is the across-group comparison that allows the researcher to parse meaning and relationships. The reader is referred to Creswell (1998) for further examples of constant comparison techniques applied across different studies. Without some process for ensuring reliability and trustworthiness, any emergent themes remain anecdotal. Three processes will be applied in this study to establish reliability and trustworthiness: triangulation, participant review, and providing excerpts of raw data.

In the triangulation process, the researcher subjects the themes and inferences he or she identifies to the analysis of other experts (Lincoln and Guba, 1985; Denzin, 1989; Schwandt, 2001; Seale, 1999), be they geologists, social scientists, or cognitive scientists. Coded data from this study and subsequent inferences drawn from them will be examined by multiple experts from multiple points of view in order to establish the validity of the inferences. In the participant review process, the original researcher presents a transcript (or summary of observations), together with the emergent themes, to the participants for their review. Usually the context of this conversation is along the lines of asking the participants if these themes correspond to what they meant to say, or their actual experience. The subsequent discussion either clarifies the themes, allows new ones to emerge, or, frequently, both. In the excerpting process, the researcher presents to his or her audience (as "results") emergent themes and interview quotes, or summarized observations, to support the identified themes.

EXPECTED RESULTS AND IMPLICATIONS

We argue that the culture of geologic fieldwork must be redefined in order to be more inclusive for learners with mobility impairments. The disadvantages of current field-based educational practices include the individualization (Lawrence, 1998), or labeling, of students who do not fit the persona of a young, able-bodied (and in good cardiovascular shape) person as being "disabled' for most field-based educational experiences. This must be transformed from the discriminatory notion that there is a problem with the student, to there being an issue with the field environment. To do so, it is necessary to include the first-hand perspective (lived experience) of students with mobility impairments in the development of supplementary, alternative field-based learning environments that will be used to accommodate them. This investigation is designed as an ethnographical needs assessment for the development of an effective, supplement to the cave and karst system course curriculum and the field-based component at Mammoth Cave National Park. Constructed to acquire the perspective of the participants, it is anticipated that this investigation will assist in obtaining a deeper understanding of the accessibility needs of students with mobility impairments. Additionally, through the direct experience of the students, this study will begin addressing the aspects of potential geographical, cognitive, and psychological barriers that students with mobility impairments may encounter. It is also expected that this work will further inform the broader research community on understanding the importance of the qualitative aspects of field-intensive coursework design for nontraditional students.

Virtual field environments have the potential to radically modify the way geoscience education is presented to all students, regardless of their physical ability. Accessibility to field-based learning environments should not prevent students from pursuing careers in the geosciences (Cooke et al., 1997). VFTs have the potential to enhance the geoscience curriculum by focusing on technology-based interpretation of new and archived geoscience data sets. Cooke et al. (1997) suggests that most modern geoscience careers utilize laboratory-based inquiry as a primary means of geologic interpretation, and do not require all members of an interpretation team to collect observational data from an external field site, but instead have an understanding of how field data are collected. With this in mind, given the potential for developing an interpretation-based geology curriculum based on preliminary geoscience data, it could therefore be argued that it is possible to become an *expert* geoscientist without direct, traditional fieldwork experience. A future vision of field-based geoscience curricula could suggest that this advanced, technology-based interpretation track be geared toward career-minded students with mobility impairments.

SUMMARY

This paper presents a research project in its early stages, as a model for conceptualizing qualitative inquiry from the ground up. The issue of accommodating mobility-impaired students is a significant area of inquiry in geoscience education, and lends itself well to qualitative inquiry. In order to begin developing an alternative learning environment that will accommodate students with mobility impairments, three specific research objectives have been proposed: (1) document how mobility-impaired students construct knowledge in the face of field barriers; (2) describe student experience and interaction within the physical environment; and (3) investigate the overall effectiveness of a VFT to virtually mimic one's interaction with the natural environment. We propose that these objectives are best addressed in the framework of critical theory by workers who locate themselves as action-researchers. The preferred methodology is phenomenology, with an excursion into empirical (quantitative) inquiry. The selected methods will be individual and focus group interviews, attitudinal and motivational surveys, student journaling, and observation; empirical methods will include statistical analysis of student outcomes. Reliability and trustworthiness will be established by triangulation, participant review, and excerpting of data together with coded themes. The expected outcome of this research is improved access to the geosciences for mobility-impaired students via adaptive technology, as a challenge to current educational praxis.

ACKNOWLEDGMENTS

This research is currently being conducted at the Ohio State University in collaboration with the Schools of Teaching and Learning and Earth Sciences, Ohio Supercomputer Center, Advanced Computing Center for Arts and Design, Ohio's STEM Ability Alliance, National Multiple Sclerosis Society Buckeye Chapter, Mammoth Cave National Park, and the National Cave and Karst Research Institute. This work is funded through the National Science Foundation GEO Directorate, Opportunities for Enhancing the Diversity in the Geosciences, Planning Grant 0939645. Chris Atchison wishes to thank Jill Karsten, Karen Irving, Garry McKenzie, and Scot Danforth for their support and thoughtful suggestions. Anthony Feig wishes to acknowledge the assistance of Alison Stokes and Steven Semken.

REFERENCES CITED

American Geological Institute (AGI), 2006, Geoscience Currents: Geoscience Workforce Program 21: http://www.agiweb.org/workforce/ (accessed 19 April 2009).

American Geological Institute (AGI), 2009, Status of the Geoscience Workforce 2009: Alexandria, Virginia, American Geological Institute: http://www.agiweb.org/workforce/reports/2009-StatusReportSummary.pdf (accessed 14 November 2009).

Americans with Disabilities Act of 1990, Public Law 101-336, § 2, 104 Stat. 327 (July 26, 1990).

Bakas, C., and Mikropoulos, T.A., 2003, Design of virtual environments for the comprehension of planetary phenomena based on students' ideas: International Journal of Science Education, v. 25, p. 949–967, doi:10.1080/09500690305027.

Barab, S.A., Hay, K.E., Barnett, M., and Keating, T., 2000, Virtual solar system project: Building understanding through model building: Journal of Research in Science Teaching, v. 37, p. 719–756, doi:10.1002/1098-2736(200009)37:7<719::AID-TEA6>3.0.CO;2-V.

Borgman, C., Abelson, H., Dirks, L., Johnson, R., Koedinger, K., Linn, M., Lynch, C., Obling, D., Pea, R., Salen, K., Smith, M., and Szalay, A., 2008, Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge: A 21st Century Agenda for the NSF: Report of the National Science Foundation Task Force on Cyberlearning: Washington, D.C., National Science Foundation, 61 p.

Clark, S.K., and Libarkin, J.C., 2011, this volume, Designing a mixed-methods research instrument and scoring rubric to investigate individuals' conceptions of plate tectonics, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(07).

Clary, R.M., and Wandersee, J.H., 2011, this volume, Geobiological opportunities to learn at U.S. fossil parks, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(09).

Cooke, M.L., Anderson, K.S., and Forrest, S.E., 1997, Creating accessible introductory geology field trips: Journal of Geoscience Education, v. 45, p. 4–9.

Creswell, J., 1998, Qualitative Inquiry and Research Design: Thousand Oaks, California, Sage Publications, 403 p.

Denzin, N., 1989, The Research Act (3rd ed.): Englewood Cliffs, New Jersey, Prentice Hall, 379 p.

Elkins, J., and Elkins, N., 2007, Teaching geology in the field: Significant geoscience concept gains in entirely field-based introductory geology courses: Journal of Geoscience Education, v. 55, p. 126–132.

Feig, A.D., 2008, Phenomenology in geoscience education: Geological Society of America Abstracts with Programs, v. 40, no. 6, p. 421.

Feig, A.D., 2011, this volume, Methodology and location in the context of qualitative data and theoretical frameworks in geoscience education research, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(01).

Freire, P., 2000, Pedagogy of the Oppressed, 30th Anniversary Edition (M.B. Ramos, trans.): New York, Continuum, 183 p.

Garrison, J.R., Jr., and Endsley, G., 2005, Field-based geoscience education: A valid experience in outdoor education: The Texas Science Teacher, v. 23, p. 29.

Glaser, B., and Strauss, A., 1965, Awareness of Dying: Chicago, Aldine Press, 309 p.

Gould, S.J., 1993, American polygeny and craniometry before Darwin, *in* Harding, S., ed., The "Racial" Economy of Science: Bloomington, Indiana University Press, p. 77–91.

Hall, T., and Healey, M., 2005, Disabled students' experiences of fieldwork: Area, v. 37, p. 446–449, doi:10.1111/j.1475-4762.2005.00649.x.

Hall, T., Healey, M., and Harrison, M., 2004, Fieldwork and disabled students: Discourses of exclusion and inclusion: Journal of Geography in Higher Education, v. 28, p. 255–280, doi:10.1080/0309826042000242495.

Haymes, S., 1995, Race, Culture, and the City: Albany, New York, State University of New York Press, 167 p.

Healey, M., Roberts, C., Jenkins, A., and Leach, J., 2002, Disabled students and fieldwork: Towards inclusivity: Planet, v. 3, p. 9–10.

Holt, T., 1996, Field trips and technology: Science Scope, v. 20, p. 54–55.

Hunter, W., 2007, Chapter 9: Action research as a framework for science education research, *in* Bodner, G., and Orgill, M., eds., Theoretical Frameworks for Research in Chemistry/Science Education: Upper Saddle River, New Jersey, Pearson/Prentice Hall, p. 152–171.

Jackson, R., and Winn, W., 1999, Collaboration and learning in immersive virtual environments, *in* Hondley, C.M., and Roshelle, J., eds., Proceedings of Computer Support for Collaborative Learning '99: Palo Alto, California, p. 260–264.

Lascara, C., Wheless, G., Cox, D., Patterson, R., Levy, S., Johnson, A., Leigh, J., and Kappor, A., 1999, Tele-immersive virtual environments for collaborative knowledge discovery: University of Illinois–Urbana: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.76.7521andrep=rep1andtype=pdf (accessed 7 December 2009).

Lawrence, D., 1998, Different territories within our communities: Disability studies in the classroom: New Zealand Journal of Geography, v. 105, p. 1–7.

Lincoln, Y.S., and Guba, E.G., 1985, Naturalistic Inquiry: London, Sage Publications, 415 p.

Locke, S., 2005, The Status of Persons with Disabilities in the Geosciences: White Paper, RASEM^2 Symposium, 13–14 October: Las Cruces, New Mexico, New Mexico State University: http://rasem.nmsu.edu/Pdfs/symposium/RASEM_white_paper9_28.pdf (accessed 31 October 2009).

Maskall, J., and Stokes, A., 2008, Designing Effective Fieldwork for the Environmental and Natural Sciences: Plymouth, UK, The Higher Education Academy Subject Centre for Geography, Earth and Environmental Sciences, 78 p.

Mason, J., 2002, Qualitative Researching: London, Sage Publications, 223 p.

Mayo, P., 2007, Chapter 14: Critical theory, *in* Bodner, G., and Orgill, M., eds., Theoretical Frameworks for Research in Chemistry/Science Education: Upper Saddle River, New Jersey, Pearson/Prentice Hall, p. 243–261.

McCarthy, C.B., 2005, Effects of thematic-based, hands-on science teaching versus a textbook approach for students with disabilities: Journal of Research in Science Teaching, v. 42, p. 245–263, doi:10.1002/tea.20057.

McKenzie, G.D., Utgard, R.O., and Lisowski, M., 1986, The importance of field trips: A geological example: Journal of College Science Teaching, v. 16, p. 17–20.

McKinney, W.J., 1997, The educational use of computer based science simulations: Some lessons from the philosophy of science: Science and Education, v. 6, p. 591–603, doi:10.1023/A:1008694507127.

Nairn, K., 1999, Embodied fieldwork: The Journal of Geography, v. 98, p. 272–282, doi:10.1080/00221349908978941.

National Center for Education Statistics, 1999, Students with Disabilities in Postsecondary Education: A Profile of Preparation, Participation, and Outcomes, NCES 1999-187: Washington, D.C., U.S. Department of Education: http://nces.ed.gov/Pubsearch/pubsinfo.asp?pubid=1999187 (accessed 5 December 2008).

National Center for Education Statistics, 2008, Digest of Education Statistics, 2007, NCES 2008-022: Washington, D.C., U.S. Department of Education: http://nces.ed.gov/programs/digest/d07/tables/dt07_179.asp (accessed 5 December 2008).

National Science Foundation, 2004, Women, Minorities, and Persons with Disabilities in Science and Engineering (NSF 04-317): Arlington, Virginia, National Science Foundation: http://www.nsf.gov/statistics/wmpd/pdf/nsf04317.pdf (accessed 28 November 2008).

Nicoli, J., Holmlund, J., Sofaer, A., Price, A., Kraybill, L., and Stone, W., 2008, Digital restoration of the Sun Dagger site, Chaco Canyon, *in* Sofaer, A., ed., Chaco Astronomy: An Ancient American Cosmology: Santa Fe, New Mexico, Ocean Tree, 174 p.

Norman, A., 2002, Mobility impaired students could face access and location problems on entering higher education: Planet, v. 3, p. 19–21.

Orion, N., 1993, A model for the development and implementation of field trips as an integral part of the science curriculum: School Science and Mathematics, v. 93, p. 325–331, doi:10.1111/j.1949-8594.1993.tb12254.x.

Orion, N., and Hofstein, A., 1994, Factors that influence learning during a scientific field trip in a natural environment: Journal of Research in Science Teaching, v. 31, p. 1097–1119, doi:10.1002/tea.3660311005.

Potter, N., Jr., Niemitz, J., and Sak, P., 2009, Long-term field-based studies in geoscience teaching, *in* Whitmeyer, S.J., Mogk, D.W., and Pyle, E.J., eds., Field Geology Education: Historical Perspectives and Modern Approaches: Geological Society of America Special Paper 461, p. 185–194.

Qiu, W., and Hubble, T., 2002, The advantages and disadvantages of virtual field trips in geoscience education: The China Papers, v. 1, p. 75–79.

Rehabilitation Act of 1973, Public Law 93-112, § 504 (1973).

Riggs, E., 2005, Field-based education and indigenous knowledge: Essential components of geoscience education for Native American communities: Science Education, v. 89, p. 296–313, doi:10.1002/sce.20032.

Riggs, E., Lieder, C., and Balliet, R., 2009, Geologic problem solving in the field: Analysis of field navigation and mapping by advanced undergraduates: Journal of Geoscience Education, v. 57, p. 48–63.

Saat, R.M., 2004, The acquisition of integrated science process skills in a web-based learning environment: Research in Science & Technological Education, v. 22, p. 23–40, doi:10.1080/0263514042000187520.

Schuchardt, P., and Bowman, D.A., 2007, The benefits of immersion for spatial understanding of complex underground cave systems, *in* Proceedings of Association for Computing Machinery Symposium on Virtual Reality Software and Technology: Newport Beach, California, p. 121–124, doi:10.1145/1315184.1315205.

Schwandt, T., 2001, Dictionary of Qualitative Inquiry (2nd ed.): Thousand Oaks, California, Sage Publications, 281 p.

Schwert, D.P., Slator, B.M., and Saini-Eidukat, B., 1999, A virtual world for earth science education in secondary and post-secondary environments: The geology explorer, *in* Proceedings of the International Conference of Mathematics/Science Education and Technology (M/SED '99): San Antonio, Texas, Association for the Advancement of Computing in Education, p. 519–525.

Seale, C., 1999, Quality in qualitative research: Qualitative Inquiry, v. 5, p. 465–478, doi:10.1177/107780049900500402.

Semken, S., and Brandt, E., 2010, Implications of sense of place and place-based education for ecological integrity and cultural sustainability in contested places, *in* Tippins, D., Mueller, M., van Eijck, M., and Adams, J., eds., Cultural Studies and Environmentalism: The Confluence of Ecojustice, Place-Based (Science) Education, and Indigenous Knowledge Systems: New York, Springer, p. 287–302, doi:10.1007/978-90-481-3929-3_24.

Shepardson, D.P., Wee, B., Priddy, M., and Harbor, J., 2007, Students' mental models of the environment: Journal of Research in Science Teaching, v. 44, p. 327–348, doi:10.1002/tea.20161.

Skinner, B.F., 1953, Science and Human Behavior: New York, The Free Press, 461 p.

Smith, G., and Williams, D., 1999, Introduction: Re-engaging culture and ecology, *in* Smith, G., and Williams, D., eds., Ecological Education in Action: On Weaving Education, Culture and the Environment: New York, SUNY Press, 244 p.

Spicer, J.I., and Stratford, J., 2001, Student perceptions of a virtual field trip to replace a real field trip: Journal of Computer Assisted Learning, v. 17, p. 345–354, doi:10.1046/j.0266-4909.2001.00191.x.

Stokes, A., and Boyle, A., 2009, The undergraduate geoscience fieldwork experience: Influencing factors and implications for learning, *in* Whitmeyer, S.J., Mogk, D.W., and Pyle, E.J., eds., Field Geology Education: Historical Perspectives and Modern Approaches: Geological Society of America Special Paper 461, p. 291–312.

Stredney, D., Hittle, B., and Sessanna, D., 2008, Evaluating elicited anxiety in a simulated environment: Studies in Health Technology and Informatics, v. 132, p. 496–501.

Tan, W.H.L., and Subramaniam, R., 2003, Virtual science centers: Web-based environments for promotion of non-formal science education, *in* Aggarwal, A.K., ed., Web-Based Education: Learning from Experience: Hershey, Pennsylvania, Idea Group, p. 308–329.

Thomas, R., and Roberts, S., 2009, Experience One: Teaching the geoscience curriculum in the field using experiential immersion learning, *in* Whitmeyer, S.J., Mogk, D.W., and Pyler, E.J., eds., Field Geology Education: Historical Perspectives and Modern Approaches: Geological Society of America Special Paper 461, p. 65–76.

Thrift, D., 1975, Field trips: A priceless ingredient: Journal of Geological Education, v. 23, p. 137–139.

Whitmeyer, S.J., and Mogk, D.W., 2009, Geoscience field education: A recent resurgence: Eos (Transactions, American Geophysical Union), v. 90, p. 385–386, doi:10.1029/2009EO430001.

Williams, D., and Semken, S., 2011, this volume, Ethnographic methods in analysis of place-based geoscience curriculum and pedagogy, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(05).

Wolcott, H., 1999, Ethnography: A Way of Seeing: Thousand Oaks, California, Sage Publications, 335 p.

Manuscript Accepted by the Society 23 June 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

A phenomenographic approach to investigating students' conceptions of geoscience as an academic discipline

Alison Stokes
Experiential Learning CETL (Centre for Excellence in Teaching and Learning), University of Plymouth, Plymouth PL4 8AA, UK

ABSTRACT

Phenomenography is an empirical approach to identifying the qualitatively discrete ways in which individuals experience and understand aspects of the world around them. Although established for several decades, the technique is seldom applied (if at all) in geoscience education research, yet it has the potential to significantly enhance undergraduate instruction. This paper presents an overview of phenomenographic inquiry in terms of its characteristic methods and applications to education research. The value of this approach to geoscience education is then demonstrated in a study investigating conceptions of geoscience as an academic discipline. Students enrolled in undergraduate geoscience programs at a single U.K. university, together with geoscience faculty, provided brief, written responses to the question "what do you believe your chosen discipline to be about or concerned with?" Phenomenographic analysis revealed six qualitatively distinct conceptions, ranging from simple to complex, to be present within both the student and faculty populations. Although process-based conceptions dominated both the student and faculty data, simpler conceptions were more pervasive among students, and complex conceptions were more pervasive among faculty. This has implications for curriculum design and instruction since the conceptions held by faculty will influence their assumptions about students' perceptions of geoscience, and the learning strategies and techniques likely to be effective.

INTRODUCTION

Government initiatives aimed at increasing participation in higher education (e.g., DfES, 2003) have resulted in the U.K. student population becoming increasingly diverse, with recruitment to university programs being further driven by escalating unemployment linked to the current global economic crisis (Lipsett, 2009). This widening of participation means that students are entering university from a richer, more diverse range of social, ethnic, and academic backgrounds, bringing with them a variety of understandings and beliefs relating to teaching and learning that have been shaped by their prior experiences (Watters and Watters, 2007). The diversity in student experience that naturally results from these changing demographics means that the issue of student success is becoming more prevalent, and must be addressed if all students are to have a fair chance of succeeding academically (e.g., Power et al., 1987; McKenzie and Schweitzer, 2001).

A key to achieving this success is the delivery of undergraduate curricula informed by sound pedagogic research and practice (McKenzie and Schweitzer, 2001). Research to date has significantly increased our understanding of students' learning experiences and how these relate to learning outcomes, both in science disciplines and more widely. A substantial body of research now exists that investigates not only *what* students know, but the different *ways* in which they know or understand. It is the variation

Stokes, A., 2011, A phenomenographic approach to investigating students' conceptions of geoscience as an academic discipline, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 23–35, doi:10.1130/2011.2474(03). For permission to copy, contact editing@geosociety.org.

in experience and understanding inherent in learners, commonly referred to as "conceptions," that forms the basis for this paper. Identification of the variation in conceptions that exists within a group of learners (novices) can play a vital role in curriculum development by helping instructors (experts) to develop their teaching practices, and to design learning activities aimed at helping students to construct knowledge and gain understanding of specific concepts (Ebenezer and Fraser, 2001).

The purpose of this paper is twofold. Firstly, it provides an overview of "phenomenography," a qualitative research method developed specifically to identify and explore conceptions. It then presents an example of how phenomenography can be applied to geoscience education research by describing an investigation into conceptions of geoscience as an academic discipline at a single U.K. university. The following research questions are addressed:

1. What are the conceptions of geoscience as an academic discipline that are held by undergraduate geoscience students and academic faculty?
2. Are the dominant conceptions held by undergraduates similar to, or different from, those held by academic faculty?

Overview of Conceptions

In education research the term "conceptions" relates to the various ways in which individuals, i.e., learners, experience or understand a particular phenomenon, although precise definitions may vary according to context. Discussions by Wandersee et al. (1994) and Scott et al. (2007) focus specifically on learners' understandings and experiences of science and scientific concepts. Conceptions held by learners are termed "alternative," and defined as "experience-based explanations constructed by a learner to make a range of natural phenomena and objects intelligible" (Wandersee et al., 1994, p. 178). These are considered distinct from "scientific conceptions," which represent the explanations applied by expert scientists to natural phenomena. By comparison, definitions of conceptions applied by phenomenographers tend to be less specific, e.g., "ways of experiencing a specific aspect of reality" (Sandbergh, 1997, p. 203). Despite these variations in definition, there is general consensus that the formation of conceptions results from interactions between the learner and the world around them.

The rooting of conceptions in experience means that novices, i.e., individuals who are new or have had minimal exposure to a domain (Hoffman, 1998), frequently hold a range of conceptions that vary from those held by experts, i.e., individuals who have spent significant time training or solving problems in a particular domain (Petcovic and Libarkin, 2007), or presented in textbooks (Marton, 1981). A fundamental goal of instruction is to recognize this variation, and to help students acquire conceptions that are consistent with those held by experts, or with an accepted scientific model (e.g., Wandersee et al., 1994). Historically, however, we find that "accepted" conceptions of natural phenomena can become replaced by alternative ideas and understandings as paradigms inevitably shift. Consider, for example, the "replacement" of geosynclinal theory with plate tectonic theory. This means that conceptions of a phenomenon can vary both at a specific instant in time (i.e., between individuals), as well as through time (Marton, 1981). However, it is through changes in the way in which *individuals* experience and conceptualize phenomena that new scientific discoveries, or paradigm shifts, take place (Kuhn, 1995). Darwin may never have proposed the theory of evolution had specific experiences not caused him to change his original conception of a static natural order, to one of natural selection (Gruber, 1974; Marton, 1981). So, while we might consider the conceptions that we help learners to acquire today to be "scientific," their "alternative conceptions" could in fact represent the paradigm-breaking ideas of tomorrow (Marton, 1981; Kuhn, 1995; Cousin, 2009).

The conceptions of geoscience held by individuals are important, since these will influence the extent to which the discipline is considered relevant to everyday life. Bezzi (1999) reports findings from an investigation into perceptions of geoscience held by a small number of Italian faculty and undergraduates in which students were found to hold somewhat "stereotypical" perceptions of geoscientific knowledge as subjective and approximate, with apparently little concern for the societal aspects of geoscience. Other than this particular example, however, there appear to be no further published investigations focusing specifically on conceptions of geoscience as a discipline. We do find in the literature several examples of phenomenographic investigations into conceptions of other academic disciplines, which provide some interesting insight into the ways in which learners conceptualize their subject. Of particular significance to this study is the investigation by Bradbeer et al. (2004) into undergraduate geographers' conceptions of geography, in which written data from 153 students sampled from universities in the U.K., United States, Australia, and New Zealand were subjected to phenomenographic analysis. These data revealed five conceptions of geography that varied from simple to complex, and that focused exclusively on declarative knowledge (i.e., the "what" of geography). This contrasts with the five conceptions of mathematics identified by Crawford et al. (1994), which exhibited both a strong methodological focus (i.e., "how"), as well as a knowledge focus ("what"). From this, Bradbeer et al. (2004) surmised that geography students tend not to consider their discipline as having a distinctive methodological basis—which is interesting considering geography's strong historical focus on fieldwork.

Despite the apparent paucity of research into conceptions of geoscience as an academic discipline, there have been numerous investigations into conceptions of specific geoscientific concepts (see King [2008] for a comprehensive review of research to date). Many of these studies apply a qualitative, or mixed qualitative-quantitative approach, in which written, verbal, or graphic data are analyzed for dominant themes. Conceptions investigated include the structure of Earth (Nussbaum, 1985; Schoon, 1992; Libarkin et al., 2005), plate tectonics (Gobert, 2000; Sibley, 2005), rocks and rock formation (Kusnick, 2002; Ford, 2003;

Kortz and Murray, 2009), geological time (Trend, 1998, 2000), and landscape features (Harwood and Jackson, 1993; Trend et al., 2000; Mackintosh, 2005). As well as influencing curriculum development in geoscience education, these findings have a wider geocognitive application in terms of investigating how society understands and experiences aspects of geoscience. Future investigations of this type would, I argue, lend themselves very nicely to phenomenographic analysis.

PHENOMENOGRAPHY

The purpose of this section is to provide a brief overview of the aims, assumptions, and methods characteristic of phenomenography. Essentially, this is an empirical approach to identifying the various ways in which a particular phenomenon is experienced or understood within a group or collective (Marton, 1981, 1994; Marton and Booth, 1997). Phenomenographic analysis therefore involves describing the qualitatively distinct conceptions that exist about a phenomenon, and exploring the relationship between them (Cousin, 2009). The "categories of description" that emerge from phenomenographic data, and that describe the conceptions held by individuals, are collectively termed "outcome space" and frequently (but not always) form a hierarchy, ranging from simple to more complex conceptions (Marton and Booth, 1997; Trigwell, 2006). A phenomenographic study therefore enables data from a collection of participants to be reduced to a limited number of discrete categories, which together represent the "whole" of the way in which a phenomenon is conceptualized (Minasian-Batmanian et al., 2006).

The foundations of phenomenography are based in the constructivist principle that meaning is constructed from social and personal experience, and its value to education is in exposing the different ways in which learners understand a particular aspect of their subject (Cousin, 2009). The following assumptions are typical of most types of phenomenographic inquiry (e.g., Marton and Booth, 1997; Svensson, 1997; Ebenezer and Fraser, 2001; Booth, 2008):

1. Individuals can experience or understand the same phenomenon or aspect of reality in different ways, and thus hold different conceptions of it.
2. An individual's conceptions can be accessed, e.g., verbally or in writing.
3. A limited number of conceptions exists about a phenomenon.
4. These conceptions are logically related, typically by way of a hierarchy ranging from simple to complex.

Early studies by Marton and Säljö (1976) and Säljö (1979) showed that learners held a hierarchy of conceptions of "learning" that ranged from simple, such as increasing one's knowledge, to complex, such as changing as a person (for example, by adapting one's world-view in light of new experiences). These studies demonstrated the effectiveness of the phenomenographic method, paving the way for myriad subsequent investigations into conceptions of teaching and learning. These include conceptions of specific academic disciplines (e.g., Crawford et al., 1994; Reid and Petocz, 2002; Bradbeer et al., 2004), learning in specific contexts (e.g., Prosser et al., 1996; Tynjälä, 1997; Minasian-Batmanian et al., 2006; Walsh et al., 2007), and students' understandings of disciplinary concepts (e.g., Beaty, 1987; Svensson and Hogfors, 1988; Linder and Erickson, 1989; Bowden et al., 1992; Ebenezer and Fraser, 2001; Loughland et al., 2002; Åkerlind and Kayrooz, 2003; Tóth and Ludányi, 2007).

Method in Phenomenography

Early phenomenographic studies used in-depth interviews as their data source (e.g., Marton and Säljö, 1976; Säljö, 1979; Marton, 1986), and, in many respects, this continues to be the favored or "ideal" method of data collection (Marton, 1994; Ashwin, 2005). In reality, however, data can take a variety of forms, including observation, group interviews, drawings, and written text such as open-survey responses, essays, or even historical documents. Written data have been used with particular success in a number of phenomenographic investigations, including conceptions of academic disciplines (e.g., Crawford et al., 1994; Bradbeer et al., 2004), learning in specific contexts (e.g., Prosser et al., 1996; Tynjälä, 1997), and natural or scientific phenomena (e.g., Loughland et al., 2002; Tóth and Ludányi, 2007). A key advantage of this technique is that a large amount of data can be captured in a relatively short amount of time; this makes it particularly useful for investigating variation within larger samples (Minasian-Batmanian et al., 2006). It should be noted, however, that nonverbal data are typically less detailed than verbal data, and the conceptions expressed may not always be complete.

Once collected, data are then analyzed to identify the range of conceptions present within the sample population. Walsh (2000) describes two opposing approaches to the analytical process: discovery and construction. The discovery approach requires the researcher to put aside or "bracket" their own preconceptions about the phenomenon under investigation (Ashworth and Lucas, 2000; Åkerlind, 2005), thus allowing categories of description to emerge from the data naturalistically. By comparison, researchers applying the construction approach focus on establishing structure within the data, for example, by using predetermined categories (e.g., McLean, 2001) or by applying their own preexisting knowledge, thus exerting a degree of control over the categories which emerge (Walsh, 2000). Although the discovery approach is widely favored (Marton, 1994), the actual approach that a researcher adopts is likely to sit somewhere between these two end members, since it may be impossible for some researchers to bracket their preexisting knowledge (Ashworth and Lucas, 2000). For example, a geoscientist used to taking an objective stance toward the phenomenon being investigated may struggle to acknowledge the existence of multiple realities, and thus to detach themselves from their own knowledge or understanding.

The results of a phenomenographic analysis comprise both the conceptions themselves, and the relationship between them (i.e., the outcome space). It is frequently assumed that conceptions of a phenomenon are logically related in some way, e.g., in a hierarchy ranging from simple to complex (Marton, 1981; Marton and Booth, 1997; Trigwell, 2006), and many phenomenographic studies do indeed report neatly structured outcome spaces (e.g., Marton et al., 1993; Crawford et al., 1994; Prosser et al., 1994; Loughland et al., 2002; Reid and Petocz, 2002; Minasian-Batmanian et al., 2006; Walsh et al., 2007). However, this assumption has attracted some criticism, specifically, that it can limit the emergence of more nuanced or complex understandings, be influenced by the researcher's own view (Webb, 1997; Ashworth and Lucas, 2000; Walsh, 2000; Cousin, 2009), or artificially impose orders on fluid, organic processes (Feig, 2009, personal commun.). Other forms of representing outcome space, such as the nonlinear or branching formats reported by Tynjälä (1997) and Bradbeer et al. (2004), might actually be a better reflection of the characteristics of experience described by the participants, and more representative of the "messy" nature of real data (Åkerlind, 2005; Morris, 2006; Cousin, 2009).

Relationship to Other Methods

Although distinct research approaches, the similarity between the terms *phenomenography* and *phenomenology* can make them easily confused. Both of these approaches consider human experience as the research object, but they vary in the perspective from which that experience is considered (Marton, 1981). Phenomenology represents a "first-order approach," in which the researcher describes or defines a particular phenomenon as he or she perceives it (Trigwell, 2006). For example, I might explore the research question "what is geoscience about?" from a first-order, phenomenological perspective, based upon my own direct experiences of geoscience. This would provide a rich and highly individual account, but it would reflect only a limited range of the possible conceptions that might exist. By comparison, phenomenography describes how a phenomenon is experienced by others (Marton, 1981; Marton and Booth, 1997) and is therefore a "second-order approach." In this case, a more appropriate research question would be "what do people think that geoscience is about?" This would be explored using the descriptions and explanations provided by others to identify critical variations in the collective experience. The focus of the first-order (phenomenological) and second-order (phenomenographic) approach therefore varies between identifying the "essence" of a phenomenon, and exploring the different ways in which it can be experienced, respectively.

Figure 1 summarizes some of the ways in which phenomenography differs from other research approaches. Philosophically, phenomenography takes a *nondualist* stance by considering the inter-relationship between an individual and a particular phenomenon (Hasselgren and Beach, 1997). This contrasts with the dualist philosophy in which mind and matter are treated as discrete entities (e.g., Derrida, 1978). Phenomenographic data are

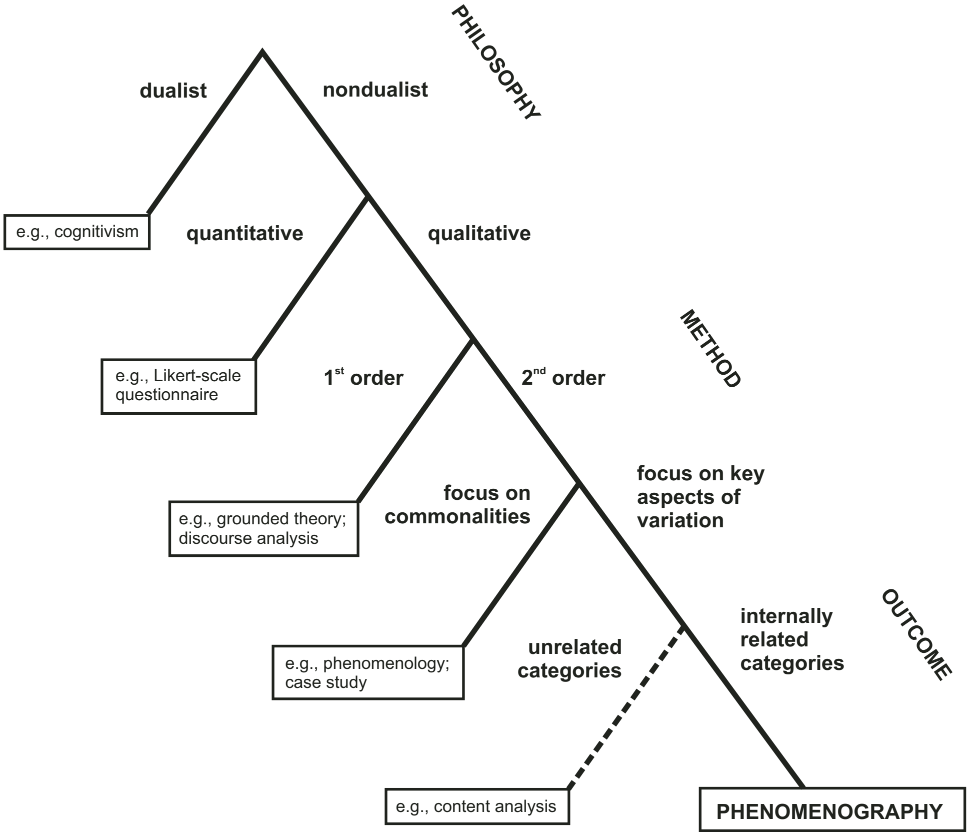


Figure 1. Defining phenomenography with respect to other research approaches (adapted from Trigwell, 2006).

generated and analyzed using *qualitative* methods, rather than the quantitative, numerical methods familiar to scientists. Data are investigated from a *second-order* perspective that focuses on *critical variation* within the collective. Unlike some other qualitative methods in which data are coded to generate a series of unrelated categories, the categories emerging from phenomenographic analysis are assumed to be *internally related* (Trigwell, 2006), at least to some degree.

Rigor in Phenomenographic Inquiry

One of the most frequently encountered criticisms of qualitative research is that it lacks the "rigor" of scientific research (Mays and Pope, 1995; Cope, 2004, Albert et al., 2008), which is typically assessed in terms of validity and reliability. Some of the key concerns around validity and reliability in phenomenography arise from the variation in phenomenographic method, particularly in terms of the construction versus discovery dichotomy (Morris, 2006). Variation is not only apparent in methods, however. A single data set can be open to multiple legitimate interpretations (Åkerlind, 2005); this means that, rather than finding the "right" answer, the emphasis is on the researcher being able to defend their chosen approach, and their interpretation of the data (Marton and Booth, 1997; Åkerlind, 2005).

In phenomenography, the researcher's findings will inevitably be influenced by their location in relation to the phenomenon being investigated, i.e., their background, role in the research, and potential biases (Feig, this volume). By taking a reflexive stance and setting out their own background and beliefs, a researcher can make explicit any variables that might bias or influence the research findings (Cousin, 2009). Indeed, there is widespread consensus that the key to establishing rigor in phenomenographic inquiry lies in providing as open and honest an account as possible of methods and procedures followed (e.g., Sandbergh, 1997; Cope, 2004; Morris, 2006). This includes presenting a "critical and honest" display of extracts from responses in order to demonstrate the trustworthiness of the data (Cousin, 2009).

APPLICATION OF PHENOMENOGRAPHY TO GEOSCIENCE EDUCATION

In this section of the paper, I describe how a phenomenographic approach has been applied to investigating the conceptions of geoscience as a discipline held by geoscience students and faculty.

Study Population and Setting

The population for this study consisted of undergraduate geoscience students and geoscience faculty at a single U.K. university. Although originally a teaching institution, this university has an expanding international research profile, and an enrolment of ~30,000 students. Unlike North America, where students follow a broad curriculum before selecting their major, students at U.K. universities are recruited directly onto subject-specific degree programs. These programs typically run for three or four years, during which time the students take classes exclusively, or almost exclusively, in their chosen discipline. A further difference between the U.K. and U.S. degree system is that U.K. students move through their degrees in distinct year groups, and hence they will take classes with the same group of peers for the duration of their studies. This contrasts significantly with the U.S. system, where classes can include students from a range of academic levels. Students participating in this study were at the beginning of the first, second, or third year of the following geoscience programs (referred to in the U.K. as "degree pathways"), and by U.S. standards would all be considered geoscience "majors":

- B.Sc. Geology,
- B.Sc. Applied Geology,
- B.Sc. Physical Geography and Geology, and
- M.Geol. (Master of Geology) Geology (4 yr program).

Students in all of these programs follow a broadly common curriculum for the first two years, but take a more diverse range of courses in the third year (and fourth year, where appropriate), which are more specific to their chosen pathway.

The number of students registered in each year group varied from 58 to 79 students, with 17 faculty (Table 1). The total student population comprised ~70% males and 30% females. Although no data relating to ethnicity were collected, the study institution is located in an area of the U.K. of low ethnic diversity, and the study population was almost exclusively Caucasian. The majority of students in all three years were aged under 21, meaning that they had come to university either straight from school or following a "gap-year." Academic backgrounds varied from those coming direct from secondary (high) school with qualifications in geology or cognate disciplines (e.g., geography), to school-leavers with no prior experience of or qualifications in geoscience, to older students returning to education, typically with no or limited prior experience. Academic faculty ranged in teaching experience from less than two years to over 20 years, and

TABLE 1. RESPONSE RATES FOR CONCEPTIONS DATA 2006/2007

	Cohort size	Usable responses (*n*)	Response rate (%)
Students: year 1	79	51	65
Students: year 2	77	45	58
Students: year 3	58	39	67
All students	214	135	63
Academic faculty	17	13	72

represented a wide cross section of geoscience subdisciplines, including petrology, structural geology, sedimentology, stratigraphy, paleontology, geophysics, geochemistry, and engineering geoscience. The majority of faculty teach most or all of the four degree pathways.

Researcher's Perspective

An explanation of my philosophical and disciplinary perspectives will identify the "lens" through which this study is conducted, and which may also act as a potential for bias. I graduated from my undergraduate degree with a B.Sc. in Geological Sciences, where I opted to take courses covering traditionally "pure" aspects of geoscience such as petrology, structure, paleontology, and stratigraphy. For my Ph.D., I investigated microstructural and chemical controls on the nucleation of mineral phases during regional metamorphism, during which time I gained some teaching experience in structure, petrology, and in the field. I therefore have significant experience in aspects of "hard rock" geoscience, but more limited knowledge and experience of "soft rock" and applied geoscience.

After my Ph.D., I spent five years working outside of academia, before returning to take up a research position focusing on teaching and learning in environmental and natural sciences. This required me to work within a completely different research paradigm, and to develop different ways of thinking and practicing than I had previously applied to my scientific research. Having no previous background in education research, I found this a difficult and sometimes uncomfortable process, but one that has broadened my perspective beyond the "scientific" method of inquiry to include more subjective, interpretive approaches. Most significantly, it has led me to question the notion of "objective truth," and to accept the concept of multiple ways of knowing.

Methods

Data Collection

The entire population of geoscience students and faculty at the study institution was selected for participation in this study. Data were collected using an open-ended survey in which participants were asked to provide brief (2–3 sentence) written responses to the question: *Think about your main academic discipline. What do you believe this to be about/concerned with?* Since all students were enrolled on geoscience programs, their "main academic discipline" would be geoscience. The phrasing of the question was informed by previous studies into conceptions of academic disciplines (e.g., Crawford et al., 1994; Bradbeer et al., 2004), and the suitability to this study was tested by seeking expert feedback from experienced education researchers, and by piloting the question on six undergraduate students and nine faculty.

All student data were collected in October 2006 during induction meetings held immediately prior to the commencement of teaching. I personally attended these meetings in order to introduce the research, and to distribute and collect the paper-based surveys. My "location" was therefore that of researcher-observer, actively generating data relating to students' conceptions. This method of data collection was considered appropriate since data could be collected from the three student cohorts within a short amount of time, and under similar conditions. Although attendance at the meetings was compulsory, not all students did attend, and there was no obligation to complete the survey. The majority of students attending chose to participate, thus generating reasonable response rates (Table 1). Data were collected from academic faculty in November 2006 using an online version of the survey. An electronic survey was easier for faculty to access, and thus more likely to generate a response than the paper survey. This would, however, have afforded faculty the opportunity to spend more time than the students thinking about and articulating their response, and so should be recognized as a potential limitation of the data collection method. All responses were assigned a unique reference beginning "06GL" (referring to data collected in 2006 from geoscience students and faculty), followed by the academic stage of the student and the survey number. Thus, reference 06GL1/23 relates to stage 1, survey number 23. The prefix "06GLF" was used to identify responses from faculty. Data were then copied verbatim into Excel spreadsheets, where they were sorted according to subpopulation ready for analysis.

Some previous studies based around written responses have used follow-up interviews to check that all possible categories have been identified (e.g., Crawford et al., 1994). Since this study aimed to collect data from the entire study population, it was assumed that all of the conceptions existing within the study population would be contained within the written data, and that follow-up interviews would not be necessary. This decision was also influenced by the time commitment required of the participants, and of me. I further assumed that the written responses provided a true reflection of the participants' actual ideas and beliefs about geoscience at the time at which data were collected. In reality, not all of the study population contributed to the data set, and it is entirely possible that interviews would reveal additional conceptions, or prompt participants to express conceptions that differ from those expressed in the written survey. Follow-up interviews should therefore be included in any future research.

Data Analysis

Previous phenomenographic studies into conceptions of academic discipline have involved multiple researchers during the data analysis stage (e.g., Crawford et al., 1994; Loughland et al., 2002; Bradbeer et al., 2004). A different approach was applied to this study, however, whereby I opted to undertake the data analysis independently. This decision was influenced by previous experience of analyzing phenomenographic data, during which I worked with a second researcher, from a different academic discipline, to identify and categorize data containing multiple conceptions (Stokes et al., 2010). Despite agreeing on the range of conceptions present within the data, and compiling detailed rubrics to guide the categorization of individual written responses, we only achieved an inter-rater agreement of around 50%. Sandbergh (1997) discusses the potential limitations of what he terms "interjudge reliability" in phenomenography, particularly where multiple researchers are attempting to categorize responses containing more than one conception. He states that confusion can often arise between researchers as to the conceptions being expressed; this appeared to be case with the study described in Stokes et al. (2010). Sandbergh (1997) also recognizes that inter-rater reliability pays no account to the procedures followed by the researchers in describing and categorizing conceptions.

Based on this experience, I opted to provide a reflexive and thorough account of my methods such that another researcher could follow the same procedure, rather than attempt to achieve inter-rater agreement with another researcher. I began by analyzing a sample of data (responses from the first-year students) using an iterative process of reading through all the responses several times in order to familiarize myself with the content, identifying and describing preliminary categories of description based upon both similarities and differences within the data, and then provisionally sorting responses into those categories. During this process I attempted to apply a discovery approach (Walsh, 2000) by bracketing my own ideas and beliefs about geoscience and allowing categories to emerge naturalistically from the data. In reality, this proved extremely challenging, and it is almost inevitable that the categories that have emerged are influenced, to some extent, by the specific lens through which the data have been interpreted, which in turn reflects my personal location and perspectives in relation to the study.

In order to reduce potential bias, both the method and the findings were scrutinized by a member of geoscience faculty previously unrelated to the study. This involved describing the process of analysis, justifying why certain decisions had been made, and how the categories had evolved. This critical examination established the confirmability of the conceptions (i.e., by demonstrating that they could be recognized by someone else), and helped to refine the criteria by which individual responses were classified. The remainder of the student and faculty data were then categorized. An individual can experience or conceptualize a phenomenon in a number of different ways (Marton, 1994), so it is not uncommon for a single response to express more than one conception. If a clear, logically inclusive hierarchy emerges whereby lower-order categories are subsumed into higher-order categories, then responses are classified according to the most "sophisticated" conception expressed. That is, it is assumed that lower-order conceptions form part of the more complex higher-order conception. As will be discussed in the following section, a strict hierarchy did not emerge from the data, so each response has been categorized according to *all* of the conceptions expressed therein (see, for example, Tynjälä, 1997).

Trustworthiness

In phenomenography it is the plausibility of the account, rather than the arrival at a "final truth," that establishes the trustworthiness of the study findings (Cousin, 2009). The criteria of Lincoln and Guba (1985) are used here to summarize how the establishment of trustworthiness was approached in this study:

1. Credibility: This is established if the study findings are judged to be credible or believable from the participants' perspective.
2. Transferability: This concerns the extent to which the study findings can be transferred or generalized between contexts, and is addressed by providing a detailed description of the research context (e.g., study population and setting, researcher's perspective), together with the assumptions made during the research.
3. Dependability: This is an alternative to "reliability" in quantitative inquiry, and is addressed through providing an open and honest account of the research procedures such that another researcher could replicate the study under similar conditions. It is *not* necessary to assume that another researcher would collect identical data, or arrive at similar conclusions.
4. Confirmability: This relates to the controlling of potential biases. Here, a geoscience expert external to the study scrutinized the method and analysis, and verified that the conceptions described could be recognized by others. The inclusion of participant quotes provides a further link between the raw data and interpretations.

Results and Discussion

Phenomenographic analysis of the combined student and faculty data revealed six qualitatively distinct conceptions of geoscience as an academic discipline, ranging from simple to complex.

Conception A: The Earth

This is the simplest conception identified; geoscience is about studying Earth as an object. Typical examples of this conception include:

> *06GL1/11: The study of the earth.*
>
> *06GL2/34: The science of the earth.*

Conception B: The Composition of the Earth

Responses expressing this conception make reference to Earth being composed of some kind of matter, e.g., rocks or minerals, or to having "structure" or features. Typical examples of this conception include:

> *06GL1/52: Study of the earths* [sic] *crust, in particular rock formations.*
>
> *06GL3/40: Geology is about rocks, geochemistry, petrology, mineralogy, sedimentology, igneous rocks and metamorphic.*

Conception C: Processes Operating upon and within the Earth

This conception is more complex than A and B, considering geoscience in dynamic rather than static terms and relating surface features to processes. Although some responses make reference to specific processes such as plate tectonics or deposition, the majority refer in more general terms to the way Earth works. Typical examples of this conception include:

> *06GL1/32: The workings of the earth and how formations, structures, rocks have formed because of this.*
>
> *06GL2/47: The study of all physical processes that operate on the planet.*
>
> *06GL3/4: Understanding the processes that shape the earth's surface and having knowledge of subsurface features and processes.*

Conception D: The Earth through Time

Although the operating of processes might *imply* the passing of time, responses categorized under this conception make specific reference to time as a fundamental characteristic of geoscience, e.g., by referring to Earth history, evolution, or to points in time such as the past, present, or future. This conception was almost exclusively expressed in conjunction with conception C. Typical examples of this conception include:

> *06GL1/31: Geology—the study of rocks and geological features and their evolution or change over time.*
>
> *06GL2/17: Looking at rocks in order to understand the history of the planet.*
>
> *06GL3/17: Discerning the history of the earth using evidence from the earth itself.*

06GL1/31 and 06GL2/17 would be categorized as both conception B and conception D, since they make reference to studying or looking at rocks, thereby recognizing that Earth has some kind of composition.

Conception E: Interacting Systems

This conception considers geoscience in terms of the action and interaction of natural systems, both terrestrial and extraterrestrial (i.e., geoscience is to do with more than just our home planet). As well as making general reference to "systems," responses may also refer to specific systems-related issues such as climate change. Typical examples of this conception include:

> *06GL1/34: Geology is the study of the earth and earth systems, as well as other planets etc.*
>
> *06GL2/25: Earth systems science.*
>
> *06GLF/30: The student gaining insight into how the Earth works and how different Earth systems integrate into the whole.*

In addition to Earth systems, 06GLF/30 also contains conception C, since it makes reference to the action of Earth processes.

Conception F: The Relationship between the Earth and Society

Responses expressing this final conception make some reference (implicit or explicit) to the relationship between geoscience and society, e.g., to Earth as providing resources, or presenting hazards to humans, or to the impact of humans on the planet. Typical examples of this conception include:

> *06GL1/20: The study of the earth, its processes and resources.*
>
> *06GL2/5: Our planet, how it works as an environment, ecosystem, interaction between all the earth sciences, the effects of global warming, outcome of man* [sic] *on the earth.*
>
> *06GL3/32: Assessment of hazards caused by natural environment and risks imposed on human race, finding a solution to resolve the problems.*

As with the previous category, responses 06GL1/20 and 06GL2/5 also make reference to processes, and would therefore be categorized as both C and F.

Outcome Space

This set of conceptions, which emerged from the combined student and faculty data set, describes the variation in ways of conceptualizing geoscience that exists within the study population. Having identified these conceptions we now need to consider how they relate to each other, i.e., determine the format of the outcome space. Although the idea of a nice, neat hierarchy of conceptions might be appealing to many, I argue that "real" data (particularly of the qualitative kind) are seldom at all tidy. That said, the emergent conceptions appear to express a certain degree of ordering. Conception A is the most naïve, considering geoscience simply in terms of an object to be studied (the Earth). Conception B also focuses on the Earth as a static object, but is more complex since it recognizes that Earth has a composition, i.e., it is made of something. A further increase in complexity, together with a change in focus from a static to a dynamic Earth, is reflected in conceptions C and D. The relationship between these two conceptions is unclear, and it could be argued that time and process are so intimately linked that they should, in fact, be considered as a single conception. However, it is common for U.K. undergraduate instruction to emphasize *either* process or time within the context of individual courses, particularly during the early stages of a degree program. For example, first year instruction at the study institution includes courses entitled "Planet Earth" and "Earth History"; the former focuses on processes which cause the planet to change (e.g., plate tectonics, the rock cycle, climate change), and the latter on recognizing and deciphering evidence for Earth changing through time (e.g., using the fossil and stratigraphic record). Process and time are therefore presented here as separate, but equally complex, conceptions. Likewise conceptions E and F are more complex than C and D, focusing on the interactions between different natural systems, and the specific interaction between Earth and society (i.e., the "human" system), respectively. Conception F could, in fact, be considered a more "applied" subset of conception E, but I argue that conceptualizing geoscience in human terms is sufficiently qualitatively distinct to warrant being considered separately.

A somewhat hybrid outcome space therefore emerges from this analysis, containing both hierarchical and branching elements (Fig. 2). Conceptions A and B differ from the remaining four not just in terms of their relative simplicity, but also in terms of conceptualizing geoscience as "static" as opposed to "dynamic."

This dichotomy of types of conception has been recognized in previous investigations into conceptions of academic disciplines. Crawford et al. (1994) classified conceptions of mathematics into "fragmented," which focus on the study of numbers, and "cohesive," whereby math is conceptualized as a logical way of thinking that can be applied to understanding the world. Similar terminology was applied by Reid and Petocz (2002), who used the terms "fragmented" and "inclusive" to describe conceptions of statistics. Bradbeer et al. (2004) categorized conceptions of geography as "structural," in which geography is seen simply as the study of the world, or "relational," where it involves the study of inter-relationships between natural phenomena, e.g., people-environment interactions, or spatial distributions. Interestingly, Bradbeer et al. (2004) also reported a hybrid outcome space, in which structural conceptions formed a hierarchy and relational conceptions stood as "alternatives." This in turn is similar to the categorization used by Loughland et al. (2002), who applied the terms "object focus" and "relational focus" to conceptions of environment that focus on the environment as an object, and on the relationship between people and the environment, respectively. The categorizations of Bradbeer et al. (2004) and Loughland et al. (2002) in particular appear to correlate broadly with the "static" and "dynamic" classifications applied to the conceptions identified in this study, in which geoscience is conceptualized in terms of the study of a static object (i.e., Earth), or the inter-relating of natural phenomena over time.

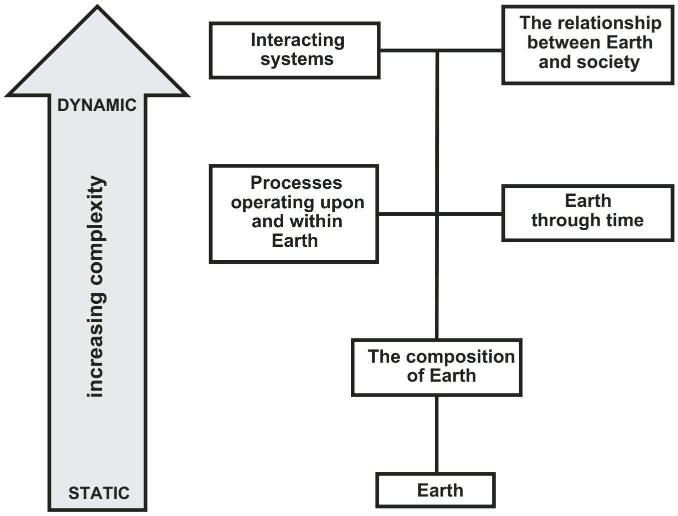


Figure 2. The outcome space for geoscience as defined by this study. Conceptions form a hybrid hierarchical-branching structure that varies from simple conceptions of geoscience as a "static" phenomenon through to more complex conceptions of geoscience as "dynamic."

Quantitative Variations in Student and Faculty Conceptions

The six conceptions described here emerged from considering all of the student and faculty responses as a single data set. However, if the data set is subdivided into undergraduate year of study we still find all six conceptions being expressed within each subgroup; new conceptions do not emerge, nor do existing conceptions disappear. I suggested earlier that we might expect academic faculty (experts) to hold different conceptions of geoscience to undergraduates (novices) based on differences in experience of geoscience. In fact, the same six conceptions are also found to be present within the faculty data set (Table 2). This suggests that the critical variation in ways of conceptualizing geoscience remains consistent regardless of the stage of study, or level of expertise. This does not mean, however, that the same conception or conceptions of geoscience dominate at different levels of expertise; to explore this further, we need to quantify the findings, and compare frequency distributions between students and faculty. Since the focus of this paper is on qualitative inquiry, I will keep the quantitative discussion brief. It is necessary to include, however, since it will determine the frequency, and hence the "strength" of each conception (Kempa and Orion, 1996) among students and faculty, which in turn has implications for instruction.

Table 2 shows the resulting frequency distributions for conceptions A–F across the student and faculty populations. It should be noted that the sum of proportions across each subpopulation equates to more than 100%; this is because many responses contained more than one conception, and are hence counted in multiple categories. For example, the response *"study of geological processes past and present"* would be counted under conception C and conception D, since it relates to both process and time. Although there are limitations to these data (e.g., varying population size and response rates), they nonetheless provide some indication of the pervasiveness of the six conceptions within the student and faculty populations (Kusnick, 2002). Most significantly, it appears that the majority of students and faculty in this study conceptualize geoscience in terms of processes operating within and upon Earth. The dominant conception (i.e., the one that is most frequently encountered) of geoscience as a discipline

TABLE 2. FREQUENCY DISTRIBUTIONS FOR CONCEPTIONS A TO F

		Conceptions (%)					
	n	A	B	C	D	E	F
Students: year 1	51	12	22	75	31	10	20
Students: year 2	45	13	29	60	20	16	9
Students: year 3	39	13	15	56	41	10	10
All students	135	13	22	64	30	12	13
Academic faculty	13	8	15	62	46	38	31

Note: A single response may contain multiple conceptions.

is therefore consistent across levels of expertise. There are some interesting variations within the data, however, which can be seen most clearly by comparing the combined student data against faculty data (Fig. 3). These indicate that the most complex conceptions of geoscience, i.e., those concerning interacting systems or the relationship between Earth and society, are more pervasive among faculty than among students, where they are the least pervasive. For faculty, the least pervasive conceptions are the most naïve, i.e., those considering geoscience in terms of "matter."

Looking at the student data in Table 2 more closely, there is a degree of variation in the frequency distributions across academic years, although conception C remains dominant throughout. This suggests that a process-focused conception of geoscience persists continuously through an undergraduate program, but also that other conceptions might become more or less pervasive over time. Although intriguing, there is a limit to which these data can be interpreted since they represent a "snapshot" of the conceptions held at a specific moment in time by students at different academic stages. A rigorous investigation into conceptional *change* will require longitudinal data, collected from the same cohort of students as they gain progressively more expertise over the course of their degree program (Stokes and Anderson, 2009).

Implications for Curriculum Development and Undergraduate Instruction

While it is not a stated aim of this study to consider the implications of the findings in terms of curriculum development and undergraduate instruction, it is worth considering how these might be applied to curriculum design. Ebenezer and Fraser described phenomenography as "the beginning of a process of curriculum change" (2001, p. 512), and this research has two findings that have important implications for undergraduate geoscience instruction. The first concerns the range of conceptions of geoscience emerging from the study data. The fact that these form a partial hierarchy is evidence that both students and faculty hold conceptions of geoscience that vary from simple and naïve, to complex and sophisticated. The second key finding concerns the dominance of particular conceptions within student and faculty subpopulations. Here, we find the dominant conception to be congruent across different levels of expertise, with students at all academic levels and faculty, conceptualizing geoscience in terms of processes operating upon and within Earth. However, we find the simplest conceptions (those focused on Earth as a static object) to be more prevalent among students, while the most complex conceptions (those focused on systems) are more prevalent among faculty.

The first stage in applying these findings to curriculum development is recognizing that learners may hold conceptions that differ, to some extent, from those held by course designers, i.e., faculty. This recognition is critical since faculty may make assumptions about students' conceptions of geoscience based upon their own conceptions, which in turn will influence the types of teaching and learning strategies assumed to be effective (e.g., Watters and Watters, 2007). Once the range of novice (student) conceptions has been identified, this knowledge can be used to structure the development of learning activities such that learners are able to acquire progressively more complex conceptions (e.g., Ebenezer and Fraser, 2001; Scott et al., 2007). Although classroom-based learning can be effective in promoting

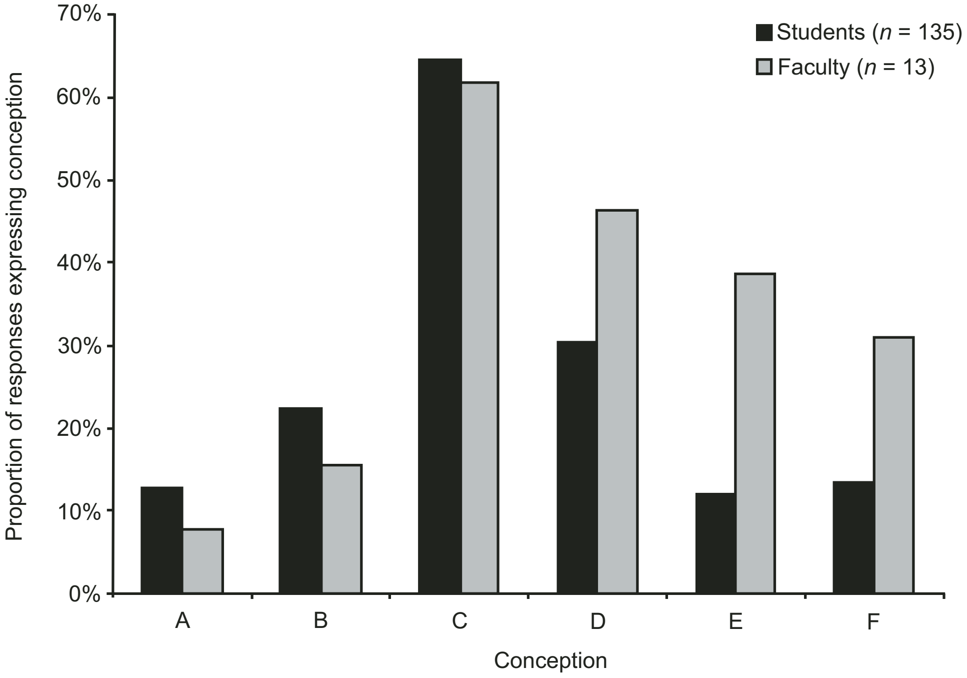


Figure 3. Frequency distributions for the six conceptions of geoscience. The process-oriented conception (conception C) is most prevalent among students and faculty. The data also suggest that complex, "dynamic" conceptions (D–F) are more pervasive among experts (faculty) than novices (students).

conceptional change (e.g., Tóth and Ludányi, 2007), active learning environments such as the field or laboratory enable novice geoscientists to engage in concrete rather than abstract experiences, and thus promote the transformation of direct experience into a deeper and more sophisticated understanding of Earth phenomena (e.g., Kern and Carpenter, 1986; Elkins and Elkins, 2007). The effectiveness of these environments in engendering conceptional change within novice geoscientists should therefore form a focus for ongoing investigation.

The potential implications of these findings for the study institution are interesting, raising questions about the existing curriculum and presenting interesting avenues for further research. Official criteria against which the knowledge, skills, and abilities of U.K. geoscience graduates are assessed (i.e., those used to determine academic "success") state that degree programs in earth sciences should include (QAA, 2007, p. 2):

- *a systems approach to understanding the present and past interactions between the processes operating in the Earth's core, mantle, crust, cryosphere, hydrosphere, atmosphere, and biosphere, and the perturbations of these systems by extraterrestrial influences;*
- *the scientific study of the physical, chemical, and biological processes operating on and within the Earth;*
- *the structure and composition of the Earth and other planets; and*
- *the history of the Earth over geological timescales.*

These criteria broadly reflect the range of conceptions of geoscience identified in this study, and, more significantly, they make explicit the expectation that geoscience graduates should possess a sound understanding of Earth systems. Findings from this study suggest that systems-based conceptions may not be widely developed among students at the study institution entering their final year of study. Earth systems are, however, covered extensively within the final year curriculum, although the context in which these are encountered varies depending upon the degree pathway chosen. The B.Sc. Applied Geology and B.Sc. Physical Geography and Geology programs have an applied focus, and students following these pathways will encounter concrete, systems-based concepts such as geohazards and georesources during their final year of instruction. By comparison, the B.Sc. and M.Geol. Geology programs focus on the purer aspects of geoscience, considering systems in a way that is much more abstract. One might therefore hypothesize that a relationship exists between the development of more complex conceptions of geoscience, and degree pathway, particularly during the final year of instruction. Since the testing of this hypothesis requires conceptions data to be collected at the end, as well as the beginning, of the students' final year, i.e., data that are longitudinal, this cannot be investigated further within the current study. However, should future investigations reveal significant variation in the conceptions of geoscience held by students graduating from different types of degree pathways, then the implications for curriculum development and undergraduate instruction will need to be carefully considered.

CONCLUSIONS AND FUTURE DIRECTIONS

A key aim of this paper has been to establish the value of phenomenography to education research in general, and to demonstrate its application to geoscience education research in particular. This approach has successfully identified a range of qualitatively distinct ways in which undergraduate students and faculty at a single U.K. university conceptualize geoscience as an academic discipline, and it has provided some indication of the pervasiveness of various conceptions at different levels of expertise. A logical next step is to collect longitudinal data to explore the extent of conceptional change resulting from an undergraduate geoscience degree. These data are currently being collected (Stokes and Anderson, 2009), and will enable the following questions to be addressed:

1. To what extent do students' conceptions change during the course of an undergraduate degree?
2. Is conceptional change uniform or variable within a cohort?
3. If variable, what are the likely parameters of variation (e.g., gender, degree pathway)?
4. Can conceptual change be linked to specific courses or learning experiences such as fieldwork?

As well as conducting a longitudinal investigation, the study should be replicated within other institutions to establish the transferability of the findings, and provide a clearer indication of their application and value. Of course, phenomenographic inquiry does not need to be restricted to educational contexts; the flexibility offered by this approach makes it well suited to wider geocognitive inquiry, and hence future investigation should be expanded to include novices and experts in educational and non-educational settings.

ACKNOWLEDGMENTS

This research was supported by funding from the University of Plymouth Experiential Learning CETL (Centre for Excellence in Teaching and Learning). I would like to extend my gratitude to all students and faculty who participated in this research, and to Mark Anderson and Ruth Weaver for helpful discussion concerning the findings. The two reviewers and Anthony Feig are thanked for their thorough and insightful comments, which greatly improved the first draft of this manuscript.

REFERENCES CITED

Åkerlind, G.S., 2005, Variation and commonality in phenomenographic research methods: Higher Education Research & Development, v. 24, p. 321–334, doi:10.1080/07294360500284672.

Åkerlind, G.S., and Kayrooz, C., 2003, Understanding academic freedom: The views of social scientists: Higher Education Research & Development, v. 22, p. 327–344, doi:10.1080/0729436032000145176.

Albert, M., Laberge, S., Hodges, B.D., Regehr, G., and Lingard, L., 2008, Biomedical scientists' perception of the social sciences in health research: Social Science & Medicine, v. 66, p. 2520–2531, doi:10.1016/j.socscimed.2008.01.052.

Ashwin, P., 2005, Variations in students' experiences of the "Oxford Tutorial": Higher Education, v. 50, p. 631–644, doi:10.1007/s10734-004-6369-6.

Ashworth, P., and Lucas, U., 2000, Achieving empathy and engagement: A practical approach to the design, conduct and reporting of phenomenographic research: Studies in Higher Education, v. 25, p. 295–308, doi:10.1080/713696153.

Beaty, E.M., 1987, Understanding concepts in social science: Towards an effective evaluation strategy: Instructional Science, v. 15, p. 341–359, doi:10.1007/BF00139620.

Bezzi, A., 1999, What is this thing called geoscience? Epistemological dimensions elicited with the repertory grid and their implication for scientific literacy: Science Education, v. 83, p. 675–700, doi:10.1002/(SICI)1098-237X(199911)83:6<675::AID-SCE3>3.0.CO;2-Q.

Booth, S., 2008, Researching learning in networked learning—Phenomenography and variation theory as empirical and theoretical approaches, *in* Hodgson, V., Jones, C., Kargidis, T., McConnell, D., Retalis, S., Stamatis, D., and Zenios, M., eds., Proceedings of the 6th International Conference on Networked Learning: Halkidiki, Greece, p. 450–455, http://networkedlearningconference.org.uk/past/nlc2008/abstracts/PDFs/Booth_450-455.pdf (accessed 20 September 2010).

Bowden, J., Dall'Alba, G., Martin, E., Laurillard, D., Marton, F., Masters, G., Ramsden, P., Stephanou, A., and Walsh, E., 1992, Displacement, velocity, and frames of reference: Phenomenographic studies of students' understanding and some implications for teaching and assessment: American Journal of Physics, v. 60, p. 262–269, doi:10.1119/1.16907.

Bradbeer, J., Helaey, M., and Kneale, P., 2004, Undergraduate geographers' understandings of geography, learning and teaching: A phenomenographic study: Journal of Geography in Higher Education, v. 28, p. 17–34, doi:10.1080/0309826042000198611.

Cope, C., 2004, Ensuring validity and reliability in phenomenographic research using the analytical framework of a structure of awareness: Qualitative Research Journal, v. 4, p. 5–18.

Cousin, G., 2009, Researching Learning in Higher Education: An Introduction to Contemporary Methods and Approaches: Abingdon and New York, Routledge, 257 p.

Crawford, K., Gordon, S., Nicholas, J., and Prosser, M., 1994, Conceptions of mathematics and how it is learned: The perspectives of students entering university: Learning and Instruction, v. 4, p. 331–345, doi:10.1016/0959-4752(94)90005-1.

Derrida, J., 1978, Structure, sign and play in the discourse of the human sciences (trans. Bass, A.), *in* Writing and Difference: Chicago, University of Chicago Press, p. 278–293.

DfES (Department for Education and Skills), 2003, Widening Participation in Higher Education: London, Department for Education and Skills, 27 p., http://www.bis.gov.uk/assets/biscore/corporate/migratedd/publications/6/6820-dfes-wideningparticipation2.pdf (accessed 4 January 2011).

Ebenezer, J.V., and Fraser, D.M., 2001, First year chemical engineering students' conceptions of energy in solution processes: Phenomenographic categories for common knowledge construction: Science Education, v. 85, p. 509–535, doi:10.1002/sce.1021.

Elkins, J.T., and Elkins, N.M.L., 2007, Teaching geology in the field: Significant geoscience concept gains in entirely field-based introductory geology courses: Journal of Geoscience Education, v. 55, p. 126–132.

Feig, A.D., 2011, this volume, Methodology and location in the context of qualitative data and theoretical frameworks in geoscience education research, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(01).

Ford, D.J., 2003, Sixth graders' conceptions of rocks in their local environment: Journal of Geoscience Education, v. 51, p. 373–377.

Gobert, J.D., 2000, A typology of causal models for plate tectonics: Inferential power and barriers to understanding: International Journal of Science Education, v. 22, p. 937–977, doi:10.1080/095006900416857.

Gruber, H.E., 1974, Darwin on Man: A Psychological Study of Scientific Creativity: New York, Dutton, 495 p.

Harwood, D., and Jackson, P., 1993, "Why did they build this hill so steep?": Problems of assessing primary children's understanding of physical landscape features in the context of the UK National Curriculum: International Research in Geographical and Environmental Education, v. 2, p. 64–79, doi:10.1080/10382046.1993.9964911.

Hasselgren, B., and Beach, D., 1997, Phenomenography—A "good-for-nothing brother" of phenomenology? Outline of an analysis: Higher Education Research & Development, v. 16, p. 191–202, doi:10.1080/0729436970160206.

Hoffman, R.R., 1998, How can expertise be defined?, *in* Williams, R., Faulkner, W., and Fleck, J., eds., Exploring Expertise: New York, Macmillan, p. 81–100.

Kempa, R.F., and Orion, N., 1996, Students' perception of co-operative learning in earth science fieldwork: Research in Science & Technological Education, v. 14, p. 33–41, doi:10.1080/0263514960140103.

Kern, E.L., and Carpenter, J.R., 1986, Effect of field activities on student learning: Journal of Geological Education, v. 34, p. 180–183.

King, C., 2008, Geoscience education: An overview: Studies in Science Education, v. 44, p. 187–222, doi:10.1080/03057260802264289.

Kortz, K., and Murray, D., 2009, Barriers to college students learning how rocks form: Journal of Geoscience Education, v. 57, p. 300–315.

Kuhn, T.S., 1995, The Structure of Scientific Revolutions (3rd ed.): Chicago, University of Chicago Press, 226 p.

Kusnick, J., 2002, Growing pebbles and conceptual prisms—Understanding the source of student misconceptions about rock formation: Journal of Geoscience Education, v. 50, p. 31–39.

Libarkin, J.C., Anderson, S.W., Dahl, J., Beilfuss, M., and Boone, W., 2005, Qualitative analysis of college students' ideas about the Earth: Interviews and open-ended questionnaires: Journal of Geoscience Education, v. 53, p. 17–26.

Lincoln, Y.S., and Guba, E.G., 1985, Naturalistic Inquiry: Beverly Hills, California, Sage Publishing, 416 p.

Linder, C.J., and Erickson, G.L., 1989, A study of tertiary physics students' conceptualizations of sound: International Journal of Science Education, v. 11, p. 491–501, doi:10.1080/0950069890110502.

Lipsett, A., 2009, Cash-strapped ministers freeze university expansion: The Guardian, London, http://www.guardian.co.uk/education/2009/jan/22/university-recruitment-cap (accessed 20 September 2010).

Loughland, T., Reid, A., and Petocz, P., 2002, Young people's conceptions of environment: A phenomenographic analysis: Environmental Education Research, v. 8, p. 187–197, doi:10.1080/13504620220128248.

Mackintosh, M., 2005, Children's understanding of rivers: International Research in Geographical and Environmental Education, v. 14, p. 316–322, doi:10.1080/10382040508668365.

Marton, F., 1981, Phenomenography—Describing conceptions of the world around us: Instructional Science, v. 10, p. 177–200, doi:10.1007/BF00132516.

Marton, F., 1986, Phenomenography—A research approach to investigating different understandings of reality: Journal of Thought, v. 21, p. 28–49.

Marton, F., 1994, Phenomenography, *in* Husen, T., and Postlethwaite, T.N., eds., The International Encyclopedia of Education (2nd ed.): London, Pergamon, p. 4424–4429.

Marton, F., and Booth, S., 1997, Learning and Awareness: Mahwah, New Jersey, Lawrence Erlbaum Associates, 224 p.

Marton, F., and Säljö, R., 1976, On qualitative differences in learning: II. Outcome as a function of the learner's conception of the task: The British Journal of Educational Psychology, v. 46, p. 115–127.

Marton, F., Dall'Alba, G., and Beaty, L., 1993, Conceptions of learning: International Journal of Educational Research, v. 19, p. 277–300.

Mays, N., and Pope, C., 1995, Education and debate: Rigor and qualitative research: British Medical Journal, v. 311, p. 109–112.

McKenzie, K., and Schweitzer, R., 2001, Who succeeds at university? Factors predicting academic performance in first year Australian university students: Higher Education Research & Development, v. 20, p. 21–33, doi:10.1080/07924360120043621.

McLean, M., 2001, Can we relate conceptions of learning to student academic achievement?: Teaching in Higher Education, v. 6, p. 399–413, doi:10.1080/13562510120061241.

Minasian-Batmanian, L.C., Lingard, J., and Prosser, M., 2006, Variation in student reflections on their conceptions of and approaches to learning biochemistry in a first-year health sciences service subject: International Journal of Science Education, v. 28, p. 1887–1904, doi:10.1080/09500690600621274.

Morris, J., 2006, The implications of either "discovering" or "constructing" categories of description in phenomenographic analysis, *in* Challenging the Orthodoxies Conference, Middlesex University, December 2006, http://www.m.ac.uk/aboutus/fpr/clqe/docs/Jennymorris.pdf (accessed August 2009).

Nussbaum, J., 1985, The Earth as a cosmic body, *in* Driver, R., Guesne, E., and Tiberghein, A., eds., Children's Ideas in Science: Milton Keynes, UK, Open University Press, p. 170–192.

Petcovic, H.L., and Libarkin, J.C., 2007, Research in science education: The expert-novice continuum: Journal of Geoscience Education, v. 55, p. 333–339.

Power, C., Robertson, F., and Baker, M., 1987, Success in Higher Education: Canberra, Australian Government Publishing Service, 76 p.

Prosser, M., Trigwell, K., and Taylor, P., 1994, A phenomenographic study of academics' conceptions of science learning and teaching: Learning and Instruction, v. 4, p. 217–231, doi:10.1016/0959-4752(94)90024-8.

Prosser, M., Walker, P., and Millar, R., 1996, Difference in students' perceptions of learning physics: Physics Education, v. 31, p. 43–48, doi:10.1088/0031-9120/31/1/022.

QAA (Quality Assurance Agency for Higher Education), 2007, Benchmark Statement for Environmental Science, Earth Science and Environmental Studies: Gloucester, QAA, 24 p., http://www.qaa.ac.uk/academicinfrastructure/benchmark/statements/EarthSciences.asp (accessed August 2009).

Reid, A., and Petocz, P., 2002, Students' conceptions of statistics: A phenomenographic study: Journal of Statistics Education, v. 10, p. 1–18.

Säljö, R., 1979, Learning in the Learner's Perspective: I. Some Common-Sense Conceptions: Department of Education, University of Göteburg, Report 76.

Sandbergh, J., 1997, Are phenomenographic results reliable?: Higher Education Research & Development, v. 16, p. 203–212, doi:10.1080/0729436970160207.

Schoon, K.J., 1992, Students' alternative conceptions of Earth and space: Journal of Geological Education, v. 40, p. 209–214.

Scott, P., Asoko, H., and Leach, J., 2007, Student conceptions and conceptual learning in science, *in* Abell, S.K., and Lederman, N.G., eds., Handbook of Research on Science Education: Mahwah, New Jersey, Lawrence Erlbaum Associates, p. 31–56.

Sibley, D.F., 2005, Visual abilities and misconceptions about plate tectonics: Journal of Geoscience Education, v. 53, p. 471–477.

Stokes, A., and Anderson, M., 2009, Undergraduate students' conceptions of geology: What are they, and do they change?: Geological Society of America Abstracts with Programs, v. 41, no. 7, p. 156.

Stokes, A., Magnier, K., and Weaver, R., 2010, What's the use of fieldwork? Conceptions of students and staff in geography and geology: Journal of Geography in Higher Education, doi:10.1080/03098265.2010.487203 (in press).

Svensson, L., 1997, Theoretical foundations of phenomenography: Higher Education Research & Development, v. 16, p. 159–171, doi:10.1080/0729436970160204.

Svensson, L., and Hogfors, C., 1988, Conceptions as the content of teaching: Improving education in mechanics, *in* Ramsden, P., ed., Improving Learning: New Perspectives: London, Kogan Page, p. 162–177.

Tóth, Z., and Ludányi, L., 2007, Combination of phenomenography with knowledge space theory to study students' thinking patterns in describing an atom: Chemistry Education Research and Practice, v. 8, p. 327–336.

Trend, R., 1998, An investigation into understanding of geological time among 10- and 11-year-old children: International Journal of Science Education, v. 20, p. 973–988, doi:10.1080/0950069980200805.

Trend, R., 2000, Conceptions of geological time among primary teacher trainees, with reference to their engagement with geoscience, history, and science: International Journal of Science Education, v. 22, p. 539–555, doi:10.1080/095006900289778.

Trend, R., Everett, L., and Dove, J.E., 2000, Interpreting primary children's representations of mountains and mountainous landscapes and environments: Research in Science & Technological Education, v. 18, p. 85–112, doi:10.1080/02635140050031064.

Trigwell, K., 2006, Phenomenography: An approach to research in geography education: Journal of Geography in Higher Education, v. 30, p. 367–372, doi:10.1080/03098260600717489.

Tynjälä, P., 1997, Developing education students' conceptions of the learning process in different learning environments: Learning and Instruction, v. 7, p. 277–292, doi:10.1016/S0959-4752(96)00029-1.

Walsh, E., 2000, Phenomenographic analysis of interview transcripts, *in* Bowden, J.A., and Walsh, E., eds., Phenomenography: Melbourne, Australia, MIT Publishing, p. 19–33.

Walsh, L.N., Howard, R.G., and Bowe, B., 2007, Phenomenographic study of students' problem solving approaches in physics: Physical Review Special Topics, Physics Education Research, v. 3, 12 p., doi:10.1103/PhysRevSTPER.3.020108.

Wandersee, J.H., Mintzes, J.J., and Novak, J.D., 1994, Research on alternative conceptions in science, *in* Gabel, D.L., ed., Handbook of Research on Science Teaching and Learning: New York, Macmillan Publishing Company, p. 177–210.

Watters, D.J., and Watters, J.J., 2007, Approaches to learning by students in the biological sciences: Implications for teaching: International Journal of Science Education, v. 29, p. 19–43, doi:10.1080/09500690600621282.

Webb, G., 1997, Deconstructing deep and surface: Towards a critique of phenomenography: Higher Education, v. 33, p. 195–212, doi:10.1023/A:1002905027633.

MANUSCRIPT ACCEPTED BY THE SOCIETY 23 JUNE 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

The qualitative underpinnings of quantitative concept inventory questions

Julie C. Libarkin*
Emily M. Geraghty Ward*
Department of Geological Sciences and Center for Research on College Science Teaching and Learning,
Michigan State University, East Lansing, Michigan 48824, USA

ABSTRACT

Concept inventory (CI) development in higher education is an active research area. As evaluation instruments, the validity and reliability of CIs should be important considerations, particularly as these tools become integral parts of individual, department, or programmatic assessments. While methods used to establish validity and reliability vary, most researchers agree that qualitative data analysis is a necessary prerequisite to writing meaningful questions. The CI's strength as a proxy for conceptual understanding depends upon the link between inventory content and ideas held by the testing population. Most commonly, CI developers utilize qualitative data about student alternative conceptions to write incorrect response options and to check student intentions with think-alouds; less frequent use of qualitative data to write question stems also occurs. The Geoscience Concept Inventory (GCI) was developed in the early 2000s in response to a growing need for a widely applicable assessment tool. The instrument is grounded in student data, following earlier efforts in other disciplines, albeit with significant modification. As with earlier efforts, semistructured interviews probed student thinking about foundational ideas in geology and were used to craft inventory questions following protocols for survey development and psychometric analysis. These common qualitative foundations for CI development have been discussed in the CI literature extensively; the significant role that qualitative data play in the question review and revision process has not been considered. This paper explores the importance of qualitative data in question development itself, providing a detailed unpacking of the review and revision process for an exemplar case.

INTRODUCTION

The design of surveys for assessment, and especially qualitative best practice, is a well-investigated and important area of research. This research is important for a number of frames of discourse, including: measurement of student achievement relative to an intervention (e.g., Gronlund, 1993), consideration of attitudinal changes in response to instruction (e.g., Schau et al., 1995), growth in conceptual understanding over time (e.g., Liu and McKeough, 2005), and large-scale sampling of populations considered nearly impossible via qualitative methods (e.g., Miller, 2004). The movement toward concept inventory (CI)

---

*E-mails: libarkin@msu.edu; wardem@msu.edu.

Libarkin, J.C., and Geraghty Ward, E.M., 2011, The qualitative underpinnings of quantitative concept inventory questions, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 37–48, doi:10.1130/2011.2474(04). For permission to copy, contact editing@geosociety.org.

development in assessment for higher education science and engineering is an attempt to take advantage of the significant benefits that quantitative surveys can provide for evaluation of learning for large groups and/or across different populations.

The development of CIs is a widely discussed subfield of research within the larger domain of survey research (e.g., Libarkin, 2008; Reed-Rhoads and Imbrie, 2008). CIs are multiple-choice instruments that target fundamental concepts within a field, and they are most commonly used to diagnose conceptual understanding and evaluate learning. Interestingly, development of CIs is not always grounded in existing survey design, psychometric standards, or educational frameworks. CI development in higher education typically emerges from within the target discipline, and experience with work on item writing (e.g., Haladyna and Downing, 1989; Frey et al., 2005) and psychometric statistics (e.g., Bond and Fox, 2001) is not always available within each CI development group. While CIs are certainly useful to users within the target domain, usefulness outside the domain will largely depend upon the extent to which attention has been paid to the broader literature emerging outside of science and engineering fields.

As evaluation and assessment instruments, the validity and reliability of CIs are important considerations, particularly as these tools become integral parts of individual, department, or programmatic assessments. While the methods used to establish validity and reliability in developing CIs vary significantly across fields, most developers agree that analysis of qualitative data, such as interviews or open-ended surveys, is a necessary prerequisite to writing meaningful CI questions. In fact, qualitative research has shown that test-takers do not always respond to multiple-choice questions for reasons intended by the test developers (e.g., Harlow and Jones, 2004). In addition, the use of qualitative techniques such as think-alouds (e.g., Zeilik et al., 1999) is a necessary component of CI development for the establishment of communication validity (Lopez, 1996).

Concept Inventories in Higher Education Science

Libarkin (2008) identified 23 distinct CIs that have been developed in higher education science; a similar number of inventories in engineering has also been documented (Reed-Rhoads and Imbrie, 2008). Multiple-choice diagnostic tests have been used for decades in science education research (e.g., Treagust, 1986), and they were adopted quickly by the physics education community (e.g., Halloun and Hestenes, 1985; Hestenes et al., 1992). Other sciences followed with concept inventories created for very particular assessment niches, including additional instruments in physics/astronomy (e.g., Hufnagel, 2002; Lindell and Olsen, 2002), chemistry (e.g., Mulford and Robinson, 2002), biology (e.g., Bowling et al., 2008), and geoscience (e.g., Libarkin and Anderson, 2005, 2007).

Interestingly, CIs in higher education science and engineering cannot be interrelated. Each instrument was developed by a different research group, using different development strategies, applying different validation standards, and targeting different populations. As a result, a score on one CI cannot be compared directly to a score on a different CI. This limits the overall utility of CIs for determining the efficacy of instructional interventions across multiple domains, settings, and populations. Development of mechanisms for interrelating CI scores is an underinvestigated, yet important area of emerging research.

Regardless of their limited use in comparing learning across disciplines, CIs provide a valuable source of information about student understanding to faculty and researchers, and they can be administered quickly and easily. Although other means of assessment, such as open-ended surveys, observations, or interviews, would provide the richer understanding possible with qualitative methods, CIs allow for testing of large participant samples. This testing generally covers a broad range of material as well. CIs in science contain between 17 and 43 items, although Libarkin (2008; after Libarkin and Anderson, 2007) recommended that the smaller sample be used to avoid subject fatigue during the testing process.

The strength of a CI to serve as a meaningful proxy for conceptual understanding depends on the link between the inventory content and the actual ideas held by the population being tested. Most commonly, CIs utilize qualitative data that provide insight into alternative conceptions (see Libarkin [2008] and Reed-Rhoads and Imbrie [2008] for discussions). These alternative conceptions are integrated into multiple-choice questions ("items") as meaningful incorrect response options ("distracters"; Fig. 1). More rarely, qualitative data can also be used to develop the question stems themselves, by exposing unexpected areas of difficulty (e.g., Libarkin and Anderson, 2007). Stems can then target authentic student understanding rather than faculty perceptions of student thinking.

The Geoscience Concept Inventory (GCI) was developed in the early 2000s in response to a growing need for a widely applicable assessment tool in the geosciences, particularly for evaluation of learning in entry-level courses. The grounding of this instrument in authentic student data followed from earlier CI development efforts in higher education science (e.g., Halloun and Hestenes, 1985; Hestenes et al., 1992), albeit with significant modification to instrument design methodology. In particular, semistructured interviews were utilized to probe student thinking about general and foundational ideas in the geological sciences, and these interviews were then used to craft question stems and distracters (Libarkin and Anderson, 2005). The use of interviews to craft both item stems and distracters was based in grounded theory (Libarkin and Anderson, 2007), and analysis of

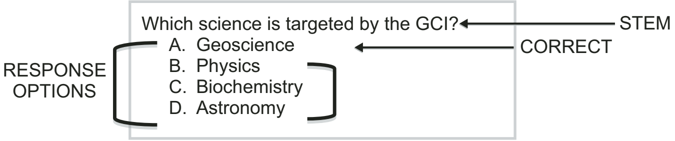


Figure 1. Components of a multiple-choice question.

the validity and reliability of GCI items was explored via expert and user review, Rasch analysis, and an ongoing cycle of revision and review. Additional discussion of the validity and reliability of the GCI can be found in Libarkin and Anderson (2006, 2007) and Libarkin (2008).

This current paper will explore the relationship between qualitative data and question development in detail, with particular emphasis on a generally underdiscussed process: question review and revision. Overall, we stress the importance of qualitative data in quantitative instrument development, and use the Geoscience Concept Inventory (GCI) as a case example. The geoscience education research community will be able to utilize this paper in review and development of new GCI, or other quantitative instrument, questions. Other CI developers targeting different content domains can also utilize this discussion to craft items that are responsive to student ideas and standards for CI development. In addition, we believe this approach can be helpful in developing more meaningful multiple-choice classroom assessments, such as those used during in-class exercises or on exams.

While most concept inventory items maintain the simple form depicted in Figure 1, two other formats have been included on some CIs, including the GCI. The first alternative format involves reflective or two-tier questions (e.g., Treagust, 1988; Tan et al., 2005) where a conceptual question is followed by a second question asking the test subject to choose an explanation for their response to the first question. A second alternative question format involves multiple-response questions where subjects are asked to "choose all that apply." Multiple-response questions are essentially a set of true/false questions, although they are often much more difficult than typical single-response questions and have been shown to be similar to free response questions (e.g., Kubinger and Gottschall, 2007). While not common in classical test approaches, item response models (such as Rasch and partial credit) can provide insight into the stability of this type of question.

METHODS

GCI question development follows a careful cycle of qualitative data collection and analysis, review, revision, and further qualitative investigation (Fig. 2). Libarkin and Anderson (2007) laid out the full cycle of GCI development through qualitative and quantitative mechanisms; the quantitative approaches were detailed in Libarkin and Anderson (2006). In this chapter, we provide a detailed picture of the steps taken during GCI question development, with particular emphasis on the qualitative changes that occur as questions evolve, as well as on the qualitative data which are the foundations of GCI validation and usability.

As noted by Libarkin and Anderson (2007), question development can begin once the writer determines the topic that the question will target, considers the specific concept that the question will focus upon, and either identifies common alternative conceptions from the research literature or performs research to reveal authentic student ideas. The target concept that will be covered by a proposed GCI question should be carefully considered prior to initiation of development; where possible, target concepts should emerge from course objectives, program goals, or learning objectives agreed upon at a community level (e.g., Earth

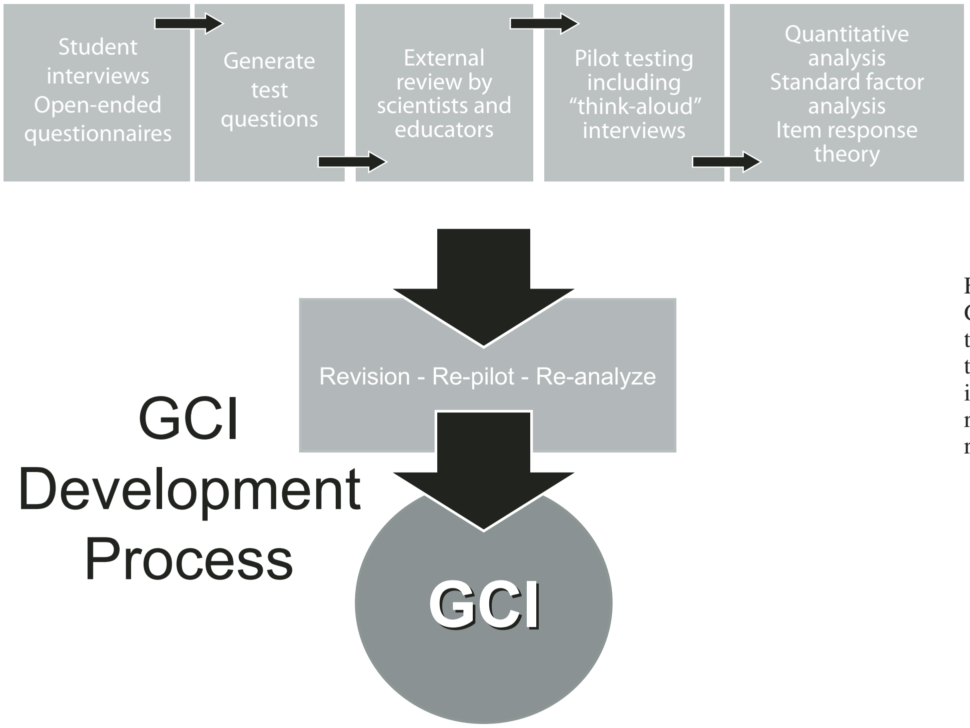


Figure 2. The development process for Geoscience Concept Inventory (GCI) questions. GCI development begins with qualitative data collection and analysis. After initial question writing, a cycle of review, revision, and qualitative data acquisition is required. Modified from Ward et al. (2010).

Science Literacy, 2009). This attention to the goal of item writing will ensure that the question is targeting the concepts that are most important for the developer, the targeted population, and the research or instructional setting.

Alternative conceptions of the topic area can either come from the writer's perceptions of the target population, or can evolve from the perspective of the study population itself. Predetermination of content by CI developers, most often faculty, is very common; many CI developers generate question stems and use open-ended surveys to elicit student responses. These stems are often then transferred directly into CI questions, with the most common four or five open-ended responses becoming question response options. Libarkin and Anderson (2007) advocate use of grounded theory for development of some CI stems; often, faculty can overlook those concepts that are most difficult for students. The collection of a wide array of alternative conceptions is crucial for development of good CIs (e.g., Hestenes et al., 1992). Although not sufficient for effective CI development alone, using alternative conceptions that are attractive to the study population is an important first step. Alternative conceptions can be identified through review of the existing research literature, or through collection of qualitative data. Qualitative sources that could be used for identification of alternative conceptions include open-ended exams, interviews with students, focus group discussions, or analysis of homework responses.

Most likely, the alternative conceptions identified from the literature or new research will cover a variety of ideas. For example, alternative conceptions could include ideas related to processes, effects, relationships, definitions, characteristics, impacts on people, and spatial/temporal conditions. Only one of these characteristics should be targeted in each question. The specific characteristic chosen will depend upon the purpose of the assessment, and the expected abilities of the testing population. As stated already, this step should be reflective of assessment needs.

The actual writing of a multiple-choice question involves careful attention to qualitative rules that have developed out of a long tradition of instrument development in fields such as psychology (cf. Haladyna and Downing, 1989; Frey et al., 2005). Questions should be developed based upon existing and research-based standards in multiple-choice question development. Some of these standards are universal, while other rules will depend upon the test construction theory driving the work. For example, identification of coexisting ideas about a single concept might be addressed through use of items with multiple response options ("choose all that apply" items). In classical test theory, items with multiple response options are generally discouraged because they are more difficult and troublesome to score than single response items. Partial credit item response theory models, on the other hand, can easily address scoring concerns related to this type of item; partial credit models have not yet been applied to the GCI and should be the focus of future quantitative work for this and other multiple-choice CIs.

The field of test development provides us with a number of "rules" for writing assessment questions (Fig. 3). These rules significantly increase the likelihood that a question will have satisfactory construct, content, and communication validity. Certainly, exceptions to these rules can be found in almost any validated test, and rules should be viewed as guidelines rather

Rules related to writing STEMS

1. Structure the stem as a question when at all possible. Use: "What is obsidian?", rather than the completion form of "Obsidian is______." If you use a completion form, keep the blank at the end.

2. Use unambiguous and simply worded stems. Use as few sentences as possible.

3. Use appropriate vocabulary. Avoid technical language for nonmajors, for example.

Rules related to writing RESPONSE OPTIONS

1. Use plausible response options. Make sure that the distracters are meaningful to the population being tested.

2. Use 3 to 5 response options. More than five options adds no psychometric value and may produce confusion for the test-taker.

3. Avoid TYPE K format questions. TYPE K: A list of statements is provided, and responses are a combination of statement choices. As in: (a) I; (b) II; (c) III; (d) I and III.

4. Avoid absolutes and complexity in response options. Do not use "All of the Above," "None of the Above," and complex response format (e.g., "a and c, but not b").

5. Keep the lengths of response options similar. The longest or shortest answer is often the correct one. (Anecdote: If you choose all of the longest answers on the Force Concept Inventory, then you will score at the national average.)

Figure 3. Rules for writing multiple-choice questions garnered from existing literature (after Libarkin, 2008).

than strict laws. Rules listed here have been collated from several sources, and a version of each of the rules in Figure 3 appears in Haladyna and Downing (1989) or Frey et al. (2005).

The development of plausible response options requires the careful choosing of alternative conceptions previously identified through literature review or collection and analysis of new qualitative data. In embarking upon a literature review, it is helpful to take advantage of existing compilations that have grouped or distilled the research literature. A comprehensive review of alternative conceptions in some areas and for some populations can be found at the Students' and Teachers' Conceptions and Science Education website (http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html). In addition, thorough literature reviews of alternative conceptions exist in many disciplines. In geoscience, the most recent such reviews (King, 2008; Cheek, 2010) provide a discussion of a range of data collected from multiple populations. Finally, the GCI WebCenter is the home of a compilation of alternative conceptions documented in the research literature.

In addition to published literature, many faculty already have rich qualitative data sets that can be used to generate a deeper understanding of alternative conceptions. For example, faculty can document existing student ideas in a number of ways:

1. They can use data already on hand by considering student responses to assignments (e.g., activities, homework, open-ended exams). Use of these data is exempt from human subjects review under federal guidelines, but will still require approval by an institutional review board as well as informed consent by study participants.
2. They can collect data from a target population by using open-ended questionnaires or conducting interviews. This must be accompanied by approval for this data collection from a human subjects review board before initiating subject recruitment or data collection.
3. Some researchers also recommend the use of a Delphi approach, wherein a group of faculty familiar with the classroom performance of students comes to agreement about the most prevalent alternative conceptions (Landeta, 2006). We note that this approach is not grounded in student experience, and should be used with caution.

Finally, the target population and assessment objectives should always be kept in mind. If researchers are interested in studying third-year geoscience majors, data collected from students in a first-year course might be useful for writing fundamental questions for the third-year students. The reverse (application of data collected from third-year students to studying first-year students) is probably not appropriate unless fundamental ideas have been targeted. For example, while an understanding of the relationship between compasses and Earth's magnetic field is an appropriate target concept for entry-level students, understanding of the source of Earth's magnetic field might not be. The target population for the developed question needs to be considered carefully; alternative conceptions identified in one population may not always be applicable to a different population or subgroup.

Our experience writing GCI questions, as well as preexisting literature from the survey development field, provides insight into steps that should be taken to develop and validate CI questions. Development of concept inventory questions, for the GCI or any CI, should follow a pathway similar to the following (we will discuss steps 1–8 in more detail):

Concept Inventory Question Development

1. Choose a topic.
2. Write a questions stem and a correct response that is as long and detailed as you like. Consult dictionary or other sources (e.g., textbooks) as needed.
3. Write preliminary response options, including the correct option.
    a. Follow rules.
    b. Attempt to coordinate structure and length.
    c. Decide if question will be single option, multiple option, or paired with another question.
4. Fine-tune stem and response options to adhere to and maximize rule use (Fig. 3) and clarity.
5. Revise question based upon informal feedback when possible.
6. Formally send question to content and testing experts (~3–6 experts per question). Experts might be from several of these groups:
    a. 1–2 context experts (e.g., volcanologists);
    b. 1–2 general content experts (e.g., geologists);
    c. 1–2 testing experts (e.g., psychometricians, geoscience educators, education psychologists); and
    d. 1–2 cultural validity experts as appropriate (e.g., if question is going to be used with Native American students, send to partners in those communities for linguistic and conceptual check).
7. Review expert comments. Decide which comments to use and which to argue against.
8. Rewrite stem and response options as needed based on comments.
9. Pilot the questions with testing population.
10. Apply quantitative analysis (see Libarkin and Anderson, 2006).
11. Employ think-alouds with testing population (e.g., introductory students) to further evaluate stem/response interpretation as well as match between actual alternative conceptions and response options.

CASE STUDY

The following case study details the development and revision of a GCI question, and highlights the important roles that qualitative data play in quantitative instrument development. We hope that this case study will provide guidance for scholars interested in understanding the work underlying construction of useful CI questions, and will be helpful to those readers interested in writing their own CI questions. We concentrate on steps 1–8 as outlined here since they represent the relevant qualitative data collection that must precede the piloting of CI questions with the target population. This aspect of CI question development is not something that has been previously published on, especially at this level of detail. The following description applies to a set of questions developed by the second author (hereafter referred to as Geraghty Ward) as an inexperienced question writer.

Topic

A categorization of existing questions appearing on the GCI v. 2.1. (see GCI WebCenter, 2009) indicated that questions related to mountains were needed. In addition, a review of initial entries into an alternative conceptions database indicated multiple documented alternative conceptions related to mountains and volcanoes that were not utilized in existing GCI questions. Four questions related to mountains and volcanoes were developed; this case study primarily focuses on one question related to volcanoes.

Review of Alternative Conceptions

The GCI WebCenter is initiating the development of a database of alternative conceptions relevant to geosciences as described previously. Perusal of this database provided examples of the range of alternative conceptions that students have about volcanoes. Specific question stems evolved from these conceptions, as described in the following sections.

Identify Target Population

Geraghty Ward's experience writing multiple-choice questions was minimal prior to the writing of the question discussed here. Any questions that she had drafted in the past for exams or other classroom purposes were not informed by the test-writing literature or grounded in student misconceptions. Hence, Geraghty Ward first explored the survey development literature, and expended extra effort in composing items with the appropriate structure and language. For example, many of the documented alternative conceptions seemed to focus on the location of volcanoes, and, therefore, "location" was the initial focus of the question stem. Initially, four response options were chosen directly from the alternative conceptions database, and the correct answer was written based upon the definition of a volcano from an online geology dictionary (*Geology Dictionary,* 2009). This definition parallels definitions found in typical introductory level textbooks.

Write Stem (First Try). **Volcanoes are found ______________**

Choose Response Options.

Specific alternative conceptions were chosen from the database:

a. Volcanoes occur on fault lines.
b. Volcanoes are only found in warmer climates.
c. Volcanoes are only found in colder areas.
d. Volcanoes occur wherever there are areas of crustal weakness.

Identify Correct Answer.

e. Volcanoes occur where ash and lava accumulate around a vent (Geology Dictionary, 2009)

The response options were actually revised versions of alternative conceptions documented in the database. These revisions were done according to "the rules" (Fig. 3), taking care to keep the sentiment of the student misconception without necessarily using the student language directly. The distracters are then expected to be attractive options for students because they were rewritten from the documented misconceptions.

Write Preliminary Question

The response option language was modified from the original alternative conceptions. Five response options were included (four distracters and the correct answer) in order to maximize the choices available, and to accommodate the existence of a number of alternative conceptions about volcano location. Answer lengths were calibrated to the content of the response. For example, two of the misconceptions focused on "temperature" of the volcano locale (options b and c), so they were written to include similar language and length. The correct answer and distracter focused on "crustal weakness" and were paired in length, leaving the fifth option (option a) somewhat isolated. Due in part to Geraghty Ward's inexperience in writing CI questions, she had difficulty reconciling the language of the student with the revised language of the responses. This difficulty is a natural part of the CI development process. The construction of test items that are responsive to authentic student ideas requires wording that will be both acceptable to the faculty/researcher and recognizable to the student. While this may seem simple on its face, the process is actually quite difficult, for inexperienced and experienced writers alike. The first draft of the question was sent out for review (see following) with the expectation that reviewer comments would assist in a major revision. The question was distributed internally for review by Geraghty Ward in this form:

Volcanoes are found:

a. Along fault lines where earthquakes cause them to erupt
b. In warmer climates
c. In colder areas
d. Where ash and lava accumulate around a vent
e. Wherever there are areas of crustal weakness

Internal Review

The initial review by a GCI team member (Fig. 4) identified issues with: stem structure, response option content and length, and use of technical language. This initial review did not attempt to rewrite the question, but rather to focus Geraghty Ward on the purpose and overall structure of the question. The review noted that the stem was not worded as a question (rule #1 in Fig. 3) and suggested that the stem read: "*Where are volcanoes found?"* Furthermore, the reviewer highlighted the fact that response option (a) included a "process as well as place" (whereas the other response options only contain a place), and could perhaps influence the student such that she/he would choose it over the others. This response option needed to be rewritten to follow the same structure of the other response options. Response options (d) and (e) contained technical language ("vent" and "crustal") that either an introductory geology student might not fully understand or that may prove attractive to a student simply because they appear to be technical. Finally, the variation in response option length was problematic.

First Revision

In Figure 5, note the revision of the stem to form a question and removal of the process in response option (a). In addition, by removing the language associated with process, the response length more closely matched those of (b) and (c). Terms were changed in response (c) (from "areas" to "climate") to better match the language of response (b). Vocabulary changes from "vent" to "crack" in response (d) and language changes from "crustal weakness" to "weaknesses in the Earth" in response (e) simplified the overall language of both response options:

Where are volcanoes found?

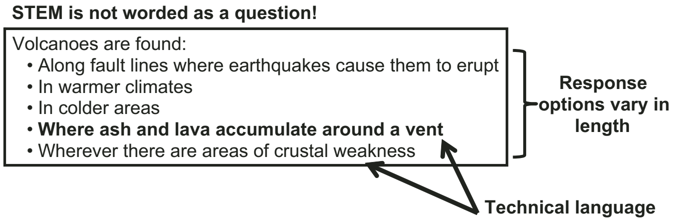
a. Along fault lines
b. In warmer climates
c. In colder climates
d. Where ash and lava accumulate around a crack in the Earth
e. Wherever there are areas of weakness in the Earth

Second Review

Though the question improved greatly after the initial review and revision, difficulty reconciling the wording from the student misconceptions and the wording of the response options still remained. To help with this, the first author (hereafter referred to as Libarkin) suggested that Geraghty Ward paste the exact wording of the documented alternative conception or definition next to the written response item in order to easily compare and contrast the language of the response option with the option's intended focus (Fig. 6).

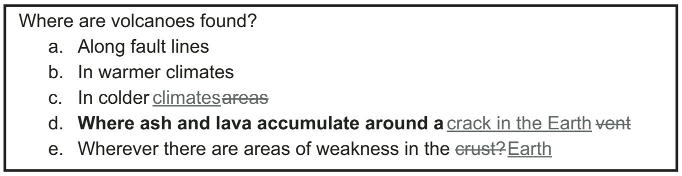
Second Revision

The results from the second review and revision were a change in the layout of the stem and response questions to show both the interpreted and original misconceptions in tabular form. The stem was rewritten to allow students to choose more than one


STEM is not worded as a question!

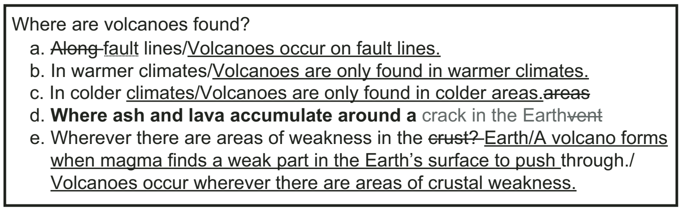
Volcanoes are found:
- Along fault lines where earthquakes cause them to erupt
- In warmer climates
- In colder areas
- **Where ash and lava accumulate around a vent**
- Wherever there are areas of crustal weakness

Figure 4. Initial review of potential Geoscience Concept Inventory (GCI) question by GCI team member.


Where are volcanoes found?

a. Along fault lines
b. In warmer climates
c. In colder climates ~~areas~~
d. **Where ash and lava accumulate around a** crack in the Earth ~~vent~~
e. Wherever there are areas of weakness in the ~~crust?~~ Earth

Figure 5. Initial revision of the Geoscience Concept Inventory question after review. Stem and response option changes are in accordance with the question-writing guidelines outlined in the previous section.


Where are volcanoes found?

a. ~~Along~~ fault lines/Volcanoes occur on fault lines.
b. In warmer climates/Volcanoes are only found in warmer climates.
c. In colder climates/Volcanoes are only found in colder areas. ~~areas~~
d. **Where ash and lava accumulate around a** crack in the Earth ~~vent~~
e. Wherever there are areas of weakness in the ~~crust?~~ Earth/A volcano forms when magma finds a weak part in the Earth's surface to push through./ Volcanoes occur wherever there are areas of crustal weakness.

Figure 6. Second review of potential Geoscience Concept Inventory question. The focus of the second review was to reconcile the response option language with the original misconception, as well as to adjust the response option length.

answer; the question thus became a “multiple response option” type question. By including “Choose all that apply” in the stem, we hypothesize that students who do not fully understand the concept may also choose response option (a), drawing out more than one alternative conception (specifically with regard to the role of fault lines and the formation of volcanoes). Response option (e) was completely removed, primarily because the item language was too difficult to reconcile for this particular question and to collapse the “areas of weakness” meaning into the “fault lines” response. Each of the four response options was modified to become a concisely written response that included only adjectives and nouns. In the final version, two options focus on locales, while the other two focus on features:

Which of the following are needed for a volcano to form? **Choose all that apply**

Response Options	Documented Alternative Conceptions from Literature
A. Fault lines	Volcanoes occur on fault lines.
B. Warm climate	Volcanoes are only found in warmer climates.
C. Cold climate	Volcanoes are only found in colder areas.
D. Melted rocks	Correct answer (lava and magma)

Expert Review

After careful internal review, the question now aligns with the stem and response option rules detailed in Figure 3. This question (along with four others) was sent out to six experts: two volcanologists (one with a geoscience education background), two geochemists (professors at small colleges who teach a variety of geology classes), and two geology professors (both of whom have a geoscience education focus to their research). Experts might include geoscientists, geoscience educators, psychometricians, and cultural validity experts; the choice of expert reviewers should be carefully aligned with the intent of the items. A solicitation letter was sent along with the questions, providing the experts with some guidance for review. These review questions were adapted from Libarkin (2008) and focus on:

1. Construct validity: Is the topic covered by this question important for geosciences understanding?
2. Content validity: From the perspective of an expert geoscientist, does the question actually measure some aspect of geoscience understanding?
3. Communication validity: Would a test-taker interpret this question, including both the stem and the response options, IN THE SAME WAY as intended by the test developer?

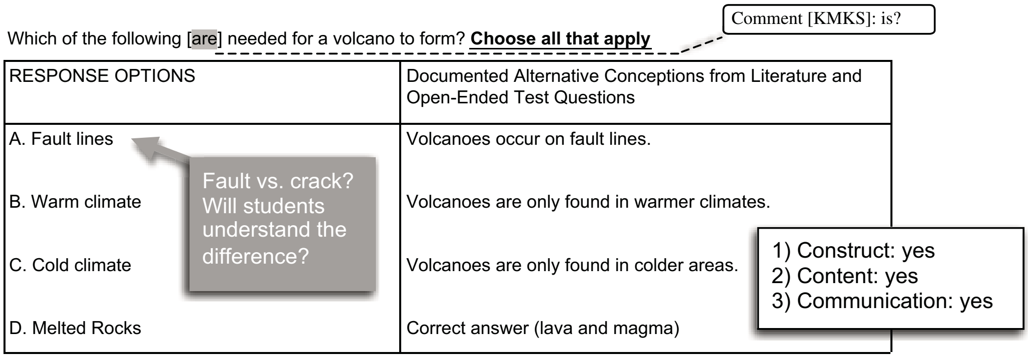
Three reviews were received from the original six requests; a response rate of 50% is greater than expected and likely results from the personal request and low level of effort required for the review of five questions. After reviewing the suggestions from these experts (Fig. 7), this question was actually left in its original form. Arguments agreeing for or against the reviewer comments were devised. The stem deliberately solicits more than one answer so that additional misconceptions can be drawn out from the students. Though the term “fault” might be considered equivalent to “crack” in the mind of some students, Geraghty Ward argued that (based on the language of the original alternative conception) students would interpret “faults” as the cause of earthquakes, and might choose this answer because they believe that earthquakes are necessary for volcanic eruptions to occur. More likely, earthquakes are associated with volcanic eruptions because of the movement of magma through cracks and conduits beneath the surface before the actual onset of an eruption. Future piloting of questions with students, as well as think-alouds, will provide further insight into the validity of this question for targeting actual alternative conceptions and will likely provide foundations for additional revisions.

Note that the expert reviews can be more extensive, and can bring about significant changes to the wording of a potential question (Fig. 8). To illustrate this, we provide expert responses to a different question:

Which of the following can result in the formation of a mountain? **Choose all that apply**

Response Options	Documented Alternative Conceptions from Literature and Open-Ended Test Questions
A. Oceans moving material	Tides wash material away from around the mountain to form it.
B. Air moving material	A mountain is formed when wind deposits material.
C. Rocks moving material	Correct answer (forces acting on plates)
D. Volcanoes moving material	A mountain forms when there is a lot of pressure underneath it and it builds up and explodes.

Expert reviewers noted problems with both communication and content validity, as detailed in Figure 8. Revision of this question is ongoing, and will involve greater thought into the choice of vocabulary and the intended communication. Another iteration of the review-revise-rewrite process will be required after this subsequent revision is completed. This example illustrates the importance of review and revision in item development; item writers must be willing to revise stems and response options, or provide careful arguments against revision. This



Comment [KMKS]: is?

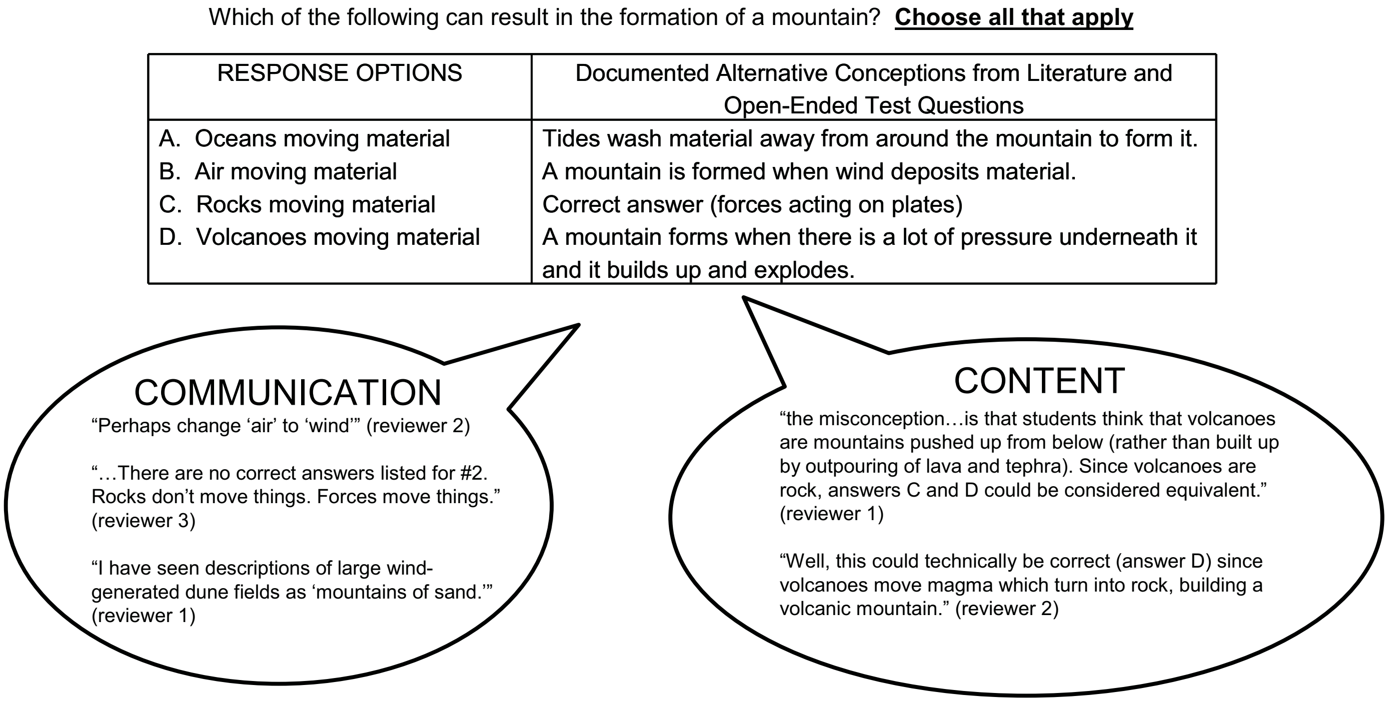
Which of the following [are] needed for a volcano to form? **Choose all that apply**

RESPONSE OPTIONS	Documented Alternative Conceptions from Literature and Open-Ended Test Questions
A. Fault lines	Volcanoes occur on fault lines.
B. Warm climate	Volcanoes are only found in warmer climates.
C. Cold climate	Volcanoes are only found in colder areas.
D. Melted Rocks	Correct answer (lava and magma)

Fault vs. crack? Will students understand the difference?

1) Construct: yes
2) Content: yes
3) Communication: yes

Figure 7. Question markup after expert review. Three of the six experts responded to the solicitation for review and suggested only minor changes to the language of the question.



Which of the following can result in the formation of a mountain? **Choose all that apply**

RESPONSE OPTIONS	Documented Alternative Conceptions from Literature and Open-Ended Test Questions
A. Oceans moving material	Tides wash material away from around the mountain to form it.
B. Air moving material	A mountain is formed when wind deposits material.
C. Rocks moving material	Correct answer (forces acting on plates)
D. Volcanoes moving material	A mountain forms when there is a lot of pressure underneath it and it builds up and explodes.

COMMUNICATION

"Perhaps change 'air' to 'wind'" (reviewer 2)

"…There are no correct answers listed for #2. Rocks don't move things. Forces move things." (reviewer 3)

"I have seen descriptions of large wind-generated dune fields as 'mountains of sand.'" (reviewer 1)

CONTENT

"the misconception…is that students think that volcanoes are mountains pushed up from below (rather than built up by outpouring of lava and tephra). Since volcanoes are rock, answers C and D could be considered equivalent." (reviewer 1)

"Well, this could technically be correct (answer D) since volcanoes move magma which turn into rock, building a volcanic mountain." (reviewer 2)

Figure 8. Expert reviews focused on content and communication validity of a potential Geoscience Concept Inventory question.

review and revision process can occur nearly continuously, and should ideally involve as many members of the user community as possible.

Pilot Testing with Student Population

After completing the review-revise-rewrite process, the question is ready to pilot with students. New questions can be included in student testing with established GCI questions, allowing for comparison with the well-tested questions from the GCI v. 2.1 and subsequent quantitative analysis (cf. Libarkin and Anderson, 2006). The Geoscience Concept Inventory Team has developed the GCI WebCenter (2009) in order to facilitate question submission by the community, GCI subtest creation, and online test administration. The goal of the WebCenter is to expand the content of the GCI question bank and to collect data on student conceptions from a variety of institutions and student populations. With a large and varied data set, researchers can begin to expand their investigation in documenting new alternative conceptions across a more diverse student population.

REFLECTIONS ON THE PROCESS

Geraghty Ward engaged in the development of new GCI questions just after beginning a new position in Libarkin's laboratory. The experience of moving from simple familiarity with the GCI as an assessment tool to becoming a core member of the GCI team provided a unique opportunity for Geraghty Ward to reflect on her experience, and allowed Libarkin to consider the importance of clarifying CI writing steps for new developers. The following reflection is provided in the first person, with explanation of the speaking voice as needed.

Reflections of Geraghty Ward

As a new researcher in the field of geocognition, I am quickly becoming aware of the importance of qualitative research methodologies. Though I have teaching experience at the undergraduate level and have explored the literature regarding teaching and learning strategies, I am new to the area of research regarding student conceptions. As I began working with the Geoscience Concept Inventory, I realized that these questions were much more than a content bank of multiple-choice questions. Before, I regarded multiple-choice questions as a poor test of student learning. The few times I had written multiple choice questions of my own, I did not read up on the test-writing literature to become informed of the proper methodology for question writing. I crafted questions that were laden with geological terms and typically included at least one nonsensical response option for students to easily dismiss. I never asked colleagues to review and comment on the construct, content, or communication validity of my questions. Changes to the questions came about only if students asked me to help clarify or interpret the question for them, alerting me to an issue regarding language use. I trusted that students who did not choose the right answer did so because they had a low level of understanding, and that those who did correctly answer the question did so because they understood the material.

I never realized how multiple-choice exams (if properly constructed) could actually be used to effectively diagnose student misconceptions, inform teaching strategies, and evaluate student learning. Proper construction of the GCI questions involves careful qualitative research into the conceptual understanding of students (through student interviews, surveys, and open-ended exams) and conscientious word choice. However, the most striking realization for me was the rigorous process of review-revision-rewrite that followed the initial construction of each question. Furthermore, I appreciate the involvement of my peers in the review process and am proud of the questions that resulted from that collaboration. The questions, in essence, act as "working models" that are revised and improved through peer review and research. Now, I put more stock into the responses that students choose, right or wrong. The student responses to the GCI questions offer meaningful insight to their conceptual understanding and a starting point for instructors.

Reflections of Libarkin

The Geoscience Concept Inventory has been a core component of my research practice for almost a decade. Although my work has necessarily moved on to other areas of importance in geocognition, I have never viewed the GCI as a completed instrument. Rather, the GCI is a living document, one that should reflect the needs, values, and expectations of the community that benefits from it. We are moving toward a community-developed GCI, one that many scholars can claim ownership of and that is responsive to the feedback of both faculty/researchers and test-takers. Explorations of the research literature in multiple disciplines, initially from science education, general education, psychometrics, and survey design, and later from fields like cognitive science and sociology, have provided me with some unique insights into best practices for CI development. Putting these best practices to work in the GCI WebCenter, however, has proven to be both more challenging and more engaging than I would initially have believed.

My understanding of CI development has become clearer as I have trained Geraghty Ward to develop concept inventory questions and as we worked collectively to provide a detailed accounting of the development process. The level of detail required to effectively explain our CI approach is extensive, as reflected in the case study presented here. I learned that a significant amount of what instrument developers do is hidden, and that in fact very few scholars have enumerated the step-by-step process in a way that is illuminating to novice question writers. This lack of clarity is not unique to CI development, of course; experts in any domain rarely explain underlying assumptions in as much detail as would be necessary to train novices. Interestingly, the collaboration with Geraghty Ward, a geosciences expert training in a new field, offered me an opportunity to reconsider the theoretical foundations of my work and to become explicit about the steps that are necessary for writing multiple-choice questions that are responsive to student ideas. This experience provided an insight into best practices for both CI development and training in the process itself.

Although I would like to say that the GCI was developed in alignment with what I now know about best practice, the emergence of my understanding of best practice was itself developed during the GCI construction, dissemination, and review process. This confirms my impression that CIs are living documents, and that authors should embrace change in response to community input, eventually viewing the community as a co-authors in the development process. The totality of my experience working on the GCI, as well as my most recent experience collaborating with Geraghty Ward, illustrates the importance of being explicit at many stages of CI development. This importance becomes obvious when one considers that developers of CIs should explicitly:

1. Consider the goals of the CI development prior to embarking on the process of investigating student ideas and writing questions. Both the targeted users, the faculty or researchers who will implement the CI in classrooms or research settings, and the targeted participants, the students or research subjects that will be assessed, need to be considered.

2. Be embedded in the survey design literature in education, psychology, sociology, and many other domains. I recommend that anyone embarking on the development of CI questions become familiar with as much of the CI-relevant literature as possible. This exploration should not be limited to a single content area, nor to the disciplinary science education communities from which CI development often emerges. Rather, literature from as many communities as possible should be synthesized and integrated into an overarching theoretical framework that acknowledges the significant effort that already exists in survey development.
3. Seek out the perspective and feedback of community members using the CI. In particular, a best practice approach to CI development would involve community feedback early in the process to ensure that the CI is meeting as many needs as possible. This feedback would be iterative, involving the same type of continual review-response interactions as documented here.

This list of explicit actions is not complete, both because specific steps taken in developing and revising CIs have been discussed before (e.g., Libarkin, 2008), and because we are still learning how to make the GCI more relevant and community-oriented. At the same time, I do believe that the willingness to change, the explicit inclusion of community in the process, and the embedding of CI development in multiple research traditions are all absolutely necessary.

CONCLUSIONS

The importance of qualitative data in CI development has certainly been discussed in previous literature. For example, most CIs incorporate alternative conceptions gleaned from student interviews or open-ended questionnaires into distracters. Validation steps applied to CI development might also include think-alouds, another form of qualitative data that is important for determining student intentions in choosing specific response options. The role of qualitative data in the question writing process, however, has until now not been explored.

The significant amount of qualitative data that should be collected during CI question writing is clearly illustrated by the case study presented here. Certainly, without a careful internal and external review, including subsequent question revision, the questions written by the second author could not be considered valid. This lack of validity would be true regardless of either the grounding of distracters in student alternative conceptions or the belief of the developer that the questions were appropriate.

We encourage all CI developers to carefully consider the way in which questions are developed and reviewed, and to include the entire community of users in both reviewing and developing CI questions. We have found the internal and external review process to be invaluable in revision of questions, and consider the growing number of GCI question writers and reviewers to be authors and co-authors in the process. We anticipate that the next version of the GCI to emerge from this cross-community discourse will be both orders of magnitude more beneficial to the community and a product of the community itself.

ACKNOWLEDGMENTS

We thank the participating question reviewers for their helpful suggestions and willingness to participate in the question development process. Faculty and students affiliated with the Geocognition Research Laboratory are also thanked for internal review of questions and suggestions during this process. We thank J. Sexton and K. Richardson for their help in developing the database of alternative conceptions. This work was funded by the National Science Foundation through grant DUE-0717790. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES CITED

Bond, T.G., and Fox, C.M., 2001, Applying the Rasch Model: Fundamental Measurement in the Human Sciences: Mahwah, New Jersey, Lawrence Erlbaum Associates, 352 p.

Bowling, B.V., Acra, E.E., Wang, L., Myers, M.F., Dean, G.E., and Markle, G.C., 2008, Development and evaluation of a genetics literacy assessment instrument for undergraduates: Genetics, v. 178, p. 15–22, doi:10.1534/genetics.107.079533.

Cheek, K.A., 2010, Commentary: A summary and analysis of twenty-seven years of geoscience conceptions research: Journal of Geoscience Education, v. 58, no. 3, p. 122–134.

Earth Science Literacy, 2009, Earth Science Literacy Principles: The Big Ideas and Supporting Concepts of Earth Science: http://earthscienceliteracy.org/es_literacy_22may09.pdf (accessed 9 November 2009).

Frey, B.B., Petersen, S.E., Edwards, L.M., Pedrotti, J.T., and Peyton, V., 2005, Item-writing rules: Collective wisdom: Teaching and Teacher Education, v. 21, p. 357–364, doi:10.1016/j.tate.2005.01.008.

GCI WebCenter, 2009, GCI WebCenter: http://gci.lite.msu.edu/ (accessed 9 November 2009).

Geology Dictionary, 2009, Geology Dictionary: http://geology.com (accessed August 23, 2009).

Gronlund, N.E., 1993, How to Make Achievement Tests and Assessments (5th ed.): Boston, Massachusetts, Allyn & Bacon, 181 p.

Haladyna, T.M., and Downing, S.M., 1989, Validity of a taxonomy of multiple choice item-writing rules: Applied Measurement in Education, v. 2, p. 51–78, doi:10.1207/s15324818ame0201_4.

Halloun, I.A., and Hestenes, D., 1985, The initial knowledge state of college physics students: American Journal of Physics, v. 53, p. 1043–1055, doi:10.1119/1.14030.

Harlow, A., and Jones, A., 2004, Why students answer TIMSS science test items the way they do: Research in Science Education, v. 34, p. 221–238, doi:10.1023/B:RISE.0000033761.79449.56.

Hestenes, D., Wells, M., and Swackhamer, G., 1992, Force concept inventory: The Physics Teacher, v. 30, p. 141–158, doi:10.1119/1.2343497.

Hufnagel, B., 2002, Development of the Astronomy Diagnostic Test: Astronomy Education Review, v. 1, p. 47, doi:10.3847/AER2001004.

King, C., 2008, Geoscience education: An overview: Studies in Science Education, v. 44, p. 187–222, doi:10.1080/03057260802264289.

Kubinger, K.D., and Gottschall, C.H., 2007, Item difficulty of multiple choice tests dependent on different item response formats—An experiment in fundamental research on psychological assessment: Psychological Science, v. 49, p. 361–374.

Landeta, J., 2006, Current validity of the Delphi method in social sciences: Technological Forecasting and Social Change, v. 73, p. 467–482, doi:10.1016/j.techfore.2005.09.002.

Libarkin, J.C., 2008, Concept inventories in higher education science: National Research Council Promising Practices in Undergraduate STEM Education Workshop 2 (Washington, D.C., 13–14 October 2008): http://www7.nationalacademies.org/bose/PP_Commissioned_Papers.html (accessed 23 August 2009).

Libarkin, J.C., and Anderson, S.W., 2005, Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory: Journal of Geoscience Education, v. 53, p. 394–401.

Libarkin, J.C., and Anderson, S.W., 2006, The Geoscience Concept Inventory: Application of Rasch analysis to concept inventory development in higher education, *in* Liu, X., and Boone, W., eds., Applications of Rasch Measurement in Science Education: Maple Grove, Minnesota, JAM Publishers, p. 45–73.

Libarkin, J.C., and Anderson, S.W., 2007, Development of the Geoscience Concept Inventory, *in* Deeds, D., and Callen, B., eds., Proceedings of the National STEM Assessment Conference: Washington, D.C., Drury University, p. 148–158.

Lindell, R.S., and Olsen, J.P., 2002, Developing the Lunar Phases Concept Inventory, *in* Franklin, S., Marx, J., and Cummings, K., eds., Proceedings of the Physics Education Research Conference 2002, Boise, Idaho, August 2002: PERC Publishing; available at http://www.compadre.org/Repository/document/ServeFile.cfm?ID=4323&DocID=1162 (accessed 29 March 2010).

Liu, X., and McKeough, A., 2005, Developmental growth in students' concept of energy: Analysis of selected items from the TIMSS database: Journal of Research in Science Teaching, v. 42, p. 493–517, doi:10.1002/tea.20060.

Lopez, W., 1996, Communication validity and rating scales: Rasch Measurement Transactions, v. 10, p. 482–483.

Miller, J.D., 2004, Public understanding of, and attitudes toward, scientific research: What we know and what we need to know: Public Understanding of Science (Bristol, England), v. 13, p. 273–294, doi:10.1177/0963662504044908.

Mulford, D.R., and Robinson, W.R., 2002, An inventory for alternate conceptions among first-semester general chemistry students: Journal of Chemical Education, v. 79, p. 739–774, doi:10.1021/ed079p739.

Reed-Rhoads, T., and Imbrie, P.K., 2008, Concept inventories in engineering education: National Research Council Promising Practices in Undergraduate STEM Education Workshop 2 (Washington, D.C., 13–14 October 2008): http://www7.nationalacademies.org/bose/PP_Commissioned_Papers.html (accessed 23 August 2009).

Schau, C., Stevens, J., Dauphinee, T.L., and Del Vecchio, A., 1995, The development and validation of the Survey of Attitudes Toward Statistics: Educational and Psychological Measurement, v. 55, p. 868–875, doi:10.1177/0013164495055005022.

Students' and Teachers' Conceptions and Science Education, 2009, Students' and Teachers' Conceptions and Science Education: http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html (accessed 23 August 2009).

Tan, K.D., Taber, K.S., Goha, N., and Chia, L., 2005, The ionisation energy diagnostic instrument: A two-tier multiple-choice instrument to determine high school students' understanding of ionisation energy: Chemistry Education Research and Practice, v. 6, p. 180–197.

Treagust, D.F., 1986, Evaluating students' misconceptions by means of diagnostic multiple choice items: Research in Science Education, v. 16, p. 199–207, doi:10.1007/BF02356835.

Treagust, D.F., 1988, Development and use of diagnostic tests to evaluate students' misconceptions in science: Journal of Biological Education, v. 10, p. 159–169.

Ward, E.M.G., Libarkin, J.C., Kortemeyer, G., and Raeburn, S., 2010, The Geoscience Concept Inventory WebCenter provides means for student assessment: eLearning Papers, no. 20 (www.elearningpapers.eu).

Zeilik, M., Schau, C., and Mattern, N., 1999, Conceptual astronomy. II. Replicating conceptual gains, probing attitude changes across three semesters: American Journal of Physics, v. 67, p. 923–972, doi:10.1119/1.19151.

Manuscript Accepted by the Society 23 June 2010

The Geological Society of America
Special Paper 474
2011

Ethnographic methods in analysis of place-based geoscience curriculum and pedagogy

Deborah Williams
School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona 85287-2402, USA

Steven Semken
School of Earth and Space Exploration and Center for Research on Education in Science, Mathematics, Engineering, and Technology (CRESMET), Arizona State University, Tempe, Arizona 85287-1404, USA

ABSTRACT

Place-based education is locally situated, experiential, and transdisciplinary. It is informed not only by scientific knowledge of places and regions, but also by the humanistic meanings and affective attachments (senses of place) that people affix to them. Enhanced sense of place is an authentic learning outcome of place-based teaching. Qualitative analyses of a student's behavior and attitudes in a place-based learning context can be used to triangulate instrument-driven psychometry of pre- to postexperience changes in sense of place and content knowledge. Two qualitative ethnographic methods, direct behavioral observation and semistructured interviews, were used formatively and summatively in a Southwest-based earth science course offered to in-service teachers in two underserved rural Arizona school districts in 2006–2007 and 2007–2008. Direct observations were obtained as field notes and video recordings, which were transcribed and coded in an ethogram to ascertain engagement with curriculum and pedagogy. Ethnographic analysis demonstrated increased engagement with place-based course elements over more globally situated components. For interviews, a questionnaire was developed to elicit cognitive and affective responses regarding the course, its curriculum and pedagogy, and the student's sense of the places studied. Verbal, text, and content analyses were applied to the interview data to uncover concepts, patterns, and relationships that were linked into thematic categories. Positive responses to the place-based approach were reported by a majority of participants in three areas: enhanced place attachment and meaning, enhanced science comprehension, and enhanced teaching. These ethnographic methods offer a means to evaluate situated, transdisciplinary teaching for which quantitative instruments may not capture all relevant outcomes.

INTRODUCTION

Place-based education is philosophically rooted in civics and environmental education. While initially practiced mostly in small, rural precollege settings, place-based teaching is now being applied to a much broader range of educational levels and contexts (Sobel, 2004; Gruenewald and Smith, 2008). Geoscience educators are part of this movement, as demonstrated by

Williams, D., and Semken, S., 2011, Ethnographic methods in analysis of place-based geoscience curriculum and pedagogy, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 49–62, doi:10.1130/2011.2474(05). For permission to copy, contact editing@geosociety.org.

a fully subscribed, far-ranging, and well-attended topical session on place-based education at the 2009 Geological Society of America Annual Meeting (GSA, 2009).

Place-based education is not simply experiential teaching, or teaching in the field or community, or the use of local examples of features or processes, although these are all important attributes of the approach. Authentically place-based teaching is fully situated in *place*, which is a social construct that has been defined as any locality imbued with meaning through human experience (Tuan, 1977). Place meanings accrue and evolve as various people and cultures occupy or otherwise interact with a place diachronically and for different reasons. Significant features among these meanings are culturally defined landscapes, heritage, and social values. For example: Grand Canyon, one of the most iconic landscapes of North America, is sacred territory to a number of Native American nations. The canyon and its environs are richly endowed with indigenous place names and stories tied to specific landforms, and populated by many places that are either inhabited or ceremonially visited by Native American peoples. This place was at the heart of one of the last and most storied regional traverses by nineteenth-century American explorers and is a landscape that has encoded and now reveals nearly two billion years of geological history. Grand Canyon has been portrayed in endless works of visual and literary art, has been subject to mining activities, was once threatened with submergence for hydroelectric power generation, and is a beloved National Park visited by millions each year.

From this perspective, scientific knowledge obtained in and of a place is seen as one subset of its meanings, on its face not necessarily more important or relevant to learners—particularly those with personal or cultural connections to the place—than any of the other forms of meaning (Semken, 2005). Ideally, place-based teaching is experiential and transdisciplinary, engaging with the full set of meanings known for the place or places under study, and integrating both scientific and humanistic inquiry into these places. This approach is intended to provide more engaging context and scaffolding for diverse learners (e.g., Kawagley and Barnhardt, 1999; Lim and Calabrese Barton, 2006), to teach global concepts through local examples (Gruenewald and Smith, 2008), to promote environmental and cultural sustainability (Orr, 1992; Sobel, 2004), and to evoke care and concern for places. In practice, a spectrum exists that extends from the use of meaningful places simply to illustrate disciplinary concepts, to a complete integration of disciplines with place as the focus (Ault, 2008). If textbooks and published or online curricula are useful indicators, most geoscience teaching is not far along this continuum.

To fully encompass the transdisciplinary qualities of authentically place-based education, defined learning outcomes should include but also transcend disciplinary knowledge and skills, so that the student's personal connection to place can be leveraged, enhanced, and assessed. This connection involves not only cognitive factors but also affective factors, as people tend to form emotional attachments to meaningful places (Shamai, 1991). The *sense of place*, defined as the set of all of the meanings and affective attachments that an individual or a community maintains for a given place (Brandenburg and Carroll, 1995), usefully represents this connection. This construct has a strong theoretical foundation in geography, environmental psychology, and rural sociology (for a review, see Semken and Butler Freeman, 2008).

Validated quantitative instruments have been developed to empirically measure and characterize sense of place in different groups (e.g., Williams and Vaske, 2003). For these reasons, Semken and Butler Freeman (2008) proposed that leverage and enhancement of the senses of place of students and instructors should be considered a valid learning outcome of place-based teaching and assessed in tandem with traditional measures such as content knowledge.

Previous work on sense of place in place-based geoscience education (Semken and Butler Freeman, 2008; Perkins, 2008; Semken et al., 2009) has been mostly quantitative and focused on (1) characterization of sense of place in diverse student and nonstudent groups, and (2) pre- to postassessment of changes in senses of place of students in experimental place-based undergraduate courses. The results of these psychometric studies show that sense of place can be measured and that significant pre- to postexperience gains in student sense of place have been observed in some classes. To date, this work has been limited by small study populations, but it is ongoing.

ETHNOGRAPHIC RESEARCH IN SCIENCE EDUCATION

Place-based education constitutes a significant philosophical shift in content and pedagogy that requires innovative assessment design. It is to be expected that instrument-driven quantitative characterization and assessment of the method, though valuable, will have limits. Qualitative research methods expand the range of assessment tools available to researchers, to include those that capture attitudes and behavior. Qualitative analyses can and are used to triangulate and refine instrument-driven quantitative assessment.

Qualitative methodologies encompass a wide range of methods, including "ethnography," a naturalistic method of inquiry rooted in both empiricism and humanism (Denzin and Lincoln, 2000; Bernard, 2006). Ethnography is the systematic description of a specific culture and is concerned with individual variation within a particular cultural group (Barfield, 1997). Both during the process and in the product of research, ethnography employs various rigorous quantitative and qualitative methods. The ethnographic methods most often employed in field-based research include observations, interviews, surveys, focus groups, field notes, and text analysis. These methods are by no means restricted to the study of cultures (Barfield, 1997; Handwerker, 2001); other fields of social science also use them to characterize and find causes for specific individual and group behaviors and events (Spradley, 1979; Wolcott, 1987; Handwerker, 2001; Bernard, 2006).

Ethnographic methodology in education research is adapted from the anthropological study of human cultures (Schensul et al., 1999) and investigates teaching and learning from the perspectives of the participants themselves (Wolcott, 1987, 1990, 1994; Cohen et al., 2007). It allows for a multilayered approach to assessment and is especially useful for formatively gauging the attitudes, behavior, and self-efficacy of students and teachers participating in a class or other type of educational intervention.

The project described here employs ethnographic techniques (Wolcott, 1987) to evaluate the perceived efficacy of place-based geoscience learning. As such, the focus was on individual and group behavior, knowledge acquired, and affective responses to the course, rather than cultural analysis or ethnography. Hence, it is most accurately described as an ethnographically informed study. Because place-based education is experiential and contextual, ethnographic methods are especially useful to identify and assess these factors and to uncover any interrelationships that may exist among them. This paper details the application of two ethnographic methods, semistructured interviews and direct behavioral observation, in assessing the place-related efficacy of an experimental place-based earth science enhancement course for in-service teachers.

RESEARCH SETTING

The project was centered in Superior, Arizona, ~80 km east of Phoenix: a small, struggling community situated in a geologically, ecologically, and culturally diverse zone where the Basin and Range and Transition Zone physiographic provinces meet. Fault-block mountains loom immediately east of town, and the steep elevational gradient locally compresses the ecological transition from the lower Sonoran Desert to piñon-juniper-oak woodlands and then to ponderosa pine forests (Ffolliott and Gottfried, 2008). Queen Creek, a major tributary of the Gila River system, originates here and descends through a spectacular canyon to the desert floor at Superior. The mountains consist of Proterozoic to Mesozoic sedimentary and igneous rocks that were extensively mineralized with copper and silver in the mid-Cenozoic (Hammer and Peterson, 1968; Manske and Paul, 2002). The region around Superior was originally part of the ancestral homelands of the Yavapai and Apache peoples, who maintain ties to the area, and is rich in archaeological and cultural sites. Mining began in the 1870s and attracted Mexican miners, whose descendants today comprise most of Superior's population, as well as smaller groups of Eastern Europeans and Chinese.

Like most extant mining towns, Superior has withstood several cycles of boom and bust, but it is currently in a period of severe socioeconomic downturn and population loss initiated by the closing of the last major copper mine in the 1990s. A new, technologically advanced mine has been proposed to tap another copper deposit east of town that is extraordinarily rich but dauntingly deep. The plan and prefeasibility studies for the new mine have stimulated some new economic growth, and both optimism and controversy (Semken and Brandt, 2010).

About half of the participating teachers traveled to Superior each week from the San Carlos Apache reservation located ~60 km to the east. The San Carlos Apache Nation is wholly rural, largely agrarian, and one of the most economically impoverished Native American communities in the United States. As noted previously, most San Carlos Apaches consider the Superior area to be part of their homeland, and many Apache families conduct ceremonial and recreational activities in the highlands east of the town.

STUDY PARTICIPANTS

Place-based science teaching has been identified as the most appropriate and inclusive pedagogy for students who have strong multigenerational cultural ties to their home landscapes (Kawagley and Barnhardt, 1999; Cajete, 2000; Riggs, 2005), such as Native Americans and Mexican Americans in the southwestern United States. These students have historically been underrepresented in geoscience studies and careers (Huntoon and Lane, 2007; Riggs et al., 2007). To investigate the potential and efficacy of this approach, a graduate-level Arizona- and Southwest-based earth science enhancement course was offered to practicing (in-service) teachers from two underserved minority-majority school districts in rural south-central Arizona.

The course was entitled *Situating Earth Science in Superior* (SESS). It was taught by the second author, an ethnogeologist and geoscience education researcher. The first author, an anthropologist, recorded behavior and conducted interviews. The authors were thus situated in the study as researcher-participants.

Study participants were in-service teachers grouped in two cohorts by academic year: 2006–2007 (cohort 1, abbreviated C1) and 2007–2008 (cohort 2, abbreviated C2). Each of the two sessions of SESS took place from late autumn to early spring, when the weather was coolest and most conducive to field trips in the desert. An important characteristic of the SESS experiment, reflective of the nature of place-based education, was that teachers from any grade level and any discipline were welcomed to participate and receive academic credit. The only prerequisites were an interest in learning about the Southwest and intent to apply what they would learn in SESS to their own teaching.

A multicultural mix of study participants (teachers) was present in both courses, which reflected the demographics of the study area. Class sizes were comparable for both cohorts: Cohort C1 had 16 participants and C2 had 15. The 15 teachers in C2 included six who completed SESS in good standing the previous year as members of C1, but returned to retake the course because they had greatly enjoyed and benefited from it. Participants self-reported their race or ethnicity as follows: five (20%) Hispanic/Latino, two (8%) American Indian, two (8%) Asian, and sixteen (64%) White, non-Hispanic. All of the participants were employed either by the Superior school district or the San Carlos school district. The majority of participants were native to the area or long-term residents, but six teachers in C1 and five in C2 had recently been brought in under 2-year contracts from the

Midwestern United States to fill teaching vacancies. All of these individuals were new to the Southwest as of 2006.

COURSE AND CURRICULUM

The SESS course models place-based, inquiry-driven practice while introducing (or reintroducing) area teachers to the surrounding physiography, rocks, structure, geologic history, hydrology, and mineral resources of their surroundings, in the context of their cultural and historical meanings.

SESS is an abridged but more hands-on and experiential version of an Arizona- and Southwest-based introductory geoscience course that was previously developed by the second author for the large-lecture format at Arizona State University. The curriculum for the latter course was described in detail by Semken and Butler Freeman (2008); most of its attributes also apply to SESS and are not elaborated upon here. Geoscientific content is organized under the theme of interacting rock, water, air, and life systems: a concept that bridges mainstream earth system science and indigenous Southwestern ethnoscience (Cajete, 2000; Semken, 2005; Semken and Butler Freeman, 2008).

A "sense of the Southwest" is leveraged and enhanced by frequent evocation (often with imagery, art, quotations, and literary excerpts) of the region's beauty; importance as a home to diverse cultures over millennia; and enduring allure to scientists, artists, and visitors. Some SESS lessons are focused on analytical "case briefs" on regionally situated economic and environmental topics: water and energy resources; copper mining; climate change and drought; and land subsidence and surface fissuring in developed areas. The curriculum also included three optional half-day field trips on Saturdays to nearby outcrops and small mines, to interpret the geological histories of and collect samples from exposed strata. The required texts included an inquiry-based introductory geology textbook (Reynolds et al., 2008), a trade book on Southwestern natural history (Wiewandt and Wilks, 2004), and the geologic highway map of Arizona (Kamilli and Richard, 1998).

In both offerings of the course, 10 three-hour interactive lectures were held over 10 weeks, in addition to the three supplementary field trips. In each session, about 2 hours were devoted to lecture, interspersed with 15- to 20-minute sessions of participant activity (e.g., investigation of local rock specimens, interpretation of maps). The course materials were essentially the same for each cohort, but certain activities (primarily field trips) followed slightly different schedules owing to the availability of the participants and instructor or the vagaries of the weather.

RESEARCH DESIGN

The research design was structured to emphasize ethnographic methodologies. Two ethnographic methods, behavioral observation and semistructured interviews, were used in this study. Behavioral observation is a mixed methodology: both scientific and humanistic (Bernard, 2006). This type of observation uncovers the actual behavioral response in a particular situation or event, yet requires interpretive coding of the data collected. Semistructured interviews feature a written set of closed and open-ended questions that shape the discussion but do not constrain the participant's response. This allows for in-depth discussion of the material presented, methodologies used, and the participant's response to it, as well as opening up new avenues of study.

Behavioral Observation

Funding and logistical constraints limited behavioral observation to the first year (C1) only. This method was utilized to ascertain engagement with curriculum elements and place-based method. The observation took two forms: participant observation during each class session, and direct behavioral observation via time-sampling (scans) of video recordings of each class. The study participants (teachers) were fully informed of the recording prior to the study, and all signed informed consent forms, though one individual did not wish to be visible on any videotapes. This was accommodated through the positioning of the camera, which placed that participant in acoustic but not visual range.

Behavior was recorded in situ via participant observation using field notes. This type of ethnographic method, focused on participant activity, can be sensitive to an individual's place meaning and place attachment because behavior is often place-specific (Bernard, 2006). Field notes detailed the behaviors of individuals and groups, and provided analytic descriptions of the research setting, structural elements, curriculum, activities, and characteristics of individuals and their behaviors. Field notes enabled the researcher to document the context of behavior and to describe the patterns and interrelationships among individuals, phenomena, and behavior. For example, it was noted that during class particular students sat up, leaned forward, and began asking questions of the instructor when geological processes were illustrated by reference to specific local places. This behavior would pull other, less engaged students into the discussion. Another student exhibited discomfort, via bodily movements and facial expressions, whenever geologic time was discussed.

This type of observation allows the researchers to assess interactions within the particular setting as well as with the material presented far more accurately than self-reported behavior (Bernard, 2006). This method is limited, however, by the constraints of the researcher's ability to observe and record multiple participants and hidden multiple meanings within behavior, to which the researcher may not have access. Hence, interviews were conducted, and video recordings of each class were also made in order to offset the limitations of the field notes.

Direct behavioral observation of individuals in a research setting is often done by means of an ethogram. Ethograms are a form of continuous monitoring that have been most often associated with primate studies, and used to catalogue specific, discrete behaviors (usually innate; Bernard, 2006). However, they have been increasingly used with humans, to record behavioral patterns and facial expressions, in such fields as behavioral psychology, anthropology, and architecture (Kneidinger et al., 2001; Bernard, 2006). An ethogram is a list of specific behaviors that are recorded focally (on a single individual) or as scans (on multiple individuals) to elicit the range, frequency, and description of behaviors. In this case, behaviors were identified through the preliminary scanning of multiple videotapes and field notes. An ethogram (Fig. 1 is a sample excerpt) was developed from preliminary scanning of the videotapes, full transcripts of the recordings, and comparative analysis of field notes. This list of behaviors was then evaluated and translated into a set of specific, easily observed behaviors to be monitored and recorded for each taped session. This permits quantitative and qualitative coding and assessment of observable behavior patterns, which may be difficult to assess otherwise (Bernard, 2006).

		Time (min)		
	Value	**4**	**8**	**12**
Body position				
Leaned forward	+	0	4	6
Relaxed	~	11	7	4
Slumped	–	0	0	0
Turned forward	+	10	11	11
Turned away	–	1	0	0
Seated	~	11	11	10
Standing	~	0	0	1
Walking	~	0	0	0
Gaze				
Forward	+	7	10	5
Away/D/ On-topic	+	2	1	6
Away/D/ Off-topic	–	2	0	0
Away/ND	–	0	0	0
		Time (min)		
	Value	**0**	**4**	**7**
Verbalization				
Teacher-D	+	0	3	4
Other-D/ On-topic	+	0	1	3
Other-D/ Off-topic	-	0	0	
Localized	+		4	7
Global	~			
Content	+			
Method	+			
Off-topic	–			
Writing				
	+	2	1	0
Use of materials				
Hand lens				X
Rocks				X
Handouts				
Movie				
Lesson content				
Global		X		
Regional				
Local			X	X

Positive behaviors (indicated by **+** in **Value** column):

Posture:	Body turned toward teacher, screen
	Body leaning forward
Gaze:	Turned directed (D) forward toward teacher, screen
	Turned away but on-topic (handouts, rocks, etc.)
Verbalization:	Directed toward teacher/on-topic
	Other directed/on-topic (classmates)

Negative behaviors (indicated by **–** in **Value** column):

Posture:	Body turned away from teacher, screen
	Body in a slumped position
Gaze:	Directed away/off-topic (other objects/persons)
	Nondirected (ND) gaze (wandering/glazed)
Verbalization:	Directed to others/off-topic
	Teacher directed/off-topic

Neutral behaviors (indicated by **~** in **Value** column):

Posture:	Relaxed
	Standing, walking (dependent upon activity in room)
Gaze:	None identified
Verbalization:	Responses to direct questions without a switch in topic

Figure 1. A sample ethogram for engaged classroom behavior, showing classification and tally of behaviors made from observation of a video recording.

Interviews and Questionnaire

Semistructured exit interviews were conducted with each class participant in both C1 and C2, following the completion of SESS, with the exception of one participant in C1 who moved out of state before he could be contacted. Interviews were held as closely as possible to the final class session in May, and most were completed within 2 weeks of that time.

The questionnaire (Table 1) was designed to elicit both affective and cognitive responses regarding the class itself and the participants' sense of place. Fifteen questions were asked of each participant, half of which were open-ended. Affective items were grouped into evaluations of the class and responses regarding the area around Superior and San Carlos (place) and the greater Southwest region. Cognitive items were centered on pedagogy and the place-based approach. These included: (1) assessments of the class content, focus, and structure, (2) strengths and weaknesses of the course, and (3) the participant's own plans for implementation of place-based teaching. The first three questions elicited residence and career patterns, including length of residence, family residence history, and the participant's career plans. These data were used to establish a baseline residence pattern. Activity patterns and affective responses to place and residence were elicited in four additional open-ended queries. Three additional open-ended questions were designed to elicit place attachment and meaning both locally and regionally. The remainder of the queries addressed course strengths and weaknesses, and the respondent's general feelings about the course.

DATA COLLECTION AND ANALYSIS

Direct Behavioral Observation

The number of participants made it impossible to record behaviors continuously for each participant. Time-sampling (also called instantaneous spot sampling) allowed for accurate recording at specific intervals to ensure a representative sample of behavior. Each scan lasted 1 in, in which the behaviors of each visible participant were recorded, and was repeated at 4 min intervals throughout the session. The content of the lecture material, or discussion, was also recorded at each interval as either global or place-based. Videotaped behavior was coded twice: initially in the summer after the course, and then after a period of 3 months using the same parameters that were established prior to coding. The coded behaviors were then enumerated, and frequency counts were tabulated. The layout of the classroom was such that only about two-thirds of the participants could be video recorded at any one time. Since the participants generally placed themselves in the same location each class, the subjects were held constant. Scans were recorded manually and included lecture and demonstrations but not student activities. This was in part due to the difficulty of accurate assessment. Participants in these activities were coalesced into dense group settings and too far away for accurate audio and video recording. In addition, the experiential nature of the activities created a contextual setting that could be confused with place-based emphasis and thus muddy results.

TABLE 1. QUESTIONNAIRE USED FOR EXIT INTERVIEWS

1. How long have you lived in the Superior/San Carlos area?
2. Do you have family living in the area? If so, who are they and how long have they been here?
3. Do you plan to continue teaching here for at least five more years? If not, where do you plan to go?
4. What do you think about the Southwest-based focus of the course?
5. How did what you learned in this course affect how you think and feel about the Southwest?
6. In what ways will what you have learned in this course impact your teaching in the future?
7. What two things about this course were the most useful or most successful, in your opinion?
8. Which two things about or in this course most need to be improved, and how would you recommend doing this?
9. How much time do you spend in the Superior area?
10. If you had a choice, would you continue living in this area?
11. If you plan to leave the area, would you return? Why or why not?
12. What types of activities do you participate in that are specific to this area?
13. Describe your feelings about this area? The Southwest in general?
14. How would you characterize your relationship, or how would you describe your feelings about the Superior area? With the Southwest?
15. What is or are the most important feature(s) or characteristic(s) of this area for you? That you think students should learn?
16. How has your view of the Southwest changed as a result of taking the course last year? Please be specific.*
17. How have you implemented what you learned or practiced in the course last year in your own teaching? Please be specific.*

*Items 16 and 17 were administered only to cohort C1 teachers who returned in C2.

Observed behaviors were categorized and coded as either states or events. States are ongoing time-dependent behaviors, while events are punctuated or short-term behaviors that intervene within a state. Three general groups of behavior were documented: body position, gaze, and verbalizations. These were further broken down into discrete, easily identifiable units in order to log them accurately. For example, body position states were classified as seated, standing, or walking. Body position events were identified as turned forward, turned away, leaning forward, relaxed, or slumped. Gaze was treated as an event (because of its punctuated nature) and was categorized by focus and position. Four types of gaze were identified: forward, away but directed on-topic, away and directed off-topic, and away but non-directed. Verbalizations were initially categorized as teacher-directed, other-directed but on-topic, and other-directed but off-topic. In addition, the content of verbalizations was coded as being local (place-based), global, or off-topic. This allowed the researcher to distinguish interest in the material presented by the instructor from interest directed elsewhere (toward other participants, personal materials, or unfocused and unengaged). Behaviors were then classified as positive, negative, and neutral. Positive behaviors were those that exhibited a direct connection to the instructor or material presented (e.g., leaning forward, speaking to the instructor or a participant about the material presented, a change in gaze-focus to the instructor, etc.). Negative behaviors were those which indicated that the participant's focus was directed at something or someone other than the instructor or the material being presented (e.g., turned away, talking on a cell phone, gazing into space). Neutral behaviors were those behaviors expected in a class setting or those that were ambiguous (e.g., sitting in a chair and turned forward, relaxed posture, changing gaze). The results of the observational studies were then compared to field notes taken in situ.

Verbalizations in the Classroom Setting

In the context of the recording of behaviors, it became apparent that student verbalization offered some tantalizing clues to engagement, interest, and efficacy. It was therefore decided to re-code the raw data and analyze both student and instructor verbalizations within the classroom setting, via continuous monitoring (rather than in 4 min intervals), in an attempt to elicit new information. This enabled the researchers to obtain both quantitative and qualitative data and allowed for more nuanced analysis. Participants were divided into two categories: instructor and students. Verbal episodes, defined as vocalization by a participant with a clearly understood meaning, were identified as the primary units of analysis. A verbal episode could manifest as either a single vocalization or a verbal interaction occurring between the instructor and a student or students.

Coding categories were developed inductively from participant observation, field notes, and interviews, and then applied systematically (LeCompte and Preissle, 1993). A subset in the behavior pattern of verbal episodes was revealed that appeared to indicate increased interest: uninitiated student responses. Student categories were thus made more specific to capture both initiated and uninitiated verbal episodes. Each verbal episode was therefore further categorized as being either instructor-initiated (IL—instructor lecture; IRP—instructor response; or IQ—instructor question) or student-initiated (SRP—student response or SQ—student question). This allowed for the identification of 10 interactive forms, depending upon which participant initiated the exchange, as well as the presence of uninitiated responses: (1) IL-SRP, (2)IL-SQ, (3) IRP-SQ, (4) IRP-SRP, (5) IQ-SRP, (6) IQ-SQ, (7) SRP-IRP, (8) SRP-IQ, (9) SQ-IRP, and (10) SQ-IQ. Instructor lecturing was held to be the baseline behavior, which removed verbal exchanges that resulted in the resumption of the lecture from consideration.

Each verbal episode, whether a single or interactive vocalization, was subsequently coded as being either global (e.g., related to worldwide plate tectonic processes) or local (place-based; e.g., related to a tectonic episode in Arizona) in content. Uninitiated student responses were then identified and tabulated to determine the global or place-based content of these verbal episodes. Frequency counts were then tabulated for all verbal episodes (as student responses and uninitiated responses), and relative frequency was calculated for each of the categories for each class session. This allowed for episodic differences in general student responses by date. Student responses and questions were then matched against instructor content to ascertain any "switches" in content, global to local or local to global (switch in focus). The actual verbalizations within the domains of "local" (place-based) and "global" (Table 2) were then analyzed via text analysis and coded. These descriptive data yielded characteristics, patterns, and properties contained within the domains, helped to refine our hypotheses, and stimulated new questions.

Questionnaire and Interviews

A multilayered qualitative approach was applied to the formation of the interview schedule and the analysis of the resulting data, encompassing verbal analysis, text analysis, and content analysis. Verbal analysis as outlined by Chi (1997) allows concepts and relationships to emerge from the texts in a form that can be easily quantified. Patterns can be elicited, and these can be grouped into thematic units that allow for further investigation (Chi, 1997). Text analysis is hermeneutic in nature; it focuses on the meanings within texts and their interconnections (Bernard, 2006).

Content analysis is a deductive coding method often concerned with determining form, substance, or trends (Berelson, 1971; Titscher et al., 2000; Bernard, 2006). This research strategy can be combined with other techniques to interpret data from any form of text (written, verbal, or nonverbal). It proceeds from the assumptions that valid inferences can be made between intent and content and content and effect, and that this study is meaningful (Berelson, 1971). Analysis may take several paths depending on the approach utilized. The research questions themselves dictated

TABLE 2. CATEGORIZATION OF PARTICIPANT VERBALIZATIONS

	Participant responses			Participant switches in focus			Uninitiated responses		
Lesson dates	Global	Local	Total	Local to global	Global to local	Total	Global	Local	Total
4 December 2006	57	29	87	0	0	0	5	7	12
11 December 2006	109	191	300	0	0	0	4	5	9
18 December 2006	56	17	73	2	2	4	7	7	14
8 January 2007	15	6	21	1	1	2	5	3	8
29 January 2007	54	32	86	0	1	1	7	10	17
5 February 2007	44	17	61	2	4	6	8	11	19
26 February 2007	69	16	85	0	1	1	6	5	11
5 March 2007	43	40	83	2	2	4	8	16	24
Totals	447	348	796	7	19	26	50	64	114
Percentages	56	44	100	27	73	100	44	56	100

whether a more quantitative or qualitative approach was used and hence the forms of data analysis that best fit the research.

Interviews were transcribed, and an initial reading was done to ascertain categories of interest and units of analysis. All interviews were completely coded and included in the analysis. Initial coding of analytic units was done by two research assistants (a science-education graduate student and a psychology undergraduate student), using coding categories agreed upon among the research assistants and the first author. Analytic units were of two types: recording and contextual (Titscher et al., 2000). Recording units are small textual units that indicate definite categories or meanings and hence are usually one-word or a phrase in length (e.g., "yes," "It hasn't changed," "beautiful"). *Contextual units* are multisentence and thematic (e.g., "360 degree panoramas that take your breath away," "brings up the story of our culture"); these explain and elaborate on the initial response. Patterns in the results were noted, and contextual coding and analysis were captured by the first author. These were dependent upon the source material and research questions.

The cognitive and closed (fixed-choice) question section of the questionnaire was designed to facilitate coding by dyadic or occasionally triadic responses. These responses took the form of yes/no, more or less than five years residence, will affect/won't affect, and so on, and were determined to be recording units. This facilitated coding of the responses because categories were easily identified and were explicit. Thus, in this section, category formation preceded analysis (Sepstrup, 1981; Titscher et al., 2000). These questions also allowed for elaboration upon the initial response, however, as each response also included a phrase, sentence, or paragraph in the participant's own words. This created larger contextual units of analysis. Category formation in this type of textual unit proceeds from the text itself and is implicit (Berelson, 1971; Titscher et al., 2000; Bernard, 2006). Open-ended and affective questions were broken down into categories as the texts presented them. These generally fell within three to four themes per question. Responses were then coded into these categories using phrases or sentences that clearly elicited the meaning as conveyed by the participant. Patterns were detected in this initial coding and subsequently quantified. The responses were then qualitatively reevaluated and coded according to the patterns and themes detected. This allowed for re-categorization into more specific and meaningful categories and also into general thematic units, which were then quantified to ascertain the results presented here.

For example, when participants were queried as to the possible impact the place-based SESS course might have on their own teaching, initial coding identified two categories: *will affect* or *won't affect* (recording unit). The elaborated responses (contextual units), however, revealed additional patterns within this dyad. Participants identified changes in the focus of their teaching, the methodology they planned to utilize, or the content of their instruction. These categories could have been grouped under the overarching theme of enhanced teaching, and quantified. However, the coded categories exhibited a richer meaning.

Participants were also asked to describe how the course affected their feelings for the Superior–San Carlos area (question 5), and, later, in a separate question (13), what their feelings for the area in general were. The first question was initially coded according to the themes that emerged from the responses. These were a greater understanding of the geology of the area, an appreciation for the area, a desire for more information, and no change. The second question was coded initially for emotional, descriptive, or not applicable responses. Affective responses might include "I feel connected to this place" or "It's kind of depressing…" Descriptive items included, "It has its own beauty," or "It's beautiful." Themes were identified within these responses and in the responses as a whole. The open-ended nature of question 13 allowed for determination of baseline affective meanings, which were then cross-checked with the responses to question 5. These two questions were then re-coded. Other items in the questionnaire also elicited responses that fell into these categories. The structure of the questions allowed participants to explain how and why they answered as they did. For example, question 15, concerning the most important feature or characteristic of the area, elicited a response of "Superstition Mountains! I have loved the Superstition Mountains since my parents first came here. And I call it my mountain."

RESULTS

Interviews

The results obtained from the interview portion of the research were grouped into three thematic categories: enhanced place attachment and meaning, enhanced science comprehension, and enhanced teaching. Response totals for each category are presented in Table 3.

Of the 15 participants in C1, 53% reported a positive affective result in place attachment or place meaning. A positive result was reported by 87% of the participants in C2. Deconstructing the larger contextual responses and then recoding them enabled the researchers to identify larger thematic domains underlying the responses. These included an increase in appreciation, connection, and sense of place, as well as an increase in understanding of and a more personal relationship with place. One participant noted "…you look at these mountains, you see them but, you don't know anything about them…It makes me appreciate them more." Inquiries into place-specific activities revealed an engagement with place, specifically mountains ("This gave me an adventurous trend of exploring the area"), but also activities that plumbed the rich historical and cultural senses of the area: "…our history is here" and "it's our culture and all our traditions connecting us." A participant who took the class twice put it most evocatively, "Last year it was like I was awakened from a sleep. Not being familiar with what was out here…When I took [the second author's] class I was just like 'Wow!'…I look at the area differently now. I knew the names of the rocks." Another noted, "It's a whole different new world that's opened up."

Cognitive results were itemized by science content and focus (as exemplified by the thematic, inquiry-based curriculum). Participants in both cohorts reported a significant increase in place-related scientific knowledge. Many participants specifically noted

TABLE 3. RESULTS OBTAINED FROM SEMISTRUCTURED INTERVIEWS

	Cohort 1 (N = 15)	Cohort 2 (N = 15)
Enhanced place attachment and meaning	8	13
Increased appreciation	7	10
Increased connection	3	7
Increased understanding	5	6
Deeper sense of place	1	1
Personal relationship	6	7
Enhanced science comprehension	10	6
Focus (inquiry, scenario-based)	2	1
Content (depth, breadth of knowledge)	9	5
Enhanced teaching	10	13
Focus (professionalism, place)	2	8
Method	5	3
Content	8	9
Implementation		
Concurrent	11	13
Intended	11	8
Actual	3	5
Subsequent		6
Actual	N.A.*	4

*N.A.—not applicable.

the volcanic activity in the area, and regional rock-forming and geomorphic processes, as valued knowledge gains. C1 documented a 67% increase while 40% of C2 participants reported increased science comprehension. This was overwhelmingly based on content rather than focus. As one participant put it, "…you're kind of tunnel-vision and now I look at how the layers are there and these cuts in the roads or a peak sticking up—looking at how it's starting to disintegrate and break down and how the canyons come together." However, for some, it was the place-specific inquiry process that they felt led them to a better understanding of the material.

Enhanced teaching was very positively reported: 67% of C1 participants and 87% of C2 participants noted gains in focus, methodology, or content. Content and focus were responsible for over half of the gains, while gains in methodology were noted by less than one-fourth of the participants. Teacher-participant comments on implementing place-based content and pedagogy within their own classroom were also elicited. These were categorized as concurrent implementation, for those teachers who implemented content or pedagogy in the same year they took SESS, and subsequent implementation, to capture comments from C1 teachers who returned in C2. These two categories were each further subdivided into intended implementation and actual implementation. Eighty-seven percent of all participants stated that they would implement either knowledge gained or procedures learned within the course. Participants spoke about the need for students to understand what rocks, landforms, and resources were present in the area, and how these developed: "They need to be aware of the aquifers and how the Southwest developed, where they are at…. They are not really aware. They just take what they see for granted."

However, only 20% of C1 participants and 33% of C2 participants reported actual implementation at the time of their interviews. These were most often those participants who taught some form of science. One science teacher reported "I just did my project on a whole earth science unit. Everything I learned was applied into what my project is going to do." Another participant who taught both geoscience and social studies implemented a lesson on plate tectonics and volcanoes in the Southwest. However, of the six C1 participants who returned in C2, four had implemented and planned to continue implementing place-based content. Half of these were not science teachers. For example, one participant who taught English at San Carlos incorporated specific content and a change in focus in her composition classes: "I had the students write about peridot in their compositions or about Point of Pines, where they like to go fishing, camping, hunting, or all that. What they write is just beautiful. I also shared Apache Leap [a culturally significant tuff ridge that looms over Superior] with the students. Lots of them weren't aware of what happened there." Another English teacher not only assigned place-specific books to be read but often brought in locally collected specimens (rocks, plants, found objects) and had the students write about these. Those who indicated that they did not implement content or pedagogy were not teaching in the classroom at the time of the interviews.

Direct Behavioral Observation

Direct behavioral observation supported the results found in the interviews. Field notes and the ethogram showed that behaviors indicative of increased interest or engagement with the material presented were documented in every class session. Differences were noted in the behavioral characteristics of the participants when they were presented with place-based content versus global content. These differences consisted of changes in body posture to a forward-leaning position, movement of the visual gaze to one which was oriented directly at the instructor, and teacher-directed verbalizations that centered upon place-based content. Coding gaze and verbalization into discrete directed behavior (i.e., teacher-directed or other-directed, but on-topic) captured a broader range of engagement. Participants might not be focused upon the teacher when specific place-based material was discussed but instead might be commenting about the example to a fellow student or looking at rock specimens. In fact, participant-participant (other-directed) comments were most frequent during place-based discussion, and these were consistently on-topic, often with some personal connection to the place mentioned. While positive behaviors were always greater during place-based material, the class periods that were predominantly place-based showed large gains in both the intensity and vibrancy of verbalizations and gaze, as well as in sheer numbers of all positive behaviors across the board. While enumerating and analyzing the behaviors yielded important data on engagement and interest, it was the qualitative analysis of the behaviors that provided even more understanding of place-based engagement. Individual participants might lean forward casually or with great vigor. Participant comments might demonstrate a process of integrating the material with the place-based example or might instead offer cultural insights and personal connections to the given place. While the sample size was too small to allow for statistical significance, these observations coupled with interview analyses provide impetus for further ethnographic and mixed-methods study of place-based or similarly situated (e.g., problem-based) teaching and learning.

Verbalizations

Qualitative (text and content) analysis of the verbalizations demonstrated intriguing results. Student responses appeared dependent to some degree on the material presented; global content generally elicited student responses that were also global in content and vice versa. However, student responses demonstrated switches in content. These switches were overwhelmingly global to local (place-based) in focus: ~73% of the time (Table 3). An example of this type of shift or switch occurred during a lecture centered upon crystallization processes. A participant interrupted to ask if the stalactites and stalagmites in Carlsbad Caverns (in southeast New Mexico) were formed by these same processes. In another episode, several participants responded to discussions of volcanism with queries and anecdotes concerning the nearby Superstition Mountains caldera complex.

The majority of uninitiated responses, which suggest a strong participant engagement, were also found to be place-based in character. These responses varied in nature, exhibiting connections from global processes to specific place-based examples (e.g., is there any evidence of glaciation in the desert landscapes in Arizona?), querying information concerning specific places (e.g., was Picket Post Mountain—an igneous butte that looms directly west of Superior—made of lava or ash?), or revealing personal experiences with places. Personal experiences often referred to visits or activities within a particular place, but occasionally revealed specific cultural knowledge. During an inquiry exercise to explore how impact craters form, a Native American participant linked the crater and impact processes under discussion to traditional knowledge. This participant revealed that traditional narratives held that a particular crater (which was not identified or located) was an emergence site from a previous world for her people, and wondered if this idea could be listed as a hypothesis to be tested. Thus, while the absolute number of responses may have remained global in character for a particular class episode, teacher-participants regularly volunteered responses that shifted the focus to place-based content.

Validity and Reliability

Variation exists on the cognitive, emotional, and behavioral scales of study participants, even those who appear to be members of the same culture (Spradley, 1979; Wolcott, 1994; Handwerker, 2001). This is magnified with diverse cultural populations. A dialectical approach, using multiple methods to document results, tacking back and forth between methods, and triangulating results, helps the researchers ensure validity and reliability.

The study population was known in advance to be a multicultural group of participants. The researchers attempted to minimize selection bias by including all the participants in the participant observation and interview portions of the study. The interviews were constructed to help identify cultural differences so these could be taken into account. Structured questions enabled collection of data that historically and culturally situated the participants, whereas open-ended questions allowed for characterization of individual life experiences that may have affected behavior and attitude. All interviews were conducted within the same time frame and in the same general physical setting (the school at which the participant taught), except for one participant. Content and verbal analyses of these interviews provided statistical measures that allowed for comparisons between individuals while text analysis allowed for an in-depth interpretation of participant meanings. Multiresearcher coding, agreement upon categories, and re-coding of contextual variables all enhanced reliability in the analysis of interview data.

Some selection bias in field studies is natural (and sometimes desired; Bernard, 2006) since movement catches the eye more than stationary behavior, louder verbal statements are more readily heard than whispers, and so on. This is the reason why field notes were supplemented with direct behavioral observation in this study. This procedure had several advantages. While field notes were recorded initially at the time of the behavior, observational studies were done approximately 4 months later. This allowed for the synthesis of initial findings and a review of conceptual categories and coding. Observational markers (distinct behaviors) were identified and tested separately. These were then applied universally to all taped participants. Coding and re-coding of raw data after a time lag helped establish reliability in the coding of behaviors, and subsequent refined coding allowed for a more nuanced analysis. While the physical setting for data collection constrained the viewing to only two-thirds of the participants at any given time, the recording camera was placed to maximize the number and diversity of participants (and hence, the data). This created a consistent set of participants for observation and incidentally excluded those participants whose attendance was less consistent.

While field notes identify behavior, subtle interactions, and meanings within the context in which they occur, evaluating behavior with an ethogram focuses attention on discrete segments of behavior isolated from the larger context. Parsing participant behavior into discrete, identifiable actions minimizes researcher subjectivity (Bernard, 2006). These actions could then be re-contextualized after curriculum materials and verbal content were identified and recorded for reference.

Observation does need to account for the random fluctuations and individual variation that occur in the naturally occurring rhythm of time and personal lives. On any given day, participants may have been fatigued, ill, or distracted by personal issues. These factors, as well as personality differences, would affect engagement with any material. However, an attempt to minimize this was made by coding multiple participants and coding each class session, creating continuity. At the same time, behavior without context, focus, and interpretation tells us little. Field notes permitted interpretation of behaviors and understanding of participant opinions, personalities, and histories. As familiarity increased, personality traits became more evident, and this enabled the researchers to distinguish between a reserved but attentive participant and one who was more outwardly enthusiastic.

DISCUSSION

Findings

Ethnographic analyses applied in this study show that in-service teacher-participants in a Southwest place-based earth science course were actively engaged in the material presented. This was documented by behavioral observation and by text and content analysis of verbal episodes. Moreover, participants made repeated connections between the material presented and specific places and experiences within those places. This was demonstrated by switches in verbal episode content from global to local (or place-based), and increased verbalizations and engagement seen in behavioral analyses. The interviews supported these findings. Participants self-reported gains in engagement, citing cultural ties, emotional bonds, the scenic beauty, and the physical familiarity with specific places as factors. Cultural ties included family residence, ancestral heritage, and community involvement. Responses in these cases took the form of: "My mom and dad live here," "…it's the culture and our traditions connecting us…the sunrise dances," "…there are places I've experienced…," or "As you grow older you tend to realize that this is where you belong." Emotional bonds were frequently noted by references to home and connections; for example: "Superior is my home! I've only been here a little while but this is my home," and "…what I want out of life is here."

Appreciation of the aesthetic beauty of the study region was often focused on mountains and sunsets, described using such phrases as: "It's exotic and unique," "The beauty of the place. It focuses and catches your attention." Physical familiarity also played a part in engagement. Participants could readily examine and identify examples of geological processes. This made it easier to understand the material but also heightened self-reported engagement; e.g., "If you've been here all your life, you pretty much know the area," "…now I look at it in terms of its geological form," "The idea of being able to introduce some of these kids, because they're Reservation kids, to a part of their home that they're not even aware of." These same interviews added depth to observed behavior. One participant who frequently appeared uncomfortable or unengaged stated, "Actually, it didn't change how I felt about the area much…Home is where the heart is, and my heart is not here." Another participant mentioned, "I don't really care about this area that much. It's pretty, but just a place to visit."

In addition, most participants self-reported gains in knowledge. The exit interviews showed these gains to be perceived as gains in the depth of content knowledge ("I can go deeper with my students"), in better appreciation of the surroundings ("I have more appreciation of the landscape…that's an enriching thing"), the acquisition of new skills and methods ("…actually working with a geologic map," "…for me that's a better way to learn, is scenario-based"), a greater sense of place ("I do have more of a sense of place…"), a change in perception or organization ("Using the [place-specific] information to organize your perception, so that it's richer…"), and in cultural grounding ("How different people think…even though it's the same place," "And the addition of the cultural [content]…").

Recommendations and Future Work

A growing body of literature documents the need to reposition and reinvigorate mainstream science teaching to better engage students and teachers through meaning, relevance, and participation (e.g., Barab and Roth, 2006; Tytler, 2007; McWilliam et al., 2008). The highly and locally contextualized and transdisciplinary nature of place-based education is well suited to this use (Ault, 2008; Gruenewald and Smith, 2008). However, authentic and comprehensive learning outcomes for place-based education are complex, overlapping the cognitive and affective domains, and possibly the psychomotor domain as well (Semken and Butler Freeman, 2008). This is a cultural shift in metacognitive teaching that necessitates a culturally informed approach to assessment.

As discussed previously, pre- to postexperience changes in sense of place, attitude, self-efficacy, and content knowledge can be measured, and recent results (admittedly still limited by small sample sizes) favor the continued use of such quantitative tools for assessment (Semken and Butler Freeman, 2007, 2008; Semken et al., 2009). However, these findings do not show how participants engaged formatively with curriculum and pedagogy, nor elucidate their interest in and satisfaction with the approach. These types of data are best ascertained through qualitative ethnographic methods such as the two demonstrated in this study. These methods provide a window into the mix of factors that underlies student behavior, and enables triangulation of quantitative results such as pre- to postexperience gains.

Ethnographic methods and analyses offer many advantages to geoscience education (both in assessment and in learning research), particularly in highly situated or transdisciplinary contexts in which quantitative tools are not sufficient to capture the full range of authentic learning outcomes. Their design is holistic and focused on relationships and patterns within the structure and distribution of events over time, and on an understanding of the social setting in which behavior occurs (Denzin and Lincoln, 2000). Geoscience educators can use these methods to pinpoint realms of tension as well as effective concurrence.

The results of this study also posed new questions for further consideration: Is place-based teaching effective for topics and subjects less locally or regionally situated than geoscience? Does it have lasting impact on K–12 curricula and teacher retention, especially in those teachers who are new to the place(s) studied? Do affirmative behaviors and responses determined through ethnographic analysis correlate with quantitative evidence of improvement in knowledge and skills obtained through place-based learning? These questions can be addressed through continued application of ethnographic methods in qualitative or mixed-methods studies of larger populations over longer times.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under award number GEO-0706653 (to Semken) and EHR-0412537 (to Center for Research on Education in Science, Mathematics, Engineering, and Technology [CRESMET], Arizona State University). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Logistical support was provided by CRESMET. Tuition costs and course materials for teachers participating in *Situating Earth Science in Superior* (SESS) were funded by an award from the Resolution Copper Mining Company, which is gratefully acknowledged. The authors also wish to express their deep appreciation to Marilyn Carlson, Kristine Wilcox, and Ginger Romaine of CRESMET; to research assistants Nievita Bueno Watts and Monica Pineda; to the Superior and San Carlos School Districts; and especially to all the enthusiastic teacher participants in this project.

REFERENCES CITED

Ault, C.R., 2008, Achieving querencia: Integrating a sense of place with disciplined thinking: Curriculum Inquiry, v. 38, p. 605–637, doi: 10.1111/j.1467-873X.2008.00438.x.

Barab, S.A., and Roth, W.-M., 2006, Curriculum-based ecosystems: Supporting knowledge from an ecological perspective: Educational Researcher, v. 35, p. 3–13, doi: 10.3102/0013189X035005003.

Barfield, T., 1997, The Dictionary of Anthropology: Malden, Massachusetts, Blackwell Publishers, 626 p.

Berelson, B., 1971, Analysis of Communication Content: Content Analysis in Communication Research: New York, Hafner Publications, 220 p.

Bernard, H.R., 2006, Research Methods in Anthropology: Qualitative and Quantitative Approaches (4th ed.): Lanham, Maryland, AltaMira Press, 585 p.

Brandenburg, A.M., and Carroll, M.S., 1995, Your place or mine?: The effect of place creation on environmental values and landscape meanings: Society & Natural Resources, v. 8, p. 381–398, doi: 10.1080/08941929509380931.

Cajete, G., 2000, Native Science: Natural Laws of Interdependence: Santa Fe, New Mexico, Clear Light Publishers, 315 p.

Chi, M.T.H., 1997, Quantifying qualitative analyses of verbal data: A practical guide: Journal of the Learning Sciences, v. 6, p. 271–315, doi: 10.1207/s15327809jls0603_1.

Cohen, L., Manion, L., and Morrison, K., 2007, Research Methods in Education (6th ed.): London, Routledge Publishers, 328 p.

Denzin, N., and Lincoln, Y., 2000, Handbook of Qualitative Research (2nd ed.): Thousand Oaks, California, Sage Publications, 643 p.

Ffolliott, P.F., and Gottfried, G.J., 2008, Plant communities and associations, *in* Ffolliott, P.F., and Davis, O.K., eds., Natural Environments of Arizona: From Deserts to Mountains: Tucson, University of Arizona Press, p. 70–119.

Geological Society of America (GSA), 2009, Earth science in place-based teaching: Geological Society of America Abstracts with Programs, v. 41, http://gsa.confex.com/gsa/2009AM/finalprogram/session_24204.htm (accessed 22 April 2010).

Gruenewald, D.A., and Smith, G.A., eds., 2008, Place-Based Education in the Global Age: Local Diversity: New York, Lawrence Erlbaum Associates, 377 p.

Hammer, D.F., and Peterson, D.W., 1968, Geology of the Magma Mine area, Arizona, *in* Ridge, J.D., ed., Ore Deposits in the United States, Volume II: New York, American Institute of Mining, Metallurgical and Petroleum Engineers, p. 1282–1310.

Handwerker, W.P., 2001, Quick Ethnography: Lanham, Maryland, AltaMira Press, 291 p.

Huntoon, J.E., and Lane, M.J., 2007, Diversity of the geosciences and successful strategies for increasing diversity: Journal of Geoscience Education, v. 55, p. 447–457.

Kamilli, R.J., and Richard, S.M., eds., 1998, Geologic Highway Map of Arizona: Arizona Geological Society and Arizona Geological Survey Map M-33, scale 1:1,000,000, 1 sheet.

Kawagley, A.O., and Barnhardt, R., 1999, Education indigenous to place: Western science meets Native reality, *in* Smith, G.A., and Williams, D.R., eds., Ecological Education in Action: On Weaving Education, Culture, and the Environment: Albany, State University of New York Press, p. 117–140.

Kneidinger, L., Maple, T., and Tross, S., 2001, Touching behavior in sport: Functional components, analysis of sex differences, and ethnological considerations: Journal of Nonverbal Behavior, v. 25, p. 43–62, doi: 10.1023/A:1006785107778.

LeCompte, M.D., and Preissle, J., 1993, Ethnography and Qualitative Design in Educational Research (2nd ed.): San Diego, California, Academic Press, 425 p.

Lim, M., and Calabrese Barton, A., 2006, Science learning and a sense of place in an urban middle school: Cultural Studies of Science Education, v. 1, p. 107–142, doi: 10.1007/s11422-005-9002-9.

Manske, S.L., and Paul, A.H., 2002, Geology of a major new porphyry copper center in the Superior (Pioneer) District, Arizona: Economic Geology and the Bulletin of the Society of Economic Geologists, v. 97, p. 197–220.

McWilliam, E., Poronnik, P., and Taylor, P.G., 2008, Redesigning science pedagogy: Reversing the flight from science: Journal of Science Education and Technology, v. 17, p. 226–235, doi: 10.1007/s10956-008-9092-8.

Orr, D.W., 1992, Ecological Literacy: Education and the Transition to a Postmodern World: Albany, State University of New York Press, 210 p.

Perkins, T., 2008, Place Attachment in Geology Students and the General Public [M.S. thesis]: Tempe, Arizona State University, 58 p.

Reynolds, S.J., Johnson, J.K., Kelly, M.M., Morin, P.J., and Carter, C.M., 2008, Exploring Geology: Boston, McGraw-Hill Higher Education, 589 p.

Riggs, E.M., 2005, Field-based education and indigenous knowledge: Essential components of geoscience education for Native American communities: Science Education, v. 89, p. 296–313, doi: 10.1002/sce.20032.

Riggs, E.M., Robbins, E., and Darner, R., 2007, Sharing the land: Attracting Native American students to the geosciences: Journal of Geoscience Education, v. 55, p. 478–485.

Schensul, S.L., Schensul, J.J., and LeCompte, M.D., 1999, Essential Ethnographic Methods: Observations, Interviews, and Questionnaires: Walnut Creek, California, AltaMira Press, 318 p.

Semken, S., 2005, Sense of place and place-based introductory geoscience teaching for American Indian and Alaska Native undergraduates: Journal of Geoscience Education, v. 53, p. 149–157.

Semken, S., and Brandt, E., 2010, Implications of sense of place and place-based education for ecological integrity and cultural sustainability in contested places, *in* Tippins, D., Mueller, M., van Eijck, M., and Adams, J., eds., Cultural Studies and Environmentalism: The Confluence of Ecojustice, Place-Based (Science) Education, and Indigenous Knowledge Systems: New York, Springer, p. 287–302.

Semken, S., and Butler Freeman, C.L., 2007, Cognitive and affective outcomes of a Southwest place-based approach to teaching introductory geoscience, *in* Proceedings of the National Association for Research in Science Teaching: New Orleans, Louisiana, National Association for Research in Science Teaching, p. 1–10.

Semken, S., and Butler Freeman, C., 2008, Sense of place in the practice and assessment of place-based science teaching: Science Education, v. 92, p. 1042–1057, doi: 10.1002/sce.20279.

Semken, S., Butler Freeman, C., Bueno Watts, N., Neakrase, J., Dial, R., and Baker, D., 2009, Factors that influence sense of place as a learning outcome of place-based geoscience teaching: Electronic Journal of Science Education, v. 13, p. 136–159.

Sepstrup, P., 1981, Methodological developments in content analysis, *in* Rosengren, K.E., ed., Advances in Content Analysis: Beverly Hills, California, Sage Publications, p. 133–158.

Shamai, S., 1991, Sense of place: An empirical measurement: Geoforum, v. 22, p. 347–358, doi: 10.1016/0016-7185(91)90017-K.

Sobel, D., 2004, Place-Based Education: Connecting Classrooms and Communities: Great Barrington, Massachusetts, The Orion Society, 96 p.

Spradley, J.P., 1979, The Ethnographic Interview: Belmont, California, Wadsworth Group, 247 p.

Titscher, S., Meyer, M., Wodak, R., and Vetter, E., 2000, Methods of Text and Discourse Analysis: London, Sage Publications, 278 p.

Tuan, Y.-F., 1977, Space and Place: The Perspective of Experience: Minneapolis, University of Minnesota Press, 235 p.

Tytler, R., 2007, Enriching science education, *in* Glascodine, C., ed., Re-Imagining Science Education: Engaging Students in Science for Australia's Future: Camberwell, Victoria, Australian Council for Educational Research, p. 17–31.

Wiewandt, T., and Wilks, M., 2004, The Southwest Inside Out: An Illustrated Guide to the Land and Its History (2nd ed.): Tucson, Arizona, Wild Horizons Publishing, 207 p.

Williams, D.R., and Vaske, J.J., 2003, The measurement of place attachment: Validity and generalizability of a psychometric approach: Forest Science, v. 49, p. 830–840.

Wolcott, H.F., 1987, On ethnographic intent, *in* Spindler, G., and Spindler, L., eds., Interpretative Ethnography of Education: Hillsdale, New Jersey, Lawrence Erlbaum Associates, p. 37–57.

Wolcott, H.F., 1990, Making a study "more ethnographic": Journal of Contemporary Ethnography, v. 19, p. 44–72, doi: 10.1177/089124190019001003.

Wolcott, H.F., 1994, Transforming Qualitative Data: Description, Analysis, and Interpretation: Thousand Oaks, California, Sage Publications, 426 p.

MANUSCRIPT ACCEPTED BY THE SOCIETY 23 JUNE 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

Developing a process model for visual penetrative ability

Matthew Alles*
Department of Geological Sciences, San Diego State University, San Diego, California 92182-1020, USA

Eric M. Riggs[†]
Department of Earth and Atmospheric Sciences and Center for Research and Engagement in Science and Mathematics Education, Purdue University, West Lafayette, Indiana 47907-2051, USA

ABSTRACT

Visual penetrative ability (VPA) is a basic skill for any geology undergraduate student and is also required in many introductory laboratory exercises designed for nonmajors. This is the ability to visualize the three-dimensional (3-D) underground structure of folded sedimentary rocks from two-dimensional (2-D) surface clues. This study seeks to understand the origins of difficulties that introductory geology students encounter with this task. Problem-solving interviews were conducted with students representing the range of performance on the GeoSAT, a semiquantitative instrument designed to measure VPA. We conducted think-aloud, discursive interviews where participants solved similar 3-D visualization tasks. The interviewers interacted with participants to probe their problem-solving difficulties and thought processes while they were working. Analysis of interviews and videotapes of student gestures yielded insight into the nature of the difficulties faced by students in solving this style of spatial problem, and explained the origin of many of the common incorrect responses seen by previous workers using the GeoSAT. Students with high VPA appear to rapidly construct a 3-D internal visual model, and readily produced gestures and physical expressions illustrating their spatial understanding. Students with poor VPA tend to view external information as merely a type of "gift wrapping" over the cubic volume and do not perceive the internal structure. They also do not typically address spatial concepts through physical expression. We construct a process model for VPA that describes the origin of commonly observed errors at crucial steps, and associated accommodation strategies used by students struggling with spatial visualization of this type.

INTRODUCTION

For a field geologist to completely understand and solve three-dimensional (3-D) spatial problems, on paper or in the field, it is often necessary to investigate a structure's internal attributes with the mind's eye. Because sedimentary and tectonic processes such as deposition, deformation, and erosion typically form geometrically predictable features, the trained eye can predict basic buried structural patterns from surface information. Geologists depend on the ability to visually penetrate a 3-D structure and

---

*Current address: Vinje & Middleton Engineering, Inc, Escondido, California 92029, USA.
[†]emriggs@purdue.edu

Alles, M., and Riggs, E.M., 2011, Developing a process model for visual penetrative ability, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 63–80, doi:10.1130/2011.2474(06). For permission to copy, contact editing@geosociety.org.

form mental images in order to envision and construct cross sections and predict the location of buried structures and layers.

Early descriptions of spatial abilities, (e.g., Linn and Peterson, 1985; McGee, 1979; Piaget and Inhelder, 1956) define spatial visualization as the ability to create a mental image from a "pictorially presented object" for the purpose of performing different mental manipulations on those images such as rotations and translations. Many discrete skills are included in the array of spatial skills required to be successful in geology, especially in solving structural problems. Kastens and Ishikawa (2006) outlined many of these requisite skills, which include object rotation, folding and unfolding skills, navigation, map reading, and visual penetrative ability (VPA), and Orion et al. (1997) suggested that spatial visualization can be improved by earth science courses, potentially because it is one of the few school sciences that employs them routinely. Spatial visualization in many forms appears to be related to aspects of geologic reasoning, particularly in the construction of physical or spatial analogies (Jee et al., 2010), and in the functioning of working memory (Shipley, 2009).

Background—The Skill of Visual Penetrative Ability

Visual penetrative ability, or VPA, was defined by Kali and Orion (1996) and is the ability to visualize the structure of a geologic object from surface clues and project elements of that structure into the interior of a block or underground to make predictions about the location of individual elements at any arbitrary location or cross section of that volume. By contrast with geologic experts, visualization of this type can often be a challenging task for many introductory geology students. Students with low VPA are at a disadvantage in many areas of geology in comparison with those with a high VPA or those who appear to be "naturals" at this type of visualization task. The importance of teaching and advancing the application of 3-D spatial abilities rests in recognizing that many introductory students' poor VPA performance hinders their overall understanding of many basic geologic concepts. Trying to teach students who do not have this ability or those who find it difficult to apply has become a significant hurdle for many instructors, which motivates continued investigation into the nature of this ability and also instructional strategies to improve VPA.

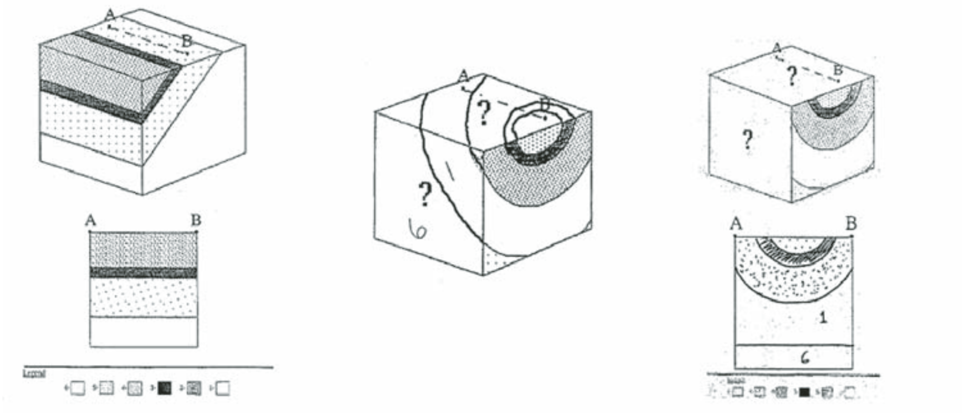
Work on visual penetrative ability is still rather limited, but a few groups are starting to build on the original work of Kali and Orion (1996). These workers studied the spatial abilities of Israeli high school geology students. Their approach led to the development of the GeoSAT, a field-specific and unique spatial ability instrument designed to reveal the internal cognition used by students during 3-D skill application. They also developed a set of codes based in recurring common responses to this instrument from students. These codes are shown in Figure 1. They interpreted that the codes show varying degrees of visual penetrative ability, from nonpenetrative responses (codes 1–3, shown in Fig. 1) where students only respond to surface information but do not project into the subsurface, to slightly and fully penetrative responses (coded as 4–6, show in Fig. 1).

Other workers discussed later herein have used this instrument in slightly modified form, or with tasks from this instrument replicated by other workers in a new setting. The tasks in the GeoSAT are very similar to many such block-diagram tasks in existence across the introductory geology laboratory curriculum worldwide. Because the original instrument was already semi-quantitative in its nature (i.e., depending on subjective coding), strict calculation of reliability was problematic, although inter-rater reliability and test-retest reliability and validity (geologic realism of the tasks) was well established in Kali and Orion (1996). The *Hidden Earth Project* (Piburn et al., 2005; Reynolds et al., 2002, 2006) yielded insight into VPA through the use of computer-generated images and interactive cross-section block models. The use of computer technology in the laboratory allowed researchers to test the similarities and differences between traditional nontechnical laboratory exercises and computer-aided learning and application. Titus and Horsman (2009) constructed new tasks and test items based on the original GeoSAT to make them more geologically challenging for geology undergraduate majors, and demonstrated the development of VPA skills from introductory to advanced undergraduate students. They showed that students given the opportunity to engage in practice at VPA tasks in a classroom setting later showed improved visualization performance when tested on similar tasks. They also found that this practice eliminated or reduced gender differences in scores. However, none of these studies has yet focused on the mechanism through which introductory geology students approach or solve (or fail to solve) problems of this nature with a focus on the underlying cognition. The aim of the present study is to address this remaining gap in our understanding of this important spatial ability in basic geology education.

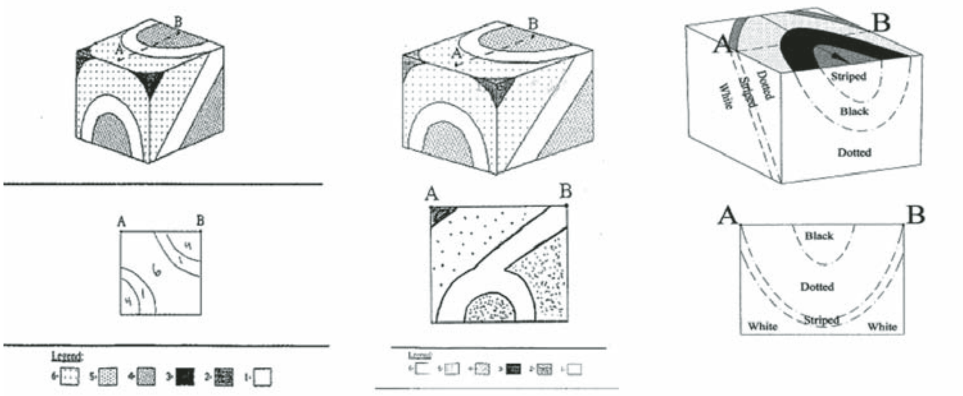
Locating the Study

The purpose of this study was to observe, test, record, and characterize the 3-D spatial abilities of participating students enrolled in college-level introductory geology laboratory courses. Our goal was to extend the work of Kali and Orion (1996) to a college-level group of nonmajors, a short extension beyond the high school audience they worked with in the content-area specialized Israeli high school curriculum. The goal for the qualitative portion of the project, described in detail in the following, was to highlight observed thought processes and the physical steps employed by nonmajors while engaged in VPA activities, to look beyond the coded final products on the GeoSAT instrument and understand how students arrived at these answers and why and where they specifically encountered visualization problems in the process.

Alles (2006), a larger study from which our data are derived, consisted of four introductory geology laboratory sections, within which a total of 37 students (evenly distributed between the sections) consented to be a part of the study. This study was



1. Face Copy/Mirror Image **2. Line Continuation** **3. Half-Face Copy**



4. Unfolding **5. Face Combination** **6. Correct Solution**

Figure 1. Samples of codes used for visual penetrative ability (VPA) question responses in this study. All faces with question marks were originally blank, as were cross sections, so all drawings in these were provided by students, with the exception of code 6, where two variants on the correct answer are supplied. These were derived directly from Kali and Orion (1996) and the original GeoSAT, and we coded our students' responses on these tasks to be as consistent with their interpretations as possible. Kali and Orion (1996) interpret codes 1 through 3 to be nonpenetrative answers, where codes 4 through 6 show partial to complete visual penetrative ability. A code of 0 was given for no response. The code numbers also became the basis for semiquantitative scoring of the GeoSAT instrument, as codes are totaled across 15 items.

reviewed and approved by the San Diego State University Institutional Review Board for compliance with all human subject research policies. All groups were given the standard laboratory exercise that requires VPA in the execution of block diagram cross sections, typical to many introductory geology laboratories at the undergraduate level. The control groups received no instructional resources or special instruction beyond the standard laboratory that had been in place for years, utilizing folding paper block models and modeling clay as instructional aids. The experimental groups were given a laboratory exercise utilizing a shortened experience with the Geo3D educational software program developed for and tested by Kali and Orion (1996). In both cases, the entire instructional experience with 3-D block diagrams and cross sections lasted only one 3 h laboratory period.

We used the same GeoSAT instrument as Kali and Orion to establish relative VPA ability, translating the instructions from its original Hebrew-language form to an English-language version. Being a nonlinguistic task, the instruction set appearing in Hebrew or English should make little difference as to the performance of visualization tasks. This issue of preverbal or nonverbal aspects of this study is discussed in the section on theoretical frameworks and assertions and methodology. The tasks used in the Alles (2006) study were the same as those presented in Kali and Orion's original study. The quantitative results showed a systematic oversaturation of the instrument based on the large number of correct responses, which were on the order of 50+%, and roughly reproduced the U-shaped distributions seen in Kali and Orion (1996). However, this work found no significant pre- to post-test changes in these two populations, and in fact shows very little change in visual penetrative ability in either group regardless of which treatment they received. This result contrasts strongly with the significant performance change documented in Kali and Orion (1996). This lack of performance change is attributed to the short duration of our intervention (one class period), relative to the very long intervention (~3 mo) in Kali and Orion (1996).

Despite the negligible performance gains seen in the original study, the study design provided an opportunity for categorization of participating students by VPA ability, based on a simple 0–6 coding designation (Fig. 1) following the scheme used in Kali and Orion (1996). The coding scheme covers the full range of errors seen in student work. The resulting coded items from each student's completed instrument were aggregated and tabulated, and a generalized rank of overall VPA competence was assigned to each applicable student. Kali and Orion (1996) argued that the low-numbered codes in their instrument correspond to absent or low visual penetrative ability. Higher codes represent progressively more penetrative responses up to fully penetrative and fully correct responses. We made the decision to produce an aggregate score for students in the study based on adding together the code values for each item. Our reasoning was that because the low number codes are interpreted to be poorly penetrative responses, low aggregate scores reflect a preponderance of these low codes, and generally lower VPA. The reverse was true at the high end.

Using the summation of the coded student scores, participants were organized into three tiers based on ability. We broke the students into ranked groups, with somewhat arbitrary numerical breaks assigned where the average response on items would have been nonpenetrative on the whole, medium, or generally highly penetrative. The categories were assigned as follows: 0–40 equals a low VPA student, 41–60 equals a moderate VPA student, and 61–90 yields a high VPA student. Because the GeoSAT is only a loosely quantitative instrument via use of categorical coding, on a stand-alone basis, these data did not characterize VPA responses. However, the instrument—as rough as these results are quantitatively—created a means for purposeful sampling of our larger student population to guide interview participant recruitment and subsequent analysis of student interview responses by demonstrated ability. The summary data for the eight participants interviewed for this study are presented in Table 1.

Problem Statement

Our work in this study was largely directed at understanding the origin of the many common visualization errors and codes shown in Kali and Orion (1996). This work was motivated particularly by the fact that many of these codes as originally reported could stem from nonunique sets of problems in visualization abilities and drafting and representation or realization of graphical solutions on the instrument. This study uses qualitative methods rooted in observation of student performance, gesture, and language to yield a more complete picture of the visualization skill that is VPA, with the goal of producing a complete process model of the ways in which students complete tasks of this nature and generate insight into spatial abilities in this area. The longer-range goal of this work is to understand how to better teach this skill and work with student difficulties in this area.

Locating the Researchers

Both of the researchers in this study are geologists by education and are experienced at conducting geological fieldwork in structurally complex areas. Both have advanced skills in VPA (“naturals” as described earlier), having had those skills in place since the earliest memories of geoscience instruction. Both authors have experience as instructors at the introductory, nonmajors course level, with the second author having years of experience as a classroom and field structural geology instructor up to the graduate level. Neither researcher, however, was involved as an instructor with the experimental or control populations in this study in any way. As described in our theoretical and methodological approach to this study (see following), our own skill sets and backgrounds were instrumental in the structure and style of our student participant interviews and in our analysis and interpretation of their words and actions.

TABLE 1. SUMMARY TABLE OF INTERVIEW PARTICIPANTS

Student name	Approximate time (min)	Interviewer	Pre- and post-test scores	Overall VPA
Cathy	22	E	29/53	L
Tom	14	E	89/89	H
Dan	20	M	79/74	M
Sandra	25	M	80/79	M–H
Cindy & Kelly*	30	E	70/65 & 68/59	M
Vini	27	M	87/87	H
Annie	16	M	39/66	L–M
Carl	15	M	81/61	M

Notes: This table identifies the interviewer (E—Dr. Eric Riggs, M—Matthew Alles) and includes the length of time of each interview and the pseudonym of the participant interviewed. Each student’s overall visual penetrative ability (VPA) (general category of low, medium, or high) is shown next to their actual pre- and post-test GeoSAT scores. The theoretical maximum on this instrument as structured for this study is 90 (15 items, each coded as a 6, fully penetrative and correct answer). All students were nongeology majors.

*Cindy and Kelly were identical twins who chose to be interviewed together and were in the same laboratory section.

Theoretical Framework

The skill under investigation in this study exists within the nonverbal or preverbal domain of human abilities, which places the theoretical basis of investigation in this area in a range of philosophical and methodological traditions spanning educational research, cognitive science research, and linguistics and semiotics investigations. Because we were interested both in how students make mental models in spatial visualization, and how they construct and work to express the meaning of their models, we place this work at the intersection of the qualitative research theoretical frameworks of hermeneutics and models and modeling. The hermeneutical tradition asserts that understanding and meaning are rooted in specific contexts, that understanding is a mediated process between a student and the object of study, that learning and understanding are mediated through the mode of communication used to express understanding, and that interpretations, reactions, and observations of researchers involved in conducting the study of meaning making are integral to the research process and are acknowledged and reported (Patton, 2002; Shane, 2007). This research tradition intertwines the researcher in the reporting and interpreting of observed actions and verbalized portions of interviews through the lens of their own experience and position in the study, and argues against the separation of observation and interpretation as philosophically untenable in the study of the construction and expression of meaning (Patton, 2002).

Models and modeling is a closely related tradition in qualitative research that is especially useful as a methodological framework, because participants are asked to reveal how they are thinking about a given problem or situation (Briggs, 2007). This forms the basis of the design of research settings and the analysis of student actions and operations. This is closely linked to the notion of “substantive theory” (Schwandt, 2007, p. 292–293), which is a localized theory focused on specific behavioral theory, rather than an overarching philosophical or epistemological theory. In this case, the specific behavioral theory is focused around the interplay of gesture formation during spatial reasoning and linguistic (i.e., verbal) expression of that same reasoning.

By adopting a hermeneutical framework, tempered by a focus on models and modeling, we have to establish that gestures and gesturing actually, in fact, have meaning that is interpretable in the context of spatial (and in this case geologic) reasoning, and they are not merely incidental bodily motion completely disconnected from the spatial/cognitive reasoning and discourse about that reasoning. Fortunately, this is established and well-documented in the field of cognitive science (both laboratory and naturalistic research) exploring “embodied cognition” (for a review, see Wilson, 2002), which developed the substantive theory that there is a strong connection between spatial thinking, gesture, action, and language referenced to spatial tasks and concepts in any area of human problem solving and communication. These conclusions from educational and cognitive research have also been supported by neuropsychological research (for a review, see Rizzolatti et al., 1997). Philosophically, embodied cognition crosses the theoretical foundations of constructivism, phenomenology, hermeneutics, and probably others, but for our purposes, with the desire to focus on how people perform this specific spatial visualization task, we have chosen to focus on the relatively localized behaviors that relate to spatial cognition.

Many workers have documented that gesture is very common in communication during spatial cognition or about spatial concepts, and in fact is likely an integral part of spatial cognition itself (Alibali, 2005; Alibali et al., 2000; Wagner Cook and Tanenhaus, 2009), and specifically in reasoning about spatial concepts in scientific and mathematical problem solving (Garber and Goldin-Meadow, 2002; Givry and Roth, 2006; Roth, 2001, 2002; Wagner et al., 2004). It has also been argued that gesture and physical actions may represent a critical step in the offloading of cognition to the body as an aid to working memory (Wagner Cook and Goldin-Meadow, 2006; Wilson, 2001), suggesting that people who are successful at spatial tasks and/or visualization are also likely to use gesture to reason and communicate about these ideas.

The use of gesture as an aid, proxy, or accompaniment to spatial reasoning has also been documented in the geosciences. Kastens and coworkers (Kastens et al., 2008a) showed that gestures are involved in reasoning about many geological concepts by students in the field and classroom, and that instruction about spatially extended concepts is also enhanced by gesture, especially when the gestures used by instructors are physically parsimonious or analogous to the concept under discussion. This is an extension of the work that argues for a strong role of analogy in geologic reasoning (Jee et al., 2010; Sibley, 2009), specifically the physical analogy presented by gesture that ultimately becomes an explanatory, or epistemic action (Kastens et al., 2008b), which in and of itself aids in teaching, learning, and cognition about three- and four-dimensional geoscience concepts.

Methodology

Using the GeoSAT results, interview questions were designed to highlight commonly observed errors or physical actions employed by the student, and to capture the real-time problem-solving approach used by each student.

We conducted semistructured, discursive, problem-solving interviews (as described in Kvale, 2007) designed to provide additional insight into some of the coded responses given by the participants on the GeoSAT. Eight participants were selected based on their GeoSAT scores, and were distributed by gender and ability (see Table 1). No students were geoscience majors, and all reported that they had never before encountered a geologic cross-section task of this nature in any prior coursework.

In keeping with a model-building and hermeneutic theoretical and methodological approach to this content area, we designed interview tasks where the researchers played an active role and were part of the interview as much as the participant. The reasoning for this was twofold. The area of visualization is difficult for many students to verbalize, being fundamentally a nonverbal skill. Therefore, the interviewer had to establish a comfortable, conversational environment, to build the "drama," "scene," and "performers" in the interpersonal interaction of this style of discursive interview (Hermanns, 2004). In this case, the setting was that of a teaching setting, similar to the location of the study and also a familiar set of roles for interaction and interpretation for students and instructors. This enabled the researchers to engage in instruction or coaching for the students through answers, to both encourage the students to explore their thinking further and to maximize the chances that they would express their dynamic model building through actions, gesture, drawing, object manipulation, and verbalization. The goal was to see how students overcame roadblocks (or not), as the aim of the study from the outset was to understand success and failure points in the mental model-building process of VPA. From a practical standpoint, this structure also most closely matches the applied teaching environment where this skill in geoscience is most frequently first encountered and taught. Our intent was that this structure would also increase the relevance and applicability of our results, observations, and interpretations to real geoscience education in classrooms.

Interview Task

Participants spent ~15–30 min with the interviewers depending on the amount of time each student felt comfortable with, and the length of time required for the completion of the two problem-solving tasks. Interview questions began with basic student information such as name, major, past experience with geology or 3-D related activities, and student's thoughts about the laboratory and GeoSAT interactions. Students then solved two problems similar to those on the GeoSAT. They were provided with all of the same learning aids they had used during in-class exercises, including paper box models, modeling clay, and access to the Geo3D program on computers directly in front of them. We interviewed students from both our experimental and control groups, and all students during the interview process were provided with all of the learning aids made available in both groups even if they did not formerly have access to them during instruction. During the process, the interviewer would routinely request midwork clarifications from students, asking them what they were visualizing, why they drew certain features, and probing or hinting if students appeared stuck in the tasks. Interviewers encouraged students to be as descriptive as possible about the difficulties they were having, and their preferences for using various physical objects or actions in solving these VPA-related tasks.

Interview Analysis

Our analysis of all data sources, transcriptions, video analysis, and the instrument and problem-solving results themselves were combined into a modified version of constant comparative analysis (Corbin and Strauss, 2008). This analytical approach is derived from the more broadly interpreted tradition of grounded theory analysis (Corbin and Strauss, 2008; Creswell, 2007). Our use of constant comparative analysis builds on that method's power as an inductive technique to draw on interview transcripts, video of gestures, instructional prompts and probes, and in-progress and final drawings generated by students. In a hermeneutical context, the traditional separation of observation and interpretation is blurred in that an observation must be interpreted by a researcher to have meaning (Shane, 2007). Formal notions of validity and reliability become harder to define or less useful constructs in this research tradition, although quality control through establishment of trustworthiness is still important, and that is discussed in the next section. The data sources from this study were used to assemble an idealized and generalized model that combines elements of how people approach problems that require VPA to solve, the indicators and elements of successful strategies that are employed, how unsuccessful students mitigate a lack of VPA, and finally how all these are combined into action in drafting cross-section solutions. The goal was to understand how these factors combine to form the characteristic responses seen by Kali and Orion (1996) and other workers in this area.

To accomplish this, we analyzed the data through grounded, a posteriori, inductive, and interpretive coding (Schwandt, 2007, p. 32–33). We transcribed the interviews for text analysis, and then iteratively coded and interpreted the videotapes and transcripts both, referencing the in-process and final drawings that students generated during problem solving. This analytical technique was selective in nature, so our open or initial coding (Saldaña, 2009) was focused on actions and verbalizations that indicated portions of mental visualization and difficulties with VPA. The selective coding was informed by the expertise of the authors in this area, instructionally and personally with this skill, and targeted at specific strategies and actions/utterances by student participants that could explain the nonuniqueness of pathways toward the prior codes generated by Kali and Orion (1996). We later performed axial coding across our eight participants to form correlated, larger-scale concepts and themes. In this study, these were the common stages where students had difficulty with this task, and their strategies to overcome them as best as possible, that led finally to the process models presented below.

Trustworthiness

Trustworthiness is the equivalent of reliability in quantitative methods, with validity being established by the expertise of the researchers in this particular content domain. For nonnumerical or nonquantifiable data such as ours, the quantitative measure of reliability and associated error ranges and statistical significance are replaced by triangulation of multiple data streams and interpretive perspectives. Triangulation seeks confirmation of similar "signals" coming from different sources, and also allows for the explanation and expression of differences in data sources (Patton, 2002). Denzin (1978) distinguished four categories of triangulation: (1) data triangulation (multiple data sources), (2) investigator triangulation (multiple researchers or evaluators), (3) theory triangulation (use of several perspectives to analyze a single data set), and (4) methodological triangulation (combination of different methods to address a single problem). All of these types of approaches are intended to add credibility to the research findings. These are all expressed as evaluation criteria that lend objectivity and dependability to qualitative data analysis (Steinke, 2004).

Our study employed primarily strategies 1 and 2: multiple data sources (interview transcript, video of gesture and sequence of actions, and student drawings from the interview task), and also independent researcher triangulation. Both researchers independently coded the data sources and agreed on interpretations where differences arose. In the inductive exercise of constructing our process models from this data, we employed a variant of strategy 3, as the final product of this study is a theoretical model grounded in the data of how students recognize, construct, and express geologic problems using VPA.

RESULTS

Due to space limitations, it is not possible to reproduce here all of the detailed observations of each of our eight participants; however, we have selected cases that illustrate gesture, language, and action that were present in students at the end members of VPA, and one case where the student was somewhere in the middle ground. For those readers wishing to see all data from these and other participants, as well as complete interview transcripts, all of this is provided in Alles (2006), which is freely available at the URL provided in the References section. What we present here instead is an excerpted version of this discussion, highlighting three representative participants, Tom, Sandra, and Cathy (all pseudonyms). Respectively, they represent the high-, medium-, and low-ability students in our study, and detailed narratives describing their interviews and problem-solving strategies are presented. We also share their drawn solutions as well as pictures from interview videotapes indicating their gesture strategy.

Participant: Tom

Tom is an example of an introductory student with an overall high VPA. Tom showed a high aptitude for VPA and scored the highest code-based ranking of 89 out of a possible 90, which he repeated in both rounds of testing.

Background

These were the first geology classes that exposed Tom, a biology major, to earth science and earth processes. Only during the pre- and post-test exercises as well as the interview did Tom encounter anything similar to the GeoSAT. He explained that he had experienced what he interpreted as a spatial task and site visualizing graphs such as those found in math and science courses.

Question 1

Question 1 (Fig. 2) began with the interviewer asking Tom to complete the face of the upper part of the cube diagram, and then follow that with a cross section on the lower part of the page. Tom immediately looked over the two-dimensional paper diagram and approached the problem with minor or no hesitation. His quickness to answer suggested that he had confidently constructed his mental model of the tilted layers seen in the cube diagram. The total time required for Tom to complete the problem was ~25 s. Tom's high aptitude for VPA was highlighted by his direct explanation of the construction. He further showed this by using his hands to show the cutting action required to expose the projection generated by cross-section line A–B–C. He swept his hand across the top, indicating that it goes straight across or bisects the cube perpendicular to the side profile provided. He explained his thought process using cross-section line B–C by showing how the resultant thickness of the tilted and bisected

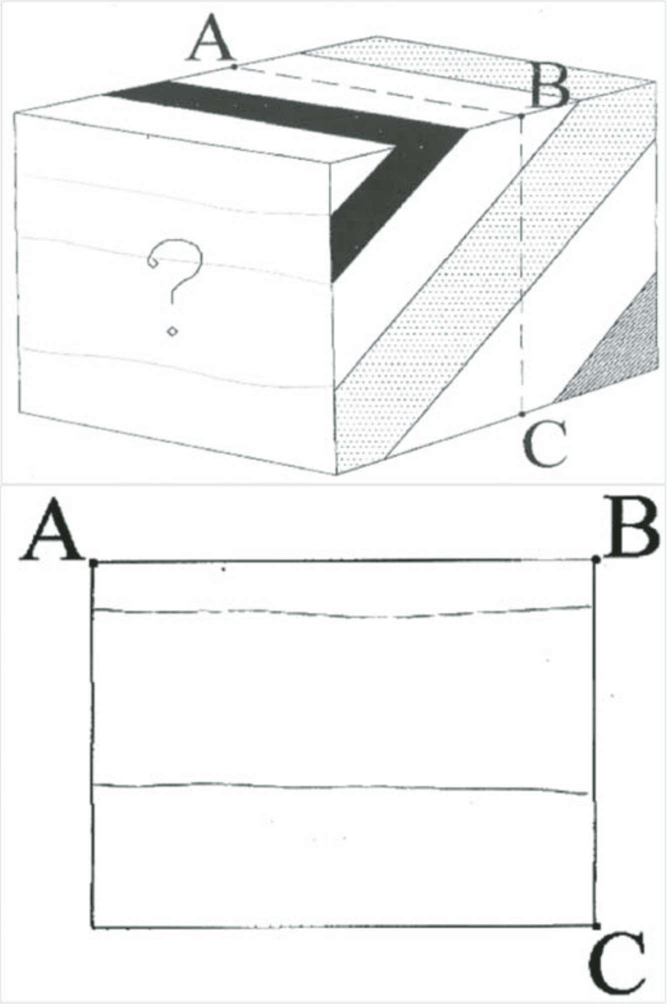


Figure 2. Tom's interview question 1. Tom does not label the layers; however, they follow the intended thickness. Notice both constructions incorporate horizontal lines yielding correct responses.

layers would be exposed in a cross-section construction. Tom further explained that the layer that was not bisected by the cross-section line would not be represented in the cross-section portion of question 1. This explanation illustrates his complete understanding of the internal structure, namely the hypothetical rock layer inside the cube.

Question 2

With task 2 (Fig. 3), Tom again appeared to be confident and well focused on the task at hand and approached the problem with little hesitation. Due to the difficult nature of the question, Tom exhibited concentration and stared at the problem for a moment (less than 3 s). Tom then used his pencil to mimic the fold axis. He laid it on the paper and tilted it in the direction of plunge. This sequence is shown in Figure 4. While Tom appeared to have visualized the object easily and quickly, he struggled with constructing the cross-section view. He picked up the paper boxes one by one looking to see if they would provide any help.

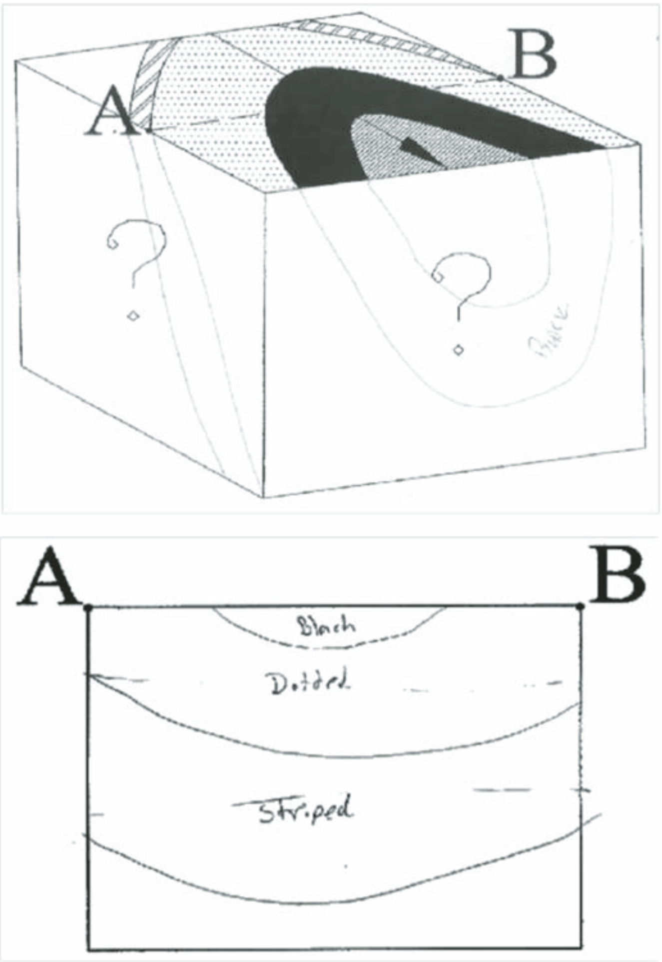


Figure 3. Tom's interview question 2. Tom's first attempt and subsequent erasure can be seen in the cross section. These were originally drawn as straight lines.

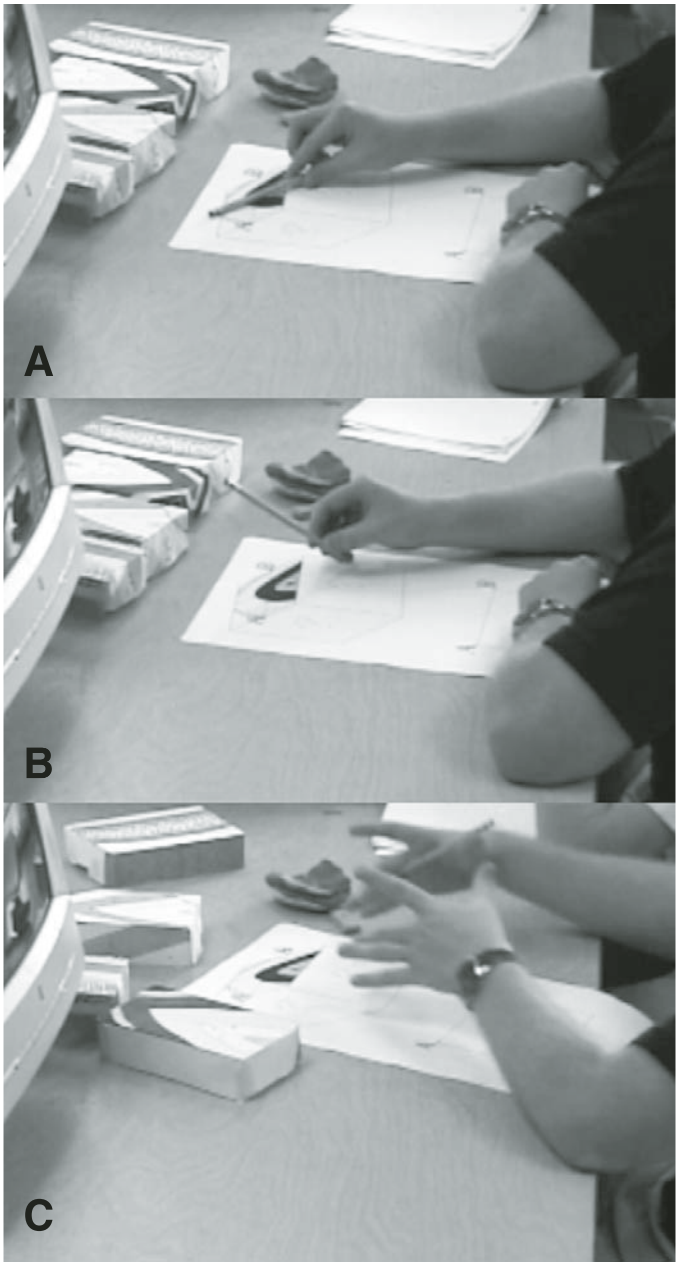


Figure 4. Video stills from Tom's interview illustrating his rapid and extensive use of gesture to aid in visualization and solution construction. In frame A, 3 s after having been presented with the problem, he laid his pencil down along the fold axis, with the pencil tip pointing in the direction of plunge. Less than one second later in frame B, he picked up the pencil to show the 3-D orientation of the fold hinge. He did both of these steps with no verbalization and no prompting from the interviewer. Shortly thereafter, when asked to describe his visualization, he used his hands as shown in C to trace out the volume occupied by the plunging syncline in the orientation of the axis illustrated in B.

He found it difficult to start and said how hard it was to connect the faces to re-create the geologic structure. He appeared to be less confident with drawing the necessary layer and he could be seen nervously drumming his fingers. He followed this by drawing in the side profile. He did not make any erasures during this step. The interviewer asked what he thought the structure was. Rather than use geologic terminology, Tom said, "It's kind of like a cylinder that has an angle… a cylinder going this way…," again using his hands to mimic the shape of the structure (Fig. 4C), moving his hands in this shape as he tried to show the three-dimensional nature of the object coming out of the paper. By performing this physical action, he finalized the correct direction of plunge and represented the correct structure projection. The time required to generate the front-face projection was substantially more than the time he needed for question 1, even though he did arrive at a correct answer.

When confronted with the cross-section task for this question, he commented that, "…this part is harder to visualize than the last one." He indicated that the black layer seen on the surface projection of the syncline will yield only a small exposure as compared to the exposure seen in the front-face projection when it is bisected by cross-section line A–B. He drew in the folded layer and labeled it black. He then drew straight layers for the deeper levels of the fold, initially contradicting his own front-face projection drawing.

Seeing this, the interviewer asked why he had drawn the black layer curved yet he drew straight lines for the other two boundaries. He was able to recognize this fact, and he admitted that he was probably thinking of the answer for the previous question 1. Otherwise unprompted, Tom changed his drawing to reflect the curved layers seen in the lower portion of Figure 3. When asked if it was pretty much similar to the front-face construction of the cube projection, he realized that they will be similar but not identical. Through further explanation, he told the interviewer how some layers will not show up in the cross section due to the location of the cross-section line seen on the map view.

Participant: Sandra

Sandra is a student with an overall medium to high VPA aptitude. Sandra's pretest score was 80, and her post-test score was 79, which is relatively good but in the middle of the range of students tested.

Background

Sandra was an international business major and considered herself aware of nature and appreciative toward it. This was her first geology laboratory, and she was concurrently enrolled in a comparable geology lecture at a nearby community college. She liked the bigger picture aspect of geology more so than the technical nature of hand specimen identification, and she said that she would recommend the class to others. Sandra reported that she liked to use her hands and model clay. She thought this would be considered a spatially oriented task.

Question 1

Question 1 (Fig. 5) began with Sandra immediately adjusting one of the paper box models to see if it resembles her question. She looked at the problem a while, and then she drew the upper and lower boundaries of the black layer and then the top boundary of the dotted layer. She shaded and filled them appropriately. She immediately followed this by constructing the correct cross section, including attention to layer thickness as observed at the bisection line B–C. The total time for Sandra to answer question 1 was approximately 1 min 45 s. When asked how she chose to draw her answers, she responded by gesturing along the side-face profile using her bent finger to mimic the layer thicknesses, describing why she drew the front face as she did. She also reported having a clear visual image of this structure in her mind at the time.

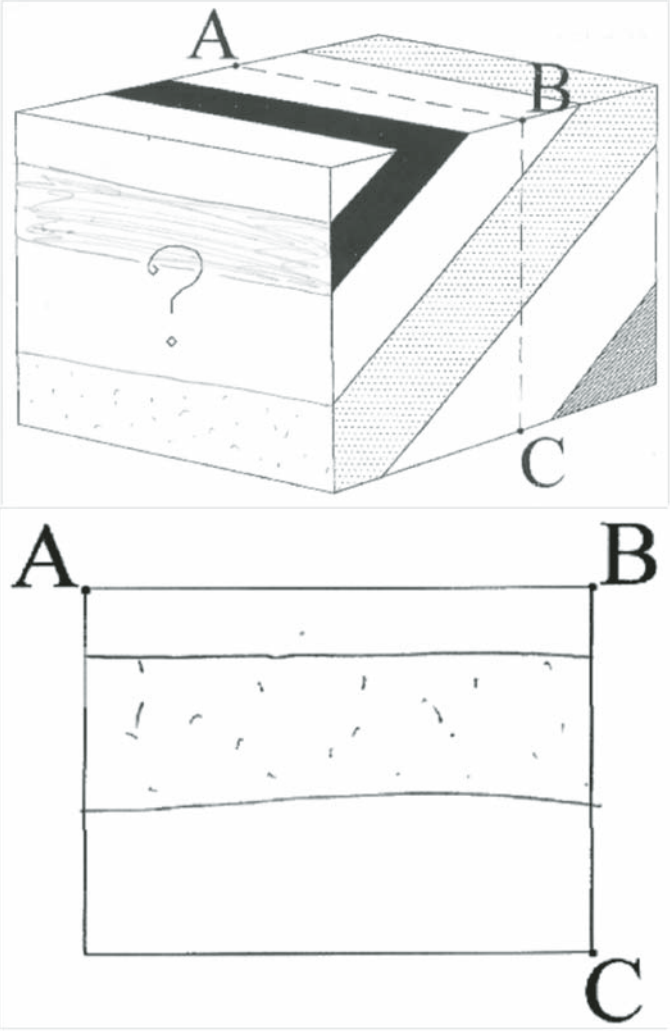


Figure 5. Sandra's interview question 1. Notice the correctly drawn horizontal layers with shading. The layers in the cross section are a little thick, but she has the general idea of matching layer thickness with the bisecting line, which is the main point of this exercise.

Question 2

Question 2 began with Sandra looking at the diagram for several moments and then adjusting the sheet 180° from upright. The way she oriented the paper made the front face of the cube projection to appear to now be the top or map view. If this were the prescribed orientation, it would totally change the structure type observed. Simply turning the paper to make it easier to draw on would not be a problem; however, the orientation she now placed the cube in described a nonplunging, upright syncline, which is not correct. She disregarded the plunging fold indicator arrow and the proper orientation of the letters A, B, and C.

She proceeded with constructing a misinterpretation of the correct orientation of the plunging synclinal structure (see Fig. 6, front-face projection). Unlike question 1, Sandra was never seen holding, moving or otherwise using the paper box models. Sandra reorientated the sheet upright after she finished constructing the layer boundaries. She shaded and stippled the layers while the page was in its proper orientation. When asked why she drew straight lines on this face, Sandra immediately rotated the page back to the orientation on which she originally constructed the fill in face to show the interviewer why she drew it as she did. She used her hands to show how the layers dipped inward (Fig. 7). She even drew strike and dip symbols indicating that the layers dipped inward, which is in conformity with a syncline fold. The interviewer continued and asked Sandra what the other side face of the cube projection would look like. Feeling committed to her answer, Sandra said, "…Maybe a continuation of it I guess…," and then drew the boundaries of the striped layer straight down like she did on the front face. The interviewer asked, "just straight down?" Sandra went into a perfect explanation of why the layers trended the way she had constructed them. Her paper was also turned in such a way as to make the front face appear to be the top face of the structure. Once she finished her explanation, she straightened the paper to the correct upright orientation.

The interviewer wanted to see which projection she felt represented the top or map view, and what type of structure she thought it was. Sandra thought it was a syncline, which is correct, but she realized during her explanation that the orientation of the paper could potentially change the structure type interpreted. The interviewer then told her how the cube was supposed to be orientated per clues on the drawing. He told her where the top, side, and front faces were located. She seemed to be surprised, but understood the short tutorial. Reexamining the structure with this new information seemed to confuse Sandra. She now thought it might be an anticline.

To end the potential confusion with technical terminology, the interviewer found one of the plunging syncline paper box models and asked her if they appeared similar. She seemed to agree. The interviewer asked her to orient the paper box model the same way she interpreted the cube projection image of question 2. He then asked her to point out, on the paper box model, her interpretation of where the top view is represented. She pointed to the front face, not the top of the paper box model. The interviewer had her interpret all of the sides as she saw them relate to the paper box model. The interviewer then pointed to the tilted layers on the side of the paper box model and asked Sandra if that matched what she called the side on her drawing. She said, "not really, this side looks a little more slanted." She drew a slight curve to the side face only. Her last attempt would have been correct for her interpretation only, but she was not considering the plunging fold represented in this question. Realizing that she did not see that she had drawn a plunging fold, the interviewer moved on by asking her to attempt the cross-section portion. The interviewer asked her if knowing a curved layer would be in the cross section would make a difference to how she would draw it.

Sandra began to see that her drawing did not match a plunging fold pattern. She drew in the curved exposure of the black layer. The interviewer asked if it would make sense compared with her first drawing on the cube front face. She saw that it would not. She was then asked to draw a basic shape of a syncline.

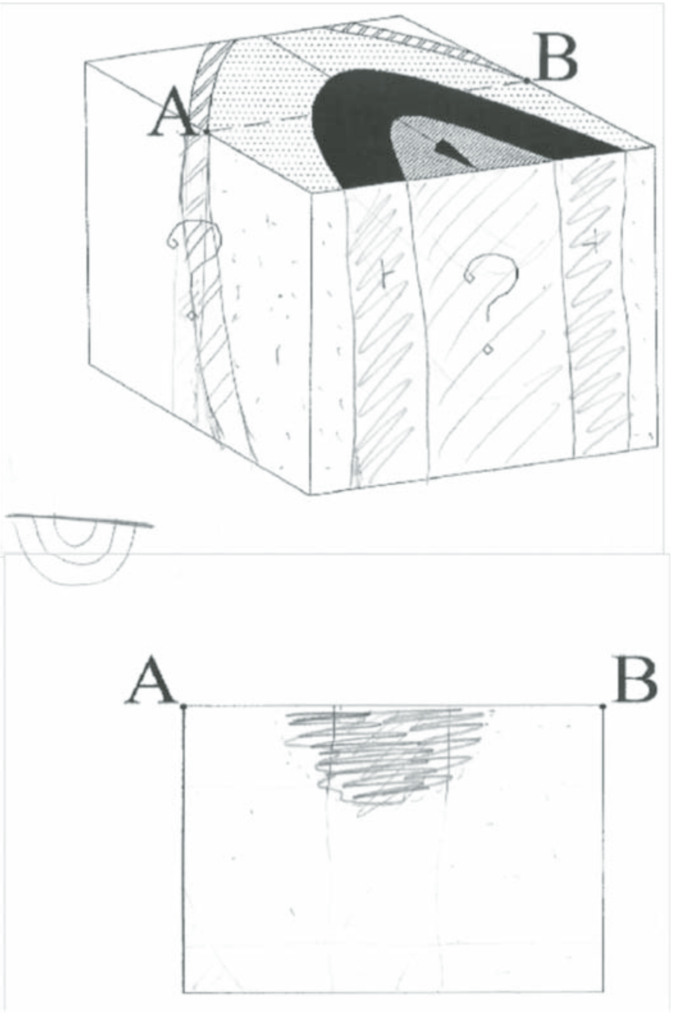


Figure 6. Sandra's interview question 2. This is Sandra's first attempt. Notice how she constructed the layers as if the fold were on its side. Her first try at the cross section consisted of vertical lines, which she later erased.

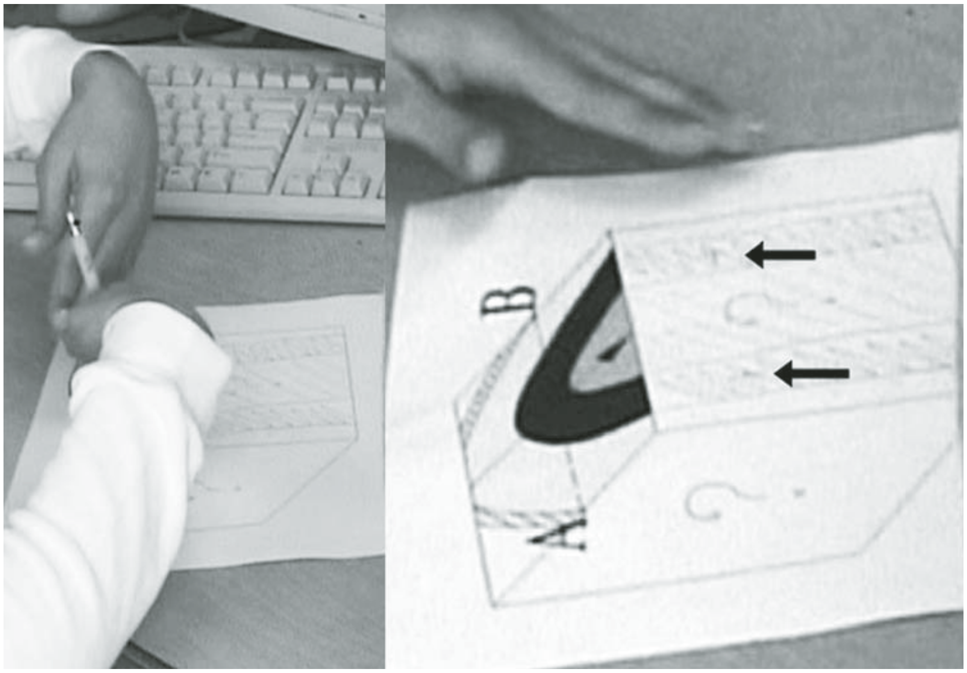


Figure 7. Video stills from Sandra's interview question 2, attempt 1. Note that she used her hands in the left image to show the dipping layers of the syncline. However, she also turned the image 180 degrees to make the surface projection shown a cross-sectional view. She then drew in map symbols (indicated by arrows), which are strike and dip symbols dipping inward. While her visualization was correct in the most basic sense, her conception of the entire geometry of the problem suffered. This is reflected in the cross-section construction in both of her attempts of this problem.

Without hesitation, she constructed a plausible cross-section profile of a syncline, which looked very different from the one she had constructed. She lightly sketched over her first drawing of the side cube face with curved lines. She continued to sweep the lines from one side of the outcrop pattern to the other side of the outcrop, making a line continuation error by sweeping the striped layers, on the side profile, right past the cube edge and on to the front face. At that point, the interviewer asked Sandra if she wanted another sheet to redraw her answers. She took a new sheet and sketched out a significantly more curved pattern than before (Fig. 8). She seemed confused with which face would become the top view. After the interviewer explained where the top was, Sandra again drew straight lines in part C; however, they were now slanting some instead of being vertical as before. Since the structure is a fold, straight lines would not be representative of the correct structure outcrop pattern. The interviewer asked her to construct the cross section. She immediately redrew the black layer as she had done on her first trial sheet for question 2. She followed this by drawing in and shading the other layers. Interestingly enough, she constructed an appropriate cross section, but it did not match the drawing in the cube projection. She chose her second attempt as her final answer. Answering question 2 took Sandra ~16 min.

Participant: Cathy

Cathy represents the end member of the low category based on her overall performance as seen by her pre- to post-test scores as well as responses provided during the interview. Cathy ranked 29 on her pre-test and 53 on her post-test, which represents a major improvement, but still indicates a very low VPA.

Background

Cathy was a hospitality tourism and management major. She had taken the lecture class the previous semester and was now in the laboratory. These two geology classes were the only experience she had up to the time of the interview. When asked if she had used spatially oriented skills before in a contextualized way, such as sculpting, computer design, or other, she said no and agreed that this was something new for her.

Question 1

When asked by the interviewer to explain what she saw after studying the question for some time (24 s) and taking no observable action, she moved directly to the modeling clay and started to shape it. She bent and tore the clay and showed the interviewer (Figs. 9A and 9B). Cathy bent the clay in a curved manner and showed the interviewer how it draped or wrapped over the external edge of the cube. She indicated none of the internal structure considerations with her model or explanations. This action has been labeled "gift wrapping" or "painted box."

While Cathy recognized the cube as a 3-D object, she did not recognize its internal structure or the individual layers as being solid and continuous. This action was further exemplified when she lightly sketched an attempt of the black unit. The construction seen in the cube face of Figure 10 details how she tried to connect the black layer, but instead of drawing straight lines, she constructed a partial line continuation on the right cube face and mirroring of the black unit by drawing a curved layer seen in the lower left corner of the left cube face. The interviewer asked if she was making the clay go around the edge of the box. She confirmed her misunderstanding and said, "It's pretty hard for me to visualize these 3-D things."

She continued to struggle with part A for over 6 min, and with much coaching and interaction with the interviewer, she began to redraw the left cube face interpretation. By eliminating potential choices via specific actions and subsequent questions for the interviewer, she ultimately drew the straight lines seen on the left cube face of Figure 10.

Her lack of VPA was further expressed when she drew the equivalent of a face copy of the side profile for the cross section. Interestingly enough, after drawing the layers slanted, she erased the lowest line, knowing that the dashed layer would not be represented. The interviewer pointed out that she drew layers slanted in the cross section A–B–C, but she drew flat layers on the left cube face completion portion. She drew the heavy lines seen in the cross section at this point with little elaboration. She took almost 14 min to complete this question.

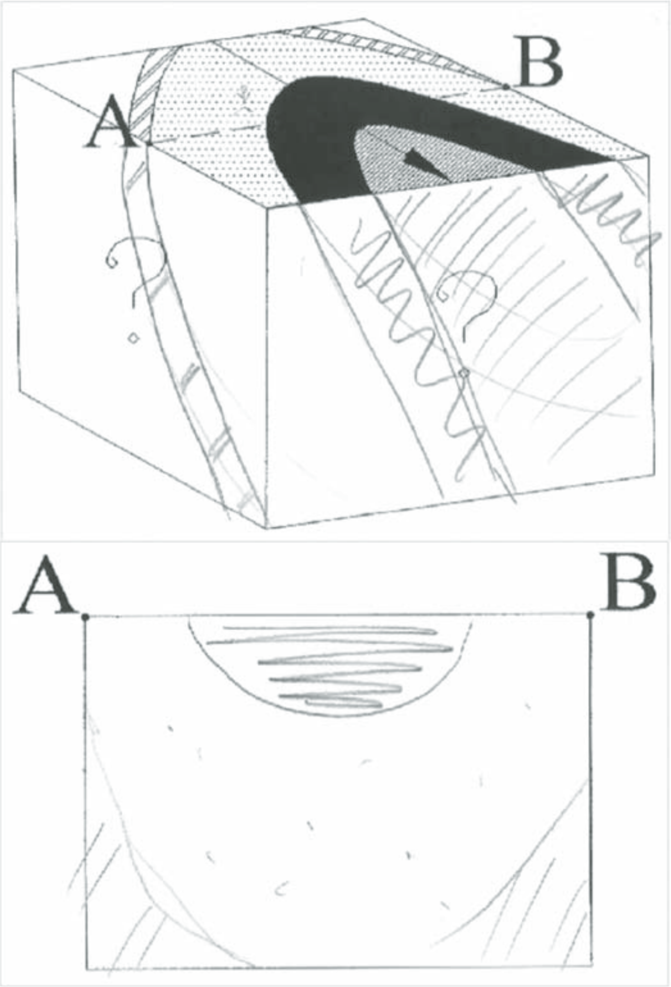


Figure 8. Sandra's interview question 2. Notice that she attempted to curve the black layer on the left cube face as she worked through the problem. This rethink may have led to the answer in the cross section showing a greater consideration for the curved nature of the layers. This is Sandra's second try at question 2 and the one she chose as her final answer.

Question 2

When presented with the second task, the interviewer made sure to review the basic terminology of plunging folds to make sure Cathy understood the problem. She examined the problem and quickly realized that the face she was required to complete was the same one missing on one of the paper box models. She looked on the reverse side to see a profile similar to the one she

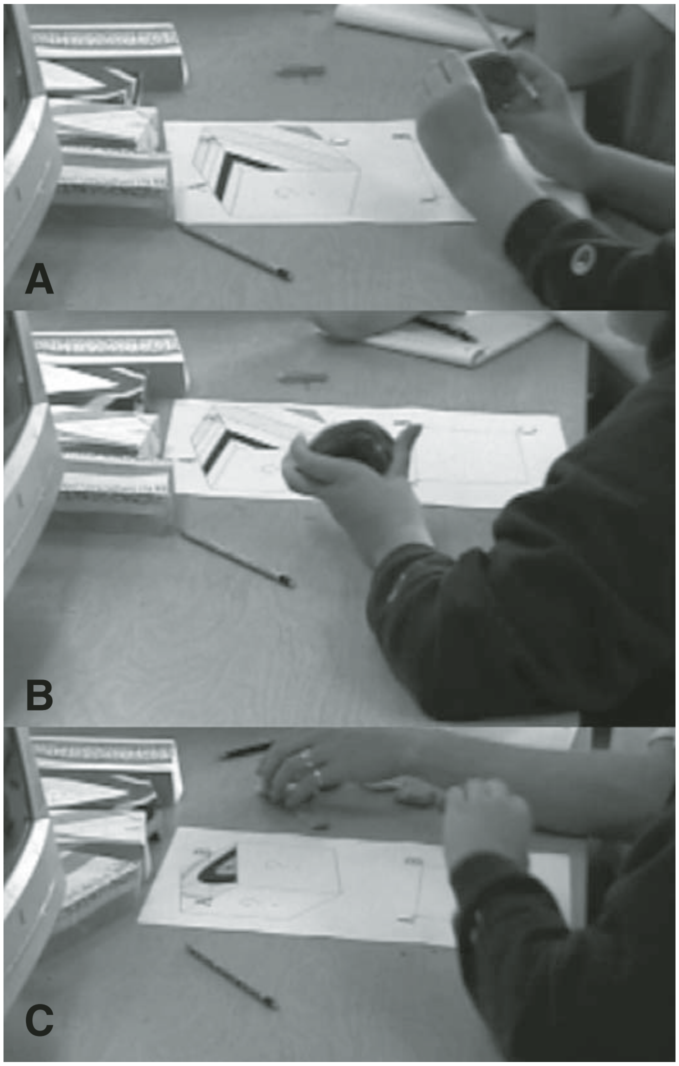


Figure 9. Video stills from Cathy's interview questions 1 (A, B) and 2 (C). 24 s after being presented with problem 1, she finally took up the modeling clay after being prompted by the interviewer to adopt this strategy, who had observed that she appeared to be stuck on this question. In frames A and B, which show a different perspective, she illustrated her "gift wrapping" visualization of this object. While the geologic object is simply tilted but otherwise undeformed flat layers, her representation with the modeling clay clearly shows that she can only visualize the outer, surface expression of this shape. She has bent the clay to mimic the apparent curvature on the diagram. In C, she showed further low VPA and a tendency to only see surface expressions, as she cupped her left hand to mimic the apparent downward curve of the fold, which in three dimensions is actually upwardly curved.

needed. She saw the similarities between the paper box profile and the one she was asked to complete. With little hesitation she constructed the black layer on the front cube face of Figure 11.

When asked about the side cube profile, she attempted to draw in the necessary striped layer. When asked about this specific layer, she drew two lines that sort of represented the correct response. The interviewer, seeing that she was finished, asked her to explain the process she had used that made it possible for her to answer this portion of question 2 significantly faster than a similar portion of question 1. Cathy responded by saying that she merely had constructed a half circle. She explained that there would not be any straight nonfolded layers exposed on the front profile. She was not sure why she knew that. The interviewer asked her if the box model helped at all. She said, "Yeah, yes definitely helps. It just has a pattern and you see it." She rotated the paper box model to show the interviewer how the curved layers copied or mimicked what she had constructed. When encouraged to move on to the cross section, she picked up the paper box model and held her pencil at the point representing the bisection line. Since the paper box model and question did not agree as far as layer configuration, she appeared to get stuck and stared at the model awhile. With the rear profile of the paper box facing her, she drew in the layers in the cross section. She started with the black layer, followed by the upper boundary of the striped layer, and then she labeled the dotted area and drew the lower boundary of the striped layer. She finished by labeling the last layer exposed, which was the white one. This process took approximately a minute and a half. The interviewer prompted her for an

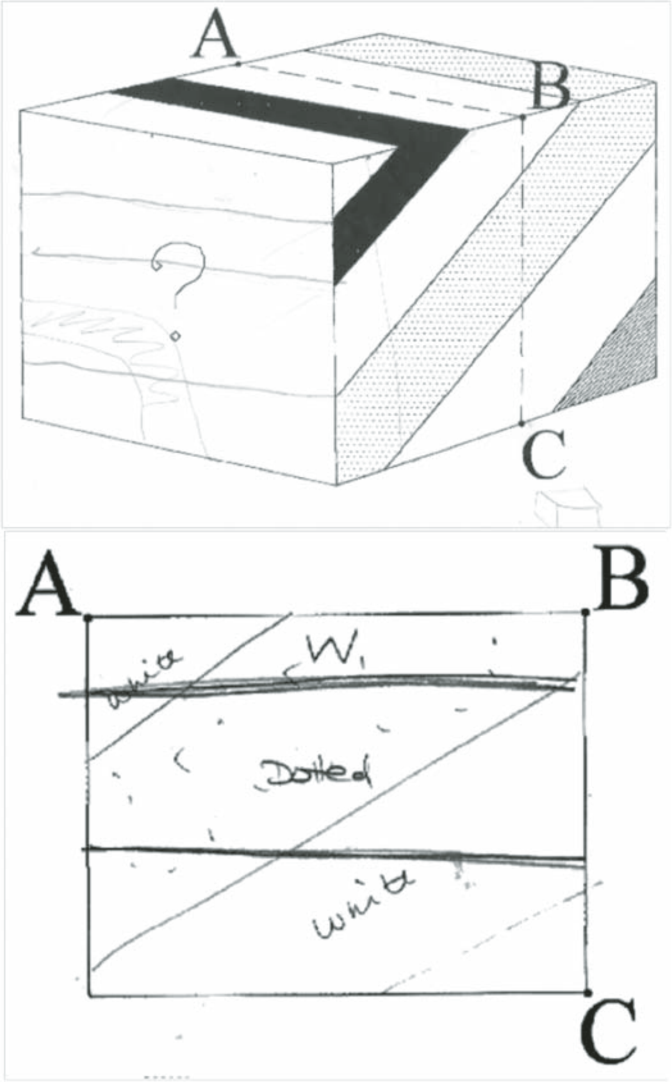


Figure 10. Cathy's interview question 1. Note the layer Cathy drew in the lower-left corner of the front face of the cube projection and the cube drawn below the "C." Cathy has constructed a combination face copy and line continuation error seen in the cross section.

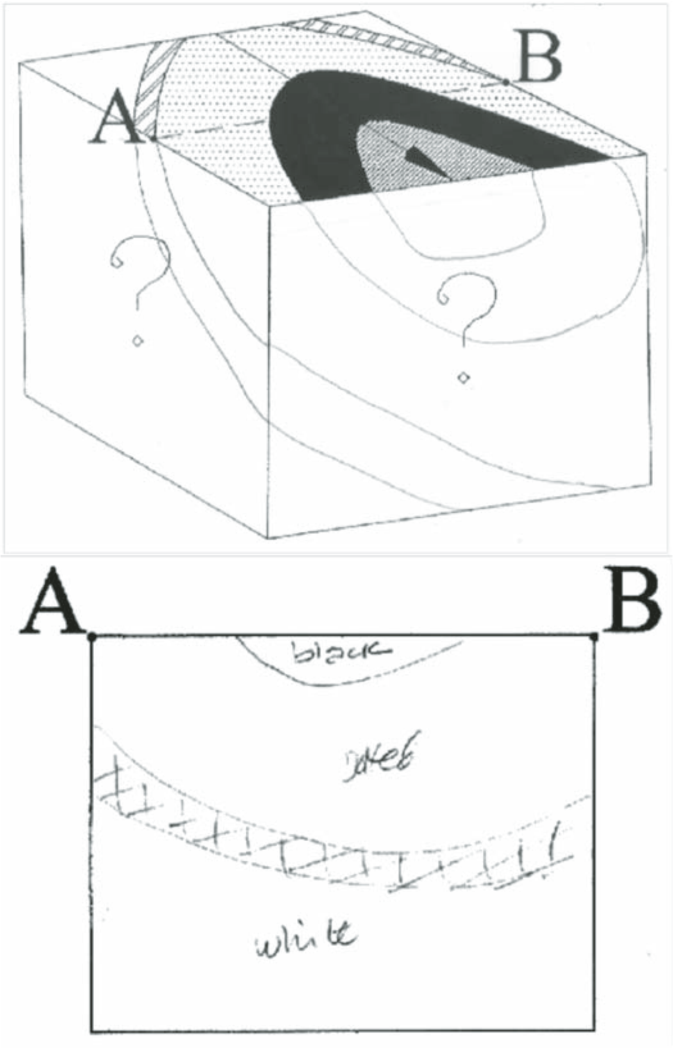


Figure 11. Cathy's interview question 2. Note the apparent line continuation error in the upper image where the striped layer transitions from the side view to the front view. Note the striped layer in the cross section and how it does not match up with the location of the cross-section line A–B.

explanation of her thought processes. She explained that the only layers exposed were those cut by the bisecting line A–B. Interestingly enough, she pointed this out in a clear and concise manner, which was opposite of her question 1 explanations.

Cathy said she envisioned looking down on the structure and drawing what she saw. She specifically used the top face projection seen in the cube projection to explain to the interviewer how she interpreted the layers she constructed. Rather than use the front face projection that describes depth, she continued to focus on the top projection that describes surface exposure. She interpreted the object as having a curved nature to it and as the reason she constructed curved lines in part D of question 2. This response indicated that some amount of VPA was present but was very dependent on existing imagery being present to mimic or borrow in order to construct a correct answer.

DISCUSSION

Our interviews and observations have led us to identify a continuum of visual penetrative ability. For those students who are not proficient at using VPA, these types of spatial problems proved to be difficult. By close observation of student problem solving, we see that time to completion seems to be a related factor, but that the time to the appearance of externalized evidence (gesture, drawing, other action) of a complete mental model was a better measure than the overall completion time. Students like Tom with very rapidly formed mental models and high VPA still had a relatively long time to completion on the drawing/construction portion of this task. The ability to represent or reconstruct internal mental imagery appears to be a separate step from the initial formation of that internal picture.

Using these student-generated gestures, language, and drawings, we constructed a generalized model that attempts to synthesize the steps and overall process one uses to work through spatial problems involving VPA, also informed by both researchers' reflections on this task as supported by our experiences working with the students in interviews. The model has two variants (illustrated in Fig. 12), an idealized process model (IPM) and a generalized version (the generalized process model or GPM), which captures the range of real student responses. The "idealized" model is nothing more than the best-case scenario of the generalized model, and it represents what we propose as the smoothest path through this style of task by someone who is highly adept at this skill. While actual experts and adepts at this skill may use different sequences and pathways of cognitive events to arrive at correct answers, we believe these to be the essential steps that must occur to generate correct, fully penetrative answers on tasks such as those seen in the GeoSAT. Figure 12 is constructed to illustrate the specific subtasks occurring at steps along the way toward successful completion of this type of problem, juxtaposed against the actual pathways observed in our interviews. Finally, the "generalized" version of this model as presented in Figure 12 incorporates the evidence from our qualitative analysis into procedural steps taken by students with varying degrees of VPA, and also proposes pathways that lead to the original six response types observed by Kali and Orion (1996). The generalized model also is constructed to illustrate the nonunique nature of pathways and issues in visualization skill that can combine to generate output on the GeoSAT and similar instruments.

Idealized Process Model (IPM)

Step one of the IPM relies on the fact that a person is able to "see" or visualize the GeoSAT style question as a three-dimensional cube perspective represented on a two-dimensional medium. This medium can be the question sheet, the computer screen, or any construct that was not specifically a tangible three-dimensional cube or model. A truly 3-D cube or structure could alternatively be substituted for the initial 3-D type representation; however, due to the nature of the GeoSAT and the available materials, it was apparent that an actual 3-D model would not be feasible or practical.

Step two, in the IPM, consists of constructing a 3-D structural image from provided clues, and is arguably the most pivotal step in the entire process. This step challenges the student to create a mental image of the external as well as internal structure of the cube perspective. This step is where the quality or continuum of VPA becomes most apparent. A successful problem solver on the idealized track would engage with the problem rapidly and would express this externally through the manipulation of tools, hands, or other object(s) as this spatial cognition is offloaded to the physical environment. This process also appears to be dynamic, in that even people who are good at this skill will refine their mental model as needed as they work through the problem. The stronger the mental image, the sooner the student can effectively proceed with the rest of the sequence.

Step three of the IPM involves recognizing how the mental model will project to the surface elsewhere in the diagram. This step is necessary in order to complete the unknown faces on the cube perspectives (shown in all other illustrated problem figures with a question mark on the blank, unknown face). A student who has a strong and correct mental image typically completed these portions of the tasks with little or no hesitation to draw and some never made any erasures during the process. This first step involving new image construction in question 2 took even the highest VPA student some time to work through the problem due to its greater reliance on the need for a good mental model.

Step four is the last step. Here, the student not only forms the mental image as in steps one and two, but along with provided information and the newly constructed face from step three, they must now attempt the most difficult step, combining all of the previous information and generating a completely new cross section. This part of the process takes the most time for many of the students because it is not only the most difficult new image to construct, but it also involves the deepest use of VPA.

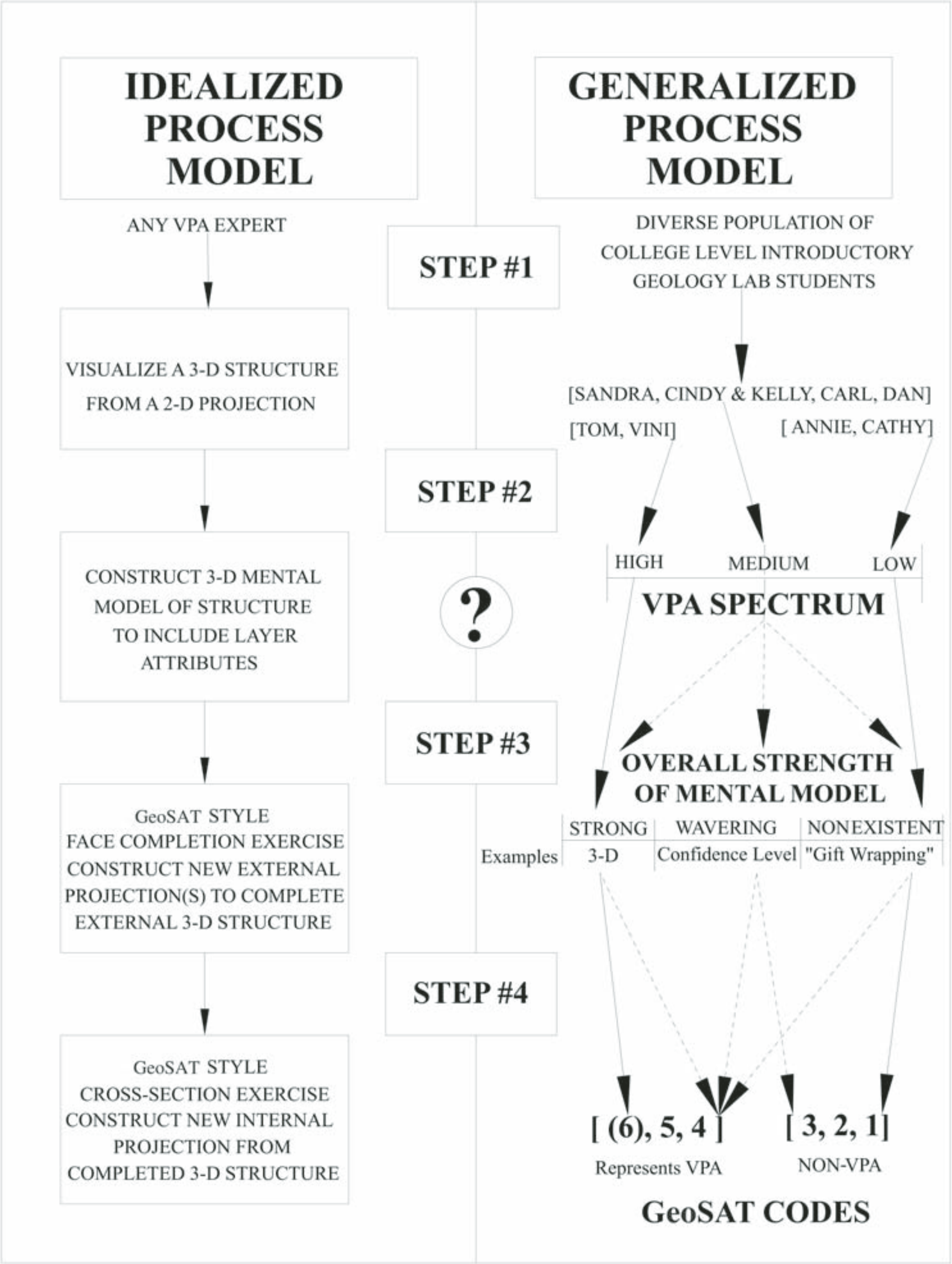


Figure 12. Idealized and generalized process models discussed in the text. Notice the question mark between steps 2 and 3; this represents the area of concern where further study is necessary. Dash-dot arrow lines represent possible path outcomes, where solid arrow lines represent a definitive path as described by this study's analysis.

Generalized Model: Incorporating Difficulties

Step one of the generalized model requires the student to use his or her own ability in order to recognize an object as a three-dimensional representation from a two-dimensional medium. This follows from the IPM, and again this step was difficult to define because no one reported having trouble seeing the representation as three-dimensional even though we suspected that some students with a low VPA did not recognize this as a three-dimensional cube perspective. We see this fact in interview two with Cathy when looking at her transcript and interview analysis. Here, the interviewer is asking for Cathy to clarify her response on question 1 of the interview task. He asks, "So you're having trouble seeing that that's actually a slanted solid object right there?" Cathy says, "Exactly!" Looking for clarification, the interviewer asks what she is specifically having trouble with. She says, "…I guess it is probably the perspective." A small cube is drawn on the question sheet and Cathy is asked if she sees it as a three-dimensional cube (see Fig. 10).

When asked the question similar to the one the interviewer asked Cathy, Tom responds much differently. He says, "Yeah I have a good visual memory. So I have no problems just like visualizing a cube in my head." This response was not prompted by the interviewer's request for clarification, it just happened naturally during the interview conversation.

Step two in the GPM involves the most distinct variation in student ability for forming a concrete mental image of the problem. The strength of the student's mental image is thought to be in direct correlation with their performance on these questions. Here, we see the role of time to problem completion, effective use of visual aids, as well as the nature of the questions asked by the student take shape and reveal the strength or stage of the student's mental image. A student exhibiting signs of a high VPA skill level, such as Tom, is able to verbally articulate their knowledge and shortcomings and also construct gestures quickly and clearly while communicating these ideas.

Tom's physical gestures indicated the plunge direction as well as layer location and orientation. Cathy, a low VPA student on the other hand, used language that led the interviewer to an inconclusive determination of whether she actually understood the topic, and resorted to accommodation strategies such as line continuation and face copying where VPA failed. The students finding themselves in the medium or average VPA category appear to have a partially formed or, in some cases, a slowly forming mental interpretation. Sandra was a student in this category, using gesture and tools to some extent, but suffering from significant conceptual and visualization errors at the same time that were quite resistant to analytic, trial-and-error problem solving to sort out difficulties.

The third step in the GPM begins to really describe the fundamental shortcomings some students have with 3-D problem-solving skills as characterized by the coded GeoSAT test results. This step is most thoroughly defined by student action due to its applied nature involving actual physical constructions rather than the strictly nontangible mental constructions described in steps one and two. A student who scored in the high VPA category would usually construct responses that would receive a code 6 (correct). This is the case for Tom, who not only completes question 1 in ~25 s, but also provides a detailed description of his actions in solving this problem.

To generate a high VPA score on instruments of this nature, one must ultimately construct the correct blank face completions. This task is almost entirely dependent on the state of the student's mental model, since the student must extrapolate from this well-formed image to the blank surface in question and project structural components onto that plane. The strength of this model is realized when they move from an introspective mental image to an externalized manifestation by consolidating their thoughts on paper and verbally explaining their mental processes. With the information provided, along with aids such as the paper box models, computer program, and instructors' help, students with a high VPA were able to complete this portion of the question set regardless of their previous knowledge base. Interestingly, the students' use of learning aids varied substantially, and not systematically. Tom used his hands to communicate his comprehension of his mental model initially and effectively, but relied on the paper box models in front of him for guidance in projecting that image on to blank faces. Sandra similarly had imagery in her mind that was expressed in gesture, and did not make much use of instructional aids. Cathy, who struggled to produce even basic mental models, used instructional aids the most in an effort to visualize the structure. Our analysis is not conclusive on the value and utility of various instructional aids in this process, and this remains an area of rich future research.

It appears that students in the low VPA category have a difficult time making sense of layer orientation as a particular unit transitions from one face to another. A student falling into the low VPA category most often generated errors associated with line continuations and mimicking or face copying one of the given cube perspectives. In Cathy's case, we see her appear to visualize the modeling clay wrapping around the edge of the cube perspective and then constructing this same perspective representation on her answer sheet. The idea of the layers being painted on or appearing as "gift wrapping" supports the notion that a low VPA student does not have a clear or usable mental image of the internal structure. The plunging fold problem is even more difficult to visualize mentally because of the multiple mental tasks required, such as rotation, bisecting layers, and tilting the model in order to yield the desired projection. Due to the need for a strong mental model, this type of problem was difficult for almost all the students, but especially hard for the low VPA students. Again, further studies are necessary to determine whether fundamental issues with different problem types cause fluctuations in a student's VPA or if the student's initial VPA skill level dictates their inability to correctly approach these problem types.

Step four of the GPM concentrates on the cross-section construction portion of the tasks. A student with a high VPA is able to quickly apply the information generated in steps one through three and typically construct the cross section without erasures. Tom is ranked as a high VPA student on the GPM and exhibits indicators showing why he ranked as he did. Tom began with a little confusion while assessing the cross-section portion. He made excellent use of improvised aids like his pencil and finger. Even for a student with a high VPA, the cross-section portion on question 2 is difficult. Tom required a few subtle hints to steer his focus in the right direction. He used the paper box models as well. Tom did a great job of explaining why some layers will not show up in the cross section due to their proximity to the cross-section line A–B. Even though Tom did not adhere exactly to the guidelines of cross-section construction in the beginning, like exposed layer thickness for the striped layer and drawing the layers curved, he still eventually constructed the correct shape and in a reasonable amount of time. A student with a high VPA will firstly be able to recognize which layers the cross-section line bisects. Secondly, they will be able to reconstruct the appropriate relative thickness of those layers bisected by the cross-section line. Lastly, they will be able to combine this information mentally and construct the new and correct cross section.

Students, who fall into the medium VPA category typically would be able to finish the cross-section portion but not entirely correctly. Thickness as well as shape and location errors with certain layers are associated with this section, especially those layers that no longer appear in the correct solution. It is failures at this stage where we also see medium-ability students draw some layers as straight lines on both the front face and side projections, even though they have already constructed other layers curved on the front face—they start to make some nongeological errors the farther into an object their mental model has to penetrate. These types of errors will result in answers likely to be coded as half-face copies, unfolding, or face combinations, i.e., partially penetrative responses.

Low VPA students such as Cathy used accommodation strategies other than VPA exclusively. In the case of the cross-section portion, she is seen heavily relying on the paper box model and its faces that were similar to the problem presented. These issues will lead to simple line continuations and face copies, or partial face copies, depending on what is available to mimic, and copy that appears correct at least superficially.

CONCLUSION

Our study of student drawings, the drawing process, and interaction with physical objects and gesturing has provided us real-time insights into the thought processes of a number of students with varying visual penetrative ability. We have illustrated this continuum of ability with a model of sequential problem-solving processes that is very closely tied to the style and sequence of tasks common to block-diagram tasks on tests like the GeoSAT. We have mapped out the behavior patterns for successful and unsuccessful students and have shed some light into the middle ground of medium-ability students at this kind of task. It is our hope that this serves to at least isolate specific areas for research as new studies move forward. We also believe we have demonstrated the utility of embodied cognition in conceptualizing studies of this nature and guiding analysis of qualitative data, as well as guiding instruction in this area. Perhaps with directed use of gesture as a teaching and learning tool, spatial abilities can be activated, thereby improving performance on the crucial middle steps of our process model. Our study expands the known types of geoscience cognition that are tightly linked to gesture, spatial cognition, and language. Further collaboration between geoscience education researchers and cognitive scientists in these fields is likely to result in further understanding of this skill. Once this mental visualization ability is even better understood, more consistently successful educational interventions can be built to enhance this geologically critical skill.

ACKNOWLEDGMENTS

The authors are grateful for the very thorough and constructive comments of two anonymous reviewers and Anthony Feig, co-editor of this volume. This manuscript and the value of this study in advancing qualitative inquiry in geoscience education were greatly improved by their insights and encouragement. We also acknowledge and thank the students and instructors at San Diego State University and colleagues at Purdue University for their open and candid participation in and discussions about this research.

REFERENCES CITED

Alibali, M.W., 2005, Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information: Spatial Cognition and Computation, v. 5, no. 4, p. 307–331.

Alibali, M.W., Kita, S., and Young, A., 2000, Gesture and the process of speech production: We think, therefore we gesture: Language and Cognitive Processes, v. 15, p. 593–613, doi: 10.1080/016909600750040571.

Alles, M., 2006, Characterizing the Three Dimensional Spatial Abilities of Introductory Geology Lab Students: San Diego, California, San Diego State University: http://www.purdue.edu/eas/riggslab/AllesThesisFinal.pdf (accessed 21 October 2010).

Briggs, M., 2007, Models and modeling: A theory of learning, *in* Bodner, G.M., and Orgill, M., eds., Theoretical Frameworks for Research in Chemistry/Science Education: Upper Saddle River, New Jersey, Pearson/Prentice Hall, p. 72–85.

Corbin, J., and Strauss, A., 2008, Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory: Thousand Oaks, California, Sage Publications, 378 p.

Creswell, J.W., 2007, Qualitative Inquiry and Research Design: Choosing among Five Approaches: Thousand Oaks, California, Sage Publications, 395 p.

Denzin, N.K., 1978, The Research Act: A Theoretical Introduction to Sociological Methods (2nd ed.): New York, McGraw-Hill, 370 p.

Garber, P., and Goldin-Meadow, 2002, Gesture offers insight into problem-solving in adults and children: Cognitive Science, v. 26, p. 817–831, doi: 10.1207/s15516709cog2606_5.

Givry, D., and Roth, W.-M., 2006, Toward a new conception of conceptions: Interplay of talk, gestures, and structures in the setting: Journal of Research in Science Teaching, v. 43, p. 1086–1109, doi: 10.1002/tea.20139.

Hermanns, H., 2004, Interviewing as an activity, *in* Flick, U., von Kardorff, E., and Steinke, I., eds., A Companion to Qualitative Research: Thousand Oaks, California, Sage Publications, p. 209–213.

Jee, B.D., Uttal, D.H., Gentner, D., Manduca, C.A., Shipley, T.F., Tikoff, B., Ormand, C.J., and Sageman, B., 2010, Commentary: Analogical thinking in geoscience education: Journal of Geoscience Education, v. 58, p. 2–13.

Kali, Y., and Orion, N., 1996, Spatial abilities of high-school students in the perception of geologic structures: Journal of Research in Science Teaching, v. 33, p. 369–391, doi: 10.1002/(SICI)1098-2736(199604)33:4<369::AID-TEA2>3.0.CO;2-Q.

Kastens, K.A., and Ishikawa, T., 2006, Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields, *in* Manduca, C.A., and Mogk, D.W., eds., Earth and Mind: How Geologists Think and Learn about the Earth: Geological Society of America Special Paper 413, p. 53–76.

Kastens, K.A., Agrawal, S., and Liben, L.S., 2008a, Research methodologies in science education: The role of gestures in geoscience teaching and learning: Journal of Geoscience Education, v. 56, p. 362–368.

Kastens, K.A., Liben, L.S., and Agrawal, S., 2008b, Epistemic actions in science education, *in* Freksa, C., Newcombe, N., Gärdenfors, P., and Wölfl, S., eds., Spatial Cognition VI. Learning, Reasoning, and Talking about Space, Volume 5248/2008: Lecture Notes in Artificial Intelligence: Berlin/Heidelberg, Springer, p. 202–215.

Kvale, S., 2007, Doing Interviews: London, Sage Publications, Ltd., 157 p.

Linn, M.C., and Peterson, A.C., 1985, Emergence and characterization of sex differences in spatial ability: A meta-analysis: Child Development, v. 56, p. 1479–1498, doi: 10.2307/1130467.

McGee, M.G., 1979, Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences: Psychological Bulletin, v. 86, p. 889–918, doi: 10.1037/0033-2909.86.5.889.

Orion, N., Ben-Chaim, D., and Kali, Y., 1997, Relationship between earth-science education and spatial visualization: Journal of Geoscience Education, v. 45, p. 129–132.

Patton, M.Q., 2002, Qualitative Research and Evaluation Methods: Thousand Oaks, California, Sage Publications, 598 p.

Piaget, J., and Inhelder, B., 1956, The Child's Conception of Space: New York, Humana Press, 490 p.

Piburn, M., Reynolds, S., McAuliffe, C., Leedy, D., Birk, J., and Johnson, J., 2005, The role of visualization in learning from computer-based images: International Journal of Science Education, v. 27, p. 513–527, doi: 10.1080/09500690412331314478.

Reynolds, S., Piburn, M.D., Leedy, D.E., McAuliffe, C.M., Birk, J.P., and Johnson, J.K., 2002, The Hidden Earth: Visualization of geologic features and their subsurface geometry, *in* Proceedings of the National Association for Research in Science Teaching 2002 Annual Meeting: New Orleans, Louisiana. Available at http://reynolds.asu.edu/pubs/NARST_final.pdf (accessed 21 October 2010).

Reynolds, S., Piburn, M., Leedy, D., McAuliffe, C., Birk, J., and Johnson, J., 2006, The Hidden Earth-Interactive, computer-based modules for geoscience learning, *in* Manduca, C.A., and Mogk, D.W., eds., Earth and Mind: How Geologists Think and Learn about the Earth: Geological Society of America Special Paper 413, p. 157–171.

Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V., 1997, The space around us: Science, v. 277, p. 190–191, doi: 10.1126/science.277.5323.190.

Roth, W.-M., 2001, Gestures: Their role in teaching and learning: Review of Educational Research, v. 71, p. 365–392, doi: 10.3102/00346543071003365.

Roth, W.-M., 2002, From action to discourse: The bridging function of gestures: Journal of Cognitive Systems Research, v. 3, p. 535–554, doi: 10.1016/S1389-0417(02)00056-6.

Saldaña, J., 2009, The Coding Manual for Qualitative Researchers: Thousand Oaks, California, Sage Publications, 224 p.

Schwandt, T.A., 2007, The SAGE Dictionary of Qualitative Inquiry: Thousand Oaks, California, Sage Publications, Inc., 322 p.

Shane, J.W., 2007, Hermeneutics and the meaning of understanding, *in* Bodner, G.M., and Orgill, M., eds., Theoretical Frameworks for Research in Chemistry/Science Education: Upper Saddle River, New Jersey, Pearson/Prentice Hall, p. 108–121.

Shipley, T.F., 2009, Spatial visualization and the role of working memory: Geological Society of America Abstracts with Programs, v. 41, no. 7, p. 195.

Sibley, D., 2009, A cognitive framework for reasoning with scientific models: Journal of Geoscience Education, v. 57, p. 255–263.

Steinke, I., 2004, Quality criteria in qualitative research, *in* Flick, U., von Kardorff, E., and Steinke, I., eds., A Companion to Qualitative Research: Thousand Oaks, California, Sage Publications, p. 184–190.

Titus, S., and Horsman, E., 2009, Characterizing and improving spatial visualization skills: Journal of Geoscience Education, v. 57, p. 242–254.

Wagner, S., Nusbaum, H., and Goldin-Meadow, 2004, Probing the mental representation of gesture: Is handwaving spatial?: Journal of Memory and Language, v. 50, p. 395–407, doi: 10.1016/j.jml.2004.01.002.

Wagner Cook, S., and Goldin-Meadow, S., 2006, The role of gesture in learning: Do children use their hands to change their minds?: Journal of Cognition and Development, v. 7, p. 211–232, doi: 10.1207/s15327647jcd0702_4.

Wagner Cook, S., and Tanenhaus, M., 2009, Embodied communication: Speakers' gestures affect listeners' actions: Cognition, v. 113, p. 98–104, doi: 10.1016/j.cognition.2009.06.006.

Wilson, M., 2001, The case for sensorimotor coding in working memory: Psychometric Bulletin & Review, v. 8, p. 44–57.

Wilson, M., 2002, Six views of embodied cognition: Psychometric Bulletin & Review, v. 9, p. 625–636.

Manuscript Accepted by the Society 23 June 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

Designing a mixed-methods research instrument and scoring rubric to investigate individuals' conceptions of plate tectonics

Scott K. Clark*
Julie C. Libarkin
Department of Geological Sciences and Center for Research on College Science Teaching and Learning, Michigan State University, 206 Natural Science Building, East Lansing, Michigan 48824, USA

ABSTRACT

Research methods and underlying theories for research designs that integrate quantitative and qualitative approaches (i.e., mixed methods) are well documented in the field of education research. What is missing in the literature is a nuts-and-bolts description of the actual practice that goes into creating a good mixed-methods survey instrument for research in the science education domain. This paper will detail the steps involved in designing, implementing, and scoring a valid and reliable mixed-methods survey instrument. This survey instrument was designed to investigate experts' and novices' conceptual understanding of plate tectonics as inferred by their answers to a series of questions related to a modified version of a commonly used cross-section schematic published by the U.S. Geological Survey. Development of the instrument involved numerous revisions with iterative inputs from local and community-based experts. After integration of expert comments, the survey instrument was piloted to a physical science for nonscience majors course. This led to further revisions in the survey instrument to improve communication validity prior to widespread distribution. Development of scoring rubrics similarly required iterative modifications based on a thematic analysis of collected data. By outlining the steps involved in designing, validating, and analyzing this mixed-methods instrument, we believe that this paper can serve as a template for future survey instrument development. In particular, we hope to illustrate the iterative and time-intensive nature of mixed-methods inquiry, both in terms of pre-investigation design and postinvestigation analysis, and to offer our empirically based insights into the instrument and rubric development process.

INTRODUCTION

The investigation of alternative conceptions held by students is a domain of research that has driven science education discourse for nearly a century (e.g., Driver, 1989; Duschl et al., 2007; Posner et al., 1982, and references therein). The approaches used to reveal student ideas range from multiple-choice survey instruments with constrained response options, to a researcher passively observing discussions in a classroom, to intensive, one-on-one interviews, to broad, open-ended survey

---

*Current address: Department of Geology, University of Wisconsin–Eau Claire, Phillips 157, Eau Claire, Wisconsin 54702-4004, USA; clarksco@uwec.edu.

Clark, S.K., and Libarkin, J.C., 2011, Designing a mixed-methods research instrument and scoring rubric to investigate individuals' conceptions of plate tectonics, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 81–96, doi:10.1130/2011.2474(07). For permission to copy, contact editing@geosociety.org.

instruments designed to illicit freeform thought. Each of these approaches provides valuable information about the range and depth of student thinking, generating tangible evidence for the missteps students can take on the pathway to scientific literacy. While interviews provide opportunities to gain a detailed understanding of thinking and reasoning for a small handful of students, multiple-choice survey instruments offer an opportunity to evaluate the prevalence of ideas across many students. Among the possible research techniques, open-ended survey instruments are highly valued as tools that offer opportunities to both collect data from many students and probe, however lightly, for explanations of ideas.

Survey instruments of all types are attractive to researchers because of their (1) perceived relative ease of development; (2) ability to acquire data from multiple population samples; and (3) possibility for simple content and semiquantitative analyses. Newly available mechanisms for web-based dissemination (e.g., www.surveymonkey.com) provide ready access to wide and diverse populations. In their broadest sense, surveys can be quantitative, such as multiple-choice concept inventories (e.g., Libarkin, 2008), semiquantitative, as in instruments that utilize a Likert scale (e.g., Adams et al., 2006), qualitative, such as open-ended questionnaires (e.g., Lederman et al., 2002), or open-ended surveys that combine components of both quantitative and qualitative methodologies, i.e., mixed-methods surveys (Creswell, 2003; Hossler and Vesper, 1993).

Quantitative surveys are particularly useful in science education research for large-scale assessments in comparison studies both nationally and internationally (e.g., Britton and Schneider, 2007). Qualitative surveys utilizing an open-ended question design are also well used. While many texts on qualitative research provide general guidelines for instrument development and analysis (e.g., Crotty, 1998; Denzin and Lincoln, 1998; Lincoln and Guba, 1985), very few works discussing the actual experience of developing a survey instrument have been published. Some very well-known qualitative surveys are supported by a literature that describes their conception, development, and use; one of the best examples is the Views of Nature of Science Questionnaire (VNOS; Lederman et al., 2002). Although the body of work documenting the development of the VNOS provides some insight into the actual process of instrument development, the actual, nuts-and-bolts process through which the VNOS was written, reviewed, revised, and piloted is never completely discussed. While tools such as the VNOS clearly required significant thought and effort for their production, the true time-consuming nature of survey instrument development and analysis is only suggested.

Mixed-methods research "combines quantitative and qualitative research techniques, methods, approaches, concepts, or language into a single study" (Johnson and Onwuegbuzie, 2004, p. 17). The mixed-methods approach has, at times, been shunned by both quantitative and qualitative research purists (see Johnson and Onwuegbuzie, 2004; Rossman and Wilson, 1985 for discussions), but has been accepted by many researchers who see the integration of qualitative and quantitative approaches as useful, (Greene et al., 1989; Johnson and Onwuegbuzie, 2004; Kidder and Fine, 1987; Rossman and Wilson, 1985; Tashakkori and Teddlie, 1998, 2003). A significant advantage to the mixed-methods approach is the ability to triangulate or corroborate findings obtained using both qualitative and quantitative techniques (Greene et al., 1989). Mixed-methods surveys are an attractive research method, but the time-intensiveness of survey instrument use in science education research needs to be explicit because scholars new to their use may be surprised as they engage in the process of survey development. This paper will detail the steps involved in designing, implementing, and scoring a valid and reliable mixed-methods survey instrument.

SURVEY INSTRUMENT AND RUBRIC DEVELOPMENT

As with any scientific endeavor, locating one's research includes a discussion of the research question, rationale for conducting the study, and a discussion of why a particular research approach (qualitative, quantitative, or mixed-methods) was chosen. However, unlike typical scientific research, one must also locate the researcher within the context of the research (e.g., Feig, this volume). This contextualization is grounded in the idea that data interpretation will be affected by the interpreter's incoming perspective (Maxwell, 2005; Patton, 2002). For example, researchers must ask: What is the researcher's position relative to the participants (Marshall and Rossman, 2006)? In a classroom setting, is the researcher also the instructor or an outside observer (Patton, 2002)? What is the researcher's perspective; that is, does the researcher view the data through the lens of a post-positivist, an interpretivist, or a naturalist (e.g., Crotty, 1998; Lincoln and Guba, 1985; Phillips and Burbules, 2000)? Addressing these questions provides insight to both the researcher and users of the research about study quality and potential limitations.

Survey Instrument Design

The overall design of a survey, as well as the design of individual questions, can have significant impacts on the quality of research (Creswell, 2003). A survey instrument that is designed without forethought of intention, or without considering the perspective of the target population, will likely yield results that are at odds with researcher expectations. Appropriate use of language and visuals, attention to page layout (e.g., Sanchez, 1992), and the limiting of distracting elements (Harp and Mayer, 1998) can all improve survey results. While the question of survey design has been discussed most extensively within the sociology or public opinion literature (e.g., Presser et al., 2004, and similar), all fields that utilize surveys in research practice adhere to similar approaches in design.

Regardless of domain, survey design follows a number of reasonable tenets (e.g., Siragusa and Dixon, 2006); these principles apply whether a survey is qualitative, mixed-methods, or quantitative in structure. These principles have been laid out in any number of good works on survey and analysis design (e.g., Creswell, 2003; Fink, 2003; Thomas, 2004) or general research methods (Trochim and Donnelly, 2007). For our purpose of creating a survey instrument that contains open-ended, guided open-ended, and fixed-response questions, three principles are most important. First, questions need to be understandable to the target population. Efforts should be made to avoid language that is outside the common knowledge of the study population, to word questions as unambiguously as possible, and to use visuals in nondistracting ways. Ultimately, we need to ensure that users are interpreting questions as intended by the developers (Lopez, 1996). Second, pilot testing and subsequent revisions should precede dissemination of the survey instrument in a larger study; this piloting should occur with both experts and a small sample of the targeted population (e.g., Presser et al., 2004). Pilot results may inform the overall study, but cannot, in and of themselves, constitute a study. Finally, the use of cognitive interviews (i.e., think-alouds and probing) to inform survey design serves to validate the researcher's postulated inferences about test-taker intent and thinking (Collins, 2003; Beatty and Willis, 2007; Presser et al., 2004). Similarly, the significant effort invested into the application of these principles to survey design needs to be duplicated during development of rubrics for survey analysis (Ambrose et al., 2004; Bresciani et al., 2009).

Rubric Design

Rubrics for scoring or analyzing qualitative survey data can be used to categorize survey responses or to rank order responses along a relevant continuum (e.g., least to most scientific). In many ways, the development of scoring rubrics mirrors the development of survey instruments, with well-established mechanisms for ensuring that scoring is as unbiased and based in reality as possible. Thematic content analysis, a form of constant comparative analysis, is a common approach used in developing scoring rubrics for qualitative data. During thematic content analysis, the researcher uncovers common themes within the data through an inductive analysis of the data itself (Denzin and Lincoln, 1998; Patton, 2002). Thematic codes are continually changing as data analysis proceeds, although most questionnaire studies in science education are simple enough for major themes to emerge very early in the analytical process. Ultimately, the most important aspect of rubric design is attention to researcher bias; a rubric must reveal, as closely as possible, the perspective of the research subjects rather than the biases of the researchers themselves (Denzin and Lincoln, 1998).

Validity, Reliability, and Trustworthiness

As with any research study, assessments of the research design, data collection, and analytical methods are important in determining research quality. We may create a survey instrument that is intended to measure a specific phenomenon, but in reality may inadvertently measure something different, fail to measure anything meaningful, or may bias results toward our intended outcomes. The challenge of designing a good mixed-methods research project is to ensure that the fallibility of the data collection and analyses is limited to the extent possible. Just as mixed-methods research incorporates aspects of qualitative and quantitative research methods during design, implementation, and analysis, multiple approaches should be used in assessing the quality of mixed-methods research.

Qualitative research is commonly evaluated based on the concept of trustworthiness (Lincoln and Guba, 1985). In quantitative research, evaluation considers the rigor of the study (i.e., validity and reliability; Litwin, 1995; Morse et al., 2002). Similarly, Johnson and Onwuegbuzie (2004) and Onwuegbuzie and Johnson (2006) have presented an approach to evaluating mixed-methods research that they term legitimization. Each of these evaluation frameworks provides a means for judging the quality of research. Utilizing multiple evaluation frameworks provides flexibility in assessing those attributes of the instrument that are pertinent for the specific goals of a given research project (see Morse et al., 2002; Lewis, 2009; Trochim and Donnelly, 2007). We have chosen to assess the quality of this project using a combination of rigor and trustworthiness. We chose this approach over that of legitimization because (1) we are familiar with qualitative and quantitative approaches to trustworthiness and rigor; and (2) rigor and trustworthiness are well established within the science education community, while the legitimization approach is relatively new and unused. A blend of components of trustworthiness with specific metrics for validity and reliability seems to us to be a reasonable approach when evaluating a mixed-methods study. Indeed, a number of researchers have recently argued for various ways to apply validity and reliability to qualitative research projects (Creswell and Miller, 2000; Golafshani, 2003; Lewis, 2009; Morse et al., 2002).

Validity generally refers to how well a measurement represents the true value of the trait being measured (e.g., Trochim and Donnelly, 2007). For example, we might consider how well a test score represents the level of understanding of an individual student being tested. In the case of a conceptual rubric designed as the filter for analyzing survey data, we need to ensure that categories of qualitative data represent, as closely as possible, the underlying conceptions of the study population. Reliability, on the other hand, is concerned with the reproducibility or repeatability of a measure or study (e.g., Trochim and Donnelly, 2007). Although very difficult to actually test, a reliable measure would generate identical test scores if taken repeatedly by a single person, and assuming no change in understanding across test implementation. For qualitative questions in surveys, we can similarly ask if different researchers looking at a single data set would reach similar conclusions, a process referred to as peer review in qualitative research (Merriam, 2002). Similar to validity and reliability, trustworthiness is the application of the concept of rigor in ways that are tailored to the qualitative research setting (Lincoln and Guba, 1985). In particular, trustworthiness considers the relationships between the researcher, the population under study, and the ways in which data are analyzed. Most importantly, sources of bias, agreement of the participants with the findings, and application of findings or the research process in other settings, for example, need to be considered (Lincoln and Guba, 1985).

We think it is useful here to provide a brief background of those forms of rigor and trustworthiness that are most important for mixed-methods instrument design and analysis (Table 1). Table 1 is not intended to be exhaustive, but rather to touch on those areas of validity and reliability that should be considered when designing a survey and scoring rubric, and that we attempted to address in the design of the survey instrument as discussed here. Finally, the forms of validity and reliability documented herein represent both standard measures and measures that we feel should be considered more routinely in instrument development and analysis.

DEVELOPMENT OF A SURVEY INSTRUMENT AND SCORING RUBRIC FOR ASSESSING CONCEPTS RELATED TO PLATE TECTONICS

We will now describe the steps we took to design a plate-tectonic conceptions survey instrument and the associated rubrics used to analyze collected data. We include details of our iterative approach, and provide a discussion of our insights and reflections on the entire process. Our survey instrument was designed with three research objectives in mind: (1) investigating people's conceptions (both scientific and alternative) of plate tectonics; (2) documenting how these conceptions might vary across the expert-to-novice continuum; and (3) investigating the role of images in communicating, and possibly miscommunicating, plate-tectonic concepts. For this study, novices are considered to be individuals with only an introductory exposure to the theory of plate tectonics, whereas geoscience faculty are considered to be experts. Other participants, such as geoscience graduate students, are positioned at intermediate levels along the expert–novice continuum. The survey instrument we created consists of questions about aspects and terminology related to plate-tectonic processes (Fig. 1). Some of these questions required respondents to view a schematic plate-tectonic cross section. Respondents were also asked to report their confidence in their answers as a measure of the role of an individual's perceived ability on performance (Bandura, 1984). In addition to broad utility, we wanted an instrument that could be widely distributed and serve as the basis for semi-structured, one-on-one interviews. We feel the resultant survey instrument has met our expectations: Novices are able to describe plate-tectonic concepts presented in the survey instrument, and the image has enough layered knowledge—especially when used in interviews—to probe the deeper conceptual understandings of both novices and experts. The time required for the iterative development of the survey instrument to move from initial conception in early October 2007 to its current form, which was attained in April 2009, was 1.5 yr.

Locating the Research

The context, including setting, in which data are collected can influence study findings (Feig, this volume). For surveys that were administered to college-level students enrolled in introductory-level earth science courses (i.e., novices), the lead author distributed all surveys with the exception of surveys administered to students at a community college in the NE United States, where the course instructor distributed the surveys. For those courses in which the lead author administered the survey instruments, he had no other connection to the students. Surveys completed at an exhibitor booth at the 2008 Geological Society of America (GSA) Annual Meeting were distributed by both authors and by colleagues. Interviews were completed in private rooms at the GSA meeting and at four institutions of higher education. All interviews were conducted by the first author, who had no relationship to research participants. All participants were given a consent form and instructions that had received approval from an Institutional Review Board.

The location of the researcher within the context of the research is possibly more important than the setting for data collection (Feig, this volume; Marshall and Rossman, 2006; Maxwell, 2005; Patton, 2002). The lead author is a geoscientist with a research background in isotope geochemistry and geocognition, which is the study of how people perceive and understand Earth and Earth processes. The second author is also a geoscientist with a research background in geodynamics and geocognition. Both authors have a postpositivist perspective, meaning we perceive knowledge not as a fixed entity, but rather as being supported by the strongest warrants, or grounds, currently available, and subject to change as new evidence becomes available (Phillips and Burbules, 2000).

Instrument Design

In designing the survey, we initially sketched a cross-section image to be developed into a colored image, but then abandoned this approach in favor of modifying a preexisting image that was commonly used in entry-level geoscience instruction. We chose to modify an existing, open-access image instead of designing an original image because we assumed that experts (i.e., geoscience faculty) would accept this modified image as a reasonable model for plate tectonics, and we would then be able to then investigate the extent to which novices perceive the image relative to how experts perceive the image. This assumption was based on the nearly ubiquitous use of the image we chose. The image that we modified is in the public domain (http://commons.wikimedia.org/wiki/File:Tectonic_plate_boundaries.png) and is part of a wall map titled *This Dynamic Planet* (Simkin et al., 1994). The wall map, which was first published in 1989 (Simkin et al., 1989), is the best-selling map in the history of the U.S. Geological Survey (USGS Education webpage).We simplified the image using the drawing software Canvas v. 9.0.4 (ACD Systems). In modifying the image for our purposes, we removed all text, the continental

TABLE 1. RIGOR AND TRUSTWORTHINESS CRITERIA IMPORTANT FOR MIXED-METHODS SURVEY AND RUBRIC DESIGN

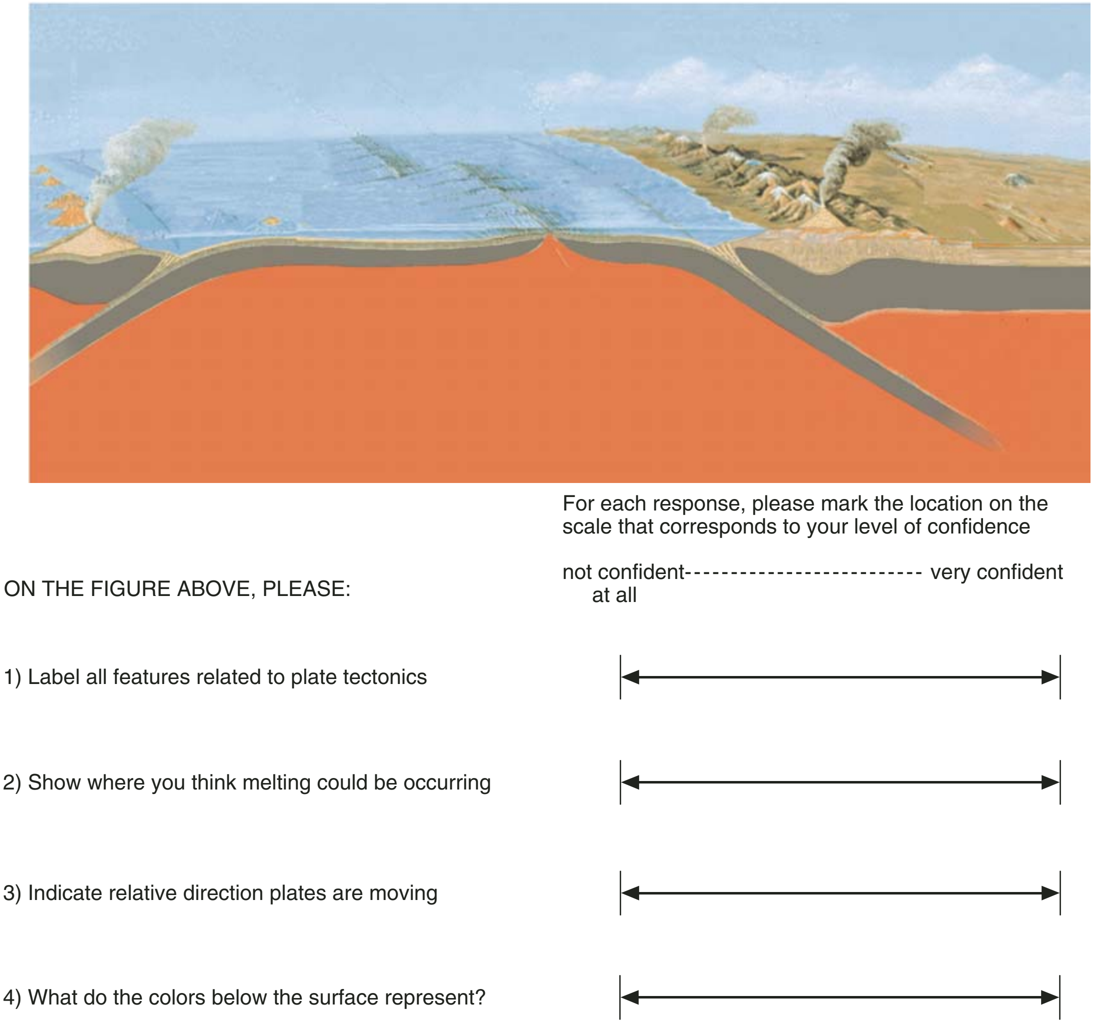
Criteria	Description and approaches	Plate-tectonics survey instrument
Content validity	A measure of whether or not items actually measure the latent trait that they are intended to measure. This is often evaluated through expert review of items and revision in response to expert opinion. Note: *Face validity* is a similar, but more casual assessment of instrument validity; we did not measure face validity per se.	Comments from five geoscientists from the Geocognition Research Laboratory and Geoeducation Research Interest Group listserv and two science educators from the Center for Research on College Science Teaching and Learning group on the pilot version of the survey instrument led to revisions. Analysis of novice responses in pilot data collection resulted in as many changes to the instrument as did expert feedback.
Conclusion validity, internal validity, credibility (see Lewis, 2009)	Conclusion validity is the measure of one's ability to determine the relationship, or lack thereof, between the variables being studied. This is a more general form of internal validity, which is most often considered when an attempt is made to determine a causal relationship between variables. In general, a researcher needs to ensure that they are not biasing study findings through personal expectations, their own actions, or failure to consider study limitations. For qualitative work, *credibility* also addresses researcher bias, and in particular the degree to which study participants agree with findings and the broader implications of the work.	We found this to be the most difficult metric of rigor and trustworthiness to evaluate. Experts exposed to our research findings during presentations at professional meetings generally agreed with the study findings and the implications for image redesign. Ultimately, we view credibility as the final step in the study validation process. As results become available for publication, we anticipate contacting interviewed experts to gauge their agreement with our general findings. In ongoing work, we are also investigating relationships, both causal and noncausal, among gender, confidence, and conceptual understanding.
Construct validity	A measure of whether or not strong support for the content of items exists. This can be estimated through both convergence and divergence of theory and reality. We expect concepts that should be related, such as expertise in plate tectonics and overall understanding of plate tectonics, to actually relate when measured by the instrument and scoring rubric. Similarly, concepts that need not be related, such as plate-tectonics understanding and attitude toward laboratory work, should not show significant correlation.	In general, participants with more expertise in geoscience received better scores on the survey instrument and provided more detailed responses. Interestingly, some misconceptions are retained until extreme levels of expertise are reached (Clark, 2009).
Criterion validity	The degree to which a measure correlates with other measures of the same latent trait (also called "concurrent" validity). Generally, qualitative measures are used to establish criterion validity for quantitative instruments, although quantitative or alternative qualitative measures (i.e., interviews) can be used to validate survey instruments.	Interviews with 61 subjects spanning the expert–novice continuum provided detailed confirmation of both the prevalence of ideas across multiple populations and our interpretations of survey results. For example, novice responses to the question of "How many tectonic plates are in the image?" are in strong agreement with novice responses from other instruments (Kortz et al., this volume).
Communication validity	Researchers develop surveys in order to generate an understanding of a study population. While researchers often assume that participants will interpret questions as intended, explicitly considering this aspect of instrument validity can generate important insights (e.g., Lopez, 1996).	Analyses of the survey instrument were enriched through comparison of researcher intentions with participant interpretations as recorded in think-alouds. The first 10 interviewees, undergraduate majors through experts, completed the survey instrument at the beginning of the interview; upon completion, they discussed their work and responded to interviewer probes about their thinking. Overall, we found that the geoscience major through expert population interpreted the survey instrument as we had intended. Communication validity for novices (nonmajors) was addressed in questions 1 and 2 as we modified the wording until nearly everyone who answered the questions was providing meaningful responses.
Cultural validity	A measure of the extent to which culture impacts participant interpretation of survey questions (Solano-Flores and Nelson-Barber, 2001). We consider this important in any effort to adopt or adapt established tools for new populations.	We do not know how culturally valid the survey instrument will be for subjects outside of the specific study population described here. Certainly, the survey instrument appears to be valid for undergraduates, graduate students and faculty affiliated with U.S. community colleges, and four-year institutions in the northeastern United States. Researchers interested in applying the survey instrument to other populations should consider whether or not cultural differences will require modification of the instrument.
Transferability	A measure of the extent to which results can be generalized to populations outside of the study. This validation is difficult to achieve, although the power of survey research lies in its ability to sample many populations, and hence generate measures of external validity.	Survey instruments were collected from 353 subjects (novices) enrolled in 5 different courses at two institutions (in Michigan and Rhode Island) and from 180 intermediate to expert subjects from an unknown number of institutions who were attendees at the 2008 GSA Annual Meeting. Interviews were conducted with 60 individuals across the expert–novice continuum from a range of universities and nations.

(*Continued*)

TABLE 1. RIGOR AND TRUSTWORTHINESS CRITERIA IMPORTANT FOR MIXED-METHODS SURVEY AND RUBRIC DESIGN *(Continued)*

Criteria	Description and approaches	Plate-tectonics survey instrument
Dependability	A measure of the extent to which other researchers would be able to replicate the study findings.	This manuscript is itself an audit trail of the survey instrument and rubric development, and it provides enough information for others to both evaluate the instrument's design and our analytical findings.
Internal consistency reliability	Although most often considered for quantitative instruments, internal consistency can provide a sense of the reliability of a mixed-methods survey. The stability of test results across samples of similar populations, consistency in test results over time, and generation of similar results using slightly different forms all provide evidence that a survey is generating reproducible findings.	Results from the piloted version through to the current version, separated by 14 mo, were similar, overall. Different forms (e.g., one-color versus two-color asthenosphere) produced the same range of responses outside of specific differences. As data analysis progresses, we will compare survey results from different universities; we would expect results to be consistent across populations once demographic or educational backgrounds are accounted for.
Inter-rater reliability	In qualitative design, inter-rater reliability can ensure that findings are reproducible. Often, this is established through an iterative process whereby multiple researchers code identical data and establish consistency in analytical results.	For the survey instrument, we utilized the inter-rater technique multiple times. Inter-rater reliability came into play at a number of analysis stages. Ultimately, we achieved 100% agreement in coding between two researchers; see text for details.

Notes: Except where noted, concepts of validity, reliability, and trustworthiness were adapted from Lincoln and Guba (1985), Litwin (1995), and Trochim and Donnelly (2007).



For each response, please mark the location on the scale that corresponds to your level of confidence

ON THE FIGURE ABOVE, PLEASE:

not confident at all ------------------------- very confident

1) Label all features related to plate tectonics

2) Show where you think melting could be occurring

3) Indicate relative direction plates are moving

4) What do the colors below the surface represent?

Figure 1. Original version of survey instrument (V1 in Fig. 2), with the one-colored asthenosphere.

rift, the hotspot, magma bodies, and the white area at the bottom of the original image. With these modifications, we then created two images: one with only an orange asthenosphere and another with the orange layer underlain by a yellow layer as seen in the original image.

The four questions initially written for the initial version (V1; Fig. 2) of the survey instrument (Fig. 1) were thoroughly discussed by the authors prior to dissemination of the survey for expert review. This version of the instrument (V1) was presented to our Geocognition Research Laboratory group in late November 2007, and to multidisciplinary (e.g., biology, chemistry, science education) members of the Center for Research on College Science Teaching and Learning at Michigan State University on 7 December 2007 for expert comments. Those comments led to changing the continuous confidence scale to a more easily quantifiable, numeric Likert scale, and making the image smaller so that the instrument could fit in a portrait alignment. This created more room under the image for questions, allowing a fifth question to be added: "Explain why melting occurs in the places you indicated in the figure" (V2).

This second version (V2) of the survey instrument was disseminated to the Geoeducation Research Interest Group listserv (geoed-research@list.msu.edu) on 5 February 2008. Feedback provided further expert validation of the instrument as well as initial ideas for the scoring rubric. On 19 February 2008, just prior to piloting the survey instrument in a nonscience majors class, the mantle lithosphere was thinned beneath the arcs so as to be more scientifically accurate (Strahler, 1998). This aspect change also aligns with the newest version of the web-based USGS image (Vigil and Tilling in Simkin et al., 2006; http://mineralsciences.si.edu/tdpmap/fom/xsection.htm). This version (V3) was pilot tested in a physical science for nonscience majors course (20 February 2008; *n* = 49) and in our initial, interview with a geoscience graduate student (26 February 2008).

The pilot testing of V3 provided a good example of how novices can notice aspects of an image that experts may not, and it illustrated how novices and experts can interpret questions differently. During the first interview, the interviewee saw and commented on an island and guyot that had not been masked in the image, and one of the students in the pilot course labeled the island as a hotspot (Fig. 1). We had previously removed the obvious hotspot feature in the image, and now recognized the need to remove the island and guyot from the survey instrument (V4). In reviewing the student responses to question 1, we realized we needed to modify how the question was worded. With the original version of: "Label anything related to plate tectonics," some respondents wrote "PT" over areas of the map that they felt were related to plate tectonics. The question was intended to probe a participant's ability to name specific features, and a response of "PT" was too generic for interpretation. Such a response could mean that (1) the respondent knows the name of the feature, but thinks that a generic label is an appropriate answer; (2) the respondent cannot remember the name of the feature; or (3) the respondent thinks the feature is related to plate-tectonic processes but is unsure. In an effort to minimize generic responses, we modified question 1 to read: "Identify anything related to plate tectonics."

Responses to question 2 in the pilot class resulted in rephrasing, as well. The original version read: "Show where you think melting could be occurring." Some respondents circled areas to indicate where they thought melting could occur; others wrote the word, "melting." When respondents used a circle, it tended to encircle an area such as a volcano, a trench, the subducting slab, or the "tip" of the subducting slab. However, when respondents wrote, "melting," it was sometimes written near a volcano, but not necessarily over the peaks of the volcanoes or below the volcano. "Melting" was also written near a subducting slab, or in the mantle next to the "tip" of the slab but not directly over it. Because the "tips" of the slabs and the volcanoes were very commonly circled responses, we felt it was likely that those who wrote "melting" near to, but not on top of, these features were likely indicating those features. However, our uncertainty in participant intentions prohibited precise coding of these "melting" data. As a consequence, question 2 was rephrased to read: "Circle areas below the surface where you think melting is occurring." This modification improved our ability to accurately code responses.

In addition to the pilot testing, we were continuously open to modifying the survey instrument in response to participant data. Throughout the study, interviews with participants whose

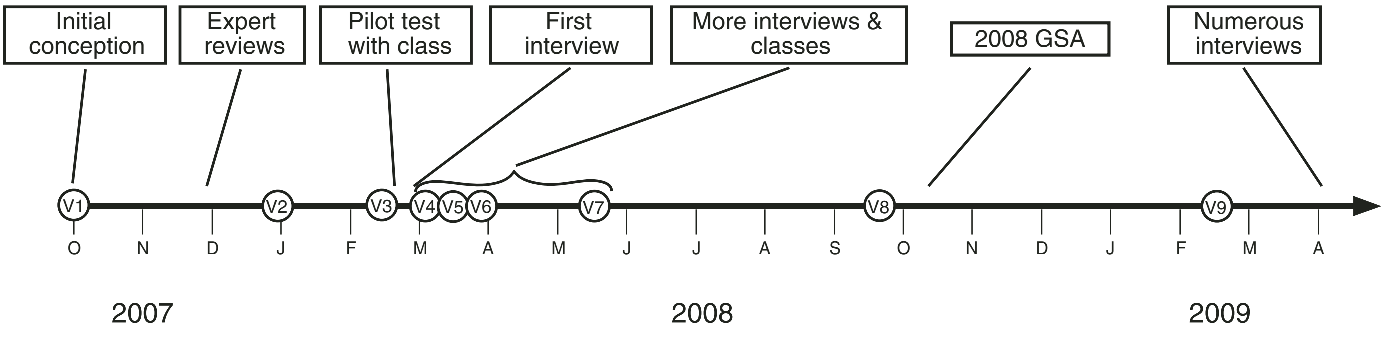


Figure 2. Time line showing the evolution of the survey instrument, indicated by version (V1–V9), and data collection events. After version 6, no changes were made to the image or to the wording of the original five questions. Versions 7 and 8 incorporate novel questions.

geoscience background ranged from novice to expert were conducted concurrently with self-administered survey instruments completed by both college-level students and attendees to a national geoscience conference (Fig. 2). Data from these interviews and completed survey instruments yielded results that both supported survey design and suggested necessary modifications. A modified version (V5) of the survey was disseminated to three courses in mid- to late March 2008. A few of the responses continued to not explicitly identify the geologically relevant features, so we again rephrased question 1 to read: "Identify by name any features related to plate tectonics." This modification further reduced the number of generic responses, and highlighted the need to be open to modifying questions to accommodate differences between our and the participants' reading (see communication validity in following). This version (V6) of the survey instrument was used through the end of April 2008, and no further changes were made to the image or to this initial set of questions.

As our data collection progressed, we obtained responses that led us to add more questions to the survey. For example, many respondents stated that the orange color represented magma. To gain additional information, we added the question, "Estimate the percentage of the mantle that is liquid (magma)." Given our focus on investigating peoples' fundamental understanding of plate tectonics, we also added the question, "Explain what causes tectonic plates to move." This seventh version (V7) was completed in late April and May 2008, while the final version (V8) was completed in September 2008 with the addition of one final question: "How many tectonic plates are in the image? (Number the plates on the figure.)" This question, which was based on a discussion between the first author and Mark Reagan, an igneous petrologist at a public Midwestern university, is in line with our research objectives, and has provided a wealth of information (e.g., Kortz et al., this volume). Both versions 7 and 8 were used during a data collection effort at the GSA Annual Meeting in October 2008. Version 7 was used in a booth where meeting attendees were invited to complete a survey; 182 attendees filled out the survey at the meeting. Version 8 was used during the 11 interviews that took place at the meeting. The current wording of the last two questions were finalized in February 2009, while the first author was working with Karen Kortz and one of her students to design a slightly modified survey instrument (see Kortz et al., this volume). These nine questions comprise the current version (V9) of the survey instrument (Fig. 3), which was subsequently used in 42 interviews between March and April 2009. Responses to the four questions that were added after the pilot testing of the survey instrument were continuously monitored for any communication validity issues. We did not detect any misunderstanding arising between the targeted concept of the questions and study participants' responses.

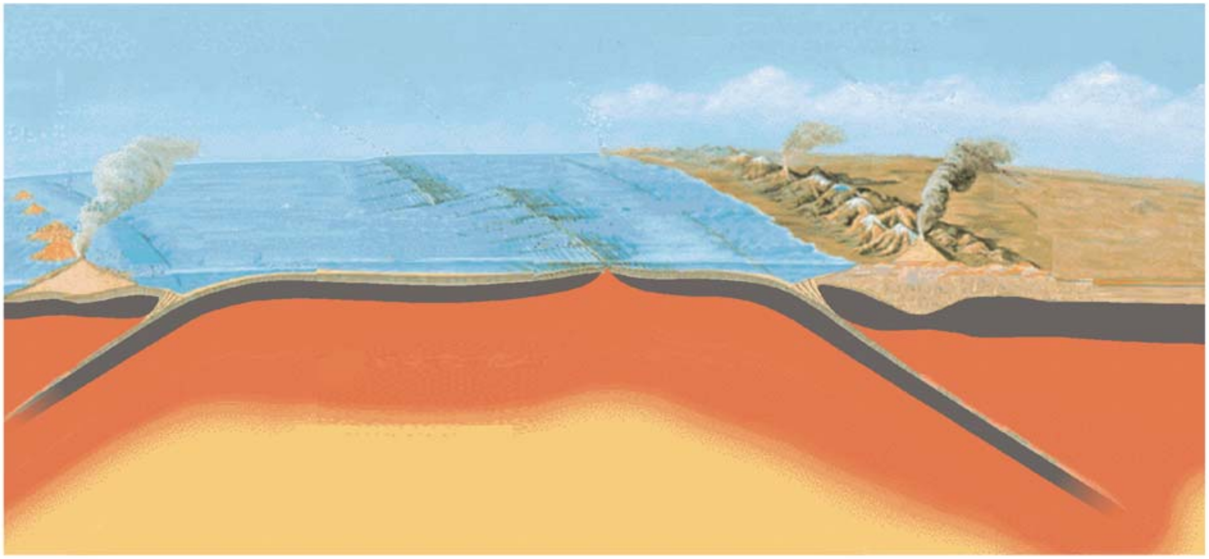
The preceding discussion illustrates how the instrument evolved concurrently with data collection. Although this does not preclude us from interpreting both early and later data, we do acknowledge that changes to questions can have an effect on subject responses. For example, responses of "melting" in answer to the question, "Show where you think melting could be occurring" cannot be interpreted as rigorously as the responses to the rephrased version, "Circle areas below the surface where you think melting is occurring." Differences in coding for these two versions of question 2 reflect this modification in wording, rather than differences in student conceptual understanding.

Rubric Design

The scoring rubrics used to analyze survey results were developed via iterative thematic content analysis of collected data (see Patton, 2002; Denzin and Lincoln, 1998). Rubric design required about 3 mo of discussion, application, and revision, with further improvements occurring as our contextual analysis continued. Initial versions of the scoring rubrics utilized codes developed during analyses of pilot data. For each question analyzed, the authors independently conducted thematic content analysis on a subset of surveys and discussed their observations. The first author then developed a preliminary scoring rubric for each question based on the analyses and points raised during discussion. Subsequent discussion between both authors yielded a scoring rubric that was grounded in the data and that could be easily explained to an undergraduate coder.

A discussion of the development of the rubrics for questions 1 and 2 can provide insights into the effort required to fully develop these rubrics. For question 1, respondents' terms were originally categorized into number of correct terms, number of incorrect terms, and total number of terms used. This approach was abandoned in favor of scoring each, individual term used by each respondent as correct, incorrect, or partially correct/incomplete. This coding scheme is effective at providing insight into participant understanding and use of language, and led to the construction of a list of commonly used terms against which each newly scored response can be compared. The list of commonly used terms grew as the study population expanded from mostly novices to include more intermediate to expert participants. For example, as one might expect, most novices do not identify features such as a forearc basin in the image. However, enough attendees at the GSA meeting in October 2008 did use the term to warrant its addition to the list. Other changes to the list dealt with how nuanced differences in term usage are handled. One example is whether or not to have a separate code for use of the word "crust" when it is used without a modifier to label the oceanic crust versus the continental crust. If a respondent writes only "crust" and uses an arrow or line to indicate that they are labeling the continental crust, then one might presume that the respondent intended the term to be understood as "continental crust." However, without a follow-up interview, ambiguity remains as to whether the term was intended to specifically label the continental crust or crust, in general. This issue was most clearly seen when a respondent labeled only one surface feature as "crust."

Responses to question 2 were the most difficult to analyze, and as with question 1, our initial scoring rubric was discarded



For each response, please circle the number that most closely corresponds to your confidence level.

ON THE FIGURE ABOVE, PLEASE:

	not confident -------- at all				very confident
1) Identify by name any features related to plate tectonics.	1	2	3	4	5
2) Circle areas below the surface where you think melting is occurring.	1	2	3	4	5
3) Use arrows to indicate the relative direction tectonic plates are moving.	1	2	3	4	5
4) Draw a line along each plate boundary and identify the type of each of the boundaries.	1	2	3	4	5
5) How many tectonic plates are in the image? Number of tectonic plates: ______ Number the plates on the image.	1	2	3	4	5

IN THE SPACE BELOW (AND ON THE BACK, IF NECESSARY), PLEASE:

6) Explain what the colors below the surface represent.	1	2	3	4	5
7) Explain why melting occurs in the places you indicated in the figure.	1	2	3	4	5
8) Estimate the percentage of the mantle that is liquid (magma).	1	2	3	4	5
9) Explain what causes tectonic plates to move.	1	2	3	4	5

Figure 3. Current version of survey instrument (V9 in Fig. 2), with the two-colored asthenosphere. A one-colored asthenosphere version is also used.

as we endeavored to accurately represent the intent of the responses. During the summer of 2008, we designed and revised a rubric (Figs. A1 and A2) to the point where we attained an initial inter-rater agreement of 80% between the two authors, and a postdiscussion, inter-rater agreement of 100% on a set of 20 randomly selected survey instruments. The majority of the nonagreement was due to missed coding of terms. This inter-rater process occurred over a number of weeks, and although this rubric did allow us to code those areas that were most frequently indicated by respondents, it was overly complicated and did not necessarily align with the circles given by the subjects' responses. In looking over our initial approach, we had not truly allowed the data to speak for itself. We were literally trying to fit round pegs, the data, into rectangular holes, our rubric (Fig. A1). We abandoned this initial rubric and created a new rubric (Fig. A3) that more closely aligned scores with how subjects marked the image. Most of this revised rubric (i.e., the first 11 groupings) was developed over 2 wk in August 2008. As analyses proceeded, three more groupings were added to account for new themes as observed in responses. We found that the protocol needed to be very explicit in order to maintain a high inter-rater agreement and for temporal consistency for individual raters (see also Ambrose et al., 2004; Bresciani et al., 2009, and references therein). For example, the diagonal lines perpendicular to the subducting slabs were added as a guide for determining whether a specific circle was to be coded as a "4" or a "5." If the center of a subject's circle was above the line, then it was coded as a "4"; if the center of the circle was below the line, it was coded as a "5."

Both authors were involved in the development of all coding rubrics. The first author and one trained, undergraduate geoscience major coded questions 1–3 for a randomly selected set of 60 completed surveys. Training consisted of a discussion of the objectives of the study, the design of the survey instrument, and intended approaches for use of the scoring rubrics. Prior to coding the data, the student rater practiced applying the rubrics. During this phase of training, both authors worked with the student rater to clarify how to apply the rubrics to the data set.

Agreement of independently obtained codes between the first author and the student rater was initially 81.5%, 83.5%, and 90%, for questions 1–3, respectively. After initial scoring, the researchers discussed their scores, and attained a consensus agreement. The majority of the nonagreement was due to missed terms. After this establishment of inter-rater reliability, the undergraduate rater scored a further 184 surveys, independently. As a further step in our validity, she flagged ambiguous responses for later inter-rater discussion. To date, we have developed reliable rubrics for the first three questions. Rubrics for questions 4–9 have not been developed with the same rigor as with the first three questions because current scoring of these responses is not sensitive to nuances in answers. That said, as we continue to analyze our data, we will continue to assess the coding for all of the questions, and will revise and even construct new rubrics if and when that becomes necessary.

SUMMARY

The often circuitous and iterative development pathways described herein provided measures of a number of forms of validity and reliability for both the survey instrument and the rubrics used to score the instrument. Although we did not necessarily set out to establish all of these measures, retrospective evaluation of our research design was made possible through careful record keeping, which allowed us to document an audit trail. The development, validation, and scoring of a mixed-methods survey instrument is difficult and nonlinear; the right-hand column in Table 1 is derived from the culmination of the piloting, revision, and analytical blind alleys described here. At this stage in the research project, we can easily articulate the forms of validity and reliability that have been addressed, intentionally or unintentionally. We also note that we have not addressed all types of validity and reliability that may be considered important for survey instrument development. Table 1 provides explicit details of how each form of validity and reliability was, or was not, addressed.

Intentional Forms of Validity, Reliability, and Trustworthiness

Several forms of validity and reliability were intentionally targeted in our research design. In particular, we knowingly established content and conclusion validity, inter-rater reliability, credibility, dependability, and transferability of our work (Table 1). Content validity was established early in our work through collection of expert feedback on the survey instrument, including both design and content. In addition to expert opinion, we utilized novice responses to early versions of the survey instrument to inform revisions (see also communication validity).

Conclusion validity and credibility are both inherently difficult to measure and should be reviewed well after a study is considered completed. Bias in our interpretations was limited through careful discussion of findings and implications within our research group. In addition, oral and poster presentation of this research at professional meetings and in seminars exposed a variety of experts to our study conclusions; in general, experts agreed with our interpretations of the data in terms of expert-novice trends and implications for knowledge representation in images. Finally, and as documented herein, we carefully considered inter-rater reliability in designing assessment rubrics. Each of these forms of rigor and trustworthiness, coupled with the detailed description of our survey and rubric design as documented in this manuscript, lends dependability to our study (see Libarkin and Kurdziel, 2002) and provides a mechanism for other researchers to evaluate their agreement with our overall conclusions.

A limitation of this study is that while we did not particularly request participation from individuals, we did target specific entry-level courses and specific levels of expertise. As a result our sample is not entirely random; this is an inherent limitation to any survey research. Therefore, although one can never completely address transferability of study findings, we sampled as broad and diverse a population as was feasible. While we cannot assume that our findings are applicable to all members of the expert-novice population, we have sampled broadly in terms of numbers and geographic distribution (Table 1) in an attempt to provide some far-reaching, and hence transferable, significance to our work. Finally, we acknowledge the importance of cultural validity to establishing transferability. Although we did not explicitly address cultural validity in our work, we encourage those interested in adapting this instrument to other cultures to consider the appropriateness of the survey design to their targeted demographic.

Unintentional Forms of Validity and Reliability

Although our intention was to construct an instrument that would provide insights into the conceptions held by individuals across the expert–novice continuum, we did not recognize the potential for documenting construct validity until we began analyzing our data and documenting the detailed responses of experts (Table 1). In particular, the most experienced experts provided more thorough and accurate responses than novices. Similarly, criterion validity was recognized through poststudy comparison of interview with survey results, as well as on a smaller scale through comparison with data collected in an unrelated study (Kortz et al., this volume). Although we did not intentionally target communication validity early in the study, some student responses to question 1 were initially so generic that they prompted us to revise the question until nearly everyone who answered the question provided feature-specific labels. Finally, the duration of our data collection and use of multiple forms provided us with a way to address internal consistency reliability. In particular, we find that the results from the survey instrument, separated by 14 mo and representing several different versions, are consistent across implementations.

REFLECTIONS ON THE PROCESS

In many ways, our research proceeded in ways that are similar to a stereotypical natural science research project. This project began with a question: the first author looked at a textbook image of plate-tectonic processes and asked himself, "Is this image confusing to students?" This led to a hypothesis: "The differences between how novices and experts view plate-tectonic representations can create barriers to learning." We felt we could investigate people's perceptions of plate tectonics, and study the role played by an image in affecting people's perceptions of plate tectonics in a well-designed survey instrument. We designed our instrument and then performed an initial check of the rigor and trustworthiness of the instrument through expert review and pilot testing. Next, we collected the bulk of our data while concurrently beginning our data analysis. Currently, we are continuing our analysis and documenting our findings for dissemination in publications. Our time line from initial conceptualization in October 2007, Institutional Review Board (IRB) approval for research with human subjects in January 2008, first implementation of the instrument in February 2008, presentation of initial findings on novices in October 2008 (Clark and Libarkin, 2008), receipt of National Science Foundation (NSF) funding in January 2009, to submitting research findings for publication in 2010, follows that of a typical research project.

One difference is that we had to create the instrument needed for measuring the traits we were interested in studying. Although instrument development is done in the natural sciences, it is not typical for most projects. Just as in the natural sciences, where an instrument's accuracy and precision must be determined, we needed to determine the rigor and trustworthiness of our instrument. For this project, rigor and trustworthiness steps required about the same amount of effort as was needed for designing the instrument. Indeed, rigor and trustworthiness testing is an ongoing process. We have asked ourselves, "When do we stop modifying a rubric?" Although we achieved 100% postinstruction inter-rater agreement on our first rubric for question 2, we felt our approach was not aligning well enough with how respondents answered the question. The revised rubric has been further tweaked at least three times, but any future potential benefits of refining our interpretations that might be gained through additional changes must be weighed against the need to be able to compare earlier scored surveys against more recently scored surveys, possibly requiring rescoring of all surveys. We feel our current rubrics are effective, while accepting that they are not perfect. At some point we have to say, "It's good enough."

In the normal course of doing research, we expected to repeatedly modify the instrument, recruit for and schedule interviews, recruit professors who would allow us access to their students, and obtain IRB approval of the instrument and study methods. During interviews, the first author encouraged subjects to provide as much detail as they wished in their explanations while trying to avoid leading questions, without coming across as didactic, and without making value judgments on responses. When subjects provided what was deemed to be an interesting explanation of a plate-tectonic process, whether scientifically valid or not, the goal was to probe deeply so as to obtain as much insight into subject's thoughts on the topic as possible (Kvale and Brinkmann, 2009).

An unexpected aspect of the research has been the unique challenge posed by interviewing experts. Whether asking experts questions that they perceived as too simple or pressing them to explain their reasoning on a topic for which they held an alternative conception, one has to be careful to not inadvertently offend the participant. Although neither author claims to be an expert in all facets of plate-tectonics research, as interviewers we needed to be well informed on the topic. Other facets that were not necessarily foreseen in the planning stages included how to handle the amount of data that quickly became quite substantial. Part of this data accumulation was due to addition of new questions to the instrument as the study progressed. This could be considered a problem of riches because those additional questions provided important insights into the ways in which many plate-tectonic concepts are perceived along the expert–novice continuum. We also learned that one needs to be willing to scrap weeks of work invested in a rubric, and to design rubric protocols that are clear and explicit. The fewer interpretations in data analysis that are left to the discretion of a coder, the more likely that coder is to score the same survey the same way each time, and the more likely two coders are to score a survey similarly.

We feel that this instrument and associated rubrics are providing a wealth of data, and we feel that we did need to create this survey instrument. However, we would encourage researchers to adopt preexisting valid and reliable research instruments, whenever possible. When it is not possible, be prepared to invest a significant amount of time and effort in creating, validating, and revising your instrument and scoring rubric.

APPENDIX

Coding rubric protocols for question 2. The original protocol (Fig. A1) was implemented in August 2008, but it was replaced by the currently used protocol (Fig. A3) starting in September 2008.

Original Protocol:

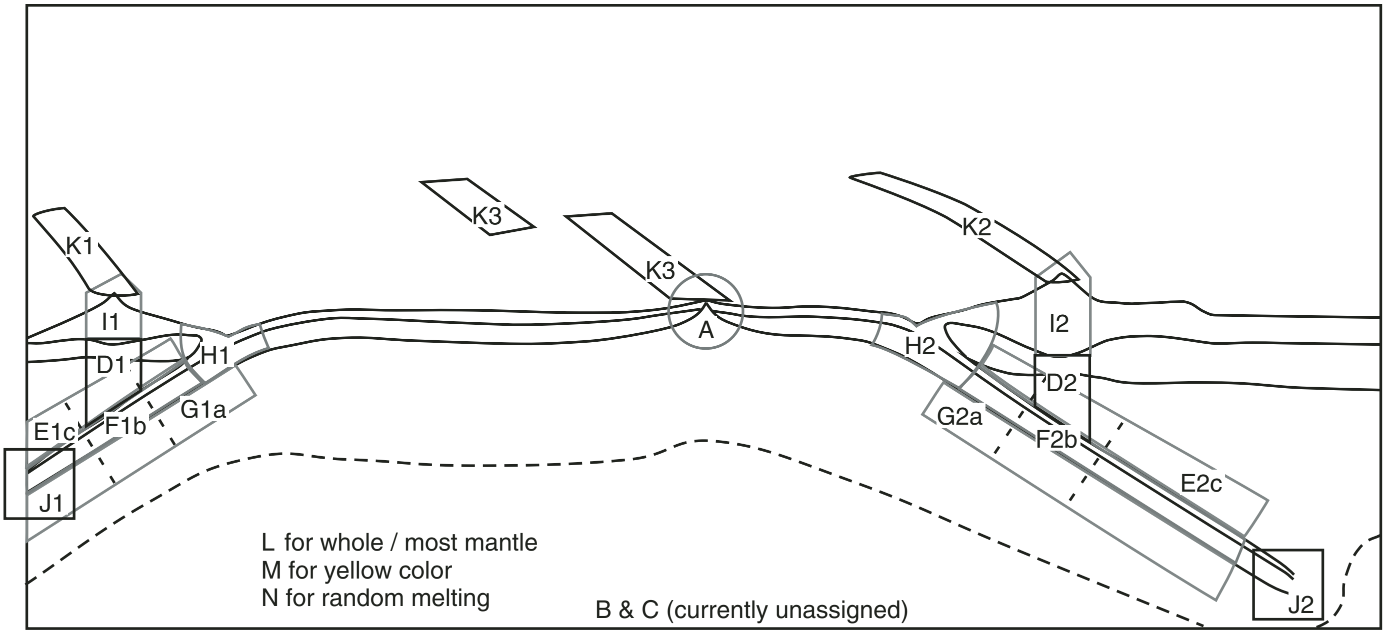


Figure A1. Early version of coding template for question 2. The template was printed on a transparency that was laid over a subject's responses.

If a circle encompasses ≈≥50% of a labeled zone, count that zone.

If ≈≥50% of a circle is within a zone, count that zone.

The subdivisions, **a, b,** and **c** of **E, F,** and **G** are designed to capture circled areas within those zones. E, F, and G are designed to capture ellipses parallel to the subducting slab. An ellipse of **E1a** and **E1b** would be **E1**, but a circle of **E1a, E1b, F1a, F1b, G1a,** and **G1b** would be listed as all of those.

D: Use for that specific area or circles in that area—do NOT include **D** in ellipses along slab.

K: Include any circled areas over volcanic peaks or mid-ocean ridges—except those responses that are centered on **A** or **I.**

J: If a circle is interpreted to represent the "end" of the slab, it should be adjudged as **J** regardless of its size.

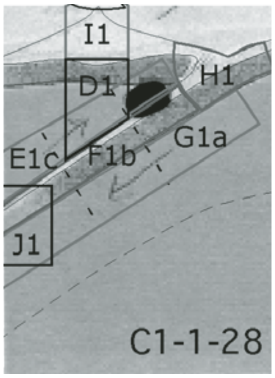
L: Use if respondent indicates all or most of the mantle.

M: Use if respondent's circle(s) or "melting indicators" are random, arbitrary, or not included within defined zones.

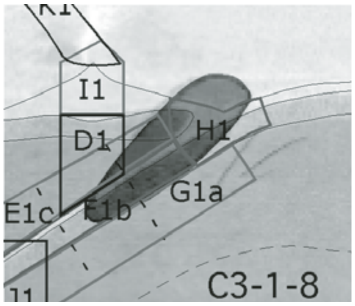
N: Use if respondent circled the orange color (of two-color mantle images).

O: Use if respondent circled the yellow color (of two-color mantle images).

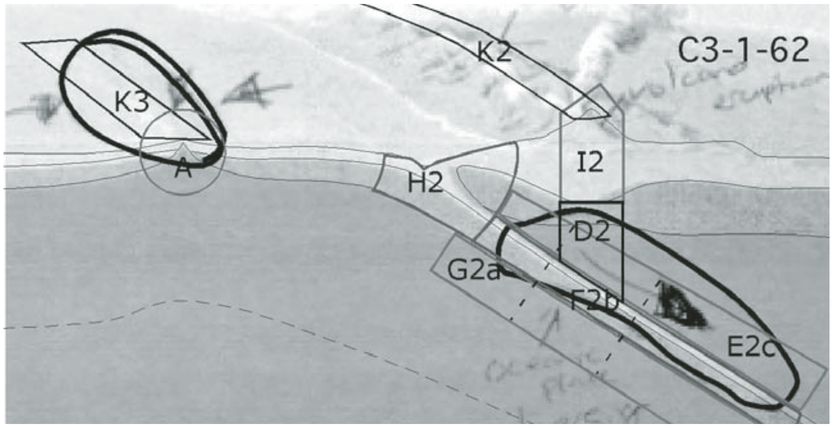
P: Use if melting is indicated by something other than circles (e.g., text or arrows).



code as: E1a



code as:
E1a, F1a, F1b, H1



code as: K3, and as E2, F2

code as: I1, K1, and as I2, K2

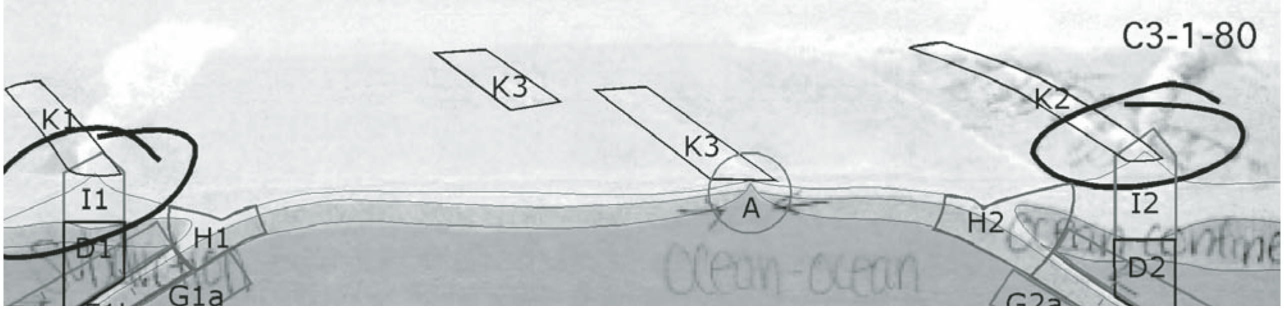
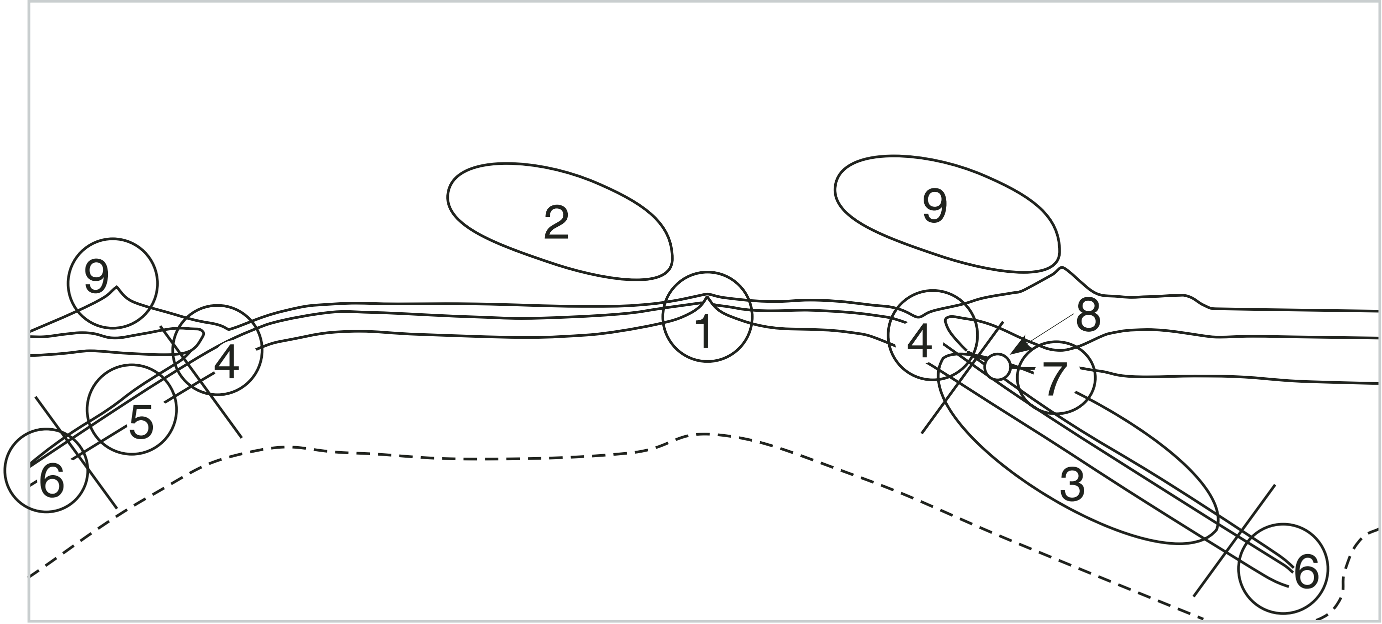


Figure A2. Images used to clarify how to apply the scoring of the template in Figure A1. Subject responses have been accentuated with a dark line or with a darkened area.

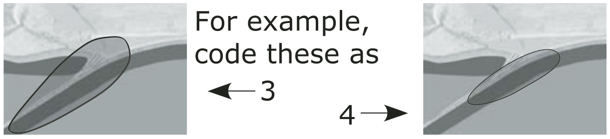


PAY ATTENTION TO ANY LABELS ON CIRCLES (Not all circles indicate melting. See code 12)

Use notes column when necessary to clarify a code, esp. useful for codes 10 and 11

Circles along the descending slabs whose center is within the diagonal lines are coded as 5

Codes apply to either or both sides of figure. For example, if a respondent circles the 'tips' of *either* or *both* plates, code this as a 6



'circle' and 'ellipse' are used in a relative, not exact sense

Every melting area should receive only ONE CODE

For a circle that covers area 4 but looks to also include 8, code that as only a 4; 8 is specifically for small circles in the corner of the mantle wedge.

CODES

1	circle at divergent boundary
2	circles over ocean ridges
3	ellipse along a significant part of subducting plate (inc. directly above &/or below plates)
4	circle over trench(es)
5	circle along middle of subducting plate(s)
6	circle at bottom 'tip' of subducting plate(s)
7	circle in mantle wedge directly below volcanoes
7b	circle centered below a volcano but above asthenosphere
8	small circle in corner of mantle wedge
9	circle over volcanoes
10	circle over area outside of codes 1 - 9
11	something other than circles indicating melting (e.g., text, arrows)
12	circles that indicate something other than melting
13	no indication of melting by circles, text, arrows, etc.
14	circle includes mantle wedge ± crust ± trench ± upper section of descending slab. This circle must be too large to be classified as 5, 7, or 8, and is not a 9. Circle may include parts of 4, 5, 7, 8, & /or 9.

Example of a #14 code

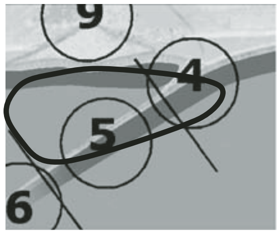


Figure A3. Current rubric for scoring question 2. Note that codes are symmetric for the subducting slabs. That is, both slabs are coded for "3," "4," and "5" circles. The rubric is printed on a transparency that is laid over a subject's responses. Subject responses have been accentuated with a dark line or with a darkened area.

ACKNOWLEDGMENTS

We thank everyone who participated in this research. Discussions with and assistance from the co–principal investigators of National Science Foundation project 0837185 and members of the Geocognition Research Laboratory at Michigan State University, especially Sarah Jordan, are appreciated. This manuscript benefited greatly from comments by Alison Stokes and two anonymous reviewers. Partial support for this work was provided by the National Science Foundation's DUE Course, Curriculum, and Laboratory Improvement program under awards 0717790 and 0837185.

REFERENCES CITED

Adams, W.K., Perkins, K.K., Podelefsky, N.S., Dubson, M., Finkelstein, N.D., and Wieman, C.E., 2006, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey: Physical Review Special Topics–Physics Education Research, v. 2, p. 010101:1–14.

Ambrose, R., Clement, L., Philipp, R., and Chauvot, J., 2004, Assessing prospective elementary school teachers' beliefs about mathematics and mathematics learning: Rationale and development of a constructed-response-format beliefs survey: School Science and Mathematics, v. 104, no. 2, p. 56–69, doi:10.1111/j.1949-8594.2004.tb17983.x.

Bandura, A., 1984, Recycling misconceptions of perceived self-efficacy: Cognitive Therapy and Research, v. 8, no. 3, p. 231–255, doi:10.1007/BF01172995.

Beatty, P.C., and Willis, G.B., 2007, Research synthesis: The practice of cognitive interviewing: Public Opinion Quarterly, v. 71, no. 2, p. 287–311, doi:10.1093/poq/nfm006.

Bresciani, M.J., Oakleaf, M., Kolkhorst, F., Nebeker, C., Barlow, J., Duncan, K., and Hickmott, J., 2009, Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines: Practical Assessment, Research, & Evaluation, v. 14, no. 12, p. 1–7. (Available online at: http://pareonline.net/getvn.asp?v=14&n=12.)

Britton, E.D., and Schneider, S.A., 2007, Large-scale assessments in science education, *in* Abell, S.K., and Lederman, N.G., eds., Handbook of Research on Science Education: Mahwah, New Jersey, Lawrence Erlbaum Associates, p. 1007–1040.

Clark, S.K., 2009, Plate tectonics as viewed by novices and experts: Geological Society of America Abstracts with Programs, v. 41, no. 7, p. 251.

Clark, S.K., and Libarkin, J.C., 2008, Postinstruction alternative conceptions about plate tectonics held by nonscience majors: Geological Society of America Abstracts with Programs, v. 40, no. 6, p. 364.

Collins, D., 2003, Pretesting survey instruments: An overview of cognitive methods: Quality of Life Research, v. 12, p. 229–238, doi:10.1023/A:1023254226592.

Creswell, J.W., 2003, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (2nd ed.): Thousand Oaks, California, Sage Publications, 246 p.

Creswell, J.W., and Miller, D.L., 2000, Determining validity in qualitative inquiry: Theory into Practice, v. 39, no. 3, p. 124–130, doi:10.1207/s15430421tip3903_2.

Crotty, M., 1998, The Foundations of Social Research: Thousand Oaks, California, Sage Publications, 248 p.

Denzin, N.K., and Lincoln, Y.S., eds., 1998, Strategies of Qualitative Inquiry: Thousand Oaks, California, Sage Publications, 346 p.

Driver, R., 1989, Students' conceptions and the learning of science: International Journal of Science Education, v. 11, p. 481–490, doi:10.1080/0950069890110501.

Duschl, R.A., Schweingruber, H.A., and Shouse, A.W., eds., 2007, Taking Science to School: Learning and Teaching Science in Grades K–8: Washington, D.C., National Academies Press, 387 p.

Feig, A., 2010, this volume, Methodology and location in the context of qualitative data and theoretical frameworks in geoscience education research, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(01).

Fink, A.G., 2003, The Survey Handbook (2nd ed.): Thousand Oaks, California, Sage Publications, 184 p.

Golafshani, N., 2003, Understanding reliability and validity in qualitative research: Qualitative Report, v. 8, no. 4, p. 597–607.

Greene, J.C., Caracelli, V.J., and Graham, W.F., 1989, Toward a conceptual framework for mixed-method evaluation designs: Educational Evaluation and Policy Analysis, v. 11, no. 3, p. 255–274.

Harp, S.F., and Mayer, R.E., 1998, How seductive details do their damage: A theory of cognitive interest in science learning: Journal of Educational Psychology, v. 90, no. 3, p. 414–434, doi:10.1037/0022-0663.90.3.414.

Hossler, D., and Vesper, N., 1993, An exploratory study of the factors associated with parental saving for postsecondary education: The Journal of Higher Education, v. 64, no. 2, p. 140–165, doi:10.2307/2960027.

Johnson, R.B., and Onwuegbuzie, A.J., 2004, Mixed methods research: A research paradigm whose time has come: Educational Researcher, v. 33, no. 7, p. 14–26, doi:10.3102/0013189X033007014.

Kidder, L.H., and Fine, M., 1987, Qualitative and quantitative methods: When stories converge: *in* Mark, M.M., and Shotland, R.L., eds., Multiple Methods in Program Evaluation: San Francisco, California, Jossey-Bass, p. 57–75.

Kortz, K.M., Clark, S.K., Gray, K., Smay, J.J., Viveiros, B., and Steer, D., 2011, this volume, Counting tectonic plates: A mixed-methods study of college students' conceptions of plates and boundaries, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(12).

Kvale, S., and Brinkmann, S., 2009, Interviews: Learning the Craft of Qualitative Research Interviewing: Thousand Oaks, California, Sage Publications, 354 p.

Lederman, N.G., Abd-El-Khalick, F., Bell, R.L., and Schwartz, R., 2002, Views of nature of science questionnaire: Toward valid and meaningful assessment of learner's conceptions of nature of science: Journal of Research in Science Teaching, v. 39, p. 497–521, doi:10.1002/tea.10034.

Lewis, J., 2009, Redefining qualitative methods: Believability in the fifth moment: International Journal of Qualitative Methods, v. 8, no. 2, p. 1–14.

Libarkin, J.C., 2008, Concept inventories in higher education science, *in* National Research Council, Promising Practices in Undergraduate STEM Education Workshop 2 (Washington, D.C., 13–14 October 2008): http://www7.nationalacademies.org/bose/PP_Commissioned_Papers.html (accessed 30 October 2009).

Libarkin, J.C., and Kurdziel, J.P., 2002, Research methodologies in science education: Qualitative data: Journal of Geoscience Education, v. 50, no. 2, p. 195–200.

Lincoln, Y.S., and Guba, E.G., 1985, Naturalistic Inquiry: Beverly Hills, California, Sage Publishing, 416 p.

Litwin, M.S., 1995, How to Measure Survey Reliability and Validity: Thousand Oaks, California, Sage Publications, 87 p.

Lopez, W., 1996, Communication validity and rating scales: Rasch Measurement Transactions, v. 10, no. 1, p. 482–483.

Marshall, C., and Rossman, G.B., 2006, Designing Qualitative Research: Thousand Oaks, California, Sage Publications, 262 p.

Maxwell, J.A., 2005, Qualitative Research Design: An Interactive Approach (2nd ed.): Thousand Oaks, California, Sage Publications, 192 p.

Merriam, S.B., 2002, Introduction to qualitative research, *in* Merriam, S.B., ed., Qualitative Research in Practice: San Francisco, California, Jossey-Bass, p. 3–17.

Morse, J.M., Barrett, M., Mayan, M., Olson, K., and Spiers, J., 2002, Verification strategies for establishing reliability and validity in qualitative research: International Journal of Qualitative Methods, v. 1, no. 2, p. 13–22.

Onwuegbuzie, A.J., and Johnson, R.B., 2006, The validity issue in mixed research: Research in the Schools, v. 13, no. 1, p. 48–63.

Patton, M.Q., 2002, Qualitative Evaluation and Research Methods (3rd ed.): Thousand Oaks, California, Sage Publications, 598 p.

Phillips, D.C., and Burbules, N.C., 2000, Postpositivism and Educational Research: Lanham, Maryland, Rowman & Littlefield Publishers, 101 p.

Posner, G.J., Strike, K.A., Hweson, P.W., and Gertzog, W.A., 1982, Accommodation of a scientific conception: Toward a theory of conceptual change: Science Education, v. 66, no. 2, p. 211–227, doi:10.1002/sce.3730660207.

Presser, S., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., Rothgeb, J.M., and Singer, E., 2004, Methods for testing and evaluating survey questions:

Public Opinion Quarterly, v. 68, no. 1, p. 109–130, doi:10.1093/poq/nfh008.

Rossman, G.B., and Wilson, B.L., 1985, Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study: Evaluation Review, v. 9, no. 5, p. 627–643, doi:10.1177/0193841X8500900505.

Sanchez, M.E., 1992, Effects of questionnaire design on the quality of survey data: Public Opinion Quarterly, v. 56, no. 2, p. 206–217, doi:10.1086/269311.

Simkin, T., Tilling, R.I., Taggart, J.N., Jones, W.J., and Spall, H., 1989, This Dynamic Planet: World Map of Volcanoes, Earthquakes, and Tectonic Plates: U.S. Geological Survey (in collaboration with the Smithsonian Institution) Map I-2800, scale 1:30,000,000.

Simkin, T., Ungar, J.D., Tilling, R.I., Vogt, P.R., and Spall, H., 1994, This Dynamic Planet: World Map of Volcanoes, Earthquakes, and Tectonic Plates (2nd ed.): U.S. Geological Survey (in collaboration with the Smithsonian Institution and U.S. Naval Research Laboratory) Map I-2800, scale 1:30,000,000.

Simkin, T., Tilling, R.I., Vogt, P.R., Kirby, S.H., Kimberly, P., and Stewart, D.B., 2006, This Dynamic Planet: World Map of Volcanoes, Earthquakes, Impact Craters, and Plate Tectonics: U.S. Geological Survey (in collaboration with the Smithsonian Institution and U.S. Naval Research Laboratory) Geologic Investigations Series Map I-2800, scale 1:30,000,000 (and companion website: http://www.minerals.si.edu/tdpmap).

Siragusa, L., and Dixon, K.C., 2006, A research methodology: The development of survey instruments for research into online learning in higher education: Issues in Educational Research, v. 16, p. 206–225.

Solano-Flores, G., and Nelson-Barber, S., 2001, On the cultural validity of science assessments: Journal of Research in Science Teaching, v. 38, no. 5, p. 553–573, doi:10.1002/tea.1018.

Strahler, A.N., 1998, Plate Tectonics: Cambridge, Massachusetts, GeoBooks Publishing, 554 p.

Tashakkori, A., and Teddlie, C., 1998, Mixed Methodology: Combining Qualitative and Quantitative Approaches: Thousand Oaks, California, Sage Publications, 185 p.

Tashakkori, A., and Teddlie, C., 2003, Handbook of Mixed Methods in Social & Behavioral Research: Thousand Oaks, California, Sage Publications, 768 p.

Thomas, S.J., 2004, Using Web and Paper Questionnaires for Data-Based Decision Making: From Design to Interpretation of the Results: Thousand Oaks, California, Corwin Press, 194 p.

Trochim, W.M.K., and Donnelly, J.P., 2007, The Research Methods Knowledge Base (3rd ed.): Cincinnati, Ohio, Atomic Dog Publishing, 361 p. (companion website: http://www.socialresearchmethods.net/kb/).

U.S. Geological Survey Education webpage, 2010, USGS Map Databases: http://education.usgs.gov/common/map_databases.htm (accessed 12 March 2010).

Manuscript Accepted by the Society 23 June 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

Qualitative methods applied in the development of an introductory oceanography concept inventory survey

Leilani Arthurs*
Department of Geology & Geography, Georgia Southern University, P.O. Box 8149, Statesboro, Georgia 30460-0002, USA

Thomas Marchitto*
Department of Geological Sciences, University of Colorado, Campus Box 399, 2200 Colorado Avenue, Boulder, Colorado 80309-0399, USA

ABSTRACT

Concept inventories are relatively new types of diagnostic instruments intended to measure student learning. Concept inventories exist for astronomy, biology, chemistry, engineering, fluid mechanics, geology, and physics. None is yet available for oceanography, and our work to construct the Introduction to Oceanography Concept Inventory Survey (IO-CIS) serves to help fill this gap. In this paper, we discuss the end members of a spectrum of test types from traditional aptitude tests to traditional achievement tests, and we suggest that concept inventory tests are hybrids of the two. An IO-CIS is constructed and validated for a specific Introduction to Oceanography course taught at the University of Colorado at Boulder. The construction of the IO-CIS is divided into development and evaluation phases, in which the development phases primarily utilize qualitative methods, and the evaluation phase uses quantitative methods. We present an efficient approach to developing a concept inventory test for a single course in the span of a single semester. The focus of this paper is on the development phases, the qualitative methods used, and the issue of validity. We discuss how classical test theory can be used to build a validity argument during test construction of the IO-CIS.

INTRODUCTION

Measurements and Test Constructs

As scientists and science educators, we are very familiar with the idea and the practice of *measuring* things. Depending on our particular subdisciplines, we measure such things as temperature, mass, distance, time, etc. These properties have several things in common. For example, these properties can be measured precisely and accurately with scientific instruments that can be calibrated. Another shared characteristic is that the measurements of these properties all have units (e.g., degrees Celsius, grams, meters, hours, etc.). A third commonality is that measuring these properties with appropriate instruments yields quantitative data that can be reasonably compared with data sets that other investigators collect.

When it comes to measuring what our students think and what they are able to do, however, we must rely on different kinds of instruments that may be more broadly referred to as tests or surveys. Tests are generally viewed as measures of "constructs." The

---

*E-mails: laarthurs@georgiasouthern.edu; Thomas.Marchitto@Colorado.edu.

Arthurs, L., and Marchitto, T., 2011, Qualitative methods applied in the development of an introductory oceanography concept inventory survey, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 97–111, doi:10.1130/2011.2474(08).

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), in their *Standards for Educational and Psychological Testing* (1999), define a construct "as the concept or characteristic that a test is designed to measure" (p. 5). Two examples of such constructs taken from the work of Briggs et al. (2006) include "the understanding of astronomical concepts" and "the ability to design a scientific experiment."

Aptitude and Achievement Tests

Two general types of tests are aptitude and achievement tests. The former measure students' intelligence (e.g., Slosson et al., 1998), personality (e.g., Edwards, 1967; Quenk, 2009), attitudes (e.g., Adams et al., 2006), ability or what *could* be achieved (e.g., Otis and Lennon, 1989), and attainment or what *has* been achieved (e.g., Ghiselli, 1973; National Foundation for Educational Research, 2009). The latter measures the accumulation of learning in terms of skills and knowledge (Kline, 2000).

Aptitude tests are designed to be univariate in the sense that there is a single cognitive trait or variable that the test measures (Kline, 1998). Qualitative methods and data play a very important role in the construction of aptitude tests (Libarkin and Geraghty Ward, this volume). Qualitative methods used in the development of aptitude tests are often ethnographic in nature and take the form of questionnaires, interviews, and focus groups (LeCompte and Schensul, 1999). Aptitude tests may take the form of a standardized test (e.g., applicable to personality) or a practical/performance exam (e.g., applicable to music, language, programming, and laboratory experimentation).

Items in achievement tests, on the other hand, are usually constructed on the basis of predefined content that experts in the subject area identify as important. These tests are often multivariate in the sense that they measure several different variables, such as knowledge of concepts, instead of cognitive traits. Achievement tests measure performance, and their results are used as indicators of apparent learning.

Concept Inventory Tests

Unlike traditional achievement tests, concept inventories are diagnostic tests that can incorporate knowledge of certain ideas and notions that students bring with them to the classroom (Treagust, 1988). This type of student thinking has been referred to as "preconceptions" (Novak, 1977), "misconceptions" (Helm, 1980), and "alternative frameworks" (Driver, 1981).

Concept inventories are multiple-choice tests. Each multiple-choice item in a concept inventory test is composed of the question or statement (called the stem) and answer options, which include the correct (i.e., most expert-like answer) and incorrect answers (Fig. 1). Concept inventories are developed on the basis of student thinking and incorporate student language and thinking. Unlike traditional achievement tests, they are not developed solely on the basis of content that subject experts predetermine. As such, incorrect answer options are not based on instructor speculation, assumptions, or anecdotal experiences but on research into student thinking (Libarkin and Anderson, 2005). The goal in crafting the incorrect answers is to produce plausible distractors (Libarkin and Anderson, 2005) that, when selected, let the instructor know that the student is unable to correctly answer the question and that also provide insight into how students think about the concepts being tested.

Question or statement goes here (i.e., the stem).
A. Distractor 1
B. Distractor 2
C. Distractor 3
D. Distractor 4
E. Correct or most expert-like answer
} Answer options

Figure 1. General format of concept inventory test items.

Concept inventory tests are relatively new kinds of instruments intended to measure student learning. The first one, called the Force Concept Inventory, was developed for physics in 1992 (Hestenes et al., 1992). Since that time, other discipline-specific concept inventory tests have been developed for astronomy (Hufnagel, 1999; Lindell and Sommer, 2004; Lindell, 2005), biology (Odom and Barrow, 1995; Anderson et al., 2002; Knudson et al., 2003; Garvin-Doxas et al., 2007; Elrod, 2008), chemistry (Tan et al., 2008), engineering (Gray et al., 2003; Wage et al., 2005), and geology (Libarkin and Anderson, 2006). None, however, is yet available for oceanography. It is our aim to develop a concept inventory test for an introductory-level oceanography course. Our approach is rooted in classical test theory and grounded theory.

Classical Test Theory and Grounded Theory

One of the most commonly invoked theories in test construction is classical test theory (National Council on Measurement in Education and American Council on Education, 2006). The key concepts of classical test theory are validity and reliability. A test is reliable when the measurements are consistent and reproducible. Reliability is estimated mathematically. Historically, a test was considered valid when it actually measured what it was intended to measure (Angoff, 1988; Lopez, 1996). However, the definition that the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education currently endorse, in their *Standards for Educational and Psychological Testing* (1999), is that validity "is a unitary concept" (p. 9) and "is the degree to which all accumulated evidence supports the intended interpretation of test scores for the proposed purpose" (p. 9).

Often, validity is thought of in terms of distinct types of validity. These include the widely referred to content validity, construct validity, concurrent validity, and predictive validity (e.g., Cronbach and Meehl, 1955; Kline, 1986, 1998, 2000), as well as face validity, which is not actual validity and deals only with the appearance of the test (Kline, 1998, 2000), and the less commonly referred to communication validity (Lopez, 1996). In designing the Introduction to Oceanography Concept Inventory Survey (IO-CIS), however, we ascribed to the notion of validity as discussed by AERA, APA, and NCME. In their *Standards for Educational and Psychological Testing*, they discuss validity in terms of types of validity evidence (1999). They outline sources of evidence that may shed light on different aspects of validity but that do not represent distinct types of validity (AERA/APA/NCME, 1999). They note that validity cannot be unequivocally proven, but that coherent arguments for validity can be made using various sources of evidence. Evidence of validity may be based on (1) test content, (2) response processes, (3) the internal structure of the test, (4) relations to other variables, (5) convergent and discriminant evidence, (6) test-criterion relationships, (7) validity generalization, and (8) consequences of testing (AERA/APA/NCME, 1999).

Evidence of validity may be gathered using a grounded theory approach. This is a data-driven approach geared at understanding a key issue (Creswell, 1998). In the case of this study, the key issue is student alternate conceptions and misconceptions about oceanographic concepts. For this study, student alternate conceptions and misconceptions were not identified a priori; instead, they emerged during the qualitative inquiry process. Student alternate conceptions and misconceptions were gathered using informal interviews, structured interviews, and surveys. Student responses were coded (i.e., categorized), and several examples are included herein.

PROBLEM STATEMENT

At present, there exist no clearly articulated and agreed upon guidelines for what a concept inventory should test, how it should be designed, and the appropriate demonstrations of validity and reliability. In fact, examination of publications about existing concept inventories as well as comparative studies of them (e.g., Lindell et al., 2007) show that existing concept inventory tests vary in terms of the types of items asked (i.e., ranging from factual content to conceptual principles), the approaches used to construct the items (i.e., from only expert input to a combination of expert and novice input), the extent to which validity and reliability are considered during individual item and overall test design, and the methods used to evaluate the tests. Furthermore, although implicit agreement that concept inventory tests should be a measure of student learning in the broad sense exists, there is no explicit agreement over whether these tools should be, for example, measures of aptitude (e.g., ability) or of achievement.

The concept inventory for oceanography that we developed using classical test theory and grounded theory may be viewed as a hybrid between traditional aptitude and achievement tests. It is designed to first provide a measure of how much students learn during the semester and, thus, possesses a key element of achievement tests. We also designed it to provide an indication of what students think with respect to critical concepts of the course; thus, the cognitive aspects of inquiry used to develop it are akin to those used in designing aptitude tests. Thus, like several other concept inventory tests, this one may be viewed as an achievement test that utilizes development methods generally used to construct aptitude tests and not traditionally used to construct achievement tests. It was validated for the specific course for which it was designed.

In this sense, we overturn Sapsford's (1999) argument stating that constructing valid achievement tests "poses no conceptual problems" (p. 136), presumably because content experts design the test items. That is, they independently write items without consideration of other factors such as student thinking. In the case of constructing this oceanography concept inventory, however, student thinking and alternate conceptions played a major role in the development process.

Locating the Study and the Researchers

With inspiration from our predecessors in other fields, our objective was to construct a concept inventory test for a specific introductory-level course, Introduction to Oceanography, which is taught at the University of Colorado at Boulder every spring semester and which enrolls ~165 students. The reason for creating such an instrument was shared in common with designers of other concept inventory tests. Namely, we wanted to use the instrument as a multiple-choice pre- and post-instruction test (e.g., Hake, 1998; Schmeiser and Welch, 2006; D'Avanzo, 2008; Smith et al., 2008; Prather et al., 2009).

The course for which the test was designed was lecture-based and has no accompanying laboratory or recitation sections. The course was designed upon the textbook titled *Oceanography: An Invitation to Marine Science* (Garrison, 2007). The course was a survey course consisting of four modules dealing with geology, physics, chemistry, and biology. The items in the concept inventory instrument for this course therefore cover a variety of critical concepts. We call the instrument the "Introduction to Oceanography Concept Inventory Survey" (IO-CIS).

The present authors, Arthurs and Marchitto, were the test developers of the IO-CIS. The instructor of the course was Marchitto, and the researcher that coordinated the construction of the IO-CIS was Arthurs. Arthurs did not teach this course, and her interactions with students in this study were limited to the administration of classroom surveys and out-of-classroom one-on-one interviews. Based on the criteria that Feig outlines (this volume), Marchitto was a researcher-participant because he was the instructor of the course, and Arthurs was a researcher-observer in this study because she engaged students via interviews and administered in-class surveys.

The IO-CIS was largely developed during the spring 2008 semester. The class that semester possessed 162 students and provided the test bed for developing and piloting items for the IO-CIS. The class make-up of the students in the course at the end of the semester included 104 males and 58 females; 4 freshmen, 32 sophomores, 54 juniors, 59 seniors, 9 fifth-year seniors, 1 graduate student, and 3 nondegree students; and 67% non–science-technology-engineering-mathematics (non-STEM) majors, 27% STEM majors, and 7% undeclared (open option) students.

Scope

Our aim was to use classical test theory and grounded theory to develop a valid and reliable instrument for the Introduction to Oceanography course that reflects student thinking about oceanography concepts in the answer options *and* assesses student achievement based on how they answer conceptual questions. The test construct that the IO-CIS was developed to measure is the understanding of oceanographic concepts, particularly as they relate to this course.

Broadly speaking, test construction consists of development and evaluation phases, with the results of the evaluation possibly leading to a reiteration of steps under the development phases. In constructing the IO-CIS, the development phases employed qualitative research methods, while the evaluation phase used quantitative methods. Thus, the overall construction of the IO-CIS utilized a mixed-methods approach.

Figure 2 illustrates the entire conceptual framework we developed for the construction of the IO-CIS. For the purposes of this paper, however, we limit the scope of discussion to *only* the development phases. Our objectives in writing this paper include presenting the conceptual framework developed for constructing the IO-CIS in the span of one semester, describing the qualitative methods used, outlining a strategy for designing a course-specific concept inventory test, and discussing the issue of validity and its relevance in different steps of development.

METHODS AND PROCESSES

Development Phase I: Design Stems for IO-CIS Items

We received Institutional Review Board (IRB) approval to administer surveys and conduct interviews from the University of Colorado IRB.

For the purpose of constructing the IO-CIS, we made a distinction between factual and conceptual test items. Test items were considered factual if providing the correct answer required only recalling relevant content knowledge. These included questions that asked for a definition, the date of a particular event, a value such as Avogadro's number, etc. On the other hand, test items were considered conceptual if they required the correct application of a "critical concept" to generate the correct (i.e., most expert-like) answer. We defined critical concepts as the first principles and foundational ideas upon which more complex ideas can be constructed and understood. Our goal in constructing the IO-CIS was to include only conceptual test items.

The goal of the first development phase was to write valid open-response questions. These open-response questions were potential stems for the final multiple-choice IO-CIS items. They were administered as short in-class exercises or surveys to solicit students' responses, which provided the data necessary to develop answer options for the final IO-CIS items. The steps involved in phase I were geared primarily at accumulating validity evidence based on test content, response processes, and internal structure of the test.

The first step in phase I was identification of the critical concepts in the course. These concepts reappeared in different parts of the course and were integral to a deeper understanding of geologic and oceanographic instances in which they apply. A second textbook, *Introduction to Ocean Sciences* (Segar, 1999), was used to aid the instructor in developing a list of 17 critical concepts.

The second step in phase I defined one to three learning goals (i.e., desired learning outcomes) for each of the critical concepts, and the instructor utilized Bloom's taxonomy (Handelsman et al., 2007) to write concept-specific learning goals. Learning goals may look like test questions, but they are not test questions. A defining characteristic of learning goals is that they are assessable. As such, it is conceivable that a learning goal could be directly converted into a test question. More importantly, however, the ability for a student to achieve a desired learning goal means that they are able to answer questions related to that learning goal. For example, learning goals written at higher levels of Bloom's taxonomy naturally subsume requisite knowledge and skills at lower levels that can be tested as a part of assessing whether a desired learning goal is achieved. In Table 1, we include seven examples of the critical concepts and one of their associated learning goals for this course.

Step 3 of phase I entailed using the critical concepts and their associated learning goals to guide the design of open-response questions. These questions were developed for inclusion in four different in-class exercises or surveys that the students would complete prior to the start of a new module during the course. These surveys were called Concept Inventory Exercises (CIEs). Each CIE was composed of four to five open-ended questions, and each question contained one to four parts. In total, 38 questions were asked through the CIEs.

Before administering the CIEs, Arthurs vetted the open-response questions in one-on-one, informal, and unstructured interviews with two experts and three novices. The experts were active researchers and geoscience faculty members with diverse expert backgrounds within the geosciences. For each question, one of the two experts was an oceanographer (Marchitto) and the second expert was selected based on meeting availability. The novices were undergraduate students not enrolled in the course and who never took an oceanography class. The same students were consulted throughout the development phases. These interviews with experts and novices provided feedback that was used to refine the language in the questions and provide validity evidence based on test content, internal test structure, and especially response processes. These interviews represent the fourth step of phase I. Steps 3 and 4 were iterative in nature and continued until questions were clear and jargon free, and interviewees interpreted them as intended.

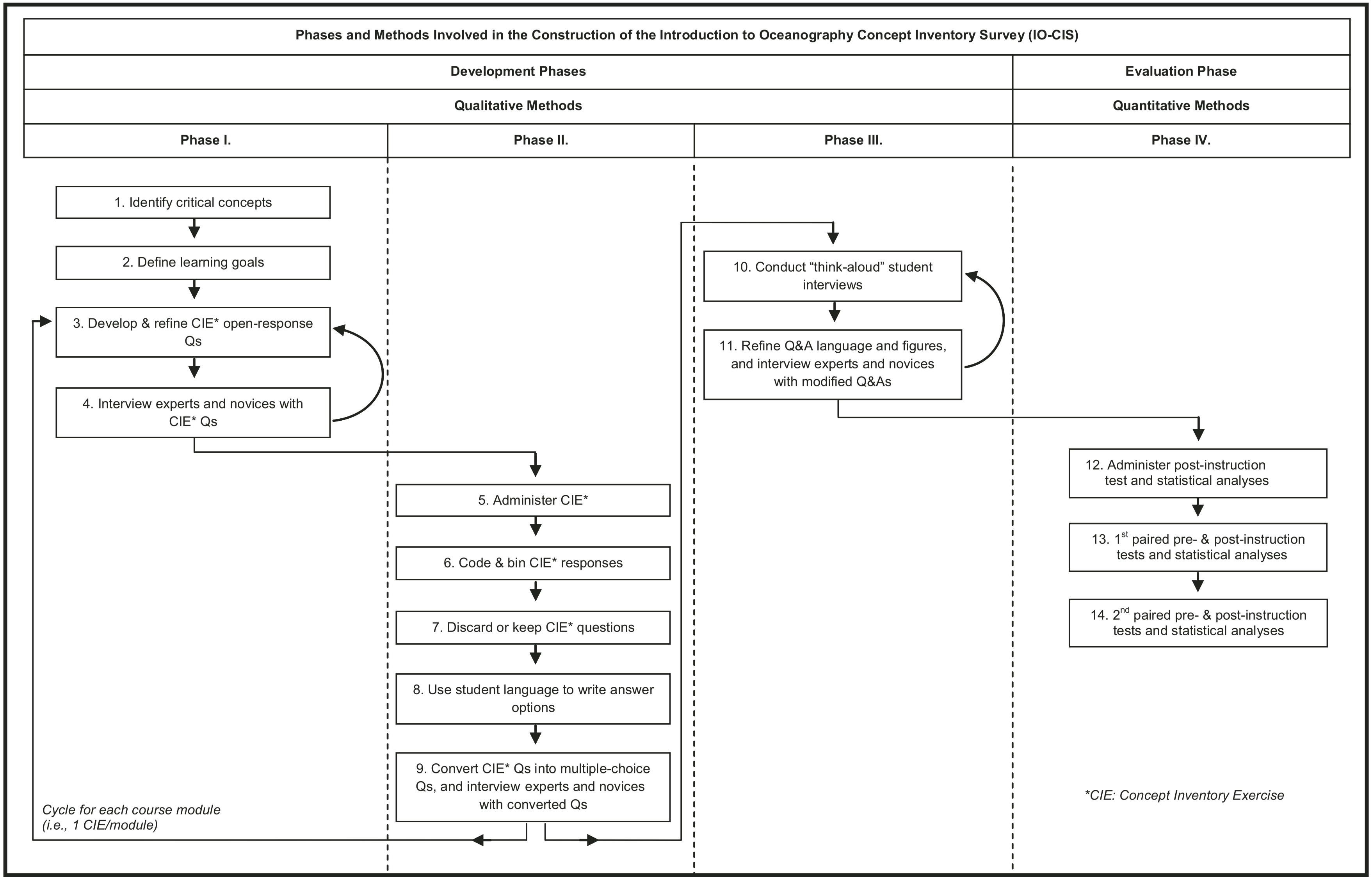


Figure 2. Flow chart of the methods and processes used in the concept inventory exercise (CIE).

TABLE 1. SEVERAL CRITICAL CONCEPTS IN THE COURSE AND EXAMPLES OF LEARNING GOALS

Critical concept	Learning goal
Density stratification	Describe the behavior of neutrally buoyant material
Isostatic equilibrium	Explain how isostatic equilibrium accounts for the existence of ocean basins
Convection	Describe the conditions necessary for the development of a convection cell
Seawater density	Compare the relative influence of salinity in warm vs. cold seawater
Deep-water vs. shallow-water waves	Distinguish between deep-water and shallow-water waves on the basis of wavelength and water depth
Chemosynthesis	Compare and contrast between photosynthesis and chemosynthesis
Maximum sustainable yield	Explain why harvesting older fish has both benefits and risks

Development Phase II: Convert Open-Response Questions into Multiple-Choice Items

Step 1 of this phase consisted of administering the CIE in class. Before the instructor began a new module in the course, Arthurs administered the exercise at the beginning of the lecture period. The CIEs were low-stakes assessments because students received no credit for their voluntary participation. Students were asked to answer the questions to the best of their ability. They were also told that their answers would help the instructor better understand what they already know and what they might still be unfamiliar with going into the new module. Finally, the students were informed that their responses would be anonymous to the instructor (i.e., he would not learn who said what, he would only learn the "what"). Students' responses were not anonymous to Arthurs, however, because students were asked to include their names on the CIEs. Students had up to 20 min to complete each CIE.

The number of students responding to each CIE declined over the course of the semester as attendance also declined. CIE 1 had 83%, CIE 2 had 57%, CIE 3 had 60%, and CIE 4 had 35% of students enrolled in the course respond. Balnaves and Caputi (2001) indicate that these response rates are excellent to good for this type of research. It is also important to note that all students who filled out a CIE did not necessarily provide a response to every question; that is, some were left blank.

The second step of phase II deals with coding and binning student responses. Student responses for each of the four CIEs were manually coded (i.e., categories were created based on the collective responses received) and binned (i.e., individual responses were placed under a specific code or category) in 16–19 h/CIE. Each CIE question solicited a range of different answers, from two to 22. During this process, student responses were sorted into categories. These categories were not determined a priori; instead, they emerged on the basis of what students wrote in their responses. After coding (categorizing) and binning (placing responses in a category), stems were either kept for inclusion in the final IO-CIS or discarded. The main criteria for keeping a stem were whether three or more categories of answer options (i.e., one correct answer and at least two distractors) emerged from the student responses, whether these categories were popular in a noteworthy fraction of the respondents, and whether the student responses associated with each category were authentic and jargon free. Criteria for discarding stems included a lack of diversity in responses and/or a majority of responses that were grounded in unexplained jargon or catch phrases. Of the 38 total questions posed in the four CIEs, 24% were discarded from inclusion as stems in the IO-CIS (see Table 2 for examples of discarded stems), 44% were kept as uncombined stems, and 32% were merged into pairs to form one combination stem.

Step 3 in this phase involved using student language and sketches to craft answer options for each stem. When multiple categories of student responses met the "keep" criteria, the top three to five most popular categories were retained. Correct answers and plausible distractors were crafted using the students' own language and/or sketches. The final step of phase II was to convert selected CIE questions into multiple-choice IO-CIS items.

Development Phase III: Conduct Structured Student Think-Aloud Interviews

Phase III consisted of two steps that were iterative in nature. The first step was to gather expert input and conduct student think-aloud interviews about the IO-CIS, and the second step entailed refining IO-CIS items based on the expert and novice input. Expert input was a means to accumulate validity evidence in terms of test content, while think-aloud interviews with novices was a way to accumulate validity evidence based on response processes.

A central factor to the think-aloud interviews was determining whether (1) the items were interpreted as intended and (2) students selected the correct answers for the correct reasons. These interviews were voluntary, and students were paid at a rate of $15/hr. These one-on-one interviews ranged in duration from 1 to 1.5 h. Prior to these interviews, the interviewer (Arthurs) referred back to the interviewee's CIEs, to get a glimpse into the level of mastery the student demonstrated in the exercises and to flag specific questions for deeper probing during the interview. These interviews were essential for obtaining a deeper understanding of how students interpreted and thought about both the

TABLE 2. EXAMPLES OF OPEN-ENDED CONCEPT INVENTORY EXERCISE (CIE) QUESTIONS DISCARDED FROM INCLUSION IN THE INTRODUCTION TO OCEANOGRAPHY CONCEPT INVENTORY SURVEY (IO-CIS)

Discarded open-ended question	Reason for discarding
The Himalaya mountains in Tibet are topographically much higher than the surrounding areas. What is it about the continental crust in the Himalaya area, compared to the surrounding areas that might explain its high topography?	*n* = 58 for this question. 54% simply wrote "continental collisions" with no further explanation. This is an example of where jargon or catch phrases are used without demonstrated understanding.
What are nutrients?	*n* = 82 for this question. 100% said that "nutrients are X, which are necessary for life." The Xs that they listed were largely correct. This is an example of an insufficient number of categories in responses (particularly incorrect responses) from which to create multiple answer options.

stems and the distractors that were developed based on student responses to the open-ended CIEs.

Upon meeting with the interviewee, the interviewer described the format of the interview, explained that student input was solicited in the development of this instrument, and solicited questions before beginning the formal part of the interview. During the interview, students were asked to think about and discuss the multiple-choice IO-CIS items. The interviews were conducted with two pass-throughs of the items. In the first pass-through, the students thought about *each* item, described how they interpreted the item, and discussed their thought processes in selecting an answer option. The interviewer's role at this stage was to listen, ask the students to verbalize their thoughts when they became quiet, and to record notes. When students posed questions to the interviewer during the first pass-through of the items, the interviewer reminded them that their questions would be discussed at the end of the interview so as not to alter their thinking or bias their responses.

In the second pass-through, the interviewer asked each student to revisit selected items in order to ask follow-up questions on what the student said during the first pass-through. This was also the time that the interviewer asked questions based on the students' CIE responses. This was done in order to obtain further clarification or elaboration on a particular aspect of student thinking that emerged in the CIEs and/or during the first pass-through of the items. Only after completion of the two pass-throughs did the interviewer answer questions and discuss items with the students (e.g., most interviewees asked whether they answered one or more specific questions correctly).

Twenty-three students volunteered to participate in think-aloud interviews, and all of them were interviewed. At least 15 students were interviewed with the final version of each IO-CIS item. Based on the results of both expert input and student think-aloud interviews, modifications were made to the language or figures in all but five of the final IO-CIS items. Modifications were made only to the stems. The student interviews provided evidence that students interpreted the items as intended and selected the correct answers for the correct reasons. Furthermore, they also indicated that students selected incorrect answer choices because of flaws in their reasoning, lack of content knowledge, or possession of interfering alternate conceptions.

During the development of the IO-CIS, myriad student alternate conceptions and misconceptions were gathered. Detailing them here, however, is beyond the scope of this paper. Nevertheless, a sampling of student responses in the next two sections will provide examples of the alternate conceptions and misconceptions that emerged. In order to illustrate the development of single IO-CIS items, the development histories of two examples are outlined in detail next.

Item Development History: Example A

Step 1. Marchitto identified a critical concept: Convection.

Step 2. Marchitto defined an associated learning goal: Describe the conditions necessary for the development of a convection cell.

Step 3. Arthurs and Marchitto together developed the two-part CIE question, shown in Figure 3.

Step 4. Arthurs conducted informal interviews with two experts and three novices to gather expert and novice input on the wording, an important part of the validation process. No revisions to the question were deemed necessary.

Step 5. The question was asked on a CIE. Arthurs coded and binned CIE responses to question a and question b, the two parts of the CIE question. In total, 134 students responded to question a and 104 responded to question b. Examples of nine out of 15 different coded categories of responses to question a are listed in Table 3, and the percentage of respondents with the same or similar response is indicated in the last column. These percentages reflect the number of student responses binned under the corresponding code or category. Each example is written in student language and is used to represent a category of student responses similar in meaning and wording.

Step 6. Arthurs used CIE responses to develop multiple-choice IO-CIS items. She then conducted interviews with two experts and three novices as part of the process of validating the multiple-choice IO-CIS items.

Question a of the CIE question was transformed into the IO-CIS question shown in Figure 4, using the most popular coded categories and the associated student responses to produce the answer options. Experts identified answer choice E as the best and most expert-like response. This was a part of the

Geologic data indicate that although Earth's mantle is mostly solid rock, it is not stationary; instead, mantle rock moves around or circulates deep inside Earth in the area between Earth's core and Earth's crust.

(a) Why does mantle rock move/circulate deep inside Earth?

(b) Using the cross-section of Earth provided below, please draw arrows to indicate a typical path of the mantle rock moving deep inside Earth.

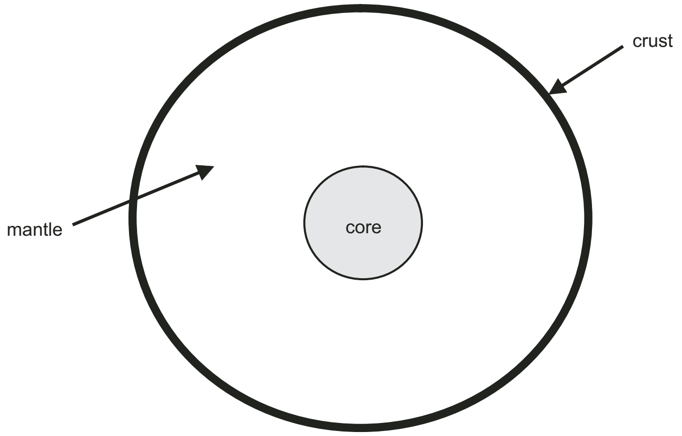


Figure 3. Questions a and b in this example remained as stand-alone or uncombined questions in the final Introduction to Oceanography Concept Inventory Survey.

TABLE 3. STUDENT REASONS FOR WHY MANTLE ROCK MOVES/CIRCULATES DEEP INSIDE THE EARTH—EXAMPLES OF CODES, CATEGORIES, STUDENT RESPONSES, AND RESULTS OF BINNING

Code*	Category[†]	Representative student response (as actual student quotes)[§]	Total student responses (%)[#]
A	Currents	Circulation currents/eddies	2
B	Liquid	All mantle rock is liquid rock	4
C	Float	Mantle rock is floating on a layer of moving liquid inside the Earth	5
D	Convection	Convection [Only the term "convection" was written, with no elaboration]	8
E	Contact	Plates at the surface are moving and because they are touching the mantle, they cause it to move when they move	5
F	Gravity	Gravitational forces of Earth cause convection within the mantle	5
G	Rotation	Earth's rotation causes convection within the mantle	10
H	Pressure	High pressure at different depths causes convection with the mantle	13
I	Heat	Heat from Earth's interior causes convection within the mantle	31

*Codes were used as a type of short-hand to mark raw student responses.

[†] These are categories that emerged from student responses to question b in Figure 3.

[§] Text in brackets represents additional information inserted by the authors.

[#] The percentages reflect the number of student responses binned under each code or category.

process of gathering validity information about test content. Neither experts nor novices indicated the need for adjustments to the wording. This was a part of gathering validity information about response process.

Question b of the CIE question was transformed into the IO-CIS question shown in Figure 5. Experts identified answer choice D as the best and most expert-like response.

Step 7. Arthurs conducted student think-aloud interviews with the transformed items. As a part of determining whether the questions were clear to the students and whether the interviewees interpreted the questions as intended, students were asked to paraphrase the questions in different words. For question b, for example, one student said,

> *"Well, it's pretty straightforward already. I guess, maybe, 'Choose the picture that best shows the direction of convection in the mantle.'"*

Responses like this were taken as indicators that the questions were clear and interpreted in the ways intended. During the interviews, students also selected the correct answers for the correct reasons. For question b, for example, most interviewees explained the connection between the temperature of mantle rock nearer to and further from Earth's core and the way in which this affects its density and therefore convection. For example, one student said,

> *"The center is hotter; so, [the mantle] is less dense, moves up and away from the center, cools off further from the center, and sinks back in to the center."*

Based on this round of student interviews, the wording of these two items did not undergo further revisions.

Geologic data indicate that although Earth's mantle is mostly solid rock, it is not stationary; instead, mantle rock moves around or circulates deep inside the Earth in the area between Earth's core and Earth's crust.

Why does mantle rock move deep inside the Earth?

(A) High pressure at different depths causes convection within the mantle.
(B) Earth's rotation causes convection within the mantle.
(C) Plates at the surface are moving and, because they are touching the mantle, they cause it to move when they move.
(D) Gravitational forces of the Earth cause convection within the mantle.
(E) Heat from Earth's interior causes convection within the mantle.

Figure 4. Student responses to an open-ended concept inventory exercise question led to the development of a multiple-choice Introduction to Oceanography Concept Inventory Survey item.

Which one of the following diagrams best describes a typical path of the mantle rock that moves deep inside the Earth?

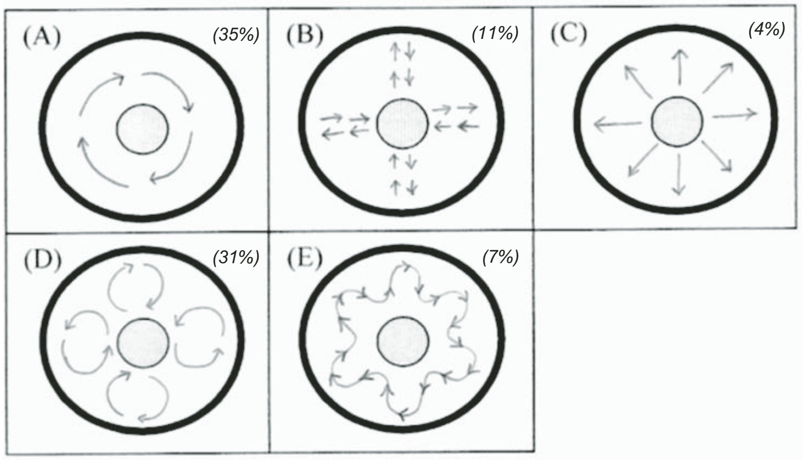


Figure 5. Student sketches to an open-ended concept inventory exercise question led to the development of a sketch-based multiple-choice Introduction to Oceanography Concept Inventory Survey (IO-CIS) item. The percentages in parentheses that appear beside each answer choice indicate the total percent of students' responses that fell into that category or type of response. These percentages are not included in the actual IO-CIS item.

Item Development History: Example B

Step 1. Marchitto identified a critical concept: Density stratification.

Step 2. Marchitto defined an associated learning goal: Explain the layering of Earth's interior and oceans as a function of composition, temperature, and pressure.

Step 3. Arthurs and Marchitto together developed the CIE question shown in Figure 6.

Step 4. As in the first example presented here, Arthurs conducted informal interviews with two experts and three novices to gather expert and novice input on the wording, an important part of the validation process. No changes were deemed necessary.

Step 5. The two-part question was asked on a CIE. Arthurs tallied CIE responses to question a and coded and binned responses to question b of the question.

In total, 92 students responded to question a and question b. For question a, 80% of respondents selected YES and 30% selected NO.

For students that answered YES to question a, examples of six out of 18 different coded categories of responses to question b are listed in Table 4. Each category is accompanied by a representative student response. Examples of three of the coded categories of responses to question b from students who answered NO to question a are listed in Table 5.

Step 6. Arthurs used CIE responses to develop multiple-choice IO-CIS items and then conducted interviews with two experts and three novices. Questions a and b of the CIE question and student responses were combined to form the IO-CIS question shown in Figure 7. Experts identified answer choice B as the best and most expert-like response. Neither experts nor novices provided cause for adjusting the item.

Step 7. Arthurs conducted student think-aloud interviews with the transformed item.

During the first round of student think-aloud interviews, students posed a number of questions that indicated the need for further elaboration in the question. Questions that students posed included:

> *"By 'stratification,' you mean different layers of water, right?"*
>
> *"By 'change or not change,' do you mean if the layering could disappear?"*
>
> *"Are you asking if it would change in the entire area between the ocean floor and the ocean surface?"*

Based on the first round of student think-aloud interviews, the stem was modified, and the answer choices remained intact. The result of the revisions was the item shown in Figure 8.

Step 8. Arthurs conducted informal interviews with two experts and three novices regarding the revised item. One of the experts suggested modifying only the stem, to read as:

> *"Could the temperature and salinity layers (i.e., stratification) of ocean water disappear over the entire water column (from ocean surface to ocean floor) in a particular region?"*

Step 9. Arthurs conducted a second round of think-aloud student interviews with a different set of students not yet interviewed, to gather validity information about process response to the modified stem in particular and the item as a whole. This round of interviews indicated that some students were still unclear about the question. These students suggested including a drawing to help illustrate the question. These students were asked to sketch a drawing that they thought would help clarify the question. As with the first round of interviews, students found the answer choices clear and plausible. When they selected the correct answer, they did so for the correct reasons. Based on the second round of student think-aloud interviews, the stem was modified, a figure was added, and the answer choices remained intact. The product of the revisions is the question shown in Figure 9.

Step 10. Arthurs informally interviewed two experts and three novices with this version of the item. No changes were suggested.

Step 11. Arthurs conducted a third round of think-aloud interviews with this item. These interviews confirmed that the stem was now clear to all students and interpreted in the intended manner. As with each item, students were asked whether there was anything unclear to them in the question, and they were asked to paraphrase the question. All students in this round of interviews said that the question was clear. For example, students made remarks such as,

> "*Yeah, the question is clear.*"
>
> "*No, I don't [have suggestions for improving the wording].*"
>
> "*[The question] is fine the way it is.*"

As such, further indication was received that the answer responses were also clear and interpreted as intended. In addition, student responses indicated that the correct answer was selected for the correct reason. For example, students who selected the correct answer indicated that the "temperature and salinity of ocean water can change" and that "that change causes ocean water density to change" and that changes in ocean water density can then "disturb ocean stratification." As a result of this round of interviews, no further revisions were made to this item.

The development history behind this particular IO-CIS item exemplifies how items were modified based on feedback received from experts and during student think-aloud interviews to further an argument for its validity based on test content and process response.

Scientists have observed that ocean water is stratified (i.e., layered) in terms of both temperature and salt content.

(a) Could the stratification (layering) of ocean water change?
Circle one: **Yes No**

(b) Based on your answer to (a), why would it change or not change?

Figure 6. Example of one multiple-part concept inventory exercise question.

TABLE 4. STUDENT REASONS FOR WHY STRATIFICATION OF THE OCEAN CHANGES—EXAMPLES OF CODES, CATEGORIES, STUDENT RESPONSES, AND RESULTS OF BINNING

Code*	Category[†]	Representative student response (as actual student quotes)[§]	Total student responses (%)[#]
A	T&S	Temperature and salinity affect density	2
B	Magnetism	Magnetic polarity changes	3
C	Disasters	Major events mix around the water [specific examples of such events students gave were hurricanes, earthquakes, volcanic eruptions, and a meteor hitting Earth]	8
D	Salt	Enough glaciers melt to change salinity of the water	14
E	Temperature	Temperature change caused by heating from the Sun	21
F	Currents	Ocean currents can disrupt and redistribute layers	58

*Codes were used as a type of short-hand to mark raw student responses.

[†] These are categories that emerged from responses to question b from students who also responded YES to question a, as shown in Figure 6.

[§] Text in brackets represents additional information inserted by the authors.

[#] The percentages reflect the number of student responses binned under each code or category, and they reflect only students who answered YES to question a.

TABLE 5. STUDENT REASONS FOR WHY STRATIFICATION OF THE OCEAN DOES NOT CHANGE—EXAMPLES OF CODES, CATEGORIES, STUDENT RESPONSES, AND RESULTS OF BINNING

Code*	Category[†]	Representative student response (as actual student quotes)	Total student responses (%)[§]
A	Inputs	Environmental inputs could result in some changes of the properties of some layers of water, but it would NOT change the overall stratification	6
B	Stability	Once the layers settle, they will not change	17
C	Density	The density of water layers does not change	34

*Codes were used as a type of short-hand to mark raw student responses.

[†] These are categories that emerged from responses to question b from students who also responded NO to question a, as shown in Figure 6.

[§] The percentages reflect the number of student responses binned under each code or category, and they reflect only students who answered NO to question a.

Could the stratification (layering) of ocean water change?

(A) Yes, because ocean currents disrupt and redistribute the layers.
(B) Yes, because changes in temperature and salinity lead to density changes.
(C) Yes, because major events (e.g., hurricanes, earthquakes, volcanic eruptions) mix around the water.
(D) No, because the density of water layers does not change.
(E) No, because environmental inputs could result in some changes of the properties of some water layers, but the inputs would NOT change the overall stratification.

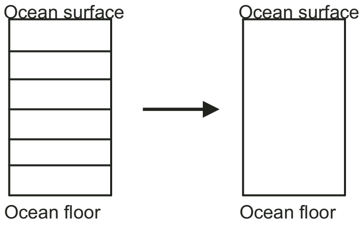
Figure 7. Example of two questions on a concept inventory exercise combined to form one Introduction to Oceanography Concept Inventory Survey item.

Could the stratification (layering) of ocean water disappear over the entire water column (from sea surface to deep seafloor) in a particular region?

(A) Yes, because ocean currents disrupt and redistribute the layers.
(B) Yes, because changes in temperature and salinity lead to density changes.
(C) Yes, because major events (e.g., hurricanes, earthquakes, volcanic eruptions) mix around the water.
(D) No, because the density of water layers does not change.
(E) No, because environmental inputs could result in some changes of the properties of some water layers, but the inputs would NOT change the overall stratification.

Figure 8. Original Introduction to Oceanography Concept Inventory Survey item revised based on student interviews.

Could the temperature and salinity layers (i.e., stratification) of ocean water disappear over the entire water column (from ocean surface to ocean floor) in a particular region? That is, could layering disappear as depicted in the figure below?



(A) Yes, because ocean currents disrupt and redistribute the layers.
(B) Yes, because changes in temperature and salinity lead to density changes.
(C) Yes, because major events (e.g., hurricanes, earthquakes, volcanic eruptions) mix around the water.
(D) No, because the density of water layers does not change.
(E) No, because environmental inputs could result in some changes of the properties of some water layers, but the inputs would NOT change the overall stratification.

Figure 9. Another version of the previous Introduction to Oceanography Concept Inventory Survey item based on the second round of student interviews.

REFLECTIONS AND IMPLICATIONS

Validity Argument

The development of the IO-CIS was divided into three phases that were grounded in qualitative methods to achieve a high degree of validity for the specific course for which it was designed. Recall that validity is the "degree to which all accumulated evidence supports the intended interpretation of test scores" and that the AERA, APA, and NCME outlined eight types of validity evidence in their *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 1999). Given that the IO-CIS was designed to be a tool that provides diagnostic information about the test construct (i.e., understanding of oceanographic concepts, particularly as they pertain to the course), a strong validity argument was needed to make the case that diagnostic interpretations of the IO-CIS results were appropriate and instructionally useful. Currently, we do not yet have sufficient evidence based on validity generalization and the consequences of testing to make such an argument.

We can, however, present components of a compelling validity argument when we draw on many other types of validity evidence nested within the approach used to develop the IO-CIS and analyze its items. Evidence based on test content and internal test structure was accumulated via the instructor's identification of critical concepts, articulation of learning goals, and input throughout the development phases. In doing so, he provided expert judgments regarding (1) the relationship between parts of the IO-CIS and the construct and (2) the adequacy with which the IO-CIS content represents the content domain for the interpretation of IO-CIS scores. Evidence based on response processes was gathered from CIE responses as well as interviews with experts and novices. These responses and interviews provided information about the extent to which the response processes were consistent with the intended interpretations.

An item not discussed in this paper, as it falls under the evaluation phase, but also important to our validity argument, was evidence derived using quantitative methods. These include evidence based on internal test structure drawn from the item difficulty analysis, which indicated that there is a range of difficulty in the IO-CIS items. In addition, evidence based on relations to other variables, convergent/discriminant evidence, and test-criterion relationships was pulled from the relationships and correlations between IO-CIS scores, average exam scores, and learning gains.

Overall, the diagnostic interpretation of IO-CIS scores follows a clear line of reasoning as listed next, and a flaw in any of the connections in reasoning would weaken the overall validity argument.

1. Item responses are connected to popular student alternate conceptions and misconceptions.
2. Popular student alternate conceptions and misconceptions are connected to critical concepts of the course.
3. Critical concepts of the course are connected to instructor-defined learning goals that describe what students ideally should understand about the construct.
4. The construct is the understanding of oceanographic concepts, particularly as relevant to the course for which the IO-CIS was developed.

Uses for the IO-CIS

The IO-CIS was designed for the Introduction to Oceanography course that Marchitto teaches. It can be used to inform the instructor (and students) about student thinking and their level of mastery as they relate to specific concepts because this tool primarily assesses students' conceptual understanding, not simply their recollection of facts. Conceptual understanding, as it is referred to here, is intimately linked with students' ability to apply first principles to answer questions not seen before.

Administered as a pre-instruction assessment, the results of the IO-CIS can inform students individually and/or collectively of their preconceptions and conceptual understanding of key course concepts coming into the course. They also inform the instructor and, therefore, can also be used in making decisions about how to approach teaching certain course material.

Administered as a post-instruction assessment, the results can provide the instructor and students useful information about the learning gains students attained during the period of instruction.

Furthermore, the results of pre- and post-instruction concept inventories can be used to inform the design, implementation, and redesign of instructional interventions intended to improve students' conceptual understanding. The IO-CIS, for example, was developed in large part to eventually evaluate potential impacts of introducing a classroom response system (also known as "clickers") on student learning.

In addition, the IO-CIS has the potential to be used in classes that share a similar classroom culture, curriculum (e.g., those also based on the same textbook), and learning goals. Prior to using the IO-CIS in such classes, however, we recommend that potential users review the test content to ensure that it is aligned with the curriculum and learning goals for that class. We also recommend that potential users conduct think-aloud student interviews to gather validity evidence for the new location and population of students. For those who cannot fund student interviews, we recommend devising other incentives because response rates are typically low when incentives are not offered.

Finally, the IO-CIS also has the potential to be further developed collaboratively with others for more widespread use. To those ends, three experts who teach oceanography courses at other institutions already reviewed the IO-CIS, and their input represents the beginning of the accumulation of evidence based on validity generalization. Others interested in developing an oceanography instrument for more widespread use are asked to contact Arthurs, to assist in identifying common learning goals in introductory-level oceanography courses that may form the basis of an instrument that has more widespread applicability and to identify existing IO-CIS items that are immediately aligned with their course curriculum and learning goals.

Although the IO-CIS is not currently a widely used concept inventory, it is worth noting that widely used assessments like it facilitate (1) discipline-based cognitive and education research (e.g., Hake, 1998) as well as (2) curricular oversight and assessment from the discipline level to the national and international levels (e.g., Wage et al., 2005).

Effectiveness of Approach to IO-CIS Development

Actions taken in earlier steps of development to achieve the highest degrees of validity possible paid off in later steps of development. The fact that the answer choice options underwent no revisions during phase III of development is a good illustration of this point. Only changes to the stems occurred during this phase. The development approach that we used made it possible to design a valid concept inventory in the span of only one semester, and it was administered as a post-instruction test that semester (spring 2008). In the second semester of the course (spring 2009), the IO-CIS was administered as both a pre- and post-instruction assessment. Further data were collected to evaluate the reliability of the IO-CIS in a third semester of the course (spring 2010), and sufficient evidence indicated that the instrument is valid and reliable.

Factors That Contribute to Development Effectiveness

Several factors contributed to the effectiveness in developing a valid concept inventory test in the span of one semester. First, the scope for which the concept inventory was developed was for a single course, and the same instructor teaches the course every year. Thus, the target population for administration was narrow in scope. Second, the instructor had experience teaching the course several times in the past and was able to identify a stable list of critical concepts of the course, and these critical concepts did not change during the period of concept inventory development. Third, there was consistently clear and timely communication between the instructor (Marchitto) and researcher (Arthurs) during the period of IO-CIS development. This was important for the relatively quick turnaround time needed in designing and administering the CIEs, which formed the basis of the final IO-CIS. Fourth, for the scope of the project, student participation in CIEs was sufficient for understanding the variability in student conceptions (this relates to the issue of data saturation in Feig, this volume). Fifth, we had access to both experts and novices for informal and formal interviews, which were integral for validation purposes. Finally, a relatively quick turnaround time (i.e., 16–19 h over one weekend) was established to convert CIE open-response questions into multiple-choice questions for the concept inventory test. This allowed for time during the semester to interview experts and novices on the converted questions, which was an important component of the validation process.

Challenges during IO-CIS Development

Although the approach used to develop the IO-CIS was very efficient and produced a valid and reliable concept inventory in the span of only one semester, the development process was not without its challenges. As is not uncommon in the design of these kinds of instruments, there was a need for great care in balancing the expert perspective with the novice perspective when designing the stems and answer options. In this regard, an iterative process that accounted for these varying perspectives was an integral component in validating the IO-CIS. A second challenge was the time and energy involved in different steps of development. In this regard, perhaps the most demanding steps in the development process were (1) converting and binning student answers to the open-response CIE questions and (2) conducting the student think-aloud interviews.

SUMMARY

At present, no concept inventory test exists for oceanography. As such the Introduction to Oceanography Concept Inventory Survey (IO-CIS) serves to fill this gap. We formulated an approach to develop the IO-CIS in the span of one semester. The IO-CIS was designed as a valid and reliable achievement test, with 23 multiple-choice items, that measures the level of student mastery of critical concepts important to the course for which it was designed. It has the potential to be administered in other similar courses and/or further developed for broader use. Our work shows that valid and useful instruments for individual classes require much shorter and reasonable time periods and workloads to develop compared to instruments designed for wider distribution and use.

For others interested in designing concept inventory tests for their own classes, we can say the following. Using our conceptual approach to constructing a concept inventory test for a course, you can expect, at the soonest, to have a valid instrument at the end of the first semester, to be able to collect matched pre- and post-instruction test data during the second semester, and to have a valid and reliable instrument at the end of the third semester.

ACKNOWLEDGMENTS

We acknowledge the Science Education Initiative for the financial support that made this work possible. We thank the Testing and Assessment Center of the University of Colorado at Boulder for conducting the item analyses. We thank Blake Redabaugh from the Office of Planning, Budget, and Analysis for providing demographic data for past GEOL 3070 classes. We thank Derek Briggs from the Research and Evaluation Methodology Program at the University of Colorado at Boulder for sharing a wealth of knowledge and for valuable feedback. Last, but not least, we thank all the students, novices, and experts for their participation throughout the development of the IO-CIS.

REFERENCES CITED

Adams, W.K., Perkins, K.K., Podolefsky, N.S., Dubson, M., Finkelstein, N.D., and Wieman, C.E., 2006, New instrument for measuring student beliefs about physics and learning physics: The Colorado learning attitudes about science survey: Physical Review Special Topics—Physics: Education Research, v. 2, p. 010101-1–14.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999, Standards for Educational and Psychological Testing: Washington, D.C., American Educational Research Association, 194 p.

Anderson, D.L., Fisher, K.M., and Norman, G.J., 2002, Development and evaluation of the conceptual inventory of natural selection: Journal of Research in Science Teaching, v. 39, p. 952–978, doi:10.1002/tea.10053.

Angoff, W.H., 1988, Validity: An evolving concept (Chapter 2), *in* Wainer, H., and Braun, H.I., eds., Test Validity: Hillsdale, New Jersey, Lawrence Erlbaum, p. 19–32.

Balnaves, M., and Caputi, P., 2001, Introduction to Quantitative Research Methods: An Investigative Approach: London, Sage Publications, 257 p.

Briggs, C.B., Alonzo, A.C., Schwab, C., and Wilson, M., 2006, Diagnostic assessment with ordered multiple-choice items: Educational Assessment, v. 11, no. 1, p. 33–63.

Creswell, J., 1998, Qualitative Inquiry and Research Design: Choosing among Five Traditions: Thousand Oaks, California, Sage Publications, 403 p.

Cronbach, L.J., and Meehl, P.E., 1955, Construct validity in psychological tests: Psychological Bulletin, v. 52, p. 281–302, doi:10.1037/h0040957.

D'Avanzo, C., 2008, Biology concept inventories: Overview, status, and next steps: Bioscience, v. 58, p. 1079–1085.

Driver, R., 1981, Pupils' alternative frameworks in science: European Journal of Science Education, v. 3, p. 93–101.

Edwards, A.L., 1967, Edwards Personality Inventory: Chicago, Science Research Associates, 54 p.

Elrod, S., 2008, Genetics concept inventory (GenCI) development, *in* Conceptual Assessment in Biology Conference II, 3–8 January, Asilomar, California: http://bioliteracy.net/CABS.html (accessed 1 September 2009).

Feig, A.D., 2011, this volume, Methodology and location in the context of qualitative data and theoretical frameworks in geoscience education research, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(01).

Garrison, T., 2007, Oceanography: An Invitation to Marine Science (6th ed.): Belmont, California, Thomson Brooks/Cole, 588 p.

Garvin-Doxas, K., Klymkowski, M., and Elrod, S., 2007, Building, using, and maximizing the impact of concept inventories in the biological sciences: Report on a National Science Foundation–sponsored conference on the construction of concept inventories in the biological sciences: CBE Life Sciences Education, v. 6, p. 277–282.

Ghiselli, E.E., 1973, The validity of aptitude tests in personnel selection: Personnel Psychology, v. 26, no. 4, p. 461–477, doi:10.1111/j.1744-6570.1973.tb01150.x.

Gray, G.L., Evans, D., Cornell, P., Costanzo, F., and Self, B., 2003, Toward a nationwide dynamics concept inventory assessment test, *in* Proceedings of the 2003 American Society for Engineering Education Annual Conference and Exposition (http://wwwtemp.asee.org/conferences/paper-search-results.cfm).

Hake, R.R., 1998, Interactive engagement vs. traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses: American Journal of Physics, v. 66, no. 1, p. 64–74, doi:10.1119/1.18809.

Handelsman, J., Miller, S., and Pfund, C., 2007, Scientific Teaching: New York, W.H. Freeman and Company, 184 p.

Helm, H., 1980, Misconceptions in physics amongst South African students: Physics Education, v. 15, p. 92–105, doi:10.1088/0031-9120/15/2/308.

Hestenes, D., Wells, M., and Swackhamer, G., 1992, Force concept inventory: The Physics Teacher, v. 30, p. 141–158, doi:10.1119/1.2343497.

Hufnagel, B., 1999, A standard survey for undergraduate non-science majors: The astronomy diagnostic test, *in* American Association of Physics Teachers, Summer Meeting: San Antonio, Texas (available at at http://www.docstoc.com/docs/43849454/A-Standard-Survey-for-Undergraduate-Non-Science-Majors-The).

Kline, P., 1986, A Handbook of Test Construction: New York, Metheun & Co., 259 p.

Kline, P., 1998, The New Psychometrics: New York, Routledge, 224 p.

Kline, P., 2000, A Psychometrics Primer: London, Free Association Books, 160 p.

Knudson, D., Noffal, G., Bauer, J., McGinnis, P., Bird, M., Chow, J., Bahamonde, R., Blackwell, J., Strohmeyer, S., and Abendroth-Smith, J., 2003, Development and evaluation of a biomechanics concept inventory: Sports Biomechanics, v. 2, no. 2, p. 267–277, doi:10.1080/14763140308522823.

LeCompte, M.D., and Schensul, J.J., 1999, Designing and Conducting Ethnographic Research: Walnut Creek, California, Altamira Press, 220 p.

Libarkin, J.C., and Anderson, S.W., 2005, Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory: Journal of Geoscience Education, v. 53, no. 4, p. 394–401.

Libarkin, J.C., and Anderson, S.W., 2006, The Geoscience Concept Inventory: Application of Rasch analysis to concept inventory development in higher education, *in* Liu, X., and Boone, W.J., eds., Applications of Rasch Measurement in Science Education: Maple Grove, Minnesota, JAM Press, p. 45–73.

Libarkin, J.C., and Geraghty Ward, E.M., 2011, this volume, The qualitative underpinnings of quantitative concept inventory questions, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(04).

Lindell, R.S., 2005, Measuring conceptual change in college students' understanding of lunar phases: American Institute of Physics, Conference Proceedings, v. 790, no. 1, p. 53–56.

Lindell, R.S., and Sommer, S.R., 2004, Using the lunar phases concept inventory to investigate college students' pre-instructional mental models of lunar phases: American Institute of Physics, Conference Proceedings, v. 720, no. 1, p. 73–76.

Lindell, R.S., Peak, E., and Foster, T.M., 2007, Are they all created equal? A comparison of different concept inventory development methodologies: American Institute of Physics, Conference Proceedings, v. 883, p. 14–17.

Lopez, W., 1996, Communication validity and rating scales: Rasch Measurement Transactions, v. 10, no. 1, p. 482–483.

National Foundation for Educational Research, 2009, Attainment testing: http://www.nfer.ac.uk/research-areas/assessment/attainment-testing.cfm (1 September 2009).

Novak, J.D., 1977, A Theory of Education: Ithaca, New York, Cornell University Press, 295 p.

Odom, A.L., and Barrow, L.H., 1995, Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction: Journal of Research in Science Teaching, v. 32, p. 45–61, doi:10.1002/tea.3660320106.

Otis, A.S., and Lennon, R.T., 1989, Otis-Lennon School Ability Test (6th ed.): San Antonio, Texas, Psychological Corp., Harcourt Brace Jovanovich, 22 p.

Prather, E.E., Rudolph, A.L., Brissenden, G., and Schlingman, W.M., 2009, A national study assessing the teaching and learning of introductory astronomy. Part I. The effect of interactive instruction: American Journal of Physics, v. 77, p. 320–330.

Quenk, N.L., 2009, Essentials of Meyers-Briggs Type Indicator Assessment: Essentials of Psychological Assessment (2nd ed.): Hoboken, New Jersey, John Wiley & Sons, 195 p.

Sapsford, R., 1999, Survey Research: London, Sage Publications, 258 p.

Schmeiser, C., and Welch, C., 2006, Test Development, *in* Brennan, R.L., ed., Educational Measurement: Westport, Connecticut, Praeger Publishers, p. 307–354.

Segar, R., 1999, Introduction to Ocean Sciences (2nd ed.): New York, W.W. Norton & Co., 525 p.

Slosson, R.L., Nicholson, R.L., and Hibpshman, T.H., 1998, Slosson Intelligence Test for Children and Adults: East Aurora, New York, Slosson Educational Publications, 45 p.

Smith, M.K., Wood, W.B., and Knight, J.K., 2008, The genetics concept assessment: A new concept inventory for gauging student understanding of genetics: CBE Life Sciences Education, v. 7, p. 422–430.

Tan, K.C.D., Taber, K.S., Xiufeng, L., Coll, R.K., Lorenzo, M., Jia, L., Goh, N.K., and Chia, L.S., 2008, Students' conceptions of ionization energy: A cross-cultural study: International Journal of Science Education, v. 30, no. 2, p. 265–285, doi:10.1080/09500690701385258.

Treagust, D.F., 1988, Development and use of diagnostic tests to evaluate students' misconceptions in science: International Journal of Science Education, v. 10, no. 2, p. 159–169, doi:10.1080/0950069880100204.

Wage, K.E., Buck, J.R., Wright, C.H.G., Welch, T.B., and Thad, B., 2005, The signals and systems concept inventory: IEEE Transactions on Education, v. 48, no. 3, p. 448–461, doi:10.1109/TE.2005.849746.

Manuscript Accepted by the Society 23 June 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

Geobiological opportunities to learn at U.S. fossil parks

Renee M. Clary
Department of Geosciences, Mississippi State University, P.O. Box 5448, Mississippi State, Mississippi 39762, USA

James H. Wandersee
Department of Educational Theory, Policy, and Practice, Louisiana State University, Peabody 223F, Baton Rouge, Louisiana 70803, USA

ABSTRACT

Unlike other informal sites, fossil parks provide visitors collecting opportunities that result in ownership of a small number of fossils. In 2003, we investigated the first three identified U.S. fossil parks at Hamburg, New York; Sylvania, Ohio; and Rockford, Iowa. Case study analyses determined the opportunities to learn geobiology at each site. Data collection proceeded through lived learning experiences, and included field notes, photographic records, informal conversations with park participants, brochures, and on-site signage. Through constant comparative methods, six variable categories converged for fossil park development: (1) informative previsit Web site, (2) authentic collecting in situ, (3) authentic collecting tools, (4) accessibility, (5) fossil identification, and (6) visitor education. These variables were optimized in a model of fossil park design. In 2005, fossil parks at Sharonville, Ohio, and Fossil, Oregon, were investigated in phase 2 of our study, and in 2006, our third case study researched fossil parks in Aurora, North Carolina, and Republic, Washington. Analysis of the seven U.S. fossil park data sets resulted in the emergence of key variables that affected the visitors' opportunities to learn geobiology concepts at fossil parks: (1) authenticity of experience, (2) age of fossils, (3) fossil-collection training and facilities, (4) availability of on-site paleontological mentors, (5) fossil identification via signage and brochures, (6) site organization and wayfinding signs, and (7) accessibility of site, including safety. The seven U.S. fossil parks were ranked against these variables according to their effectiveness as informal science education sites. We conclude that fossil parks can provide valuable informal geobiology education that can contribute to the public's geobiological literacy.

INTRODUCTION

In 2003, the international news agency CNN reported the existence of an innovative, informal geoscience educational experience: the Fossil Park at Sylvania, Ohio (Cable News Network, 2003). Sylvania was one of three U.S. fossil parks that embraced a unique educational mission. Whereas most museums, National Parks, and other informal education sites display fossils within locked cases or exhibit them in their native strata for visitor viewing only, a fossil park is developed on the concept that visitors search for and retain the fossils they find, within the guidelines set by the park. In 2003, Sylvania was identified as one of these new, pioneering U.S. fossil parks, along with Penn-Dixie Paleontological Park (New York) and Rockford Fossil and Prairie Park (Iowa).

Clary, R.M., and Wandersee, J.H., 2011, Geobiological opportunities to learn at U.S. fossil parks, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 113–134, doi:10.1130/2011.2474(09). For permission to copy, contact editing@geosociety.org.

We began our investigation of fossil parks in 2003 (Fig. 1), when we began to research the first identified U.S. fossil parks. Our second case study in 2005 extended this research, and in 2006, we initiated a third case study with on-site investigations of two additional locations. Table 1 lists the fossil parks we investigated in our three phases of case study analysis and details the order of study and the year in which we conducted each site visit.

Beginning in 2003, we surveyed the facilities at the three originally identified fossil park locations (Penn-Dixie Paleontological Park, the Fossil Park, Rockford Fossil and Prairie Park), collected fossils at each site, observed park participants, and conversed with employees, volunteers, and visitors. Next, we coded and analyzed original field data, and key findings emerged. From our initial investigation, the data and analyses informed our development through grounded theory of an optimal fossil park design.

In 2005, our fossil park research was expanded through the identification of two additional fossil park sites: Wheeler High School Fossil Beds (Oregon), and Trammel Fossil Park (Ohio). We visited the sites, collected fossils, and interacted with the employees, volunteers, and visitors present. We analyzed data from these sites, both of which are located within small towns, against our fossil park model. Finally, we compared the fossil parks to determine their geobiological opportunities to learn as outdoor teaching laboratories, and linked our research with the findings of the National Research Council study, *America's Lab Report* (Singer et al., 2005).

The third case study in the fossil park research investigation incorporated two additional U.S. fossil parks: Aurora Fossil Museum and Park (North Carolina), and Stonerose Interpretive Center (Washington). We conducted site visits in 2006 that included fossil collecting, interviews, and observations. In 2007, we returned to our seven data sets and determined the key variables that optimized visitors' educational experiences. We then ranked and analyzed the seven fossil parks for their opportunities to learn geobiology in the field.

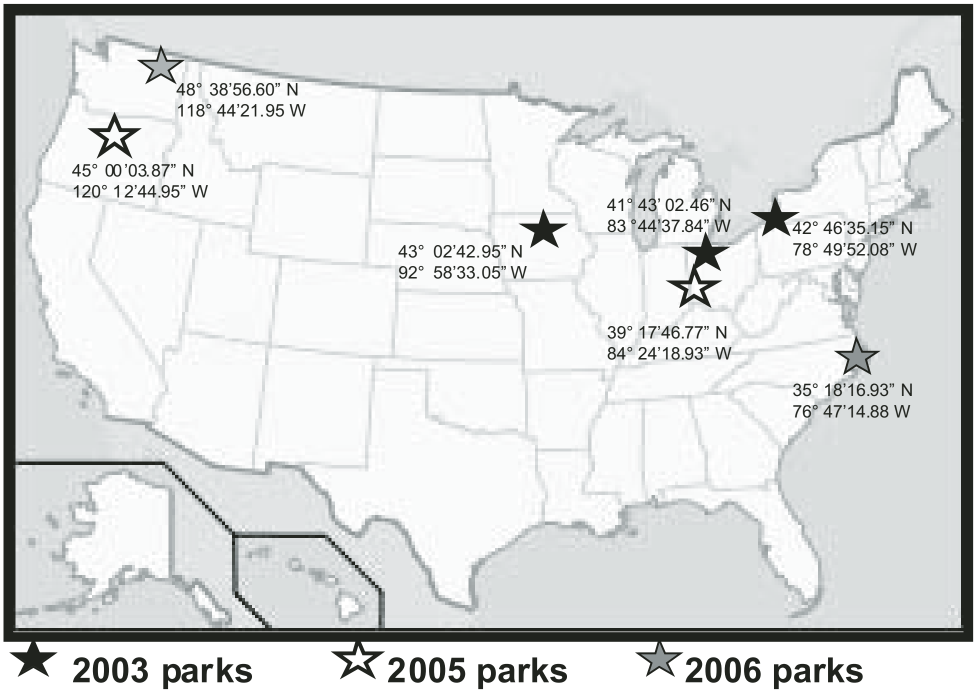


Figure 1. The geographic locations of the seven U.S. fossil parks involved in this longitudinal qualitative research investigation. Black stars show the sites of the first three U.S. fossil parks we investigated (Penn-Dixie Paleontological Park in New York, the Fossil Park at Sylvania in Ohio, and Rockford Fossil and Prairie Park in Iowa). White stars locate the fossil parks in phase 2 of our 2005 study (Trammel Fossil Park in Ohio, and Wheeler High School Fossil Beds in Oregon). Gray stars denote the locations of the last fossil parks investigated in phase 3 of our 2006 study (Aurora Fossil Museum in North Carolina, and Stonerose Interpretive Center in Washington).

TABLE 1. SEVEN DIFFERENT FOSSIL PARKS RESEARCHED DURING THREE CASE STUDY INVESTIGATIONS (PHASE 1, PHASE 2, PHASE 3) FROM 2003 THROUGH 2006

Fossil park name	Location	Research phase/year	Order of study
Penn-Dixie Paleontological Park	Hamburg, New York	Phase 1—2003	1
The Fossil Park at Sylvania	Sylvania, Ohio	Phase 1—2003	2
Rockford Fossil and Prairie Park	Rockford, Iowa	Phase 1—2003	3
Trammel Fossil Park	Sharonville, Ohio	Phase 2—2005	4
Wheeler High School Fossil Beds	Fossil, Oregon	Phase 2—2005	5
Aurora Fossil Museum	Aurora, North Carolina	Phase 3—2006	6
Stonerose Interpretive Center	Republic, Washington	Phase 3—2006	7

Problem Statement

This research focused on the identification of U.S. fossil parks, a relatively new venue of paleontological informal education for the general public. The role of fossil parks within more traditional information education sites (e.g., museums, National Parks, U.S. state parks) was ascertained through both literature searches and site visits. The primary research focus was to determine the informal opportunities to learn geobiology at U.S. fossil parks through case study. Guiding the research investigations were the principles of active, meaningful and mindful learning, as established within the learning theory of human constructivism.

Fossil Parks as Unique Informal Learning Venues

Hose (1995, p. 16) published an early definition of geotourism as the "provision of interpretative service facilities to enable tourists to acquire knowledge and understanding of the geology and geomorphology of a site beyond the level of mere aesthetic appreciation." By 2005, the state of Arizona was investigating geotourism as a method to sustain and enhance their region, protecting it from harmful tourist expansion (Long, 2005). Undoubtedly, geotourism is popular. In 2002, approximately one third of Americans (55 million) were interested in geotourism (Stueve et al., 2002). An especially encouraging statistic to science educators was that 53% of U.S. travelers acknowledged that learning enhances their travel expectations (Stueve et al., 2002).

However, visitors' learning opportunities at informal educational sites vary in both quality and quantity of instructional materials and experiences. The combination of factors considered the "best practice" in interpretive design at geotourist sites has not been empirically examined (Patzak, 2000). Science centers typically depict science out of context, especially if free-standing exhibits are disconnected from the world in which they originate (Persson, 2000). Conversely, Mir (2003) noted that outdoor science parks add a dimension to informal science education and can appeal to visitors without the cultural message that buildings project.

The Conference on Earth Heritage: World Heritage in Wareham, Dorset, UK, produced interesting results and suggestions for geotourism (Larwood and Durham, 2005). Experts in informal geotourism noted that poor interpretation counteracts a geotourist site. Since fossils represent a specialized interest outside most people's general knowledge, the interpretation of a site is crucial (Larwood and Durham, 2005). Another important factor is local community involvement in a geotourist site, which can greatly influence the site's success (Larwood and Durham, 2005). Patzak (2000) noted that there was no demonstrable conflict between tourism promotion and geoconservation.

The United Nations Educational, Scientific, and Cultural Organization (UNESCO) unveiled its Geopark initiative to promote a worldwide network of extraordinary examples of Earth's geological diversity. The Geopark initiative emphasizes the use of unique geological sites in educating the general public, the use of these sites to ensure sustainable development through geotourism, and the conservation of the world's geological heritage for the future (Patzak, 2000). Although the United States does not have a designated Geopark, the National Park System (NPS) encompasses a variety of National Parks and monuments that contain fossil remains for public informal education. The NPS published guidelines for paleontological resources, defined as including both organic and mineralized remains in body or trace fossil form, specifically requires that the fossils be protected, preserved, and managed for public education, interpretation, and scientific research (National Park Service, 1991). The National Park Service further mandates that the fossils be protected from harm, theft, or destruction. Therefore, where necessary, the NPS will guard locations of fossil resources if pilfering and removal of fossils is suspected to result from disclosed locations.

Unfortunately, the concept of "ownership" of fossils, even protected ones, has resulted in vandalism and desecration of paleontological sites. The Petrified Forest National Park in Arizona exhibits signage with "Your heritage is being vandalized every day by theft losses of petrified wood of 14 tons a year, mostly a small piece at a time." Vandalism and collecting are not limited to easily removed bits of materials, nor are they limited to U.S. National Parks. When we conducted field research in 2005 along the Lyme Regis coast, UK, we unwittingly stumbled onto dinosaur tracks for sale. We learned later that the facility we visited was raided by authorities, and the tracks were confiscated. The tracks were illegally quarried from Bendrick Rock, resulting in obvious damage to a protected site in Wales (BBC News, 2006).

Fossil parks differ substantially from U.S. National Parks and protected global sites in that visitors can collect and keep the fossils they find. Therefore, the U.S. fossil parks fill a unique niche by permitting the visitor ownership of a small, limited number of personally collected fossils, for individual study and enjoyment. Notably, the U.S. fossil parks are not lagerstätten, or sites with extraordinary preservation or diversity of fossils. Fossil park sites are established in locations that are not only fossiliferous, but have been extensively researched and collected. These informal education sites serve to bridge "traditional" informal sites that only display fossils with unstructured field opportunities where visitors collect fossils without site and/or fossil education. While fossil parks expect visitors to leave with personally collected fossils, sustainability of the site is considered in the park design.

Locating the Researchers

Our backgrounds in geology, biology, and science education undoubtedly influenced the types of data we collected, as well as the analyses. As EarthScholars Research Group, we have served as consultants for signage at the Doris I. Schnuck Children's Garden: A Missouri Adventure (Missouri Botanical Gardens, St. Louis, Missouri), and as informal education consultants and trail designers at Barton Arboretum (Burden Research Center, Baton Rouge, Louisiana). We previously analyzed signage systems within informal education sites and designed a template for optimized science signage construction (Wandersee and Clary, 2007). Our previous research also includes analysis of informal educational design (Clary et al., 2009), as well as optimized opportunities to learn in informal science field experiences (Wandersee and Clary, 2006).

Theoretical Frameworks

Learning Theory of Human Constructivism

We utilize the learning theory of human constructivism for educational research. This learning theory was originally proposed by science educator Joseph Novak (1977) and has been extensively researched and elaborated (e.g., Mintzes et al., 2000, 1998). Human constructivism is a relatively new synthesis based upon psychologist David Ausubel's previous work (Ausubel et al., 1978; Ausubel, 1968, 1963) and developed through Novak's (1963) pioneering work that proposed fundamental principles of research in science education. This learning theory has been advanced through research in cognitive science, epistemology, and the nature of science. Research investigations utilizing human constructivism have been reported in numerous science education venues, including the special issue of the *Journal of Research in Science Teaching*, devoted entirely to human constructivism–driven research (Novak and Wandersee, 1991).

Principles of Active, Meaningful, and Mindful Learning

The following theoretical principles taken from the learning theory of human constructivism are relevant to this fossil park investigation: (1) Humans seek to make meaning; (2) learning results when the meaning of experience changes; (3) knowledge is conceptual; (4) concepts are those patterns that humans identify and label; (5) concepts are used in semantic sets of propositions for thinking and expansion of learning; (6) meaningful learning occurs when new concepts are connected in a substantive, nonverbatim way to prior knowledge and experiences; resulting in (7) cognitive restructuring. Therefore, the best teaching is learning-driven, and the goal of science education is to foster conceptual change. When conceptual change occurs, learners form increasingly powerful knowledge representations that reflect contemporary scientific thought.

When learners monitor and take control of their learning, meaningful learning occurs. This is identified by learners' ability to plan, monitor, and regulate their learning, which is in turn responsible for conceptual change (Novak, 1998; Novak and Gowin, 1984). Metacognition, which can facilitate conceptual change, is the knowledge, awareness, and control of the learning process by the learner (Gunstone and Mitchell, 1998). Therefore, science education is successful when learners develop and exhibit new thoughts and feelings about the natural world. Gowin (1981) noted that successful science education scaffolds the integration of thinking, feeling, and acting within the learner.

Meaningful learning is nonverbal and nonarbitrary and results in a substantive incorporation of new knowledge within an existing conceptual framework (Ausubel et al., 1978; Ausubel, 1968, 1963). Therefore, meaningful learning will occur when learners have integrated and organized conceptual knowledge frameworks (Novak, 1998). Additionally, when learners are aware of the context of information and recognize that knowledge is not static, they can engage in mindful learning (Langer, 1997). Conversely, when learners memorize facts without context and an awareness of the changing nature of information, they have frameworks with limited uses.

DeBoer noted, "If a single word had to be chosen to describe the goals of science educators during the last 30-year period that began in the late 1950s, it would have to be inquiry" (1991, p. 206). Several research studies have affirmed the benefits of active learning (Lawrenz et al., 2005; McConnell et al., 2003; Michael and Modell, 2003). Active learning can provide authentic research experiences, resulting in better learner understanding of the research process (Felzien and Cooper, 2005; Hemler and Repine, 2006). Paralleling Gowin's (1981) science education integration, Lord and Orkwiszewski (2006) reported increased learning alongside positive affective outcomes with inquiry-based exercises. Our previous research (Clary and Wandersee, 2008) reported significant learning outcomes with active learning investigations within informal educational sites.

Role of Informal Science Education

Informal science education and free-choice learning are well-established as important venues for learners (McComas, 2006, 1996; Wandersee and Clary, 2006). Not only does informal education provide the default learning environment for most of the adult population, but school-age students also typically engage in informal learning more often than learning in traditional environments (Falk and Dierking, 2002). Several researchers have investigated motivators, assessments, and the theoretical bases behind learning in informal environments (Anderson et al., 2003; Falk, 2001; Falk and Dierking, 2000; Meredith et al., 1997; Orion and Hofstein, 1994; Rennie and Johnston, 2004). Roy and Doss (2007) reported that informal educational programs can engage citizens in data collection, independent research, and global problems. Informal learning environments can supply an interdisciplinary science "big picture" for students (Clary and Wandersee, 2009) and provide holistic experiences that are retained (Bernstein, 2003).

METHODS

This research investigation utilized a mixed methodology design (Creswell, 1994; Tashakkori and Teddlie, 1998). However, we focus on the qualitative research aspect in this chapter. Phenomenology provides a philosophical affiliation for our ongoing research on U.S. fossil parks. With a phenomenological methodology, the research question is stated broadly (Nieswiadomy, 1993) and analysis proceeds as we examine the detailed, thick descriptions recorded during our lived experiences at the fossil parks. Although the results from the fossil park research have been triangulated (Denzin, 1978) through several types of data, the products of our research investigations provide one rendering of the experiences available by visitors who engage these parks (Bogdan and Biklen, 1998).

We used a constant comparative method (Glaser, 1978; Glaser and Strauss, 1967; Strauss, 1987) for data generation at the first three U.S. fossil parks. Each fossil park investigation contributed data to identification and emergence of the key factors that helped to make a fossil park an effective informal educational design. As additional fossil parks were identified in 2005 (phase 2) and 2006 (phase 3), these were investigated to contribute new sources of data to the optimized model. Once we collected and analyzed data, the individual case study investigative results of fossil parks were generalized through grounded theory to optimize both fossil park designs and educational experiences for those who visited the facilities.

Trustworthiness and Generalizability

Our on-site visits to fossil parks were limited to short contacts within each. Therefore, our participant sample was small and should not be considered to be representative of all visitors to any of the individual U.S. fossil parks. Alternatively, the investigation of fossil parks was purposive and inclusive. We researched every informal site that we could locate that was specifically dedicated to the informal education of the visitor through the collection and retention of fossils.

Our previous experiences and backgrounds as science education researchers undoubtedly influenced our perceptions and objectivity during data generation. While complete objectivity is unattainable in qualitative research (Harper and Kuh, 2007), we utilized Lincoln and Guba's (1985) guidelines for establishing quality criteria, and achieved external validity through thick descriptions of our lived experiences. Denzin and Lincoln's (2000) criteria of credibility, transferability, dependability, and confirmability helped us to establish the quality and trustworthiness of our study. We further utilized the guidelines in geoscience qualitative research, as outlined by Stokes (this volume). Qualitative triangulation (Golafshani, 2003), internal auditing (Manning, 1997), and critical subjectivity (Lincoln, 1995) helped us to meet these criteria, establish reliability and validity, and reach authentic conclusions in our analyses.

Through the analysis of our thick description, our credibility is established through the believability of our results, judged through the lens of our lived experiences at fossil parks. Because our backgrounds influenced our perspectives, and our interactions within each fossil park were limited, the transferability of this research to other informal educational settings may be restricted in its specific applicability.

While the limited hours of collecting and lived experiences, the restricted interactions with fossil parks' visitors, and the specific days and seasons of our research visits are a small sample, we make the general assumption that our interactions were typical of visitor experiences within each fossil park. The thick descriptions gathered from each fossil park site provide an overall guideline for future researchers to replicate this research study, though undoubtedly the data collection and conclusions will not be in total agreement with our experiences and results. Thick descriptions are used to achieve dependability of our results.

It is also through thick descriptions that we achieve confirmability. Visitors' comments and interactions, notes from our lived experiences, and data we collected via Web sites and printed material are provided to support the conclusions we made in our research, as well as manage our potential biases as science educational researchers. Our research and data also have been presented and made available to science education research communities at the International Geological Congress in Florence, Italy (2004), the American Geophysical Union (2004, 2005), the National Association for Research in Science Teaching (2005), and the Geological Society of America (2006) meetings.

Therefore, while this research is generalizable to informal educational sites in the United States that allow collection and retention of fossils, the application of our results to broader informal educational science sites or sites beyond the United States is yet to be determined.

Investigative Techniques and Data Reduction

In 2003, the three U.S. fossil parks that were identified by the international news agency CNN (the Fossil Park at Sylvania, Penn-Dixie Paleontological Park, Rockford Fossil and Prairie Park) were subjected to case study analysis (Yin, 2003) as the first phase of our fossil park research. Results were triangulated through multiple data-generating techniques. When additional fossil parks were identified in 2005 (phase 2, $n = 2$) and again in 2006 (phase 3, $n = 2$), we utilized case study analyses and triangulated the results.

In each of the seven fossil park investigations, we utilized naturalistic lived learning experiences by visiting each park and collecting fossils at each site. Our goal for each site visit was 8 h minimum of site interaction time. Rockford Fossil and Prairie Park in Iowa and Trammel Fossil Park in Ohio involved additional site hours because of limited participants. We extended our observations in order to encounter and observe more visitors/collectors at these fossil parks.

We observed visitors to ascertain the effect of the fossil park landscape and available experiences within it on participant behavior. Additionally, other participants at the site, including volunteers and employees of the fossil parks, were watched and monitored. We also engaged in informal, unstructured conversations (Wolcott, 2005) with participants where appropriate. There was no predetermined interview protocol, but conversation proceeded based on the site facilities and available collecting experiences. Because we were participating in fossil recovery alongside park visitors, our manner was unobtrusive. Our field notes on these conversations were made after the conversation concluded.

Conversations can be described as "natural," leading to the probability of low observer effect. No risks were identified for visitor participation with these on-site conversations, and we informed visitors of our fossil park investigative research and secured their permissions prior to recording their comments.

We collected literature for each fossil park, acquiring both on-site paper handouts as well as posted information on the internet. The literature was analyzed for relevance to visitors' educational experiences through coding and content analysis. Additionally, we photographed signage at each site and analyzed the contents. Our impressions of participants' behaviors and the fossil park itself (physical site, geology, facilities available, fossil-collecting experiences, paleontological mentors) were recorded as field notes with thick description. Fossil park sites also were documented through photography. However, the thick descriptions constituted the bulk of our data generation for all seven fossil park site investigations. These thick descriptions of our lived experiences at fossil parks provided data through which our findings emerged. We include narrative selections of these fossil park case studies to document our qualitative inquiry process.

For each case study investigation (phase 1: *n* = 3 fossil parks in 2003; phase 2: *n* = 2 fossil parks in 2005; phase 3: *n* = 2 fossil parks in 2006), we determined the categories to be examined. Field notes, photographic records, brochures, and interviews were coded and analyzed. Through the constant comparative method (Glaser, 1978; Glaser and Strauss, 1967; Strauss, 1987), data were reduced and interpreted (Marshall and Rossman, 1989). The analyses involved coding and sorting data into categories. For content analysis of brochures and signage, we utilized Neuendorf's (2002) guidelines, while for conversational, unstructured interview analysis, we employed the methods outlined by Chi (1997).

For the content analysis of published literature, Web site information, and our collected thick descriptions, we first identified the important variables for optimal learning environments at fossil parks and defined these concepts using the applicable published research literature as a guideline. We then identified the categories that would best reflect the variable parameters. Categories in our coding scheme emerged to include accessibility, biodiversity of fossils, brochures, collecting tools, data collection procedures, ease of site location, educational activities, fees, fossil density, fossil identification through charts and signage, fossil information (genus identification, characteristics), fossil preparation stations, hours of operation, museum or visitor center on-site, paleontological mentors, public awareness of site, safety, stratigraphic context, vertical relief, visitor population, and Web site information.

We coded field notes, literature, and interviews independently, and then recoded randomly selected samples of literature, researcher descriptions, and interviews. We determined the interrater reliability at 95%. Further analysis of categories resulted in the identification of themes and patterns (Marshall and Rossman, 1989), and through the themes and patterns, a consolidated model emerged for an idealized fossil park design in 2003, and for optimized visitor experiences in 2007 (Tesch, 1990).

Phase 1: The First Three U.S. Fossil Parks (2003)

Prior to site investigation, we researched the first three fossil parks identified by CNN, which included identifying and accessing any online materials, including advertising. In August 2003, we conducted on-site research at all three sites. Our first on-site visit was to Penn-Dixie Paleontological Park in Hamburg, New York. We proceeded next to the Fossil Park at Sylvania, Ohio, and lastly to Rockford Fossil and Prairie Park, Iowa.

Penn-Dixie Paleontological Park

Penn-Dixie Paleontological Park, once utilized as a quarry for an aggregate cement operation, was established through a partnership between the Hamburg Natural Society and the Town of Hamburg (Bastedo, 2000). This outdoor informal educational site is geologically situated within a 380-m.y.-old, highly fossiliferous exposure of the Devonian Windom Shale. Other units represented on the site include the Genundewa Limestone, North Evans Limestone, Tichenor Limestone, and Wanakah Shale. Fossils collected here attest to the paleoecology of the Devonian when warm tropical seas covered New York within 20°–30° south of the equator (Bastedo, 2000). Corals, brachiopods, bryozoans, trilobites, crinoid columnals, gastropods, bivalve mollusks, cephalopods, and fish remains have been recovered.

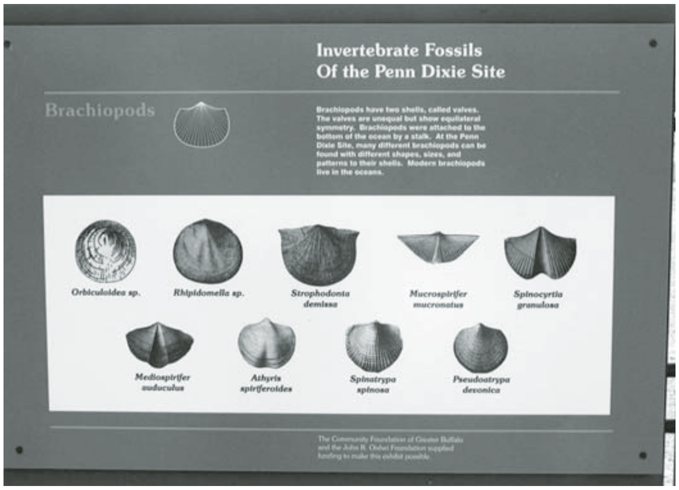
Although a Web site existed for Penn-Dixie, it was not fully developed. A brief description, hours, admission fees, and associated web links were included. The posted map provided good directions to the site, and we had little trouble locating the park.

We arrived at Penn-Dixie shortly after it opened for the day and paid the nominal collecting fee (US$4). Upon admission, we were provided with a fossil-collecting card with eight ink-drawing renderings of fossil corals (solitary rugose and colonial tabulate), bryozoans, trilobites (flattened and enrolled), crinoid stems (lateral and cross-section views), bivalve mollusks, brachiopods, cephalopods (coiled and straight forms), and gastropods (planispiral and trochospiral species). Additional handouts were provided that briefly described fossil formation and preservation, the Penn-Dixie quarry site schematic cross section, and a pen-and-ink paleoenvironmental reconstruction. We also received a small flyer advertising the "Dinosaur Days!" at the Science and Nature Store in Blasdell, New York. We observed no group reading any handouts at the site other than the fossil-collecting cards.

At the fee station, we were assigned to "Michael" (not his actual name), an undergraduate geology student who would serve as our paleontological mentor. Michael was enthusiastic and friendly. Although sunny and warm, we did not find the conditions overbearing and moved directly to the collecting site. Most of the old quarry site is level, although there are areas of vertical relief, naturally created through erosion and through exposure with earth-moving machinery. We first collected at one of these small vertical exposures and were allowed to use the rock hammers and chisels we brought. There were no requirements for personal safety equipment. We retrieved partial trilobites (*Phacops rana*, *Greenops boothi*), brachiopods (*Mucrospirifer*, *Athyris*, and *Spinocyrtia* species), rugose corals (?*Sterolasma*), crinoid columnals, bryozoans, trace fossils, and a possible partial straight-chambered cephalopod (?*Michlenoceras*). Fossil recovery was fairly easy, and we did not utilize rock hammers often because the fossils were typically eroded out of their shale matrix. During our fossil recovery, we conversed with Michael, utilizing an informal and unstructured style.

Other groups were collecting, including students participating in a summer workshop through a local college ($n = 14$), three solitary individuals, and a mother and son team. We observed each group and collected field notes. Additionally, we moved our personal collecting operations to within talking distance of groups and engaged in informal, unstructured conversations. Within the larger student group, we observed that students typically brought their "finds" to the adult supervisors or the paleontological mentor, or checked their identifications against four large information boards at the site (Fig. 2). We found the information boards to be helpful, but we were not brought to them by our mentor. We independently recorded that this was a missed opportunity to educate the visitor, given that the fossil descriptions on the information boards were more detailed than the handout. Another option available at the site was an open-air classroom, but it was not in use during our site visit.

After collecting, observing, and conversing with fellow collectors, we ended our site visit by returning to the admission shelter. A middle school speech teacher was one of the volunteers who collected money at the park's entrance, along with the park director, Jerry Bastedo. We explained our interest in the fossil park concept and discussed typical attendance at the Penn-Dixie park, programs within the park, and their perceptions of park activities.



Brachiopods have two shells, called valves. The valves are unequal but show equilateral symmetry. Brachiopods were attached to the bottom of the ocean by a stalk. At the Penn Dixie Site, many different brachiopods can be found with different shapes, sizes, and patterns to their shells. Modern brachiopods live in the oceans.

The Community Foundation of Greater Buffalo and the John R. Oishei Foundation supplied funding to make this exhibit possible.

Figure 2. Penn-Dixie Paleontological Park has several large information signs at the collecting site, including this board on brachiopods. Collecting groups were situated at a distance from these signs, but younger visitors often returned to the figures to check their fossil identifications against the diagrams.

Other pamphlets were available at the admissions area. The Penn-Dixie Paleontological and Outdoor Education Center trifold brochure incorporated information on fossil collecting, astronomy, ornithology, and the future of the Hamburg Natural History Society (HNHS). This brochure included a membership application to the HNHS. Other literature available at the station included flyers on the Saturday evening astronomy programs (two flyers), information on special occasion outings and groups (birthday parties, graduations), a Halloween special event at the park, and Western New York Earth Science Day. Some copies of the *Penn-Dixie Chronicle*, a monthly newsletter for the HNHS, were available at the station as well. We collected and reviewed five newsletters (April through May 2003), which primarily contained information on upcoming events, park updates, and membership.

Fossil Park at Sylvania

Our next stop in our first phase of fossil park case study was the Fossil Park at Sylvania, Ohio, which was the fossil park featured in CNN's 2003 article on the new fossil park concept. Although our methods for on-site study were similar to those we utilized at Penn-Dixie Paleontological Park, our visit to Penn-Dixie undoubtedly influenced our examination of Sylvania in that it helped to provide a focus and define the boundaries and variables for our on-site fossil park investigations. Following our Penn-Dixie site research, we discussed our perceptions of the benefits of collecting fossils from areas with topographic relief, the helpfulness of the site mentor, and the organization and information contained on the display boards for fossil identification at the site. We also noted the fairly long walk between some fossil-collecting areas and the display board and felt that this distance could deter visitors from accurately identifying their fossils as they recovered them. Therefore, in our future fossil park investigations, we searched for these helpful elements (topographic relief, on-site mentor, identification boards) and hindering factors (distance between fossil identification information and collecting sites).

The Fossil Park at Sylvania was developed on an abandoned quarry site. Encompassing only 5 acres (1.6 square hectometers), it is much smaller than the Penn-Dixie site in New York. The Fossil Park is geologically situated upon the bottom of the Devonian Silica Formation with the Dundee Limestone exposed at the upper quarry ledges. Within the 375-m.y.-old Silica Formation, over 200 species of fossils have been recovered, including rugose corals, bryozoans, brachiopods, bivalve mollusks, crinoids, edrioasteroids, fish, and trilobites (Stoll, 2001). Fossils collected from the Silica Formation comprise the major portion of the Devonian collection at the Smithsonian Institution, and they are representative of a marine environment. The formation is known to paleontologists and amateur collectors around the world for its excellent preservation of Devonian invertebrates.

Fossil collecting at the fossil park is not done in situ. The Hanson Aggregate Midwest mining company approached the local park system with an idea of creating a fossil park situated away from their main quarry. For safety reasons, visitors are not allowed to collect on-site in the active quarry, although Hanson Aggregate reported that numerous requests were made each year. The Fossil Park at Sylvania, managed by The Olander Park System (TOPS), opened in 2001 as the result of the quarry and park partnership. Visitors are provided spoil piles on two concrete pads at the bottom of the abandoned 25-ft-deep (7.6 m) quarry, through which they may hunt for fossil treasures.

We did not uncover a lot of information about the park prior to our arrival, although the CNN (2003) article and a visitor's guide (Ohio.com) were posted online. Both sources were descriptive and weighted toward tourism. The Fossil Park at Sylvania did not have a Web site, and the map we found posted for the fossil park proved very confusing. When we stopped for directions, the local residents could not direct us to the park location. We resorted to another map and eventually found our way to the site.

Upon arrival, we found the park had not yet opened. This gave us an opportunity to investigate the surrounding area before a potential crowd arrived. When "Andy," the resident naturalist, opened the site, we quickly drove to the parking lot. The facility was new and had access for wheelchair users. There was no admission fee, although plans for the park's future included a nominal fee (Downing, 2003).

We walked down the ramp from the parking lot into the abandoned quarry. There, water-saturated shale spoil piles awaited the geotourists on their concrete pads (Fig. 3). Covered tables were available for examining fossils, along with large plastic wash stations for cleaning the sticky weathered shale residue from the fossils collected. No tools were allowed for collecting, and hammers, chisels, shovels, and screwdrivers were banned at the site. Although the wash station provided the means with which to remove the mud from the fossils, it was underutilized. We observed children playing with the water and in the mud pile at the station's base.

Figure 3. The Fossil Park at Sylvania in Ohio did not offer fossil-collecting opportunities within strata. Instead, spoil piles were available on two concrete pads. The facility accommodated handicapped visitors, and tables were available for sorting fossils.

Upon our arrival, we signed the guest book and were given a handout. One side of the handout listed the fossil park policies that no tools were allowed for safety reasons, and fossil hunting was restricted to the enclosed areas. On the reverse side of the handout were some diagrams featuring a trilobite, brachiopod, horn coral, crinoid, and bryozoans. Very basic descriptions of these invertebrates accompanied the figures as the "Commonly Found Devonian Fossils." We did not observe anyone on-site utilizing this identification handout.

The shale of the Silica Formation, referred to as paper shale, was soft and very easy to break apart. However, we found that the spoil piles were well collected, and even after an hour of collecting, the EarthScholars team could only claim one decent, small *Phacops* trilobite. After several hours of collecting, we were able to procure a few trilobite portions, primarily the thorax portions of individuals or molts, and a few bryozoans, but the fossils were extremely small. The spoil piles were slippery, and suspected fossils were encased in mud.

We collected elbow-to-elbow with other visitors. No organized school or club groups were present, but several families and extended families visited the site. We counted a total of 71 visitors. Where appropriate, we engaged visitors in informal, unstructured conversation. We observed that within family units, children turned to adults for guidance in what they were uncovering. Often, the object in question was not a fossil. When adults could not answer questions or suspected a fossil find, they typically sent their children to the on-site naturalist for further identification.

We observed some groups collecting on the side of the quarry near the naturalist's station. We then moved to the naturalist's station and spoke with him about the park. During our conversation, several children came to ask about the fossils they uncovered. The best specimen we observed was a 3 cm brachiopod collected by an 8-yr-old boy. Although Andy had reference notebooks available to him, there were no additional copies or posted identification diagrams within the naturalist's area.

Within the naturalist's area, there was a kiosk with a wall devoted to "Fossil Finders" with photographs of particularly good fossil finds at the park. Another wall was devoted to "Fossil Facts," but specific taxonomic genera were not identified. Instead basic invertebrate groups were described, similar to the handout we received upon entry. A "Tourist ? Info" wall had plastic containers for trifold brochures, but it was empty at the time of our visit. We found that the most interesting side of the kiosk featured an interactive fossil quiz. The "Fossil Challenge" posed 15 questions for visitors. The answers were obtained by lifting the hinged wooden doors.

We retrieved an additional pamphlet for the Fossil Park at Sylvania from the naturalist's station. This trifold brochure included basic descriptive information about the fossil park. A general description, safety features and handicap amenities, and brief descriptions of the partners (TOPS and Hanson Aggregates Midwest) were included, but fossil types were absent.

Rockford Fossil and Prairie Park

Our final stop in phase 1 of the fossil park research was Rockford Fossil and Prairie Park in Iowa. We approached this fossil park in a similar manner to the other two U.S. fossil parks we previously investigated, but our iterative process and earlier experiences influenced our perceptions of this final site. The Fossil and Prairie Park is a 400 acre site (162 square hectometers) that became public land in 1990. A new visitor center opened in 2001. The site encompasses a wetlands area and native prairie in addition to the fossil-collecting quarry. The area was originally the quarry site of the Rockford Brick and Tile Company.

The Lime Creek Formation was deposited in a shelf-margin shallow-marine setting and is 385–375 m.y. old (Frasnian). The park interpreted the paleoenvironment of the area to be subtropical deltaic and estuarine during the Late Devonian. At the Fossil and Prairie Park, the Cerro Gordo Member of the Lime Creek is exposed. It is exceptionally fossiliferous and consists of fossiliferous calcareous shales, argillaceous limestones, and bedded argillaceous limestones (Drewes, 2005; Anderson, 1998). Several species of brachiopods, bivalves, gastropods, rugose and tabulate corals, and bryozoans are common, with cephalopods less common (Anderson, 1998). During the first geological survey of Iowa, spirifer brachiopods were collected in the area (Drewes, 2005). Their identifications and illustrations were included in the 1858 geological survey of Iowa (Anderson, 1998).

We obtained initial information from the Rural Iowa Community Web site prior to our arrival. Although the fossil beds were briefly mentioned, only the minimum directions to the site were provided. We accessed the information that the park was open from "sunrise to sunset." Unbeknownst to us, this did not include the visitor center hours, which were not posted. We found within a Fossil and Prairie Park Web site a general map of the collecting area. The Web site (www.fossilcenter.com) featured more information on the Prairie Heritage Days than any other topic.

We drove from Rochester, Minnesota, to access the site. The trip took longer than we anticipated, and we first accessed the site on a warm summer August afternoon. No fees were required to enter the park, and no park personnel were present. This fossil park was the most remote of the three initially investigated in phase 1 of our research.

We proceeded to the quarry, which offered an authentic in situ collecting environment with fossils, similar to the Penn-Dixie Paleontological Park experience. Unlike Penn-Dixie, the site was not primarily level but offered collecting along fairly steep walls of the abandoned quarry (Fig. 4). There was no handicap access available to the quarry's edge. Only one other group was present, a local father and son team.

Although the Web site noted that tools were not required, we found that our rock hammers were appropriate with some of the harder matrices. Fossils were plentiful. Some were weathered from the strata, and others were easily visible within the calcareous shale. We collected multiple crinoid stems, bryozoans, brachiopods (spirifers, *Atrypa*, *Dielasma*, *Cyrtina*), gastropod molds (both trochospiral and planispiral forms), rugose corals, and bivalve mollusks. The limiting factors to collection were the slope of the quarry walls and loose material, but not the number of fossils available.

Figure 4. The fossil-collection opportunities at Rockford Fossil and Prairie Park are abundant. The old quarry site has multiple exposures. However, there is no handicap access. We found the steep quarry walls to be a limiting factor in our fossil-collecting opportunities (6-ft-tall researcher on top of quarry for scale).

We observed the other group collecting fossils. They worked in a systematic manner and moved over part of the quarry slope in a slow, deliberate fashion, stopping every few minutes to pick up an object of interest. They did not consult a fossil manual, nor did they return to the posted information boards at the top of the quarry for identifications. We moved near their collecting area and engaged them in informal, unstructured conversation.

At the top of the quarry, we observed a basic, unlabeled map of the site, a fossil washing station. and posted material for fossil identification. The identifications provided basic descriptions of the types of invertebrates that could be recovered from the site, but genera were not identified. Typical specimens of the invertebrates were included alongside the descriptions.

We walked the entire Fossil Park and Prairie site, investigating the reconstructed sod house and the native prairie environment. Historic beehive kilns were also present. We were disappointed to find that the Fossil and Prairie Center was not open. We walked around the building and glanced into the facility through the windows. We spotted several hands-on exhibits with animals' antlers and skins, and identification posters of what appeared to be native birds and insects. We did not see any fossil exhibits in the center, but we were not able to view all of the interior facilities through the windows.

Because no park personnel were available, we could not retrieve any pamphlets or additional handouts that might have been present in the visitor center. We attempted to follow up with the facility for handouts but did not receive any information. Because we had engaged only one group, we returned the following day in an attempt to observe more visitors. We were unsuccessful, and after spending the second day at the fossil site, we concluded our 2003 fossil park investigation, or phase 1 of our research.

Phase 2: Case Studies of Two Additional U.S. Fossil Parks (2005)

Following the phase 1 one case study investigation of the first three U.S. fossil parks and the subsequent development of an optimized fossil park model through our identified key categories, we expanded our research in 2005 (phase 2) to include two additional U.S. fossil parks that we located: Trammel Fossil Park (Sharonville, Ohio) and Wheeler High School Fossil Beds (Fossil, Oregon). Prior to our on-site research, we investigated each site and accessed information about the fossil parks. In August and September of 2005, we visited both sites. We searched for fossils as part of an authentic fossil park experience, photographed the signage at the site, collected available literature, and interacted with the fossil park visitors and personnel. Our previous research contributed to the site visits of these new parks, and we evaluated them within our optimal fossil park model. We also performed a comparative analysis between these two new fossil parks as geotourist sites within small communities.

Trammel Fossil Park

The first stop for our phase 2 case study of fossil parks was Trammel Fossil Park in Sharonville, Ohio. The fossil park was established through a gift of the R.L. Trammel family to the City of Sharonville, to be developed as an educational site where school children and fossil hunters could explore and collect. This 10 acre site (4 square hectometers) includes a kiosk in the shape of an edrioasteroid, the fossil logo of the park.

Trammel Fossil Park rocks were deposited in the Late Ordovician and are of the Cincinnatian Series. There are four formations exposed at Trammel: Fairview (interbedded limestones and shales, with some ripple marks and annelid burrows), Miamitown (shales), Bellevue (fine-layered limestone), and the lower Corryville (shales). Marine fossils attest to a paleoenvironment in which the state was 20° from the equator, situated in warm, tropical seas. Some of the 445-m.y.-old fossils that are found here include brachiopods, bryozoans, crinoids, gastropods, bivalve mollusks, edrioasteroids, and trilobites.

A Web site provided basic information on the founding of Trammel Fossil Park as well as directions to the site. We also accessed basic geology site information on the University of Northern Kentucky Geology Department's Web site (2004).

We arrived on a warm August day in 2005 using the posted directions on the Web site as our guide. However, there was absolutely no signage leading up to the park. On the day of our visit, it was sunny with no breeze, and temperatures rose to 96 °F (35.6 °C). No park personnel or other collecting groups were present. The facilities were newly developed, which we described as in perfect shape. There were brightly painted benches, kiosks, a wash station, a huge parking lot, bathrooms, and interpretive signs. The kiosk area was shaped like an edrioasteroid (Fig. 5), and each of the four formations was described within the "ambulacrum." Not only were the formations' characteristics listed, but the types of fossils that each formation yielded were included with genera identified. Photographs accompanied the descriptions.

Each formation was identified by a fossil symbol as well as a different color. Whereas the Fairview was represented by a red background and a *Rafinesquina* brachiopod, the Miamitown was a blue edrioasteroid, the Bellevue was a turquoise *Herbertella* brachiopod, and the Corryville was an orange trilobite. This was extremely helpful for visitors as each formation was identified through signage displaying the characteristic color and fossil. Any visitor could quickly locate himself or herself within the formation containing the fossil assemblage that he or she desired. Even the contacts between formations were labeled.

In addition to the formation and fossil information, signage also explained the dedication of the park, the characteristics and sources for limestone and shale, geologic time with the placement of the Cincinnati rocks within the time scale, and the Ordovician Period events. We did not find any paper handouts at the park.

We collected within each formation and quickly retrieved more fossils than we had from any of the first three U.S. fossil parks we visited during phase 1 of our research. Many fossils were small and fairly easy to recover within the shale, although we did use our rock hammers and assorted tools. We found fossil retrieval from the Bellevue Limestone to be more difficult, however. We recovered some varied invertebrate assemblages,

Figure 5. Trammel Fossil Park's kiosk is shaped like an edrioasteroid with each of the four formations of the site described within the "arms."

and collected *Rafinesquina, Damanella*, and *Herbertella* brachiopods, bryozoans, crinoid columnals and holdfasts, and the gastropod *Cyclonema*.

Wheeler High School Fossil Beds

Our second fossil park investigation in phase 2 of our 2005 case study research was Wheeler High School Fossil Beds within the small community of Fossil, Oregon. Although the area behind the high school has been accessible to fossil collectors for many years, a small interpretive center opened in April 2005, and the park began to charge a nominal admission fee (US$3). With the $3 fee, visitors are allowed to find and keep three personal fossils.

The Wheeler High School Fossil Beds feature the 33-m.y.-old Bridge Creek Flora within the John Day Formation. These Oligocene shales record a paleoenvironment indicative of a temperate deciduous forest. Leaves, twigs, and an occasional fish and salamander were preserved in a lake bed within volcanic sediments (Meyer and Manchester, 1997). The Bridge Creek Flora at Wheeler High School Fossil Beds is also present at John Day Fossil Beds National Monument, but collecting fossils at the national monument is strictly prohibited.

We accessed some information on Web sites prior to our arrival at Wheeler High School Fossil Beds. Information about the geology of the area was available also on the John Day Fossil Beds National Monument Web site. The Oregon Paleo Lands Institute Web site provided an online overview of the area, including some photographs of fossil plant material that could be recovered there.

We arrived at Wheeler High School Fossil Beds in the early afternoon of a pleasant September day of the Labor Day holiday. The town of Fossil is small, and we encountered no unusual difficulties in finding the site. As the name implies, the fossil beds are located behind the school, and visitors were required to check in at the interpreter's station and pay the required fee. At our arrival, we observed several families and extended family groups digging for fossils within the weathered shale.

We moved to the hillside and began our own search. Collecting tools were allowed, including rock hammers, shovels, and picks. It was fairly easy to dig into the weathered shales, and we were able to retrieve fossils fairly easily at the collecting site. However, the fossil density was not as great as that at Trammel Fossil Park. We uncovered mostly fossil leaves and sticks of plants and identified our collected specimens as *Metasequoia*, *Sequoia*, *Quercus consimillis*, and *Acer*.

There were multiple family groups on the hillside, but no school or club groups were present. We counted a total of 64 visitors in the field and engaged groups in informal conversations. Most groups were focused upon collecting, although children occasionally investigated other aspects of the area, or ran up the hill. We engaged the other visitors in informal conversation and asked them which fossil they felt was their most interesting find. We identified most of the fossils being recovered by other visitors as Dawn Redwood or *Metasequoia*.

While a few visitors consulted personal fossil identification guides they brought to the site, several visitors converged at the station where Karen Masshoff served as the paleontological mentor. There was also a canvas-covered area with a table by some bleachers that could also serve as an interpreter's station. Additional facilities included a very small gift shop. The only signage at the site was on the board at the site's entry, which explained the basic plant fossils that could be recovered. This board provided good pictures as well as the genus names of the common fossils. A mural, painted on the side of a building, provided a stratigraphic context for the John Day Group (Fig. 6).

We left at the end of the day when the park closed, but returned the following morning to interact with additional visitors and the on-site interpreter, repeating the general routine for data collection.

Phase 3: Case Study of Two Additional Fossil Parks (2006)

In 2006, we uncovered two additional parks that fit into our definition of a fossil park. While neither Aurora Fossil Museum in North Carolina nor Stonerose Interpretive Center in Washington referred to itself as a fossil park, visitors were allowed to search for and keep fossils within the informal education sites. Although both sites were in existence in 2003 during our initial fossil park case study research, neither was initially identified in our exploratory research or by international news agency CNN. Therefore, we planned on-site visits to add data from these two facilities to our fossil park research data sets. Our prior research and analyses influenced our perception of the final two fossil parks. Our optimized model for fossil park design provided guiding parameters, as opposed to the raw emergence of variables that we experienced in phase 1 of our first 2003 fossil park investigation. We also used the National Research Council study

Figure 6. This mural depicts the stratigraphy of the area through the John Day Group and the underlying Clarno Group. Fossil plants at the Wheeler High School Fossil Beds are from the John Day Group.

America's Lab Report (Singer et al., 2005) in our analysis of the previous fossil parks in phase 2, in 2005. This report's content influenced our focus upon the new fossil parks' educational potential as outdoor laboratories.

Aurora Fossil Museum

Aurora Fossil Museum was established in 1976 as a nonprofit informal education venue highlighting the geology and paleontology of North Carolina. The town of Aurora, North Carolina, partnered with phosphate mines, government institutions, including East Carolina University, fossil clubs, and interested individuals to open the museum in 1978 as part of a program to bring tourism to the area.

Potash Corporation of Saskatchewan (PCS) currently operates a phosphate mine in the vicinity. During the retrieval of phosphate nodules, PCS encounters some intercalated fossiliferous layers and provides this material for the spoil pile at Aurora Fossil Museum. Called the Pit of the Pungo, the spoil pile is located outside the museum building across the street. Visitors can search through the spoil pile for shark's teeth and other fossils from sunup to sundown. There is no charge to access the site.

The Miocene-aged Pungo River Formation is composed of interbedded phosphatic sands, limestones, dolostones, and diatomaceous clays (Gilmore, 2006). During the Miocene, this area was part of the Abermarle Embayment and is interpreted as an outer continental shelf environment. In addition to 15-m.y.-old shark teeth, other fossils include early whale and teleost fish remains. Invertebrates are represented by barnacles, corals, echinoids, bivalve mollusks, and gastropods.

We accessed information online prior to our visit in June 2006. There were several Web sites with material posted about the Pungo River Formation, and a Web site devoted to the Aurora Fossil Museum. The map proved accurate, and we encountered no difficulty in finding the site.

We arrived at Aurora Fossil Museum on a pleasant but overcast day. The museum was open and housed local specimens. In addition to the Miocene material of the Pungo River Formation, Pleistocene fossils and archaeological artifacts were on display. However, since our primary research focus was the fossil spoil pile, we moved our attention outside.

There were several family and extended family groups collecting in the spoil pile on the day of our visit (Fig. 7). While we were collecting within the spoil pile, we counted 23 other individuals at the site. The material, similar to the spoil piles at the Fossil Park at Sylvania, was without stratigraphic context. There was no fossil identification signage at the site and only one sign posted at one end of the collecting area with the "Park Rules." We sorted through the material and collected several species of sharks' teeth, corals, and broken bivalve shells.

We observed the other collectors and engaged them in informal unstructured conversation. The visitors' fossil searches were focused upon finding sharks' teeth, particularly teeth from the extinct giant *Carcharocles megalodon*. On the day of our visit, none of the collectors was successful at finding a *C. megalodon* tooth, but groups did retrieve several small sharks' teeth. No one focused upon the invertebrate fauna.

Stonerose Interpretive Center

The Stonerose Center and Interpretive site was established in Republic, Washington, in 1989. The Boot Hill Fossil site was discovered in 1977 by Kirk Johnson, paleontologist, and Wesley Wehr, artist. Wehr collaborated with Republic City Councilman Bert Chadick to purchase a house within walking distance of the fossil site. This was transformed into the Stonerose Interpretive Center. Both sites—the interpretive center and the fossil-collecting hill—are owned by the nonprofit Friends of Stonerose Fossils.

Boot Hill Fossil Site preserves Eocene tuffaceous shales within the Klondike Mountain Formation. Fifty million years ago, the Republic, Washington, area was the site of a lake in a temperate highland region. Volcanic sediments trapped and entombed a variety of plant material, as well as insects, fish, and even bird feathers. The site's logo is *Florissantia quilchenensis*, a flower that resembles a rose but is an extinct relative of cocoa plants. However, the Stonerose site is famous for the earliest rose and maple fossils (Wehr, 2008). We accessed information prior to our visit, including a feature on the Stonerose fossil logo within *National Geographic Magazine* (Klesius, 2002).

We arrived at Stonerose Interpretive Center on a sunny day in August 2006. We did not have great difficulty in finding the site. With our admission (US$3), we had access to the Boot Hill Fossil site. Visitors are allowed to search for and retrieve up to three fossils each day. Visitors may also rent tools or bring their own to the site. Chisels are allowed, but hammers are only allowed to hit

Figure 7. Visitors at Aurora Fossil Museum may collect fossils in the outdoor collecting area across the street. The spoil piles come from the local phosphate mine. Fossils recovered are without stratigraphic context.

the chisel handle and not for splitting rocks. All fossil finds must be shown to the Stonerose staff.

On display at the Stonerose Interpretive Center, there are three glass cases with “conifer,” “leaves,” and “unique finds” fossils. We walked to Boot Hill Fossil site to experience the collecting opportunities ourselves (Fig. 8). Boot Hill Fossil site is accessible during normal operating hours, but the site is closed one hour earlier than the Stonerose Interpretive Center.

Collecting at the site was in situ and fairly easy. However, similar to the Wheeler High School Fossil beds, fossils here were not as prolific, and the visitor success rate for fossil recovery was not as high as those fossil park sites with marine strata. We collected several bits of Dawn redwood twigs and some unidentified plant stem material.

At the site, we observed two other groups collecting. One group was a local mother and daughter team, while the other was a group of local young adults. We engaged the groups in informal, unstructured conversation. There was only one sign available at the Boot Hill Fossil site, and this provided basic information about the site’s ownership and collecting policies. No information was posted for easy identification of fossils.

RESULTS AND DISCUSSION

For content analysis, we utilized the methodology outlined by Neuendorf (2002), while for conversational, unstructured interview analysis, we employed Chi’s (1997) guidelines. Through all three phases of case study investigations, we explored the educational opportunities available at fossil parks, and uncovered via grounded theory and data comparisons the most effective methods for meaningful informal learning at fossil park sites.

Figure 8. At the Boot Hill Fossil site, visitors may search through the tuffaceous shales for Eocene fossils. The most common fossils are plant stems and leaves. Only the collecting rules are posted at the site. No identification charts or geological information signage were posted.

Phase 1: The First Three Investigated U.S. Fossil Parks

Conversations with Penn-Dixie volunteers revealed that most of their visitor attendance is a local population, although they remembered visitors from Canada, Japan, and Lebanon. The primary visitor population results from school groups. The teacher volunteer reported that the New York high school curriculum required a mandatory earth science course in grade 9. Volunteers estimated 30–40 individuals typically visit the site for fossil collection on a daily basis, while the Saturday night astronomy programs can result in attendance of over 600 individuals. The site has hosted a “science camp” as well. There were plans for ongoing development, including interdisciplinary science (climatology, ornithology, astronomy) and additional signage, which was described as a “painstaking process.” The site relies heavily on its volunteers, which at the time of our site visit were estimated to number between 175 and 200.

Experiences at the Penn-Dixie site could vary, based on our volunteers’ responses. When large school groups visit the area, most of the larger specimens are removed, resulting in a diminished collecting experience for later groups. For our site visit, there was a large acreage exposed for the productivity we encountered. Exposure of the site through earth-moving processes was not conducted on a regular, scheduled basis. A volunteer remarked that the “kids get bored easily” without positive feedback from regular fossil recovery. The same volunteer noted that the smaller kids “collected fossils, played in the mud, and then wanted to go home.”

In Ohio, the Fossil Park at Sylvania also noted plans for expansion, including interpretive signage and running water at the site. The Olander Park System hopes that a Fossil Interpretation Center will bring national and international recognition to the geotourist site. Until 2004, the park reported an annual visitor count of ~20,000, in addition to 2000 annual students and scout groups (Jones, 2004). The on-site naturalist mentioned that a visitor count averaged 250–500 per day. On the day of our visit, we determined that most of the visitors were from local family units based on our informal conversations. We did not encounter the family from Nevada at the fossil park site though they signed the first visitor entry for the day.

The naturalist claimed that 90% of visitors were from out of state. This is a very different visitor composition from Penn-Dixie Paleontological Park. The out-of-state population may be partially attributed to CNN’s international news article that featured the park earlier that summer. The naturalist at the site expressed disappointment with the CNN news article’s wording, noting that “bucketfuls” of trilobites were an exaggeration. He further stated that 1 out of 100 visitors will find a complete trilobite.

When we visited the site, it had been approximately 3 wk since the last load of materials was deposited. This did not make for authentic or rewarding collecting. We observed that visitors had a very low success rate, and there were very few fossil treasures in the spoil piles. The spoil piles consisted primarily of slippery mounds of weathered shale. Additionally, the site was crowded.

Collecting at the Fossil Park at Sylvania was not done in situ and was without stratigraphic context. Authentic collecting tools were not allowed. Although we noticed brushes in a bucket at the naturalist's station, they were never offered to the visitors. However, the handout was correct in acknowledging that the shale was easy to break apart without tools.

We observed children playing in the mud as often as we spotted them searching for fossils. Some adults appeared to be struggling to find a "fossil." The kiosk and handouts did not describe the different species, and no explanations for evolution, geologic time, or biodiversity were offered in any format. Collecting within the weathered spoil piles can best be described as a "trial-and-error" process.

In the most remote fossil park, located in Iowa, attendance records for 2002 to 2003 documented visitors from 35 states, 6 countries, and 160 cities within the state. School groups from Iowa (21) and Minnesota (2) visited the park, contributing to the recorded annual visitation of 6400 people (Fossil and Prairie Center Foundation, 2003). However, it is highly probable that more visitors experienced the site without registering, as we did. There are special events at the fossil park, including the Rockford Fossil Days, an annual event held during the Labor Day weekend.

We could not retrieve any information from volunteers or park personnel at Rockford Fossil and Prairie Park, because no one was present when we visited. However, we can attest to the authentic collecting experience at the park. The procurement of fossils at Rockford Fossil and Prairie Park was the most realistic experience and most like a paleontological field excursion of the three U.S. fossil parks. With the quarry exposures, however, came a higher field risk. There were no fences or warning signs keeping visitors off the slopes, and even though we have experience and wore proper field boots, we found the site more difficult to navigate than other public informal sites we investigated. While the density of fossils and the large collecting site were positive attributes, we discussed the safety issues and wondered how safe a fossil experience would be for families with younger children.

There were no brochures available for visitors at the quarry site, although the identification boards were helpful in offering descriptions as well as specimen examples. However, with the size of the quarry, it would be unrealistic for a fossil hunter to trek back and forth to the signage to identify the retrieved specimens.

Content Analysis Key Findings

We analyzed each fossil park site for the geobiological opportunities to learn. Through content analysis of the coded field notes, on-site signage and brochures, and unstructured conversations of the first three U.S. fossil parks, three key findings emerged: (1) Fossils presented and recovered without stratigraphic context can hinder paleontological understanding; (2) on-site geological mentors can facilitate an authentic geological collecting experience; and (3) a fragile balance exists between safety considerations and adventure during field experiences. We diagram some of the characteristics of the first three U.S. fossil parks that contributed to these key findings in Figure 9.

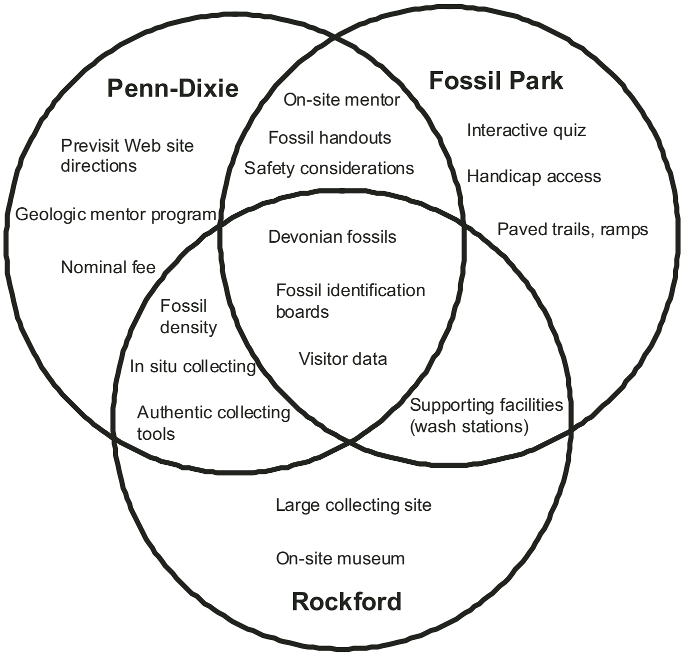


Figure 9. The shared and unique characteristics of the first three U.S. fossil parks we researched in phase 1 of our research study (2003). Represented are Penn-Dixie Paleontological Park (Hamburg, New York), the Fossil Park (Sylvania, Ohio), and Rockford Fossil and Prairie Park (Rockford, Iowa).

Although the paper shales at the Fossil Park in Sylvania facilitated easy collection of fossils, the spoil piles at the bottom of the quarry were without geological context. The concepts of geologic time, biodiversity, extinction, evolution, and paleoenvironments were not visible to the average visitor. It was easier for a park to convey stratigraphic context in situ. When visitors can retrieve fossils within vertical exposures, the principle of superposition is more apparent. Furthermore, the retrieval of fossils within the strata makes apparent the original deposition of the organism. One man within a family unit at the Fossil Park at Sylvania noted that the fossils were laid down during a "great flood," and the jumbled nature of the spoil pile did not easily counteract that misconception.

The paleontological mentor was an important resource at both Penn-Dixie Paleontological Park and the Fossil Park at Sylvania. At Sylvania, the on-site naturalist resided in a station within relatively close access to the collecting piles. This was a feasible arrangement because the site was relatively small, and children were often at the station with their latest find for identification. At Penn-Dixie, the mentor was assigned to our collecting group, and after leaving us, traversed the field site to see whether other collecting visitors had questions. A central site at Penn-Dixie, stationed with a mentor, might not have worked as well as it did at Sylvania because of the size of the Penn-Dixie site. At both locations, however, the mentor was a primary source of fossil information for collectors. Although identification boards and handouts were available, our notes detailed how visitors approached the mentor for fossil identification and information much more often than they referenced other sources of fossil information.

Although we initially questioned the number of rules and the absence of tools at the Fossil Park in Sylvania, we experienced the other extreme in conditions with Rockford Fossil and Prairie Park. We are familiar with general field protocol, but we found ourselves slipping within the quarry on several occasions. The quarry site at Rockford provided a grand adventure for the collector, but safety considerations for young children or novices were not apparent.

Fossil Park Design Model

Following the first case study of the original three U.S. fossil parks in phase 1, we compared data from the three site visits. Broad groups of similar concepts emerged. These formed the basis of six categories: informative previsit Web site, authentic in situ collection, authentic collection tools, accessibility, fossil identification, and visitor education. These categories, along with their variables, were analyzed. Our optimal fossil park design model emerged through grounded theory. This design model incorporated superior features of each category, as determined through the researchers' participation in the fossil park collection experiences and observation of participants at the fossil parks.

Informative previsit Web site. In addition to providing a basic overview of a fossil park facility, a Web site should have accurate directions for the visitor to quickly access the facility without the need to ask further directions from the local residents. The Web site can also serve as a form of wide-range advertising. In order to attract geotourists to a fossil park, the general public must be aware it exists.

Authentic in situ collection. A fossil park site should be large enough to accommodate the projected number of visitors without crowding of participants. Sites should be conducive to fossil recovery with a visitor success rate at least moderate in scope. This applies to both the number of fossils specimens to be collected as well as variety. If earth-moving machinery is required to expose new strata, the facility should ensure that this is done on a periodic basis to maximize visitor collecting experiences. Collecting in authentic sites, whether done within abandoned quarries, natural exposures, or road cuts, is preferable to spoil piles. Vertical relief provides a stratigraphic context and contributes to the opportunities to learn at the park.

Authentic collection tools. Authentic collecting tools enriched the experience for the visitor. However, if rock hammers, chisels, and screwdrivers are allowed, or even supplied at the site, the fossil park should also provide guidance and instruction on safe fossil-collecting techniques. Recommended personal safety equipment may include goggles, and visitors should be encouraged to wear sturdy, closed-toe shoes. Supporting facilities also enhance a fossil-collecting experience. Wash stations for cleaning recovered specimens and tables to sort the collected items assist the visitor in identifying his or her retrieved specimens.

Accessibility. A public fossil park should be accessible to all visitors. Therefore, handicapped access, including a paved trail and ramps, are required at the site. Daily hours should be maintained by the park, for both weekday and weekend fossil excursions. Although a fee may be charged for entrance, it should be nominal and reasonable. Detailed, accurate directions to the site should be readily available.

Fossil identification. Instead of brochures for fossil identification, we recommend semipermanent signage that can be accessed, hands-free. Because signage should be available at the collection sites, portable signs provide a good option. A high density of signs at the eighth-grade reading level is recommended (Wandersee and Clary, 2007).

Visitor education. For meaningful learning experiences, visitors should be exposed to learning opportunities with an integrative geology and biology field. The signage and site interpretation should emphasize evolution, biodiversity, and geologic time. Interactive quizzes provide good opportunities for self-testing. A paleontological mentor should be available for assisting in fossil identification as well as scientific understanding of the site. A training program for these individuals would ensure the continuation of the mentor program. Visitor centers and museums within a fossil park site can provide additional learning opportunities of geobiology through free-choice, self-paced learning. Exceptional specimens of fossils recovered from the site can be displayed.

Public safety considerations are important, and visitors should be apprised of safety risks and protective measures. If a site possesses different slopes and/or levels of difficulty, coded levels of collecting access can quickly communicate associated complexities through signage. In order for a fossil park to improve its presentation and learning experiences for the visitor, it should keep records on visitors' origin, usage, and fossil success rate. Additionally, methods should be in place for feedback from visitors.

Phase 2: Comparisons and Geobiological Opportunities of Two Additional Fossil Parks (2005)

The hot August day we visited Trammel Fossil Park in Sharonville, Ohio, may have kept local families away. However, schools, families, and geology clubs appear to make regular trips to the site. The local fourth-grade classes also travel regularly to the site. Trammel Fossil Park reported school visitors from as far away as New York. Without park personnel, volunteers, or guest books, accurate visitor counts are problematic.

Trammel Fossil Park was not staffed and is continuously open. However, we observed zero vandalism at the site. Although no park personnel were present, workers from a small factory alongside the park took their work breaks on picnic benches and seemed to keep an eye on the park. The park is handicap accessible, and there is adequate parking.

Information on geologic history of the site, biodiversity, and fossil identification was available at the kiosk. We noted that the fossil density and diversity were good when compared with other fossil sites we visited; this led to a good fossil-collecting experience. Extraction of the fossils from limestone was difficult, and even the shale extraction resulted in bits of rock flying from our rock hammer. Some visitors would probably benefit from personal protective equipment.

Conversely, Wheeler High School Fossil Beds (Fossil, Oregon) had numerous visitors, but this may be as a result of our site visit over the Labor Day weekend. The majority of the population appeared local, based on our informal conversations. However, the facility has the potential to attract other visitors, especially those who are interested in John Day National Monument and who would like to procure similar fossil plants for themselves. We could not access any data on visitor populations because the park had only been open for 4 mo.

Because Wheeler High School Fossil Beds charges an admission and relies upon an on-site interpreter for geobiological education and fossil interpretation, the site is open from 9 a.m. to 5 p.m., Tuesdays through Sundays. This fossil park is in its early stages of development, and the interpreter's station and guest shop were quickly constructed. Handicap access was not fully developed at the site.

Basic information on geologic formation was available on the entrance board, but the site relied primarily on the interpreter for fossil identification. Although our collecting experiences did not result in the same number of fossils that we recovered at Trammel, this is not inconsistent with plant fossil collection and the nature of the organisms' paleodeposition. We did not have trouble collecting fossils, and most of the groups we observed at the site were engaged in collecting processes.

Key Comparisons

Both Trammel Fossil Park and Wheeler High School Fossil Beds are located in small towns: Sharonville, Ohio, had a population of 13,000, while the population of Fossil, Oregon, was 430. Each town developed a collaborative, partnership-driven, pedagogically innovative field-based geotourism venue. While Trammel Fossil Park's educational partnership included R.L. Trammel, the University of Cincinnati, and the City of Sharonville, Wheeler High School Fossil Beds' collaboration involved town, county, state, federal, and foundation-based economic development and a field-based paleontology education program. Following the collapse of the timber industry, the economic recovery plan for Fossil, Oregon, is predicated on geotourism and geoscience education activities centered upon its fossil park as part of a rural renewal. In 2005, the Wheeler High School Fossil Beds only received 2% of Oregon's tourism dollars. In Sharonville, Ohio, Trammel Fossil Park has become an enticing geoscience education leader of the town's park system.

Both parks had future plans for expansion. While Trammel Fossil Park's plans included a future college dormitory on-site, Wheeler High School Fossil Beds anticipated a future Paleo Learning Center and on-site short courses.

The parks differed in several variables. While Trammel Fossil Park's collecting opportunities were from Ordovician-age marine strata, the Wheeler High School Fossil Beds provided collecting in an Oligocene-age lake bed and featured the state fossil of Oregon (*Metasequoia*) and the Bridge Creek Flora. The Trammel Fossil Park's site was small, with room for ~150 collectors at a time. Conversely, the Oregon Paleo Lands Institute addressed a 10,000 square mile area (26,000 km^2) radiating from Fossil, Oregon. The Wheeler High School Fossil Beds site could accommodate ~200 collectors comfortably.

The two parks were much in contrast by the way each provided fossil identifications and on-site assistance. Trammel Fossil Park utilized an effective, centralized interpretive signage system that aims to be self-teaching. Conversely, Wheeler High School Fossil Beds employed a helpful on-site geological interpreter.

Fossil Parks as Outdoor Geobiology Laboratories

Both Trammel Fossil Park and Wheeler High School Fossil Beds were noteworthy in specific ways for their geoscience education potential as outdoor teaching laboratories. The Trammel Fossil Park excelled in its focus of geological time and the fossil identification via the interpretive signage system designed by a university geology department. Wheeler High School Fossil Beds is slated to be the focus of a large paleontology education effort and will integrate on-site short courses as part of its paleontology education activities. While Trammel Fossil Park was guided by the *City of Progress* model, Wheeler High School Fossil Beds was guided by a 50-page *Master Plan for Learning* designed by the Oregon Paleo Lands Institute (2005).

We linked the findings of the two new fossil parks case study investigations, along with our results from our original three U.S. fossil park case study in phase 1, to the National Research Council (NRC) study, *America's Lab Report: Investigations in High School Science* (Singer et al., 2005). Specifically, we correlated the missions of the fossil parks as informal geoscience education venues that provided an outdoor geobiology laboratory setting. The NRC study did not exclude outdoor experiences from their definition of laboratory, but instead noted that their definition included student research and observation in outdoor settings including "nearby geological formations" (Singer et al., 2005, p. 35). The NRC study concurred that field work is an effective way to provide alternative learning experiences to laboratory research. We found it disappointing that ~50% of teachers reported that they *never* take their classes on field trips (Singer et al., 2005). Singer et al. (2005) concluded that teachers should conduct open-ended field research in order to better understand inquiry. Data can originate from either the laboratory or the field.

Geologists often rally around the slogan that "geology is best taught in the field." Notably, the National Research Council concurred. Their report affirmed that there is no equivalent technological substitute for direct interaction with the real world. We propose that fossil parks can serve as authentic field sites and outdoor laboratories to address fossilization processes, biodiversity, evolution, geologic time, biostratigraphic correlation, and environmental change over time. Furthermore, fossil parks can offer scientific experiences to the public that contribute to the nation's scientific literacy. Through public understanding of past environmental changes, including numerous transgressions and regressions of the sea, glaciations, and past warming events, modern issues such as ozone depletion and climate change are provided an Earth context. Fossil parks can successfully convey Earth's past environmental changes through effective geobiology education at their sites.

Phase 3: Case Study of Additional Fossil Parks (2006)

We discovered at Aurora Fossil Museum that one reason we did not identify the site sooner was because of its organization: Aurora Fossil Museum is an *indoor* informal education site that has a complementary spoil pile for visitors to search for fossils they may keep. Unlike the previous fossil parks we identified, the museum was not built to support a fossil park, but the fossil "park" (spoil pile) was added as a hands-on exhibit to the museum.

The collecting opportunities at the spoil pile were without context. Family groups who were collecting the day of our visit were not focused upon identification of fossils, nor did they perceive the geological context of the fossils' deposition. Younger visitors were intent on finding the "big shark's teeth" (*Carcharocles megalodon*). Although other fauna were present in the spoil pile, none of the fossils were identified at the site. Visitors could identify their fossils by comparing them to the displayed specimens in the museum, but we did not observe any visitors doing this.

Similar to the Fossil Park at Sylvania, the collecting experience of a group can vary greatly depending upon when the latest material was added to the spoil pile. Although the site was not completely picked over, we did not observe anyone collecting large fossil items.

Stonerose Interpretive Center may have also escaped our early identification as a fossil park because of the separation of the interpretive center and the fossil-collecting site: Stonerose Interpretive Center is located away from Boot Hill Fossil site, although it is within walking distance and owned by the same nonprofit group. Unlike Aurora Fossil Museum, collection at the Eocene outcrop is in situ. However, quick identification of fossil finds at the collecting site is not facilitated. The only signage posted rules for fossil recovery, but no helpful pictures and descriptions of the common fossils retrieved at the site were available. Visitors must return to the Stonerose Interpretive center to show their fossil finds to a staff member before claiming them, and at the center they may seek identification assistance from the staff, or use the posted specimens for comparison. However, we think that this delay works against the educational potential of the fossil-collecting experience.

Resultant Rankings on Key Variables (2007)

After our final on-site fossil park case study at Stonerose Interpretive Center, we returned to our field data of lived learning experiences for all seven identified U.S. fossil parks. Our data sources included field notes, photographs, visitors' comments, and park documents. Our focus was not upon an optimized fossil park model as it had been in 2003, but on our interpretation of how each of the seven U.S. fossil parks ranked as informal learning facilities of geological and biological concepts.

We searched our data for meaning and understanding as to how a fossil park could best offer informal geoscience education experiences to a visitor. Our data were filtered through the learning theory of human constructivism, particularly the principles of meaningful, mindful, and active learning. Through four years of fossil park investigations and data sources, seven variable categories emerged through grounded theory. Our focus was upon those variables that we interpreted through our authentic collecting experiences that contributed to the fossil park visitor's overall learning experience of biodiversity, geologic time, evolution, and environmental change. These variable categories included (1) the authenticity of the outdoor experience with genuine collecting tools and in situ fossil recovery; (2) the geological age of the strata and ease of fossil retrieval; (3) fossil-collection training and facilities for the novice; (4) the availability of an on-site paleontological mentor; (5) the availability of signage and brochures to aid in the visitor's fossil identification; (6) the fossil park's organization and wayfinding signage; and (7) accessibility, including fossil park awareness through publicity and posted directions to the site, as well as visitor safety considerations. We ranked the seven U.S. fossil parks on a scale from 1 (best) to 7 (worst) in each category.

Whereas we were on complete agreement with our independent rankings for the fossil park sites for the quality of the mentor (4), fossil identification (5), site organization (6), and safety and accessibility (7), we encountered differences in rankings for authenticity (1), age (2), and training (3). We resolved these issues by thoroughly defining "authenticity," and removing safety considerations from the category. We also compromised on geologic age rankings by scoring sites with similar-aged fossils identically. However, we agreed on the exception of the Fossil Park at Sylvania, which had excellent preservation of fossils that were easily extracted from the shale matrix. Therefore, we categorized Sylvania's fossils as equivalent in desirability with the earlier-aged fossils at Trammel Fossil Park (Sharonville, Ohio). Finally, we resolved our researcher disparities by defining "training" (3) to include the total training experience of the visitor through brochures, signage, and a fossil mentor. Even before compromise, we were in agreement on the top two U.S. fossil park sites as well as the two lowest ranked fossil park sites (Table 2).

Authenticity of site and collection tools. When we considered the seven U.S. fossil parks on their authenticity of field experience, the top fossil parks were Trammel Fossil Park in Ohio, and Penn-Dixie Paleontological Park in New York. Both sites offered collecting in situ with more than one formation at the park site. The Trammel Fossil Park site labeled the different formations through signage that was color-coded, for superior, directed collecting opportunities. It also had good outcrop exposure. Both fossil parks allowed visitors to utilize authentic collecting tools.

Rockford Fossil and Prairie Park offered stratigraphic context and collecting in situ, but the large exposed area without

TABLE 2. FOSSIL PARK RANKINGS

	Hamburg, New York	Sharonville, Ohio	Fossil, Oregon	Sylvania, Ohio	Republic, Washington	Aurora, North Carolina	Rockford, Iowa
Authenticity	2	1	5	6	4	6	3
Age	3	2	6	2	6	7	3
Training	1	2	3	5	4	6	6
Mentor	1	6	2	4	3	5	7
Fossil ID	2	1	3	5	4	6	7
Site organization	2	1	5	3	6	4	7
Accessibility	1	4	3	5	6	2	7
Overall score	1.6	2.3	3.9	4.3	4.6	5.1	5.6
Rank	1st	2nd	3rd	4th	5th	6th	7th

Notes: Parks were ranked from 1 through 7 in the following categories, with 1 being the best rating a fossil park could achieve, and 7 being the rating in most need of improvement: authenticity of the outdoor experience; age of fossils at the site; fossil-collection training and facilities; availability of on-site paleontological mentor; signage and brochures to aid in fossil identification; site organization and wayfinding signs; and accessibility through publicity, directions, and safety features.

signage and safety considerations detracted from the experience. The lowest-scoring U.S. fossil parks for authenticity were the Fossil Park at Sylvania and Aurora Fossil Museum, both of which utilized spoil piles instead of in situ collecting.

Age of strata and ease of fossil retrieval. We observed that visitors enjoy collecting the specimens that are the oldest and the most distant from present-day organisms. Our notes detailed greater excitement and exuberance from visitors when they found fossil specimens (such as trilobites) that are markedly different from modern life forms. Visitor responses were more subdued when they recovered specimens that more closely resembled modern organisms (such as leaves and smaller sharks' teeth). Therefore, we identified the age of fossils that could be collected at each site and ranked them accordingly. Trammel Fossil Park, Penn-Dixie Paleontological Park, the Fossil Park at Sylvania, and Rockford Fossil and Prairie Park are situated on Paleozoic sites. All of these fossil parks showcase Devonian fossils except for Trammel, which features Ordovician fossils; the younger Cenozoic sites are Aurora Fossil Museum (Miocene Epoch), Wheeler High School Fossil Beds (Oligocene Epoch), and Stonerose Interpretive Center (Eocene Epoch).

Fossil-collection training and facilities. The complete program offered by Penn-Dixie Paleontological Park earned it the best marks for fossil park training through signage and on-site assistance. Although Trammel Fossil Park did not have an on-site paleontological mentor, the *content* of the on-site signage provided good instructions for fossil retrieval for the novice visitor and specifically noted where various fossil types would be located on the site. Additionally, Trammel provided a wash station. The lowest ranked U.S. fossil parks in fossil-collection training and facilities were the Rockford Fossil and Prairie Park and the Aurora Fossil Museum. The Rockford site had a wash station, but it was inadequate given the area of the quarry. Additionally, we encountered no signage on how to collect at the site. The Fossil and Prairie Center may provide some assistance and training, but it was removed from the collecting area and was not open when we visited. Aurora Fossil Museum received a low score in this category as well. The spoil pile was a distance from the museum and visitors simply searched for fossils without instruction.

Availability of on-site paleontological mentor. Volunteers or park personnel who are knowledgeable with a fossil site can make a large impact on visitor success and satisfaction. The Penn-Dixie Paleontological Park provided a college student as a paleontological mentor and also had volunteers manning the entrance station. The Wheeler High School Fossil Beds had a knowledgeable mentor at the booth. At Stonerose Interpretive Center, personnel were available for assistance, although no one was at the Boot Hill Fossil site. The lowest ranked fossil parks in this category were Trammel Fossil Park and Rockford Fossil and Prairie Park, which had no personnel or volunteers at the site to offer visitor assistance.

Availability of signage and brochures for fossil identification. Signage and brochures are especially important if a paleontological mentor is not available. The best signage system we encountered at the U.S. fossil parks was undoubtedly at Trammel Fossil Park. Not only was it scientifically correct, but it was placed within easy visitor access. Coded signage also marked the various formations at the site. Penn-Dixie Paleontological Park scored second on signage. The signage was removed from the collecting areas of several groups, but it was scientifically correct and informative. Penn-Dixie also provided a fossil identification card to each visitor group. Ranking the lowest in signage and fossil identification were Aurora Fossil Museum and Rockford Fossil and Prairie Park. Aurora's signage and information were in the museum and not at the spoil pile, while Rockford Fossil and Prairie Park had too little signage for such a large quarry area.

Fossil park organization and wayfinding signage. The organization of the site was extremely important for enjoyable fossil-collecting experiences and informal science education opportunities. When sites were large, collecting areas were at a distance from park personnel or signage. Additionally, some sites did not provide signs within the outcrop, and we anticipated that some small visitors may easily become distracted and disoriented. For this reason, Rockford Fossil and Prairie Park scored worst in this category. The distance from the Stonerose Interpretive Center to the collecting site at Boot Hill Fossil site also earned low marks. The best site organization was displayed by Trammel, which had a well-organized site that could accommodate collectors without crowding, but which maintained a reasonable distance from the centralized signage to the outcrops. Likewise, Penn-Dixie Paleontological Park was well-organized.

Accessibility: Park awareness and visitor safety considerations. Our final variable category considered the accessibility of the fossil park. Visitors should know about the park's existence, find the fossil park easily, and interact with the site in a safe fashion. Penn-Dixie Paleontological Park scored highest in this category, followed by the Aurora Fossil Museum site. Our on-site experience with Rockford Fossil and Prairie Park resulted in our questioning the safety of the site, especially with younger visitors or guests who may underestimate the difficulty of accessing fossils on the quarry slopes.

Overall fossil park rankings. By taking an average of the fossil park's rating in each of the seven categories, we determined a numerical value for each of the 7 U.S. fossil parks. In first place was Penn-Dixie Paleontological Park, which we determined provided the best visitor experience and the greatest opportunities to learn geobiological concepts in an informal fossil park environment. In second place was Trammel Fossil Park. Although this site did not have any fossil park personnel present or a paleontological mentor to aid in fossil identification, the superior signage system, authentic collecting experiences, and site organization contributed to a fossil park site that facilitated geobiology education for the visitor.

Identification of Geobiological Ideas That Personal Fossils Can Teach

Throughout our case study research investigations, we sought to determine the opportunities to learn geology and biology within U.S. fossil parks. At the conclusion of our final case study investigation at the Aurora Fossil Museum and Stonerose Interpretive Center sites, we determined some of the important geobiological concepts that personal fossils, collected at the U.S. fossil parks, can teach. The outdoor collecting experience and the ownership of fossils can serve as the platform from which to address the geologic age of Earth, evolution, how environments change over time, and biostratigraphic correlation. Because the fossils recovered at any of the U.S. fossil park sites are millions of years old, the fossils' ages facilitate the visitor's comprehension of the enormity of geologic time and the way in which this relates to our own historical time frame. In addition, the fossils collected were often from extinct organisms, such as trilobites, rugose corals, and some extinct species of plants and sharks. Fossil ownership of these extinct organisms provides a springboard to origin, extinction, and evolution of life forms over Earth's history.

Fossils provide information for reconstructing past environments and are the only direct evidence of past life on Earth. Marine fossils can be retrieved from five of the U.S. fossil parks. These collecting sites are terrestrial today, and can serve to initiate a discussion on how environments change over time, the relative amounts of time required for these environmental changes in the fossil record, and how the human-induced climate changes of the present fit into an Earth framework. Biostratigraphy, utilized for correlating rock units of the same age over wide areas, was first used in the early 1800s. Fossil succession is an important tool utilized for relative age dating and can be more specific than radiometric techniques. Additionally, the other principles utilized in relative age dating—including superposition and lateral continuity—can be introduced in the field in many fossil parks.

IMPLICATIONS FOR GEOSCIENCE EDUCATION

Our longitudinal case study research of U.S. fossil parks demonstrated that fossil parks can serve as informal education sites for meaningful science learning experiences. However, through constant comparison, our research revealed that the U.S. fossil parks differ in their authenticity, collecting experiences, and visitor education programs. Our optimized model for fossil park design provides guidelines for others contemplating the establishment or instructional use of fossil parks. Similarly, our ranking of the seven U.S. fossil parks researched in our longitudinal investigation revealed individual parks' strengths and weaknesses. Our assessment of the opportunities to learn geobiology at fossil parks can serve as a guideline for teachers or others planning a trip to one of more of these sites. Using the identified weaknesses, future visitors can better prepare their excursions to optimize the opportunities to learn within this unique informal venue. Furthermore, previous research has indicated that directed field excursions can be more conducive to learning because some students can have difficulties focusing upon individual exhibits or activities within a large informal site (Clary and Wandersee, 2009). The fossil park rankings on the seven key variables should be helpful for planning a directed informal science field experience.

We utilized lived learning experiences, observation of visitors, and casual informal conversations with visitors, volunteers, and park staff during our case study investigations. However, we were unable to interact with visitors at each site, and many of our conversations were brief. Therefore, future qualitative investigations should sample learner impact of fossil park signage, brochures, collecting methods, and on-site mentors. More intensive ethnographic studies at each park site will generate additional data through more in-depth interviews and multiple observations of park participants, volunteers, and personnel over longer periods of time and through seasonal variations. While our thick descriptions and lived learning experiences testify to the authenticity of our research, our optimized fossil park design model and the key variables we designated for ranking fossil parks may not be similarly weighted by other researchers, park personnel, or visitors. Additional qualitative investigations can determine the reception of our optimized fossil park model, and whether different groups of people value different key variables for fossil park learning opportunities. Furthermore, future research investigations may determine that key variables at one fossil park site may not be transferable to all U.S. fossil park sites.

Therefore, our research investigation into the quality of informal education offered by U.S. fossil parks is ongoing. More research is needed to fully elucidate the effects of the fossil parks' signage systems, on-site mentors, and fossil-collecting opportunities on informal geoscience education, and the ways in which informal learning experiences at fossil park sites can be optimized for the visitor.

CONCLUSIONS

For all seven on-site case study investigations of U.S. fossil parks, our focus was to identify the opportunities to learn and the features available to learn geobiology via active learning experiences for both students and families, and to identify those that subsequently led to a sound informal geoscience education program. The original data in phase 1 of our research investigations were collected at the first identified three U.S. fossil park sites through lived learning experiences. Through grounded theory, our research progressed toward the development of an optimal fossil park design, a tool that can be used in the development or improvement of a fossil park and the optimization of instruction within these outdoor learning venues.

Following the model development, we identified and investigated four other informal education sites (phases 2 and 3) that allowed the collection and retention of personal fossils, and that considered their primary mission to be advancing geoscience education. Comparative investigation of these additional fossil parks through the lens of the optimal fossil park design resulted in the identification of the available opportunities for visitors to learn important concepts in geobiology in outdoor settings. We affirm that fossil parks can serve as outdoor teaching laboratories to address geologic time, fossilization, evolution, biostratigraphy, biodiversity, and environmental change (Singer et al., 2005).

Our 7 yr qualitative research study concludes that fossil parks can offer scientific experiences to the public that contribute to geoscience literacy, although the individual effectiveness of each park is dependent upon key variables. There is no equivalent virtual substitute for direct interaction with Earth, and fossil ownership sparks thought about deep time, evolution of life forms, and environmental change over geologic time.

We caution that our interpretation is unique and reflects investigations of fossil park facilities and their analyses over a 4 yr period. Undoubtedly, some of the fossil parks may have since modified their designs, signage, and facilities resulting in improved geobiological education opportunities for the visitor. Our qualitative work does not aim to describe all parks in which fossils are displayed, but rather provides an illuminating flavor of the U.S. informal sites we identified that permit fossil collection and retention. We acknowledge that our comparative ranking of the seven U.S. fossil parks may or may not be the same if we visited each site again. Replication of this study may be affected by the season in which investigations are conducted, the economic conditions of the area, and the conditions of the site. Our interpretation and research serve as a snapshot analysis in time.

Multiple data sources were utilized in all of our case studies, including posted materials, brochures, and on-site signage. Additional data sources included unstructured informal conversations with fossil park personnel and visitors, and field notes from our lived learning experiences. Inter-rater reliability was good and was established within 95% agreement. Therefore, the internal validity of the research investigations was established through verification.

Our lived learning experiences at the seven U.S. fossil parks were unique, and therefore the external validity and the generalizability of our research to other informal educational sites outside of the fossil park realm are limited. However, we confirmed that the optimal fossil park design that emerged from the initial case study of the first three U.S. fossil parks (phase 1) had application and relevance to the subsequent four fossil park sites we researched in phase 2 (2005) and phase 3 (2006).

We note that our findings are in concordance with the principles of active, meaningful, and mindful learning and the learning theory of human constructivism. Visitors to fossil parks can have optimized geobiology education experiences when they are actively engaged in fossil collecting with awareness of both the purpose and context of collecting, when the geobiological information presented at the site accesses the learner's existing knowledge framework, and when information and experiences at the fossil park have a context to the learner. The successful fossil park designs provide the context of geologic time, biodiversity, and environmental change through signage or a helpful paleontological mentor, and aid in the construction of visitors' geobiological knowledge by scaffolding upon experiences and general knowledge.

We continue to search for additional U.S. fossil parks that offer informal geoscience education to the visitor and that fill a gap between the U.S. National Park System and undeveloped field experiences.

ACKNOWLEDGMENTS

We are thankful for the assistance provided by volunteers and mentors at the U.S. fossil park sites involved in this research study. We also thank Anthony Feig for his helpful suggestions and editorial comments, which greatly strengthened the presentation of our research.

REFERENCES CITED

Anderson, D., Lucas, K., and Ginns, I., 2003, Theoretical perspectives on learning in an informal setting: Journal of Research in Science Teaching, v. 40, p. 177–199, doi:10.1002/tea.10071.

Anderson, W., 1998, Iowa's Geological Past: Three Billion Years of Earth History: Iowa City, University of Iowa Press, 440 p.

Ausubel, D., 1963, The Psychology of Meaningful Verbal Learning: An Introduction to School Learning: New York, Grune and Stratton, 255 p.

Ausubel, D., 1968, Educational Psychology: A Cognitive View: New York, Holt, Rinehart, & Winston, 685 p.

Ausubel, D., Novak, J., and Hanesian, H., 1978, Educational Psychology: A Cognitive View (2nd ed.): New York, Holt, Rinehart, and Winston, 733 p.

Bastedo, J., 2000, Penn-Dixie Paleontological and Outdoor Education Center: Visit to a Classic Geological and Outdoor Education Center: New York State Geological Association Meeting Guidebook, v. 71, p. A1–A19.

BBC News, 2006, Stolen dinosaur fossils recovered: British Broadcasting Corporation: http://news.bbc.co.uk/2/hi/uk_news/wales/south_east/5299016.stm (accessed 30 August 2006).

Bernstein, S., 2003, A limestone way of learning: The Chronicle of Higher Education, v. 50, no. 7, p. B5.

Bogdan, R., and Biklen, S., 1998, Qualitative Research for Education: An Introduction to Theory and Methods (3rd ed.): Needham Heights, Massachusetts, Allyn & Bacon, 276 p.

Cable News Network (CNN), 2003, Finding trilobites by the bucket, for free: http://www.cnn.com/2003/TECH/science/06/26/fossil.park.ap./index.html (accessed 26 June 2003).

Chi, M., 1997, Quantifying qualitative analyses of verbal data: A practical guide: Journal of the Learning Sciences, v. 6, no. 3, p. 271–315, doi:10.1207/s15327809jls0603_1.

Clary, R., and Wandersee, J., 2008, Earth science teachers' perceptions of an autonomous fieldwork assignment in a nationwide online paleontology course: Journal of Geoscience Education, v. 56, no. 2, p. 149–155.

Clary, R., and Wandersee, J., 2009, Incorporating informal learning environments and local fossil specimens in earth science classrooms: A recipe for success: Science Education Review, v. 8, http://www.scienceeducationreview.com/open_access/index.html (accessed 21 November 2009).

Clary, R., Brzuszek, R.F., and Wandersee, J.H., 2009, Students' geocognition of deep time, conceptualized in an informal educational setting: Journal of Geoscience Education, v. 57, no. 4, p. 275–285.

Creswell, J., 1994, Research Design: Qualitative and Quantitative Approaches: Thousand Oaks, California, Sage Publications, 228 p.

DeBoer, G., 1991, A History of Ideas in Science Education: New York, Teachers College Press, 269 p.

Denzin, N., 1978, The Research Act (2nd ed.): New York, McGraw-Hill, 379 p.

Denzin, N., and Lincoln, Y., 2000, Handbook of Qualitative Research: Thousand Oaks, California, Sage Publications, 1232 p.

Downing, B., 2003, Time travel at Fossil Park: The Beacon Journal: http://www.ohio.com/mld/ohio/entertainment/visitors_guide-3545870.htm (accessed 30 June 2003).

Drewes, C., 2005, Discovering Devonian microfossils, ABLE 2005 Workshop: http://www.eeob.iastate.edu/faculty/DrewesC/htdocs/microfossilsABLE.doc (accessed 3 November 2005).

Falk, J., 2001, Free-Choice Science Education: How We Learn Science Outside of School: New York, Teachers College Press, 216 p.

Falk, J., and Dierking, L., 2000, Learning from Museums: Visitor Experiences and the Making of Meaning: Walnut Creek, California, AltaMira Press, 272 p.

Falk, J., and Dierking, L., 2002, Lessons without Limit: How Free-Choice Learning Is Transforming Education: Walnut Creek, California, AltaMira Press, 189 p.

Felzien, L., and Cooper, J., 2005, Modeling the research process: Alternative approaches to teaching undergraduates: Journal of College Science Teaching, v. 34, p. 42–46.

Fossil & Prairie Center Foundation, 2003, The Fossil & Prairie Park and Center: http://www.fossilcenter.com/area.html (accessed 23 August 2003).

Gilmore, T., 2006, Lee Creek Geology, Aurora Fossil Museum: http://www.aurorafossilmuseum.com (accessed 3 November 2006).

Glaser, B., 1978, Theoretical Sensitivity: Advances in the Methodology of Grounded Theory: Mill Valley, California, Sociology Press, 164 p.

Glaser, B., and Strauss, A., 1967, The Discovery of Grounded Theory: Strategies for Qualitative Research: Chicago, Aldine, 279 p.

Golafshani, N., 2003, Understanding reliability and validity in qualitative research: Qualitative Report, v. 8, p. 597–607.

Gowin, D.B., 1981, Educating: Ithaca, New York, Cornell University Press, 210 p.

Gunstone, R., and Mitchell, I., 1998, Metacognition and conceptual change, *in* Mintzes, J.J., Wandersee, J.H., and Novak, J.D., eds., Teaching Science for Understanding: A Human Constructivist View: San Diego, California, Academic Press, p. 133–163.

Harper, S., and Kuh, D., 2007, Myths and misconceptions about using qualitative methods in assessment: New Directions for Institutional Research, v. 136, p. 5–14, doi:10.1002/ir.227.

Hemler, D., and Repine, T., 2006, Teachers doing science: An authentic geology research experience for teachers: Journal of Geoscience Education, v. 54, p. 93–102.

Hose, T., 1995, Selling the story of Britain's stone: Environmental Interpretation, v. 10, no. 2, p. 16–17.

Jones, M., 2004, Sylvania: Olander seeks to improve fossil park: http://www.toldeoblade.com/apps/pbcs.dll/article?AID=200410070342 (accessed 7 October 2004).

Klesius, M., 2002, The big bloom: http://ngm.nationalgeographic.com/ngm/0207/feature6/index.html (accessed 6 August 2006).

Langer, E.J., 1997, The Power of Mindful Learning: Reading, Pennsylvania, Addison-Wesley, 167 p.

Larwood, J., and Durham, E., 2005, Earth Heritage: World Heritage: Involving People in Geodiversity: Spalding, UK, Status Design, 16 p.

Lawrenz, F., Huffman, D., and Appeldoorn, K., 2005, Enhancing the instructional environment: Journal of College Science Teaching, v. 3, p. 112–115.

Lincoln, Y., 1995, Emerging criteria for quality in qualitative and interpretive research: Qualitative Inquiry, v. 1, p. 275–289, doi:10.1177/107780049500100301.

Lincoln, Y., and Guba, E., 1985, Naturalistic Inquiry: Beverly Hills, California, Sage, 416 p.

Long, L., 9 December 2005, Arizona promotes 'geotourism': The Arizona Daily Star. Available at http://enn.com/biz/PF.html?id-1266 (accessed 9 December 2005).

Lord, T., and Orkwiszewski, T., 2006, Moving from didactic to inquiry-based instruction: The American Biology Teacher, v. 68, p. 342–345, doi:10.1662/0002-7685(2006)68[342:DTIIIA]2.0.CO;2.

Manning, K., 1997, Authenticity in constructivist inquiry: Methodological considerations without prescription: Qualitative Inquiry, v. 3, p. 93–115, doi:10.1177/107780049700300105.

Marshall, C., and Rossman, G.B., 1989, Designing Qualitative Research: Newbury Park, California, Sage Publications, 175 p.

McComas, W., 1996, Ten myths of science: Reexamining what we think we know about the nature of science: School Science and Mathematics, v. 96, p. 10–16, doi:10.1111/j.1949-8594.1996.tb10205.x.

McComas, W., 2006, Science teaching beyond the classroom: The role and nature of informal learning environments: Science Teacher (Normal, Ill.), v. 73, no. 1, p. 26–30.

McConnell, D., Steer, D., and Owens, K., 2003, Assessment and active learning strategies for introductory geology courses: Journal of Geoscience Education, v. 51, p. 205–216.

Meredith, J., Fortner, R., and Mullins, G., 1997, A model of affect in nonformal education: Journal of Research in Science Teaching, v. 34, p. 805–818, doi:10.1002/(SICI)1098-2736(199710)34:8<805::AID-TEA4>3.0.CO;2-Z.

Meyer, H., and Manchester, S., 1997, The Oligocene Bridge Creek Flora of the John Day Formation: Berkeley, University of California Press, 364 p.

Michael, J.A., and Modell, H.I., 2003, Active Learning in Secondary and College Science Classrooms: A Working Model for Helping the Learner to Learn: Philadelphia, Lawrence Erlbaum Associates, 171 p.

Mintzes, J.J., Wandersee, J.H., and Novak, J.D., eds., 1998, Teaching Science for Understanding: A Human Constructivist View: San Diego, California, Academic Press, 360 p.

Mintzes, J.J., Wandersee, J.H., and Novak, J.D., eds., 2000, Assessing Science Understanding: A Human Constructivist View: San Diego, California, Academic Press, 386 p.

Mir, R., 2003, Natural attraction: Implementing your science park: http://www.astc.org/pubs/dimensions/2003/mar-apr/index.htm (accessed 14 December 2003).

National Park Service, 1991, Natural Resource Management Reference Manual 77: http://www.nature.nps.gov/rm77/ (accessed 28 October 2009).

Neuendorf, K., 2002, The Content Analysis Guidebook: Thousand Oaks, California, Sage Publications, 301 p.

Nieswiadomy, R., 1993, Foundations of Nursing Research: Norwalk, Connecticut, Appleton & Lange, 448 p.

Novak, J., 1963, What should we teach in biology?: Journal of Research in Science Teaching, v. 1, no. 3, p. 241–243.

Novak, J., 1977, A Theory of Education: Ithaca, New York, Cornell University Press, 296 p.

Novak, J., 1998, The pursuit of a dream: Education can be improved, *in* Mintzes, J.J., Wandersee, J.H., and Novak, J.D., eds., Teaching Science for Understanding: A Human Constructivist View: San Diego, California Academic Press, p. 3–29.

Novak, J., and Gowin, D., 1984, Learning How to Learn: Cambridge, UK, Cambridge University Press, 199 p.

Novak, J., and Wandersee, J., eds., 1991, Perspectives on concept mapping: Journal of Research in Science Teaching, v. 28, no. 10, Special issue, p. 921–1074.

Oregon Paleo Lands Institute, 2005, Oregon Paleo Project: http://www.oregonpaleoproject.org (accessed 5 December 2005).

Orion, N., and Hofstein, A., 1994, Factors that influence learning during a scientific field trip in a natural environment: Journal of Research in Science Teaching, v. 31, p. 1097–1119, doi:10.1002/tea.3660311005.

Patzak, M., 2000, Tourism and geodiversity: The case of GEOPARKS, United Nations Educational, Scientific and Cultural Organization: Paris, UNESCO Division of Earth Sciences, 11 p.

Persson, P.-E., 2000, The changing science center: Sustaining our mission into the 21st century: http://www.astc.org/pubs/dimensions/2000/jan-feb/changing.htm (accessed 5 January 2006).

Rennie, L., and Johnston, D., 2004, The nature of learning and its implications for research in learning from museums: Science Education, v. 88, p. S4–S16, doi:10.1002/sce.20017.

Roy, M., and Doss, L., 2007, Building migratory bridges: The Science Teacher, p. 74, no. 8, p. 56–63.

Singer, S.R., Hilton, M.L., and Schweingruber, H.A., eds., 2005, America's Lab Report: Investigations in High School Science: Washington, D.C., National Academies Press, 236 p.

Stokes, A., 2011, this volume, A phenomenographic approach to investigating students' conceptions of geoscience as an academic discipline, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(03).

Stoll, J., 2001, Fossil park—One for the ages! Fossil Park: Sylvania, Ohio, Citizens for the Olander Park System, 24 p.

Strauss, A.L., 1987, Qualitative Analysis for Social Scientists: Cambridge, Cambridge Press, 319 p.

Stueve, A., Cook, S., and Drew, D., 2002, The Geotourism Study: Phase I Executive Summary: Washington, D.C., Travel Industry Association of America, 24 p.

Tashakkori, T., and Teddlie, C., 1998, Mixed Methodology: Combining Qualitative and Quantitative Approaches: Thousand Oaks, California, Sage Publications, 185 p.

Tesch, R., 1990, Qualitative Research: Analysis Types and Software: New York, Falmer, 340 p.

University of Northern Kentucky Geology Department, 2004, Regional Geology: Trammel Fossil Park: http://physics.nku.edu:16080/geology/NKY_Geology.htm (accessed 1 August 2005).

Wandersee, J.H., and Clary, R.M., 2006, Fieldwork: New directions and examples in informal science education research, *in* Mintzes, J., and Leonard, W., eds., National Science Teachers Association (NSTA) Handbook of College Science Teaching: Theory, Research, and Practice: Arlington, Virginia, NSTA Press, p. 167–176.

Wandersee, J.H., and Clary, R.M., 2007, Learning on the trail: A content analysis of a university arboretum's exemplary interpretative science signage system: The American Biology Teacher, v. 69, no. 1, p. 16–23, doi:10.1662/0002-7685(2007)69[16:LOTTAC]2.0.CO;2.

Wehr, W., 2008, What is Stonerose? Stonerose Interpretive Center: http://www.stonerosefossil.org/aboutus.htm (accessed 31 August 2009).

Wolcott, H.F., 2005, The Art of Fieldwork: Walnut Creek, California, AltaMira Press, 292 p.

Yin, R., 2003, Case Study Research: Design and Methods (3rd ed.): Thousand Oaks, California, Sage Publications, 240 p.

MANUSCRIPT ACCEPTED BY THE SOCIETY 23 JUNE 2010

The Geological Society of America
Special Paper 474
2011

What college-level students think: Student alternate conceptions and their cognitive models of geoscience concepts

Leilani Arthurs*
Department of Geology & Geography, Georgia Southern University, P.O. Box 8149, Statesboro, Georgia 30460-0002, USA

ABSTRACT

Interviews, paper-and-pencil (PNP) exercises, and class observations were the qualitative research methods used to investigate student alternate conceptions and their cognitive models of geoscience concepts. Three categories of geoscience concepts guided the research: rocks, density and convection, and water. A taxonomy of alternate conceptions is presented for the purpose of discussing the different ways that students conceptualize geoscience concepts, and examples of student-held alternate conceptions are listed herein. Coherent cognitive models about (1) rocks and their origins, (2) mantle dynamics, (3) the storage of groundwater in the subsurface and its connections to drinking water, and (4) the origin and movement of groundwater were generated from the data about alternate conceptions. This study also contains an evaluation of the three methods used in terms of their effectiveness at revealing student thinking, and different models for conceptualizing students' alternate conceptions are discussed. These include the discrete correct-incorrect model, the continuous unscientific-scientific model, the continuum model, the radial model, and the simplified web model. The findings of this research can be used to facilitate constructivist student-centered learning when they are taken into consideration and factored into (1) the practice of teaching, (2) course curriculum development, and (3) the development of formative and summative assessments that might include tests and in-class activities, respectively.

INTRODUCTION

Language and Cognition

Determining what people think is tricky business and does not lend itself to traditional models of scientific and quantitative inquiry. However, the study of language and, specifically, linguistic expressions can provide qualitative insights into what and how humans think (i.e., human cognition and cognitive models). Lakoff (1987) discussed the categorization of linguistic expressions of concepts, such as "anger," and argued for an experiential view of cognitive models and human reason. His work was built largely off the work of others such as Rosch and Dixon. Rosch (1975, 1981) demonstrated that the study of categories provides a means through which human inferences, reasoning, and cognitive structures can be studied. Dixon (1982) described the Dyirbal (an aboriginal language of Australia) classification system of objects in the universe based on what Lakoff coined the "domain-of-experience principle." This principle states that "if there is a basic domain of experience associated with *A*, then it is natural for entities in that domain to be in the same category as *A*"; Lakoff discussed kinds of categories based on linguistic

---

*laarthurs@georgiasouthern.edu

Arthurs, L., 2011, What college-level students think: Student alternate conceptions and their cognitive models of geoscience concepts, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 135–152, doi:10.1130/2011.2474(10). For permission to copy, contact editing@geosociety.org.

evidence. He did so with the recognition that language is a part of the human experience and argued that linguistic categories are kinds of cognitive categories. His work goes into great depth to describe four categories of cognitive models. For the purposes of this paper, however, there are three main points from Lakoff's and his predecessors' works that provide the foundation for and frame the work described herein. The first is that human cognition lies in the domain of experience. The second is that the domain of experience includes interactions with the physical world as well as language. The third is that examination of the domain of experience and linguistic evidence provides insights into human cognition.

Constructivism

The notion that human cognition lies within the domain of experience has important implications for the constructivist model of education. Unlike the traditional, transmissionist model of education, the constructivist model emphasizes the importance of students' preexisting knowledge and experience in the learning process (National Research Council, 2000). Constructivism acknowledges the fact that students come into the learning environment with a wealth of prior experience and knowledge that may help, hinder, or have no impact on their integration of new knowledge and skills into what they already know (Mayer, 1998; National Research Council, 2000; Redish, 2003; Handelsman et al., 2007). Because of their impact on student learning, a constructivist approach to teaching aims to scaffold new learning in the context of current student thinking and existing ideas (The National Society for the Study of Education, 1995). Scaffolding involves instructional practices that bridge or build connections between what students already know and what the instructor would like them to know, and, in this way, students construct knowledge for themselves. This is very different from the traditional, transmissionist model of education where the student is perceived as a blank slate or empty vessel waiting for their teacher, the font of knowledge, to fill them up with information.

The prior experience and knowledge that students carry with them contribute to folk theories (Lakoff, 1987) about how the world around them operates, and these are often not in agreement with the equivalent scientific theories held by experts in a given field. Folk theories are assumptions based on everyday life experiences that people generally believe to be true. For example, pervasive geologic folk theories of the past include "the world is flat" and "the Earth is the center of the universe." For the purposes of this paper, we are going to refer to student-held ideas influenced by prior experience and knowledge as "alternate conceptions" and their folk theories as "cognitive models."

Alternate Conceptions

The idea of alternate conceptions is not new, and others have referred to the idea as naïve (Clement, 1993; Kinchin et al., 2000), preconceptions (Novak, 1977; Clement 1993), misconceptions (Helm, 1980), and alternate frameworks (Driver, 1981; Dal, 2007). They have also been discussed in terms of a spectrum from nonscientific to scientific ideas (Mortimer, 1995; Izquierdo-Aymerich and Aduriz-Bravo, 2003; Libarkin et al., 2005) or novice-like to expert-like ideas (Bereiter and Scardamalia, 1986; Lindblom-Ylänne and Lonka, 2001; Wieman and Perkins, 2005; Wieman, 2007). The terminology found in the literature to discuss this idea is used congruently but also inconsistently and ambiguously from one source to the next. This is, in part, due to a lack of a common language with which to discuss these ideas. Thus, for clarity, I define the key terminology about student thinking that was used during this study, and I provide their definitions here. Figure 1 outlines my classification system of the terms used to discuss students' alternate conceptions. I recognize that my definitions and classification system are not universal and that they may evolve over time, but it is important to clarify how they were defined and classified in this particular study.

1. Concept: (a) content concepts, which commonly are not derived knowledge but memorized knowledge such as the age of Earth or the definition of water table; (b) conceptual principles, which can be applied to solve and interpret a variety of problems and situations such as convection or Gibbs free energy.
2. Perception: idea or thought formed through direct connection to an immediate observation or other sensory experience.
3. Conception: idea or thought formed through inference, extrapolation, or imagination, which may or may not have been informed by previous perceptions.

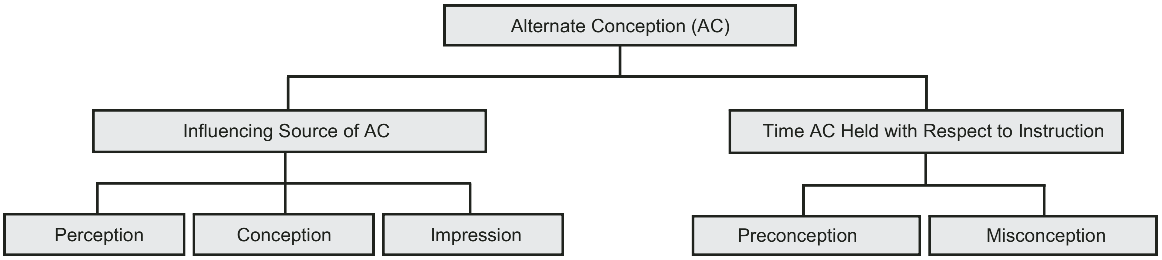


Figure 1. Taxonomy of alternate conceptions.

4. Impression: idea or thought held because a person heard it from (or read it in) a believed-to-be credible, although scientifically inaccurate, source, such as a documentary, a movie, a parent, or a teacher; may also be called a "scientific myth" (e.g., toilet flushes clockwise in one hemisphere and counterclockwise in the other).
5. Preconception: idea or thought that is held by a student prior to formal instruction during the period of interest; may be more or less scientifically accurate and complete.
6. Misconception: persistent idea or thought that is strongly held even after formal instruction and that is not accepted as scientifically accurate. A misconception could be a preconception that was not corrected, or it could be a new idea that the student took away from the course.
7. Alternate conception: idea or thought held by a student at any point in time relative to the instructional period of interest, formed by direct or inferred experience, and one that is more/less scientifically accurate and complete. This is an overarching term that encompasses the terms described previously, as illustrated in Figure 1.

Uncovering what students think is no trivial matter, and science education research plays an important role in this regard. Such research has been traditionally conducted at the K–12 levels and in the disciplines of physics, chemistry, and biology. The field of geology and earth science education research (GER) is relatively new and also has its roots in the K–12 levels. In recent years, however, GER at the college level has also seen notable growth (e.g., Delaughter et al., 1998; McKenney and Webster, 2004; McConnell et al., 2005; Dickerson et al., 2005; Libarkin et al., 2005; McNeal et al., 2008; Kastens et al., 2009). Based on a review of the literature, Sibley (2009) discussed the ways in which all scientific models are relational analogs; that is, they are "analogies based on shared relationships between the analog and target, maybe equivalent to scientific models" (p. 255). In this sense, one might argue that scientific models are extensions of mental models.

Cognitive Models

The field of mental models research is new in the area of GER, and there is an emerging body of work in the literature describing it (e.g., Vosniadou and Brewer, 1992; Sell et al., 2006; McNeal et al., 2008; Kastens et al., 2009). Mental models researchers distinguish between conceptual models and mental models. According to Norman (1983), "conceptual models are devised as tools for the understanding or teaching of physical systems" (p. 12) and "[m]ental models are what people really have in their heads and what guides their use of things" (p. 12). In this article, I use Norman's definitions of conceptual and mental models. Furthermore, I also use the term "cognitive model" synonymously with Norman's "mental model" to describe how student ideas or thoughts hinge together in a coherent framework.

Here, the method used by Norman to discuss mental models is adopted and modified. Let the target geological concept of interest be called c. In order to cogently discuss student thinking, an accepted conceptualization of the concept is needed as a reference. For this study on student thinking of geoscience concepts, the accepted conceptualizations are defined by experts in the scientific community and constrained by the learning goals defined by an instructor. Let the expert conceptualization of c that is generally accepted by the scientific community be called $E(c)$, let $E(c)$ constrained by the instructor for instructional purposes be called $I(E[c])$, and the student or novice conceptualization be called $L(c)$. Finally, to distinguish student thinking of c from what the researcher thinks the student thinks about c, let the researcher's conceptualization of $L(c)$ be called $R(L[c])$.

This distinction is critical because of the challenges inherent in discovering the cognitive models of individuals. For example, the researcher is privy to only parts of student cognitive models that are expressed either in verbal, written, or drawn form. Naturally, such self-reported student data possess limitations (Stone et al., 2000). For example, students might think one way and not be aware enough of it to express it; they might be aware enough of their thinking to satisfactorily express it but do not do so for a variety of reasons, including wanting to provide the response they believe is expected or correct (Norman, 1983). Furthermore, not all student conceptualizations are stable over time (Libarkin et al., 2003) and coherent relative to other personally held ideas (Delaughter et al., 1998; Mark et al., 1999).

Qualitative inquiry plays an important role in researching language, alternate conceptions, and cognitive models. This type of research may utilize different approaches, such as a critical theory approach or a grounded theory approach (for a comparison of the two approaches, see Feig, this volume). The methods used are diverse and include the use of, for example, open-ended questionnaires (e.g., Geer, 1991; Libarkin and Kurdziel, 2001; Libarkin et al., 2005), multiple-choice questions and questionnaires (e.g., Nield and Wintre, 1986; Scouller, 1998; Libarkin and Kurdziel, 2001), sketches or drawings and concept maps (e.g., Libarkin and Kurdziel, 2001; Dickerson and Dawkins, 2004; Sibley, 2005; Dahl et al., 2005; Shepardson et al., 2005; Dal, 2007), interviews (e.g., Bogdan and Biklen, 1998; Libarkin and Kurdziel, 2002; Libarkin et al., 2005), and observations (e.g., Bogdan and Biklen, 1998; Libarkin and Kurdziel, 2002).

Within GER, much of the research into students' alternate conceptions involves identifying them in sample populations and listing and describing them (e.g., Dove, 1998; Ford, 2003; Dickerson and Dawkins, 2004; Dahl et al., 2005; Dickerson et al., 2005; Henriques, 2005; Rule, 2005; Shepardson et al., 2005). Many such studies focus on how students' alternate conceptions are incorrect, partially correct, or mismatched compared with those of experts (e.g., Kusnick, 2002; McKenney and Webster, 2004). There exist fewer examples of GER studies that attempt to investigate students' cognitive models as they relate to geoscience concepts as well as students' cognitive development (e.g., McConnell et al., 2005; Libarkin and Kurdziel, 2006).

Objectives

The research behind this study was based on qualitative inquiry and a grounded theory approach. It was exploratory in nature, aimed at providing a preliminary understanding of what college-level students enrolled in the introductory-level geology and environmental science courses think about several geoscience concepts, and it was intended to assist with the implementation of a constructivist approach to teaching and learning in the Department of Geological Sciences at the University of Colorado at Boulder. Three broad categories of geoscience concepts were selected for this study. They include: rocks, density and convection, and water. The geoscience concepts investigated were selected because they are fundamental to several of the introductory college-level geology and environmental science courses. The research methods included surveying students through interviews and questionnaires (i.e., written exercises) and conducting observations (during class). These methods are common practice in qualitative research (LeCompte and Schensul, 1999; Schensul et al., 1999; Sapsford, 1999; Willis, 2008).

The objectives of this paper include: (1) report a summary of findings in terms of existing alternate conceptions of geoscience concepts and coherent cognitive models that emerged; (2) present models for conceptualizing how students conceptualize and think about geoscience concepts; and (3) outline avenues of further research into student cognition of geoscience concepts. This paper contributes to the growing catalogue of students' geoscience alternate conceptions, specifically at the undergraduate college level of 4-yr-degree–granting universities. It also highlights examples of cognitive models based on the domain of experience not only in the physical world but also with respect to language.

METHODOLOGY AND METHODS

The goal of qualitative inquiry is not for the findings of research to be applied universally; instead, the findings are intended to provide deeper insight into a particular issue (Mason, 2002) through the generation of data and the researcher's identification of patterns and themes (i.e., categories) in that data (Wolcott, 1994). A grounded theory approach to qualitative inquiry is a data-driven approach geared at understanding a key issue (Creswell, 1998), and this approach is applied to this study. For more in depth discussion of qualitative inquiry and grounded theory, the reader is directed to Feig (this volume).

The issue of interest in this study is student thinking as it pertains to three categories of geoscience concepts. Factors of specific interest are students' alternate conceptions and cognitive models as they relate to these concepts. Research on student thinking is not immediately amenable to quantitative inquiry; however, it is directly amenable to qualitative inquiry. Any quantification of aspects of student thinking first requires the generation of qualitative data from, for example, interviews and questionnaires. Possible ways to quantify the qualitative data generated include, for example, counting the number of categories emerging from the data and counting the number of students who were involved with each category.

For this study, student alternate conceptions and cognitive models were not identified a priori; instead, they emerged as a result of the qualitative research conducted and on the basis of a grounded theory approach.

Locating the Researcher and the Study

I received approval from the University of Colorado's Institutional Review Board to administer questionnaires and conduct interviews. My past training and education provided me with a background in cognitive science and linguistics and expertise in geoscience (specifically, low-temperature geochemistry) and pedagogy (specifically, applications of constructivism in teaching practices and assessment). I conducted this research with a variety of assistance from nine instructors and two undergraduate research assistants. Four instructors assisted with this research in consenting to my interviewing their students. Two instructors assisted via their consenting to my observing their classes on a regular basis and their consistent collaboration in developing questions for student questionnaires. Over the course of this study, at different times, all nine instructors provided expert insights on the questions developed for questionnaires and shared their expert views regarding specific concepts. The two undergraduate research assistants helped me to digitize my field notes (i.e., class observations) and transcribe audiotaped interviews. In addition, they provided student-perspective feedback on questions developed for interviews and questionnaires as a part of vetting the questions prior to their implementation.

The department in which this study was conducted is located in a large west-central research university. The campus is located at the junction of the Colorado Eastern Plains and the foot of the Rocky Mountains. The most recent demographic information publicly available for this campus is for 2004 (Office of Planning, Budget, and Analysis, 2009). It indicates that of the 24,719 undergraduates enrolled, 47% were women, 53% were men, 67% were Colorado residents, 33% were nonresidents, 4% were international students, 87% were Caucasian, and 13% were students of color.

Students enrolled in three different geology courses and one environmental science course during the period of fall 2007 to spring 2009 were the subjects sampled for the exploratory investigations described in this article. The courses included Introduction to Geology I, Introduction to Environmental Studies, Environmental Geology, and Introduction to Oceanography. All four courses satisfy the natural science requirement and are largely populated by non–science-technology-engineering-mathematics (non-STEM) majors.

Validity and Reliability

Triangulation of information derived from different data sources about the same topic is critical to the validity and reliability of qualitative research. Thus, three different methods of data collection—interviews, paper-and-pencil exercises, and class observations—were used in this research. When common themes in student thinking emerged using multiple methods, triangulation and redundancy were possible, and I could confirm and cross-check the accuracy of data generated from one method with data generated from another. This was a valuable approach to indicate whether the data were valid and reliable. In addition, the use of all three methods provided the opportunity to assess the effectiveness of each one in revealing student thinking relative to another.

Data Collection

Interviews

Interviews in fall 2007 and 2008 and spring 2008 and 2009 were conducted during the time period that coincided with the final third of the semester. These interviews were voluntary, and all volunteers were interviewed. Interviews in summer 2008 were conducted during the first week of the summer session. In total, 9 students were interviewed from Introduction to Geology I, 4 from Introduction to Environmental Studies, 11 from Environmental Geology, and 24 from Introduction to Oceanography. Of the 48 students interviewed, 65% were male, 35% were female, 10% were freshmen, 33% were sophomores, 19% were juniors, 35% were seniors, and 2% were graduate students. Fifty-six percent of interviewees were non-STEM majors, 35% STEM majors, and 8% undeclared or open-option students. Ninety-four percent of them were Caucasian Americans, and 6% were students of color.

Invitations for student interviews were extended both by personal announcements in class and by instructor-facilitated e-mail communications with the students. The interviews were conducted on a voluntary basis, were not a part of the course requirements, and did not impact students' course grades. Students who volunteered for an interview were compensated at a rate of $15/hr. All students that volunteered were interviewed. The interviews were 1–1.5 h in length and were conducted in private on a one-on-one basis in a classroom.

The interviews were semistructured in the sense that a set of scripted and sequenced questions were used for the interview, and, yet, there was flexibility within the interview protocol to ask follow-up questions for clarification and elaboration. There are at least two advantages to using a semistructured interview protocol over other types of interviews. First, using the protocol ensures that all students are asked the same questions in the same way. This minimizes possible omissions, biases, or artifacts in the data that might arise in casual, informal interviews. Second, the protocol allows the interviewer to ask unscripted follow-up questions to gain more information about student thinking. This allows the interviewer to pursue unanticipated lines of student thinking. The lists of scripted and sequenced questions that were asked varied from one interview period to the next (e.g., from fall 2007 to spring 2008) and consisted of new and/or previously asked questions. The questions varied from one interview period to another because new questions were formulated based on findings from previous interviews, class observations, and instructor input.

Two classes of questions were asked. One class involved a broad exploratory approach to specific geological concepts. These included, for example:

1. Have you heard of (insert X geological concept here)?
2. Where have heard about X?
3. Can you describe or explain what X is to me? Imagine that I am a relative or a friend, and you are trying to teach me about X. What would you say?
4. If you could draw something to help teach me about X, what would you draw?
5. Could you tell me more about what this is (while pointing to some aspect of their drawing)?

I devised these questions based on my interest in learning about what students knew, how they knew it, and how they might explain what they know. These questions were vetted by two undergraduate research assistants for clarity of meaning (i.e., whether questions were interpreted in the manner intended).

Another class of questions was designed to probe student thinking about a particular concept in a more targeted way. For example: What is eutrophication? What causes eutrophication? What are the environmental effects of eutrophication? These questions were devised in collaboration with instructors who were interested in learning more about what their students knew about specific aspects of different geoscience and environmental science concepts.

These questions were the same questions used in paper-and-pencil exercises (see following section) and were vetted by two undergraduate research assistants and a minimum of two geoscience experts. During the interviews, I took verbatim notes of student responses and also audiotaped the interviews for future reference if needed. Interviews conducted earlier in this study were all manually transcribed. The practice of transcribing interviews was not continued throughout the study because the interview notes captured verbatim the information sought, and, therefore, transcription was viewed as not only labor intensive and costly but also nonessential. The interviews conducted earlier in the study were audiotaped and transcribed. Textual analysis was performed on the earlier transcriptions and later verbatim interview notes. Drawings were retained and analyzed for commonalities, patterns, and groupings.

Paper-and-Pencil Exercises

In fall 2007, students were asked to complete six paper-and-pencil (PNP) exercises throughout the course of the semester. Each exercise contained three to five questions. These questions were developed based on instructor input, informed by documented or suspected alternate conceptions, and occasionally drawn from Libarkin's Geocognition Research Laboratory's Geoscience Concept Inventory version 1 (2004, https://www.msu.edu/~libarkin/gci.html). The PNP exercises were a source of both qualitative and quantitative data. Qualitative data came in the form of student responses to open-ended questions, and quantitative data came from students' selections of multiple-choice answers.

A copy of one of the PNP exercises is included in Appendix 1. This PNP exercise was devised based on the instructor's interest in learning more about how students think about nuclear waste repositories and their connection to geology. After discussing her interest, I formulated the PNP exercise, vetted it with the two undergraduate research assistants, and discussed it with the instructor. She then approved it for administration in her class. For other PNP exercises, sometimes revisions were made based on feedback received from the undergraduate research assistants and experts (including the instructor) before administering it in class.

During fall 2007, 26 questions were asked; 16 of the questions were multiple-part questions (i.e., a question that consisted of multiple embedded questions, such as a part [a] question and a part [b] question), 18 of the questions were multiple choice, 2 were open ended (not including a frequent follow-up question to explain the reasoning behind one's answer), 15 contained a follow-up question to explain the reasoning for the answer they provided, and 3 asked students to draw something. The class composition of the 52 students completing the course included 60% males and 40% females; 46% freshmen, 27% sophomores, 19% juniors, and 8% seniors; and 65% non-STEM majors, 19% STEM majors, and 35% undeclared (or open-option) students.

During spring 2008, students were asked to complete four PNP exercises. Each exercise contained four to five open-ended questions, and each question contained one to four parts. These questions were based on instructor input, informed by the student learning outcomes for the course, and they were vetted by three undergraduate research assistants and two experts (the instructor of the course plus a second expert from the department; the identity of the second expert depended on availability) prior to administration in an exercise. In total, 18 questions were asked; 12 of the questions were multiple-part questions, 2 of the questions were multiple choice, 16 were open ended, 1 was a follow-up question to explain the reasoning to an answer to a multiple-choice question, and 5 asked students to draw or sketch something. The class composition of the 162 students completing the course included 64% males and 36% females; 2% freshmen, 20% sophomores, 33% juniors, 36% seniors, 5% fifth-year seniors, 1% graduate students, and 2% non-degree students; and 67% non-STEM majors, 27% STEM majors, and 7% undeclared (open-option) students.

All exercises were administered prior to the start of a new module or during a natural break in the course schedule. Students had 10–20 min to answer the questions. Participation in answering the questions was voluntary. Neither their answers nor their participation had an effect on their course grades. However, students were encouraged to answer the questions as a means for the instructor to learn more about what they already knew coming into a new module and to, therefore, shape the coming lectures around what they knew. The responses to the PNP exercises were collected and systematically reviewed, scored (if multiple choice), and binned and coded (if open ended).

Class Observations

Class observations were conducted on a daily basis in the Environmental Geology course during fall 2007 and 2008 (for a total of ~76 h) and in the Introduction to Oceanography course during spring 2008 and 2009 (for a total of ~76 h). The Introduction to Geology I and the Introduction to Environmental Studies courses were not observed. The observation protocol involved unobtrusively sitting in a location that provided a clear view of most of the students, the instructor, and the instructional technologies (e.g., blackboard, projector screen, and visual aids) and from which discussion could be heard. Field notes were made to record the instructor-student and student-student interactions, specifically, the questions they asked, the answers they gave, and the comments they made. See Appendix 2 for an excerpt of the field notes. This field observation protocol was considered appropriate because it allowed me, the researcher, to focus on making and recording observations (rather than, for example, also assisting with instruction), to observe the class with minimal "external" interference that a nonstudent and noninstructor might introduce into the class, and to collect linguistic expressions uttered by students that shed light on their thinking about particular geoscience concepts.

Data Analysis

The purpose of analyzing qualitative data is to determine what patterns, themes, or categories the data can reveal and what coherent stories the data can tell. In order to do this, I performed textual content analyses (Sapsford, 1999; Libarkin and Kurdziel, 2002) of my interview notes and transcriptions, student responses to open-ended PNP exercises, and my field notes from class observations. These textual analyses involved searching for key words and key ideas that pertained to specific geoscience concepts (qualitative) and counting their occurrences (quantitative). In the case of the interviews and open-ended questions on PNP exercises, the textual analyses were coupled with coding the student responses to questions.

Coding involved the manual identification (i.e., no coding software package was used) of common categories of responses and assigning a code to each student response (e.g., Libarkin and Kurdziel, 2002). Coding made it possible to separate responses into bins (i.e., sorting student responses into different categories or different combinations of categories). Binning makes quantitative analysis of the data possible. The reliability of the results of coding and binning is strengthened when another person uses the same coding scheme to code and bin student responses. This is called intercoder reliability (e.g., Yeaton and Wortman, 1993; Kurasaki, 2000). Given the preliminary nature of this study, another person has not yet applied the coding scheme to the data. Thus, for the purposes of this paper, the discussion of the results is limited to only the qualitative data gathered and the qualitative aspects of data analysis.

In addition to using textual content analysis, I also analyzed student drawings that were collected during interviews and through PNP exercises. These drawings were another source of data used to understand student thinking expressed in verbal and written form (i.e., triangulation). They were also analyzed for commonalities, patterns, and groupings. Others, such as Gobert (2000) and Sibley (2005), have also used sketch or drawing analysis to further their understanding of some aspect of student thinking.

RESULTS

Student Ideas and Thoughts

The following section presents student alternate conceptions that were recorded during the period of the investigations. Unless otherwise indicated, student thinking that is listed is largely written in the language used by one or more students (i.e., raw student quotes or raw data). Brackets [] indicate author-inserted language, text, or comments. The findings presented emerged mainly from student interviews and PNP exercises and less so from class observations, and they were cross-correlated between methods. The findings are presented in tabular form. This is not an exhaustive collection of the findings but represents the most popular (i.e., in terms of frequency mentioned by students) and the most clearly expressed conceptions. Each alternate conception is designated as a certain type of alternate conception. The designations reflect the time held and origin. These designations were based on their most common occurrence among students in the investigations and were influenced by the timing of the interviews.

Rocks

A range of alternate conceptions dealing with rocks was found. Prevalent alternate conceptions focused on rocks existing at Earth's surface. It was notable that students did not think about rocks existing at deeper levels. This kind of thinking, at least for some students, was linked to the fact that they envision (1) the uppermost surface of the Earth as a hard very thin crust and (2) below that surface the Earth is liquid magma (this is discussed further in the section titled Student Cognitive Models and Ways of Thinking). A sampling of rock-related alternate conceptions is presented in Table 1.

Density and Convection

With respect to density and convection, the student responses to three specific questions are presented in Table 2. They deal with the formation of the ocean basins, the movement of mantle rock, and Earth's internal stratification. These data represent a wide range of student-held ideas about questions dealing with density and convection.

When asked to explain and draw a sketch illustrating how the Himalaya Mountains formed, the majority of students said "continental-continental collisions" but their sketches indicated that the term *continental-continental collisions* did not mean the same thing to everyone. Figure 2 shows reconstructions of several types of student drawings. The reconstructions contain the characteristics that were shared in common and that defined each type of drawing. Sibley (2005) also examined student thinking about continent-continent collisions using their drawings. Sibley also found, for example, that sketch C in Figure 2 was also popular among his sample of students.

When asked what causes mantle rock to move, students provided a number of very popular alternate conceptions. Many students replied with simply "convection" and no further elaboration. Students that did give responses or provide follow-up insights often connected mantle rock movement with high pressure, Earth's rotation, and gravitational forces. These are summarized in Table 2 under Question B.

When asked about Earth's internal structure and the reasons why it might be stratified or layered students often made connections with high pressures and/or high temperatures at depth. Very popular alternate conceptions in this regard are listed in Table 2 under Question C.

Water

A fair number of different alternate conceptions dealing with water were recorded during this study. They dealt with the sources of groundwater, the location of groundwater, and the subsurface storage conditions and movement of groundwater. Key

TABLE 1. ALTERNATE CONCEPTIONS DEALING WITH ROCKS

Alternate conceptions about rocks (as actual student quotes)	Type of alternate conception
Hard rocks exist only at [Earth's] surface.	Preconception, perception
Rocks are solid pieces of Earth's crust that were broken off.	Preconception, perception
Rocks are solid clumps of accumulated and hardened dirt.	Preconception, perception
Rocks are hardened and broken up forms of magma, [which] might have minerals in them.	Misconception, perception
Igneous rocks only form at the surface by volcanoes.	Misconception, perception
Igneous rocks are pushed up from the core.	Misconception, perception
Magma is all underground [i.e., below the crust is a region filled with molten rock].	Preconception, conception
[Mantle rock] behaves like lava, like thick liquid or honey.	Preconception, conception
[Mantle rock] moves ~10–15 mph.	Preconception, conception

TABLE 2. ALTERNATE CONCEPTIONS DEALING WITH DENSITY AND CONVECTION

Alternate conceptions about density and convection (as actual student quotes)	Type of alternate conception
<u>Question A. Ocean basins are topographically lower than continents because:</u>	
Plates that collide against each other to form the continents are higher than the plates that form the basins.	Preconception
The mass of ocean water compacts the ocean crust and depresses the ocean floor.	Preconception
Basins are voids that formed when plates spread apart.	Preconception
Erosion of land masses by flowing water or water formed the ocean basins.	Preconception
<u>Question B. Mantle rock moves because:</u>	
High pressure at different depths causes convection within the mantle.	Preconception
Earth's rotation causes convection within the mantle.	Preconception
Plates at the surface are moving and, because they are touching the mantle, they cause it to move when they move.	Preconception
Gravitational forces of Earth cause convection within the mantle.	Preconception
<u>Question C. Earth is stratified because:</u>	
Pressure causes a physical state of Earth's rocks to change.	Preconception
The temperature differences inside Earth cause certain layers to be liquid or solid.	Preconception
Minerals with different chemical compositions combine at different pressures and temperatures to form different layers.	Preconception

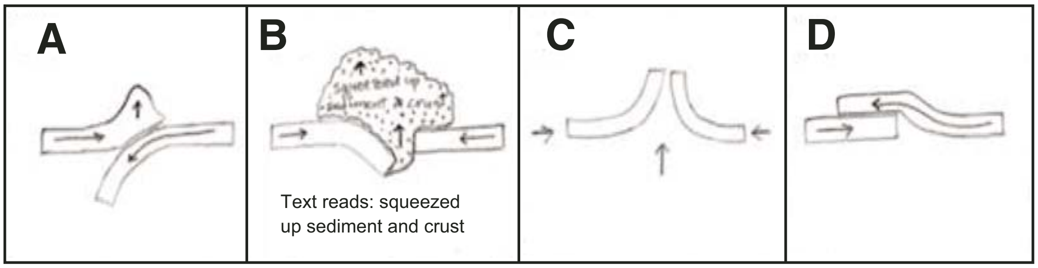


Figure 2. Reconstructions of student drawings depicting continental-continental collisions. (A) The right continental plate subducts below the left one, while the left one is pushed upward by the subducting plate and thus forms mountains. (B) Two colliding continental plates squeeze up sediment and crust, which forms the mountains. (C) Two plates colliding force each other to bow upward, thus forming mountains. (D) When two continental plates collide, one plate slides over the other one.

findings on alternate conceptions about water are summarized in Table 3.

Student Cognitive Models and Ways of Thinking

Through the course of this study, it became apparent that students hold myriad alternate conceptions about the geological concepts explored. More often than not, it was not immediately apparent how many of these ideas might be connected to a coherent cognitive model. Nevertheless, several coherent cognitive models were revealed. These are described next. Each model was derived from aggregate data and represents how different alternate conceptions hinged upon each other in student responses. The models are attempts to illustrate student thinking and reasoning that are tied to their alternate conceptions. Often one alternate conception led to or supported another. It is important to note that the cognitive models represent what was revealed by students and are not necessarily complete. It is also possible to illustrate cognitive models in a number of different ways. Three different ways were used to do so in this section—flow chart, written line of reasoning, and graphical representation.

Rocks

Many of the student perceptions and conceptions about rocks are connected in a coherent framework upon which logical lines of reasoning can be followed to inaccurate or incomplete conclusions. The framework or cognitive model is grounded in the fact that students (1) see hard rocks on the surface of Earth and (2) have seen lava emerging from the ground and flowing across Earth's surface (e.g., on television and in movies). The model is depicted as a flow chart in Figure 3.

Line A reflects incomplete conclusions about how rocks form. For example, Line A.1 indicates that students have a nascent grasp of sedimentary rocks that encompasses clastic origins (i.e.,

TABLE 3. ALTERNATE CONCEPTIONS DEALING WITH WATER

Alternate conception about water (as actual student quotes)	Type of alternate conception
Groundwater comes from oceans.	Conception
Groundwater comes from rain in the mountains.	Perception
Groundwater separates out from magma.	Conception
On a plain, you'd be closer to groundwater than [you would be] on a mountain.	Misconception, conception
Aquifers are caves with water.	Preconception, conception
Aquifers are underground rivers, like tunnels with rivers. Water pools in these areas, and these areas are good for wells.	Preconception, conception
Overpumping an aquifer leaves a [large] empty bubble or hollow space. This increases the pressure and [it] collapses in on [it]self.	Misconception, conception

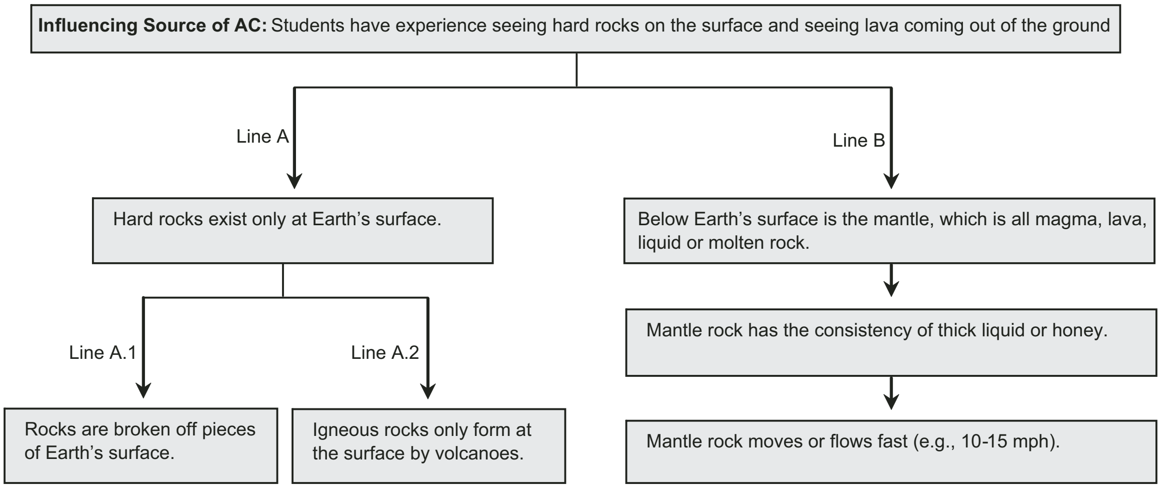


Figure 3. Cognitive model for what rocks are and their origins. Each of the alternate conceptions captured in a box is a raw student quote. Each arrow between alternate conceptions represents a direct connection made by one or more students.

broken off pieces) but not biochemical origins (e.g., carbonates, chlorides, and halides). Line A.2 indicates that students have a nascent grasp of igneous rocks that encompasses extrusive rocks but not intrusive rocks. Line B reflects logical reasoning and conclusions in a novice conceptualization, but these are incomplete compared to an expert conceptualization.

Density and Convection

One common line of reasoning regarding density and convection of the mantle that emerged during this study is summarized as follows:

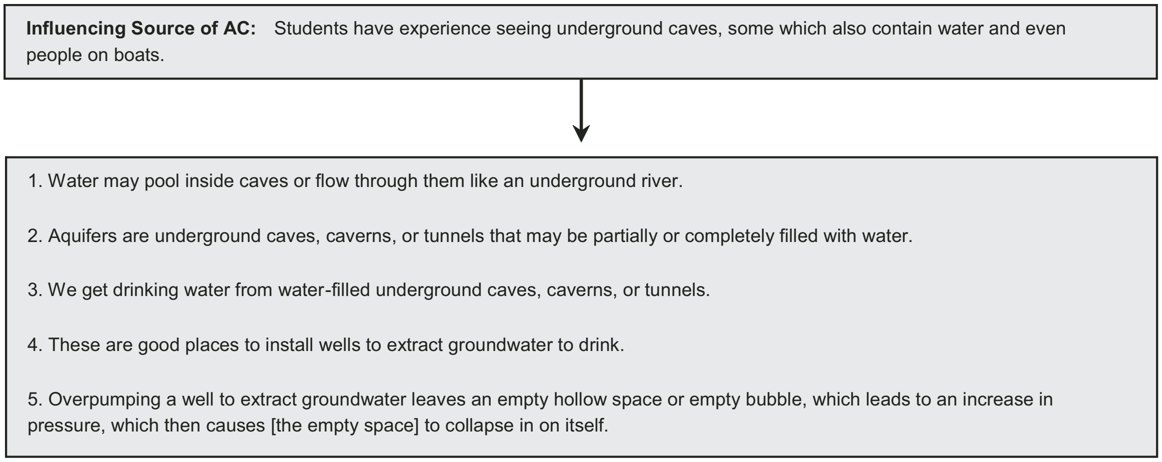
1. Students generally know that convection within the mantle means that something is moving in the mantle.
2. That something is molten rock.
3. Molten rock has the consistency of honey or thick liquid, and it moves or flows about as fast as honey does.
4. Most students know that convection is connected to something called density.
5. Although most students can say that convection and density are connected, they cannot explain exactly what density is and how it is connected to convection.

This line of reasoning likely resulted from a combination of previous experience and knowledge (as discussed previously) as well as an inadequate scientific grasp of the terms *convection*, *mantle*, *molten rock*, and *density*. For example, no students were able to correctly define *convection* and *density*. Furthermore, many students used *mantle*, *mantle rock*, *molten rock*, *magma*, and *lava* synonymously.

Water and Aquifers

Many student conceptions about rocks were influenced by their experiences, as were their conceptions about groundwater and aquifers. Most students have seen caves either in situ or in captured images (e.g., photos, television, etc.). This experience is extended into a common and comprehensive model about groundwater and aquifers, which is summarized as a flow chart in Figure 4.

Elements of this cognitive model are certainly scientifically accurate. For example, water can pool and flow through caves, and caves can collapse in on themselves, particularly so in karst environments. In the context of aquifers and drinking water, however, this cognitive model does not account for the



Influencing Source of AC: Students have experience seeing underground caves, some which also contain water and even people on boats.

1. Water may pool inside caves or flow through them like an underground river.
2. Aquifers are underground caves, caverns, or tunnels that may be partially or completely filled with water.
3. We get drinking water from water-filled underground caves, caverns, or tunnels.
4. These are good places to install wells to extract groundwater to drink.
5. Overpumping a well to extract groundwater leaves an empty hollow space or empty bubble, which leads to an increase in pressure, which then causes [the empty space] to collapse in on itself.

Figure 4. Cognitive model for the storage of groundwater in the subsurface and its connections to drinking water. The alternate conceptions numbered inside the figure are raw student quotes.

fact that (1) aquifers are composed of rock, not large cavernous open space, and (2) rock can actually contain water, via primary and secondary porosity.

Two sources of potential challenge to correcting this cognitive model emerged during this study and are connected to student thinking. The first was that many students did not, prior to instruction, think of rocks as continuous units below Earth's surface. Instead, they thought of rocks as simply being fragments of Earth's surface. Second, many students conceptualized rocks as being solids and, more specifically, solids without void spaces in them that can hold water.

Sources of Groundwater

Although a variety of ideas about the origins of groundwater was expressed, by far the most popular idea was that groundwater originates in the mountains. A notable omission was the idea that rain anywhere (not only in the mountains) can contribute to groundwater. Another popular idea was that "on a plain you'd be closer to groundwater than on a mountain." Integration of these student conceptions suggests the mental model depicted as a graphical representation in Figure 5A, where rain in the mountains flows straight down and fills the subsurface in such a way that the water table remains horizontal, as if a cup were being filled with water. For introductory-level geology courses, a standard conceptual framework that an instructor might target through learning goals includes (1) the water table generally follows the topography, and (2) groundwater flows in many directions and not just straight down (as depicted in Fig. 5B).

DISCUSSION

The investigations described herein were exploratory in nature and were intended as a means for obtaining qualitative data and a preliminary understanding of the alternate conceptions that college-level students hold. The qualitative findings of these investigations provide the foundation for further and more quantitative research into determining the prevalence of the specific alternate conceptions identified.

Data Collection Methods

The interviews yielded the most detailed data about students' alternate conceptions, and they were also the most time-intensive data collection method.

Student responses to the PNP questions were mixed with respect to their effectiveness at revealing not only alternate conceptions but also the reasoning behind them. Multiple-choice PNP questions with predefined answer choices followed by an open-ended question that asked for the reasons that the students chose a particular answer (MC-PNP-R) did not yield many responses regarding student reasoning (i.e., left blank). Furthermore, the responses that were given were generally not useful (e.g., "because"). PNP questions that proved more effective in probing student thinking were purely open-ended (OE-PNP). Overall, responses to PNP questions were less detailed than interview responses, and the detail and clarity of students' written responses ranged from being very informative to not informative at all. When responses were, for example, in the form of complete sentences that used the students' own language, they were very informative. On the other hand, when the responses were, for example, composed of a single geologic term (such as "convection"), then they were not as revealing about student thinking. Compared to interviews, however, more data on student thinking could be collected in a shorter amount of time using PNP questions. For an in-depth discussion and examples of material constituting useful and nonuseful responses, the readers are directed to Arthurs and Marchitto (this volume).

Class observations proved to be the least revealing method for identifying student ideas and thinking, mainly because there were relatively few instances where students volunteered such

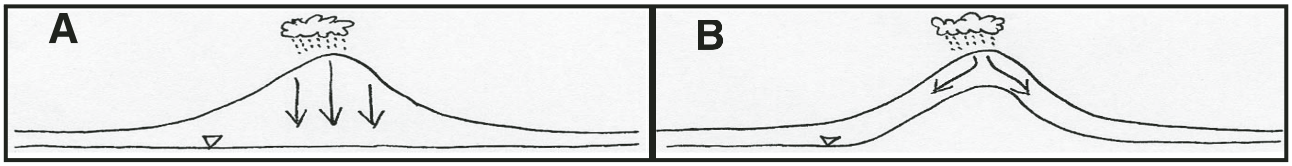


Figure 5. Cognitive models that explain the origin and movement of groundwater with respect to the water table. (A) Representation of hypothetical student mental model. (B) Representation of hypothetical instructional mental model. ∇ = water table.

information. As such, of the methods employed, their effectiveness at revealing insightful and high-quality data into student thinking can be summarized as follows, from the most effective to the least effective:

> interviews > OE-PNP questions
> > MC-PNP-R > MC-PNP > class observations.

Overall, however, OE-PNP questions proved to be the most fruitful for probing student thinking with respect to the quality of the answers provided, the insights gained into student thinking, and the number of students who could be surveyed. Interviews were even more fruitful with respect to answer quality and insights gained; however, much smaller numbers of students could be surveyed in this way. The main limiting factor on the number of interviews that were conducted was the number of students that volunteered to be interviewed.

Student Ideas and Thoughts

The designations of “preconception” and “misconception” were based on the most common occurrence among students and were also influenced by the timing of the interviews and PNP exercises with respect to the concepts that had already been covered in the course from which the students were surveyed. On a student-to-student basis, it was possible for preconception $L(c)$ to be corrected through instruction, for preconception $L(c)$ to remain intact despite instruction, which then resulted in it being referred to as misconception $L(c)$, and for misconception to $L(c)$ to emerge during the course of instruction (e.g., due to the student misinterpreting what the instructor intended to convey about a particular concept, c). The designation of “impression” was not applied to the concepts explicitly targeted in this paper; however, examples of impressions that arose during other work involved the Coriolis effect (e.g., direction in which a toilet flushes in different hemispheres) and the age of Earth.

The designations of “preconception” and “misconception” were based on data collected through interviews where students were asked where they had heard about c. Data about student thinking about c that were based only on responses to PNP questions did not include information about the origin(s) of student-held alternate conceptions. Further studies into the sources of student perceptions about geological concepts would facilitate curricular development for instructional approaches that utilize the elicit-confront-resolve strategy (Mazur, 1997) and contrasting cases (Schwartz and Bransford, 1998) to facilitate student learning. The elicit-confront-resolve strategy involves first eliciting student ideas and thinking, which may be alternate conceptions of one form or another. Next, it involves confronting these student-held conceptions with new or additional information, which may contradict their original ideas and cause cognitive dissonance. Finally, the cognitive dissonance is resolved through discussion to facilitate the integration of new knowledge. The use of contrasting cases is based on presenting students with different examples of the same concept in order to highlight integral features of that concept.

The qualitative findings of this study also indicate that college-level students hold some of the same alternate conceptions as students in elementary schools. For example, like some of the sixth graders in Ford’s study (2003), some of the college-level students defined rocks as being “made up of minerals” or as being “hard.” Like most of the sixth graders, these college-level students also primarily mentioned sedimentary and igneous processes to explain rock formation. Alternate conceptions about how rocks form were also discussed by Kortz and Murray (2009). Conceptualizations of groundwater are another area where these college students hold alternate conceptions similar to those of elementary or middle school students. For example, like the eighth graders in Dickerson and Dawkins’ study (2004), our students described groundwater as occurring in large underground open spaces. Our students used descriptions such as underground “caves,” “tunnels,” and “rivers,” and the eighth graders used descriptions that included “underground stream” and “underground pool.” Longitudinal studies within GER to track the conceptual development of individuals from early childhood into mature adulthood have not yet been done. Such studies, however, could yield findings useful in conceptualizing the evolution of cognition with respect to geoscience phenomena and informing instructional approaches.

The literature further indicates that alternate conceptions held by college-level students in this study are also held by college-level students elsewhere in the United States (e.g., Delaughter et al., 1998; McKenney and Webster, 2004; McConnell et al., 2005; Dickerson et al., 2005; Sibley, 2005). For example, in the study by Dickerson et al. (2005), students enrolled in introductory-level geoscience courses at a liberal arts university in the southwestern United States described groundwater as occurring in large underground open spaces such as “pools,” "lakes," "streams," and "rivers." These descriptions are similar to the descriptions of our students enrolled in the introductory-level geoscience courses at a research university in the west-central United States (as described previously). It is worth noting here that there is a dearth of interinstitutional studies into student alternate conceptions within the geosciences, and it is an area ripe for collaborative and interinstitutional research.

Although Delaughter et al. (1998) suggested that alternate conceptions may be common across regions, further research into the regional origins of college-level alternate conceptions is needed to (1) identify the alternate conceptions that appear to span regions (both nation-specific and internationally), (2) determine whether the apparent homogeneity in the existence of alternate conceptions at the college level across regions is due to student migration from their places of origin to hubs for college education, and (3) identify regionally defined or unique alternate conceptions that students might carry with them from their place of origin to their place of college education.

Student Cognitive Models and Ways of Thinking

Qualitative research into alternate conceptions often yields lists of identified misconceptions (e.g., Kusnick, 2002; Dahl et al., 2005; Rule, 2005; Dal, 2007), similar to the approach taken in the section of this paper titled Student Ideas and Thoughts in the Results section. Studies that investigate the completeness of $L(c)$ with respect to $E(c)$ are less common, and one such example within the geosciences comes in the work of McKenney and Webster (2004), where they discussed the completeness of student reasoning about magnets and magnetism. In their study, McKenney and Webster found that even students with what they considered robust cognitive models of magnetism had difficulty applying them correctly to account for the origin of seafloor banding.

In the Results section of the article, under the section titled Student Cognitive Models and Ways of Thinking, I presented cognitive models of student thinking about rocks, density and convection, water and aquifers, and sources of groundwater. These cognitive models were constructed based on the alternate conceptions identified during this study and the connections that students made between them. The cognitive models emerged from the data acquired. This was possible because of the qualitative approach used in this research. The data acquired presumably reflect only a portion of student-held ideas about *c* and also relate to only a limited number of connections with *c*. As such, the presented cognitive models are inherently incomplete. Nevertheless, they do act as vehicles to illustrate, at least, partial frameworks of student thinking that affect how students reason about geological concepts. The work done through this study lays the foundation for research that aims to more comprehensively describe and model ways of student thinking about specific geologic concepts and how they are interconnected.

Conceptualizing Students' Alternate Conceptions

Given that the ~40 yr discussion of students' alternate conceptions has taken place in diffuse forums from psychology (e.g., Vosniadou and Brewer, 1992) to education (e.g., Treagust, 1988) and to, increasingly, STEM disciplines, there does not yet exist a common agreed upon language with which to think about and discuss students' alternate conceptions. This paper presents a toolbox of key terminology and definitions useful for conceptualizing students' alternate conceptions, particularly with respect to their origins and the period of time that they are held relative to formal instruction during a given period of interest. In the following, I also discuss different ways of conceptualizing students' alternate conceptions. For the purposes of constructivist instruction, preferred models for conceptualizing student alternate conceptions (1) acknowledge the scope and variety of these alternate conceptions and (2) can be used to develop novice-like student thinking, $L(c)$, toward more expert-like ways of thinking, $E(c)$.

Traditionally, students' alternate conceptions have been conceptualized as being either correct or incorrect or falling on a spectrum from unscientific to scientific (Fig. 6) (e.g., Dahl et al., 2005; Rule, 2005). "Incorrect" alternate conceptions would naturally fall somewhere along the spectrum of "unscientific

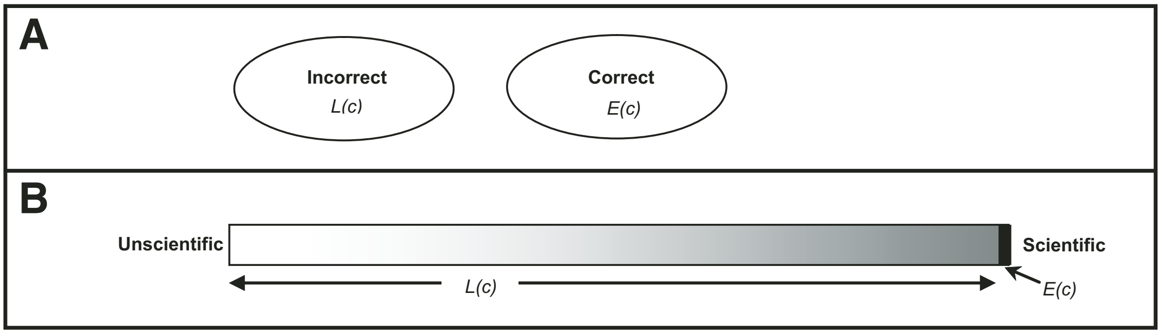


Figure 6. Discrete and continuous models for conceptualizing students' alternate conceptions. (A) Discrete correct-incorrect model. If $L(c) \neq E(c)$, $L(c)$ is incorrect. (B) Continuous unscientific-scientific model. Although $L(c) \neq E(c)$, $L(c)$ may nevertheless consist of some scientific aspects.

to scientific." However, a partially scientific although perhaps incomplete alternate conception would neither be a good fit for the "incorrect bin" nor the "correct bin." Thus, a discrete model for conceptualizing students' alternate conceptions is overly simplistic for understanding the relationship between $L(c)$ and $E(c)$. From an instructional standpoint, such an approach bars opportunities for scaffolding upon and bridging that which the student knows and that which the teacher would like him to know. Two limitations associated with conceptualizing student alternate conceptions with a continuous "unscientific to scientific" gradations model exist. The first is that it contains an element of "correct" and "incorrect" that could preclude instructional bridging and scaffolding efforts. The second lies in the fact that it is not readily apparent how unscientific a student's alternate conception is and where it might fall on such a continuum.

An additional limitation to these two models is the lack of recognition regarding evolution of thought: That which is conceived of as scientifically accurate and complete is not static and changes over time. (In this volume, Stokes also points out that scientifically acceptable ideas change through time.) Thus, it is possible for an alternate conception at a given reference point in time ($t = 0$) to have been a scientific conception held at a time prior to that reference point ($t = -1$). Geological examples of this can be found in the idea of continental drift and the theory of plate tectonics, which have evolved over time in light of additional information. Thus, a student today who envisions that the continents were torn apart by catastrophic events such as earthquakes and massive floods in fact holds one of the earliest scientific conceptions to explain the puzzle-like outlines of the continents (Dolgoff, 1998). In such cases, $L(c)_{t=0} = E(c)_{t=-1}$. The goals of instruction are for the student to develop conceptions where $L(c)_{t=0} = E(c)_{t=0}$ and to perhaps even begin imagining $E(c)_{t=1}$ (of the future).

The process of conceptualizing student ways of thinking as right-or-wrong or more-or-less scientific ignores the possibility that a student's alternate conceptions might be related to the history through which $E(c)$ evolved. I therefore propose conceptualizing student ways of thinking as being more or less ***scientifically accurate and complete*** with respect to $E(c)$ as defined not only by that generally accepted to be accurate and complete by the scientific community at $t = 0$, but also, perhaps more importantly, by the learning goals established by an instructor for a given course. For example, for an introductory-level course, the instructor might decide that it is not necessary for students to conceptualize certain phenomena in the most scientifically accurate or complete way currently ascribed to by the geologic community at large. However, the level of conceptualization will be sufficiently accurate and complete for achieving the desired learning goals. Thus, it is possible to constrain the extent to which $L(c)$ approaches $E(c)$.

A rudimentary model of conceptualizing student ways of thinking in this way can be illustrated along a spectrum (Fig. 7A). The more accurate and complete $L(c)$ is with respect to $E(c)$ aimed at through instruction, the closer to $E(c)$ it would fall. Thus, one might imagine the development of student thinking and their conceptual change to progress in a linear fashion. However, this, too, is overly simplistic given what we know about how students learn. That is, they create and integrate new knowledge in the context of the knowledge and experience that they already have, and, in the process, they create new links between ideas and develop new lines of reasoning. A radial model that (1) accounts for comparably accurate and complete but different lines of reasoning and alternate conceptions with respect to $E(c)$ and (2) represents a spectrum of accuracy and completeness would be an improvement (Fig. 7B). However, I hypothesize that a web model is a more useful representation because it accounts for the possibility of student thinking to jump from one line of reasoning to another in perhaps unexpected ways as they develop more accurate and complete conceptions of c. Figure 7C illustrates the web model in simplified form because not every link between every line of reasoning is illustrated.

Implications for Learning and Teaching Interventions

The investigations that were the subject of this paper were intended to provide a preliminary understanding of what college-level students think within the courses surveyed. However, the findings will also be relevant to geology and environmental science instructors at other institutions. Given that there exist documented similarities in several alternate conceptions at different colleges and even at different grade levels K–12, previously undocumented alternate conceptions that were presented here may aid instructors elsewhere. The findings from this study can inform instructors on how to improve their own teaching practices by using the findings to help: (1) identify these alternate conceptions in their population of students, (2) develop curriculum based on a constructivist approach that builds and scaffolds upon existing student experiences and knowledge (e.g., using the findings to develop elicit-confront-resolve strategies and examples of contrasting cases), and (3) develop formative and summative assessments, particularly of the multiple-choice type (e.g., concept inventory tests). In addition, the taxonomy of alternate conceptions terminology and the models for conceptualizing students' alternate conceptions presented here may be useful to instructors and researchers alike. Furthermore, if students are introduced to these ideas, then that knowledge might facilitate their metacognitive processes during learning.

CONCLUSION

Qualitative research methods including classroom observations and surveying techniques (interviews and written exercises) were used to conduct exploratory investigations into what college-level students think about geoscience concepts, their alternate conceptions, and their cognitive models. The findings provided a preliminary understanding of what students

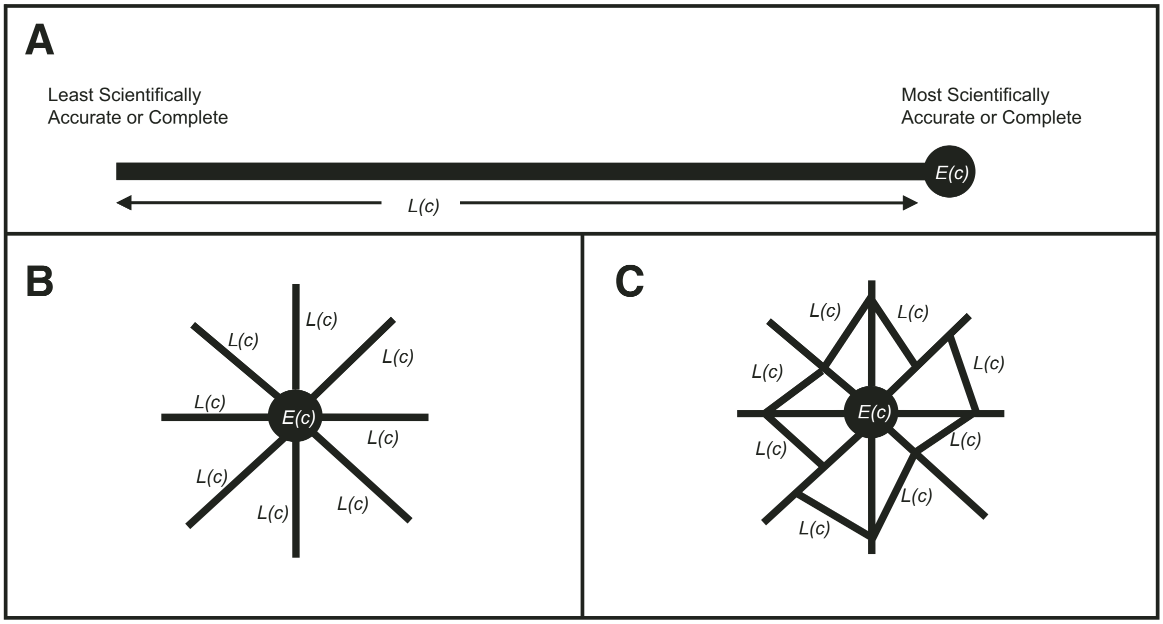


Figure 7. Continuum, radial, and web models for conceptualizing students' alternate conceptions. (A) Continuum model: Single line of reasoning upon which an alternate conception might fall. As student thinking evolves to become more scientifically accurate and complete, $L(c)$ moves toward $E(c)$. (B) Radial model: Multiple lines of reasoning upon which various alternate conceptions that are different but equally inaccurate and incomplete might fall. (C) Simplified web model: Connected lines of reasoning upon which a student's alternate conception might fall and jump from a place on one line to another place on any other line.

think with respect to three broad categories of geoscience concepts: rocks, density and convection, and water. The findings and preliminary understanding into how students think about these concepts and how we might conceptualize the evolution of student cognitive models of geoscience concepts have immediate applications to classroom instruction and curriculum development, particularly for those that employ a constructivist approach to teaching and learning. Furthermore, the findings provide the foundation for continued research at the college level in the area of student cognition of geoscience concepts to include, as previously mentioned, studies into the sources of students' alternate conceptions, the prevalence of specific alternate conceptions that were identified, the regional origins of alternate conceptions, the connections between various conceptions and the coherent cognitive models in which they rest, and research to test the robustness of the web model for conceptualizing student conceptualizations of geoscience concepts and the evolution of their conceptualizations toward higher accuracy and completeness.

ACKNOWLEDGMENTS

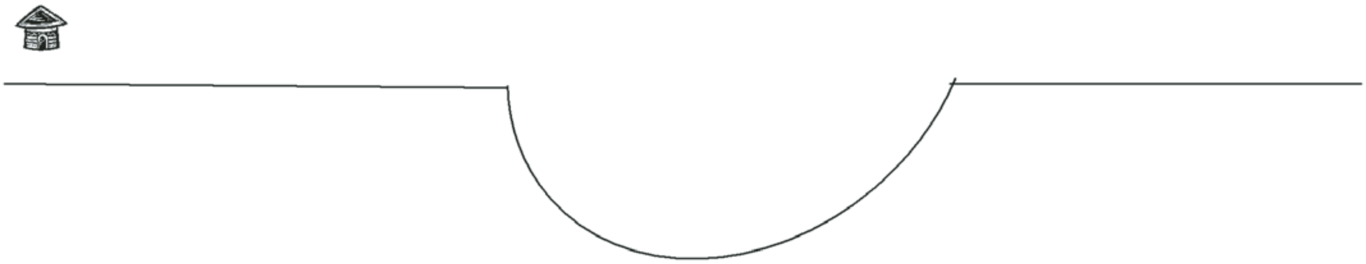
Funding and support for this project were provided by the Science Education Initiative at the University of Colorado at Boulder. Many thanks go to the instructors of the four courses from which the students were surveyed through interviews and/or paper questionnaires as well as to all the instructors that provided their expert input on various questions and concepts. Special thanks go to Alexis Templeton and Thomas Marchitto for welcoming me into their classrooms on a regular basis, our discussions about classroom observations and student thinking, and our collaborations in developing PNP exercises. A warm mahalo is extended to Barbara Fox for her linguistics expertise and scintillating discussion about language. I would also like to acknowledge and thank our undergraduate research assistants and the many students with whom I interacted during the course of this study. Finally, I am grateful to George Lakoff, whose work I was first introduced to more than a decade ago, for the inspiration behind this study.

APPENDIX 1. EXAMPLE OF PENCIL-AND-PAPER EXERCISE

Student Name ________________________________

Student Number______________________________

Question: The county of Appleton has decided to go through with building a new repository for nuclear waste. As a resident of the county, you have been given the opportunity to provide input into acceptable guidelines for its construction. Below you will find a blank cross-section of a possible site for the repository. The hole for the repository has already been dug, and a small temporary shelter has been built.



(1) Imagine that you can choose the kind(s) of rock formation(s) in which the hole was dug.
  a. **Draw** and **label** the layers that you think would be ideal to have dug the hole in.
  b. In the space below explain why you chose the rock type(s) that you did.

(2) Imagine that you can choose where the water table is.
  a. **Draw** and **label** a line to indicate where you'd place it.
  b. In the space below, explain why you placed it where you did.

(3) Imagine that you are responsible for designing the construction of the repository.
  a. **Draw** and **label** the components that you think would make the repository least likely to cause environmental harm.
  b. In the space below, explain why each component is important. If you need more space, feel free to write on the back side of this sheet.

APPENDIX 2. EXCERPT FROM FIELD NOTES

DATE: Tuesday, 15 January 2008.

Time 11:30 a.m.
Slide: Origins
Earth formed by accretion, pulling particles toward in a gravity, bombarded by meteorites. Early earth was at least semi-molten.
Water outgassing from volcanic vents → atm
Some water from comets
TQ: So, we have all this water. How do we get oceans?
MA: condensation
TQ: What do we need for condensation?
MA: a change in temp
TC: As Earth cooled, water condensed.

Time 11:33 a.m.
Slide: Origin of Life
TQ: Why was water good for early life?
MA: lack of gamma in atm not good
TQ: What about temp?
MA: too hot
MA: [Organisms] don't need solid structure in water.
TC: Water dissolves lots of nutrients.
Inorganics + sun energy → organics
We'll learn more about chemosynthesis later as well as photosynthesis.

NTS: Do they understand *chemical fractionation isotope*?

Key for Shorthand Used in Field Notes:

Slide—indicates which PowerPoint slide is being displayed at the time indicated above it
Q—indicates a question was asked
A—indicates an answer was given
C—indicates a comment was made
T—indicates the teacher is the source of the Q/A/C
M—indicates a male student is the source of the Q/A/C
F—indicates a female student is the source of the Q/A/C
NTS—indicates note to self

REFERENCES CITED

Arthurs, L., and Marchitto, T., 2011, this volume, Qualitative methods applied in the development of an introductory oceanography concept inventory survey, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(08).

Bereiter, C., and Scardamalia, M., 1986, Educational relevance of the study of expertise: Interchange, v. 17, no. 2, p. 10–19, doi:10.1007/BF01807464.

Bogdan, R.C., and Biklen, S.K., 1998, Qualitative Research for Education: An Introduction to Theory and Methods: Boston, Allyn and Bacon, 276 p.

Clement, J., 1993, Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics: Journal of Research in Science Teaching, v. 30, no. 10, p. 1241–1257, doi:10.1002/tea.3660301007.

Creswell, J., 1998, Qualitative Inquiry and Research Design: Choosing among Five Traditions: Thousand Oaks, California, Sage Publications, 403 p.

Dahl, J., Anderson, S.W., and Libarkin, J.C., 2005, Digging into earth science: Alternative conceptions held by K–12 teachers: Journal of Science Education, v. 6, no. 2, p. 65–68.

Dal, B., 2007, How do we help students build beliefs that allow them to avoid critical learning barriers and develop a deep understanding of geology?: Eurasia Journal of Mathematics: Science and Technology Education, v. 3, no. 4, p. 251–269.

Delaughter, J.E., Stein, S., Stein, C.A., and Bain, K.R., 1998, Preconceptions abound among students in an introductory earth science course: Eos (Transactions, American Geophysical Union), v. 79, no. 36, p. 429–436, doi:10.1029/98EO00325.

Dickerson, D., and Dawkins, K., 2004, Eighth grade students' understandings of groundwater: Journal of Geoscience Education, v. 52, no. 2, p. 178–181.

Dickerson, D., Callahan, T.J., Van Sickle, M., and Hay, G., 2005, Students' conceptions of scale regarding groundwater: Journal of Geoscience Education, v. 53, no. 4, p. 374–380.

Dixon, R.M.W., 1982, Where Have All the Adjectives Gone?: Berlin, Walter de Gruyter, 256 p.

Dolgoff, A., 1998, Physical Geology: Boston, Houghton Mifflin Company, 638 p.

Dove, J.E., 1998, Students' alternative conceptions in earth science: A review of research and implications for teaching and learning: Research Papers in Education, v. 13, no. 2, p. 183–201, doi:10.1080/0267152980130205.

Driver, R., 1981, Pupils' alternative frameworks in science: European Journal of Science Education, v. 3, p. 93–101.

Feig, A.D., 2011, this volume, Methodology and location in the context of qualitative data and theoretical frameworks in geoscience education research, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(01).

Ford, D.J., 2003, Sixth graders' conceptions of rocks in their local environments: Journal of Geoscience Education, v. 51, no. 4, p. 373–377.

Geer, J.G., 1991, Do open-ended questions measure "salient" issues?: Public Opinion Quarterly, v. 55, no. 3, p. 360–370, doi:10.1086/269268.

Gobert, J.D., 2000, A typology of causal models for plate tectonics: Inferential power and barriers to understanding: International Journal of Science Education, v. 22, p. 937–977, doi:10.1080/095006900416857.

Handelsman, J., Miller, S., and Pfund, C., 2007, Scientific Teaching: New York, W.H. Freeman and Company, 184 p.

Helm, H., 1980, Misconceptions in physics amongst South African students: Physics Education, v. 15, p. 92–105, doi:10.1088/0031-9120/15/2/308.

Henriques, L., 2005, Children's ideas about weather: A review of the literature: School Science and Mathematics, v. 102, p. 202–215; available online at http://www.cascience.org/csta/pub_misconceptionsjournal.asp, doi:10.1111/j.1949-8594.2002.tb18143.x.

Izquierdo-Aymerich, M., and Aduriz-Bravo, A., 2003, Epistemological foundations of school science: Science and Education, v. 12, p. 27–43, doi:10.1023/A:1022698205904.

Kastens, K.A., Liben, L.S., and Agrawal, S., 2009, How students and field geologists reason in integrating spatial observations from outcrops to visualize a 3-D geological structure: International Journal of Science Education, v. 31, p. 365–393, doi:10.1080/09500690802595797.

Kinchin, I.M., Hay, D.B., and Adams, A., 2000, How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development: Educational Research, v. 42, no. 1, p. 43–57, doi:10.1080/001318800363908.

Kortz, K.M., and Murray, D.P., 2009, Barriers to college students learning how rocks form: Journal of Geoscience Education, v. 57, p. 300–315.

Kurasaki, K.S., 2000, Intercoder reliability for validating conclusions drawn from open-ended interview data: Field Methods, v. 12, no. 3, p. 179–194, doi:10.1177/1525822X0001200301.

Kusnick, J., 2002, Growing pebbles and conceptual prisms—Understanding the source of student misconceptions about rock formation: Journal of Geoscience Education, v. 50, no. 1, p. 31–39.

Lakoff, G., 1987, Women, Fire, and Dangerous Things—What Categories Reveal about the Mind: Chicago, University of Chicago Press, 614 p.

LeCompte, M.D., and Schensul, J.J., 1999, Designing and Conducting Ethnographic Research: Walnut Creek, California, AltaMira Press, 220 p.

Libarkin, J.C., and Kurdziel, J.P., 2001, Research methodologies in science education: Assessing students' alternative conceptions: Journal of Geoscience Education, v. 49, no. 4, p. 378–383.

Libarkin, J.C., and Kurdziel, J.P., 2002, Research methodologies in science education: Qualitative data: Journal of Geoscience Education, v. 50, p. 195–200.

Libarkin, J.C., and Kurdziel, J., 2006, Ontology and the teaching of earth system science: Journal of Geoscience Education, v. 54, no. 3, p. 408–413.

Libarkin, J.C., Beilfuss, M., and Kurdziel, J.P., 2003, Research methodologies in science education: Mental models and cognition in education: Journal of Geoscience Education, v. 51, no. 1, p. 121–126.

Libarkin, J.C., Anderson, S.W., Dahl, J., Beilfull, M., and Boone, W., 2005, Qualitative analysis of college students' ideas about the Earth: Interviews and open-ended questionnaires: Journal of Geoscience Education, v. 53, no. 1, p. 17–26.

Libarkin's Geocognition Research Laboratory and collaborators, 2004, The Geoscience Concept Inventory, version 1: https://www.msu.edu/~libarkin/gci.html (accessed 29 November 2009).

Lindblom-Ylänne, S., and Lonka, K., 2001, Students' perceptions of assessment practices in a traditional medical curriculum: Advances in Health Sciences Education: Theory and Practice, v. 6, p. 121–140, doi:10.1023/A:1011422517238.

Mark, D.M., Freksa, C., Hirtle, S.C., Lloyd, R., and Tversky, B., 1999, Cognitive models of geographic space: International Journal of Geographical Information Science, v. 13, no. 8, p. 747–774, doi:10.1080/136588199241003.

Mason, J., 2002, Qualitative Researching: London, Sage Publications, 223 p.

Mayer, R.E., 1998, Learning and Instruction: Upper Saddle River, New Jersey, Merrill Prentice Hall, 516 p.

Mazur, E., 1997, Peer Instruction: A User's Manual: Upper Saddle River, New Jersey, Prentice Hall, 253 p.

McConnell, D.A., Steer, D.N., Owens, K.D., and Knight, C.C., 2005, How students think: Implications for learning in introductory geoscience courses: Journal of Geoscience Education, v. 53, no. 4, p. 462–470.

McKenney, R., and Webster, J., 2004, Magnetism, the Earth as a magnet, and seafloor banding—How much magnetism is enough?: Journal of Geoscience Education, v. 52, no. 4, p. 352–362.

McNeal, K.S., Miller, H.R., and Herbert, B.E., 2008, Developing non-science majors' conceptual models of complex Earth systems in a physical geology course: Journal of Geoscience Education, v. 56, p. 201–211.

Mortimer, E.F., 1995, Conceptual change or conceptual profile change?: Science and Education, v. 4, p. 267–285, doi:10.1007/BF00486624.

National Research Council, 2000, How People Learn: Brain, Mind, Experience, and School: Washington, D.C., National Academy Press, 374 p.

National Society for the Study of Education, 1995, Improving Science Education: Chicago, National Society for the Study of Education, 233 p.

Nield, A.F., and Wintre, M.G., 1986, Multiple-choice questions with an option to comment: Student attitudes and use: Teaching of Psychology, v. 13, p. 196–199, doi:10.1207/s15328023top1304_6.

Norman, D.A., 1983, Some observations on mental models, *in* Gentner, D., and Stevens, A., eds., Mental Models: Hillsdale, New Jersey, Lawrence Erlbaum Associates, p. 7–14.

Novak, J.D., 1977, A Theory of Education: Ithaca, New York, Cornell University Press, 297 p.

Office of Planning, Budget, and Analysis, 2009, Quick Facts: Demographics: http://www.colorado.edu/pba/qfacts/demog.html (accessed 6 November 2009).

Redish, E.F., 2003, Teaching Physics with the Physics Suite: Hoboken, New Jersey, John Wiley & Sons, 216 p.

Reiner, M., Slotta, J.D., Chi, M.T.H., and Resnick, L.B., 2000, Naïve physics reasoning: A commitment to substance-based conceptions: Cognition and Instruction, v. 18, no. 1, p. 1–34, doi:10.1207/S1532690XCI1801_01.

Rosch, E., 1975, Cognitive reference points: Cognitive Psychology, v. 7, p. 532–547, doi:10.1016/0010-0285(75)90021-3.

Rosch, E., 1981, Prototype classification and logical classification: The two systems, *in* Scholnick, E., ed., New Trends in Cognitive Representation: Challenges to Piaget's Theory: Hillsdale, New Jersey, Lawrence Erlbaum Associates, p. 73–86.

Rule, A., 2005, Elementary students' ideas concerning fossil fuels: Journal of Geoscience Education, v. 53, no. 3, p. 309–318.

Sapsford, R., 1999, Survey Research: London, Sage Publications, 258 p.

Schensul, S.L., Schensul, J.J., and LeCompte, M.D., 1999, Essential Ethnographic Methods: Observations, Interviews, and Questionnaires: London, AltaMira Press, 318 p.

Schwartz, D.L., and Bransford, J.D., 1998, A time for telling: Cognition and Instruction, v. 16, no. 4, p. 475–522, doi:10.1207/s1532690xci1604_4.

Scouller, K., 1998, The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay: Higher Education, v. 35, no. 4, p. 453–472, doi:10.1023/A:1003196224280.

Sell, K.S., Herbert, B.E., Stuessy, C., and Schielack, J., 2006, Supporting student conceptual model development of complex earth systems through the use of multiple representations and inquiry: Journal of Geoscience Education, v. 54, p. 396–407.

Shepardson, D.P., Harbor, J., and Wee, B., 2005, Water towers, pump houses, and mountain streams: Students' ideas about watersheds: Journal of Geoscience Education, v. 53, no. 4, p. 381–386.

Sibley, D.F., 2005, Visual abilities and misconceptions about plate tectonics: Journal of Geoscience Education, v. 53, p. 471–477.

Sibley, D.F., 2009, A cognitive framework for reasoning with scientific models: Journal of Geoscience Education, v. 57, p. 255–263.

Stokes, A., 2011, this volume, A phenomenographic approach to investigating students' conceptions of geoscience as an academic discipline, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(03).

Stone, A.A., Turkkan, J.S., Bachrach, C.A., Jobe, J.B., Kurtzman, H.S., and Cain, V.S., eds., 2000, The Science of Self-Report: Implications for Research and Practice: Mahwah, New Jersey, Lawrence Erlbaum Associates, 380 p.

Treagust, D.F., 1988, Development and use of diagnostic tests to evaluate students' misconceptions in science: International Journal of Science Education, v. 10, no. 2, p. 159–169, doi:10.1080/0950069880100204.

Vosniadou, S., and Brewer, W.F., 1992, Mental models of the Earth: A study of conceptual change in childhood: Cognitive Psychology, v. 24, p. 535–585, doi:10.1016/0010-0285(92)90018-W.

Wieman, C., 2007, Why not try a scientific approach to science education?: Change, v. 39, no. 5, p. 9–15.

Wieman, C., and Perkins, K., 2005, Transforming physics education: Physics Today, v. 58, no. 11, p. 36–41, doi:10.1063/1.2155756.

Willis, J.W., 2008, Qualitative Research Methods in Education and Educational Technology: Charlotte, North Carolina, Information Age Publishing, 371 p.

Wolcott, H., 1994, Writing up Qualitative Research: Newbury Park, California, Sage Publishing, 94 p.

Yeaton, W.H., and Wortman, P.M., 1993, On the reliability of meta-analytic reviews: The role of intercoder agreement: Evaluation Review, v. 17, no. 3, p. 292–309, doi:10.1177/0193841X9301700303.

MANUSCRIPT ACCEPTED BY THE SOCIETY 23 JUNE 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

Understanding, evaluation, and use of climate forecast data by environmental policy students

Toru Ishikawa*
Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

Anthony G. Barnston
International Research Institute for Climate and Society, Columbia University, Palisades, New York 10964, USA

Kim A. Kastens
Patrick Louchouarn†
Lamont-Doherty Earth Observatory & Department of Earth and Environmental Sciences, Columbia University, Palisades, New York 10964, USA

ABSTRACT

This study empirically examined understanding, evaluation, and use of climate forecasts by prospective policy makers through a quantitative analysis of their interpretation of forecast information and a qualitative analysis of their decision making in a situated condition. Concerning the understanding of forecast data, results showed that people had difficulty in understanding the probabilistic nature of three-category forecasts. In particular, a misunderstanding relating to the category having the highest likelihood versus the other two categories was observed. Concerning evaluation, results showed that forecast users demanded high accuracy, in fact accuracy higher than is possible at the current state of the science in climate forecasting. Many participants did not positively evaluate the quality of forecasts or were reluctant to use the forecasts until the level of agreement between the forecasts and observations was better than is typically possible. In addition, different attitudes toward forecasts among different individuals led to conflicts during group decision making about water allocation for farming. Many participants disregarded the forecasts of precipitation and relied more on historical data showing changes in the reservoir levels for the past 20 yr. Furthermore, people's decisions about whether to consider the forecasts in water allocation tended to be influenced by whether the forecasts supported or undercut their perceived self-interest or predetermined positions. In using the results of this study to improve instruction, the concept of probability and the inevitable existence of uncertainty in forecasts emerged as two key issues.

---

*Current address: Graduate School of Interdisciplinary Information Studies & Center for Spatial Information Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; e-mail: ishikawa@csis.u-tokyo.ac.jp.
†Current address: Department of Oceanography, Texas A&M University Galveston, 1001 Texas Clipper Rd., Bldg 3029, OCSB 280D, Galveston, Texas 77554, USA.

Ishikawa, T., Barnston, A.G., Kastens, K.A., and Louchouarn, P., 2011, Understanding, evaluation, and use of climate forecast data by environmental policy students, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 153–170, doi:10.1130/2011.2474(11). For permission to copy, contact editing@geosociety.org.

INTRODUCTION

Weather and climate have long been of major concern to people because of their potential impact on human activities. People may suffer from food shortages owing to droughts, lose their lives by floods, and have their health threatened by climate-dependent infectious diseases such as malaria. People's lives may also be affected economically by water shortages and surpluses, in terms of agriculture and infrastructure (e.g., flooding of the Mississippi in 1993, hurricanes in Florida in 2004 and in Louisiana in 2005). It would be beneficial if climate conditions having such large effects on human lives could be forecast sufficiently far in advance to allow time for appropriate planning and preventive action. Weather forecasts do this but only on a short time scale (up to 10 d). Scientists in climate prediction and diagnostics have recently had some success in forecasting on a longer time scale, up to several months in advance.

Climate forecasters now routinely issue seasonal forecasts for temperature and precipitation up to several months ahead of time, so that people can make important decisions that are sensitive to future climate conditions. Potential benefits of seasonal to interannual climate forecasts have been widely documented in various domains, such as agricultural production (Hammer et al., 2001), water management (Kirshen, 2002), and epidemic intervention (Thomson et al., 2000, 2005). For example, Cane et al. (1994) showed that more than 60% of variance in maize yield in Zimbabwe was explained by an index of the El Niño–Southern Oscillation (ENSO), and that model-derived predictions of ENSO provided fairly accurate forecasts of maize yield. Jones et al. (2000) estimated that potential economic benefits of adjusting crop mix or maize management to ENSO-based seasonal climate forecasts in the Southeast U.S. could be $5–10 million per year.

Such potential benefits of climate forecasts, however, cannot be fully realized unless the climate forecasts are understood and used appropriately by the people whose lives and livelihoods are affected by climate variability. Some researchers discussed the difficulty that people have in interpreting and applying climate forecasts in decision making and argued for the necessity of systematically examining how to communicate climate forecasts and their expected accuracy, so that people can take advantage of their potential benefits (e.g., Hansen, 2002; Hartmann et al., 2002; Nicholls, 1999; Patt and Schröter, 2008; Pfaff et al., 1999). In fact, Changnon et al. (1995) and Rayner et al. (2002) reported that seasonal climate forecasts have not historically been very widely used in actual decision making by policy makers and stakeholders.

The process of climate forecasting for societal use consists of several distinct elements or stages: Earth's climate system, one or more computer-based climate prediction models, forecasters, forecast products, and forecast users. Forecasters issue seasonal climate forecasts on the basis of their scientific understanding of the climate system, and users interpret the forecasts and make various decisions that are sensitive to future climate conditions, such as maize or water-resource management. This paper focuses on the latter half of this climate forecasting process—that is, the ways in which potential users understand and evaluate the forecasts and decide whether or not to use forecast information for practical decision making.

Due to the chaotic nature of the atmosphere, seasonal climate forecasts often have only fair to moderate accuracy (Barnston et al., 2010; Goddard et al., 2003) and are therefore necessarily given in probabilistic terms (Goddard et al., 2001). In other words, climate forecasters feel it imperative to convey, in a probability format, the existence of uncertainty inherent in climate forecasts. A currently standard format of seasonal forecasts is tercile-based, showing above-, near-, and below-normal probabilities (Hansen et al., 2004). Although this three-category format is now employed by many forecast centers, it does not seem to be readily interpretable or usable by all of the target audience, as suggested by Hansen et al. (2004), Ishikawa et al. (2005), and Patt and Gwata (2002). Pagano et al. (2002) discussed four characteristics of forecast products that affect their use in decision making: accessibility (extent of dissemination), interpretability (ease of understanding), credibility (assessment of accuracy), and relevance (meaningfulness to decisions to make). This study focuses on interpretability and credibility, using these known challenges as a vehicle to probe how students incorporate complex and incomplete information into their decision making.

Specifically, we conducted empirical studies aimed at examining the extent to which students in training for environmental policy-making professions understand seasonal climate forecasts, and the ways in which they use climate forecasts in the process of decision making as individuals and as a group. To do that, we used precipitation forecast maps issued by the International Research Institute for Climate and Society (IRI), and examined the ways in which students understand, evaluate, and use them. Our earlier study (Ishikawa et al., 2005) found that environmental-policy students had difficulty in understanding and evaluating probabilistic forecasts correctly. This paper reports on further studies that quantitatively and qualitatively examined students' understanding and use of climate forecasts, in the contexts of forecast-quality evaluation and group decision making.

Participants in this research were studying hydrology and water resources in a professional master's degree program that prepares practitioners for careers in environmental policy. Most geoscience-education research to date has focused on K–12 settings, undergraduates enrolled in general-education courses, or undergraduate earth and environmental science majors. However, there exist a suite of programs in which students study Earth and the environment in preparation for their intended profession without the goal of becoming earth scientists, including environmental law, environmental health sciences, environmental engineering, environmental journalism, and environmental policy. The motivation structure, life experience, and prior knowledge of students in such programs are likely to differ from K–12 or undergraduate students. Moreover, the education of practitioners involves not only cognitive learning goals (what students should know and understand after instruction), but also connotative learning goals (how students should act or strive to act after instruction). Thus, the thinking and learning of such students are fertile ground for educational research.

As methods, this research used both quantitative and qualitative approaches. Our previous study (Ishikawa et al., 2005) had revealed some difficulties in understanding probabilistic climate forecasts and some reluctance to use them for societally important decision making. The quantitative component of this research probed more deeply into the question of students' willingness or reluctance to use climate forecasts in decision making as a function of the accuracy of the forecasts. Many students indicated that they would require the forecasts to be of higher accuracy than can be achieved by state-of-the-art climate forecasting before they would feel inclined to use them in decision making. This is an alarming finding for those who wish to see environmental decision making grounded in science. Yet, these findings could be questioned because the circumstances of the forecast evaluation were artificial, with hypothetical, paper-and-pencil questions that were disconnected from an actual decision-making situation.

Therefore, in the qualitative component of this research, we strove to create a more realistic decision-making situation. Specifically, we ran a role-playing exercise, in which participants were asked to conduct group discussions about agricultural water allocation wherein each member took a position of a farmer, local representative, or federal representative. We then conducted follow-up group interviews and analyzed their verbal protocols to evaluate their thought processes in undertaking a set of environmental decision-making tasks, both as individuals and as a group of stakeholders.

QUANTITATIVE STUDY: JUDGMENT OF FORECAST QUALITY

Method

Participants

Fifty-three students (20 men and 33 women) in the Master of Public Administration (MPA) program in Environmental Science and Policy at Columbia University's School of International and Public Affairs participated in the study. This intensive and multidisciplinary 1 yr program brings together people who seek leadership positions in local, state, and federal government agencies, as well as in nonprofit organizations and the environmental divisions of private corporations. Throughout the course of the program, the students are trained as environmental professionals who have an understanding of the science of environmental issues and use this knowledge to develop better management, analytical, and communication skills in the process of policy formulation.

Every year, the program attracts students from a diversity of backgrounds; approximately half of them come from traditional science and technology (chemistry, biology, environmental sciences, or engineering), and half come from social sciences and humanities (political sciences, economics, literature, health sciences, or journalism). At least 2 yr of relevant professional experience is normally required for admission to this program. A typical age distribution of the students for this program has a median age of 26 yr, with first and third quartiles of 24 and 30 years, respectively. In general, the student body consists of 25%–30% foreign students coming from a diverse suite of countries (Europe, Asia, Africa, or Latin America). No specific data were collected on ethnicity, but at least two thirds of the students were Caucasian.

After completing the program, the students are now working as active and responsible policy makers. Of the ~100 graduates of the program so far, 34% are now working in government (e.g., Environmental Protection Agency, National Aeronautics and Space Administration, Department of Transportation, Bureau of Land Management); 30% work in the private sector (e.g., director of environmental services, water-resources engineers, environmental policy analysts and consultants); 24% work in national and international nonprofit organizations or nongovernmental organizations (e.g., water-resources manager, environmental project manager and coordinator, development director, operations manager); and 12% are involved in education. These data show that the vast majority of students in the program enter into policy- or decision-making careers right after graduation.

This audience is important because they will be using geoscience to inform policies aimed at societally important problems. They differ from students in the typical upper-level geoscience course because their intended career path will position them as users rather than creators of geoscience data and insights. Thus, we wished to understand not only to what extent students understand geoscientists' visualizations, but also the ways in which they evaluate the credibility and usefulness of the presented information.

Materials

Participants viewed a set of precipitation forecast maps issued by the IRI, printed in color on 8.5 × 11 in. (21.6 × 27.9 cm) paper, one map per page. The first map (Fig. 1A) conveys a forecast of precipitation for a 3 mo season beginning with the month following the issuance of the prediction. For areas where the model does make a prediction, color is used to indicate the category having the highest likelihood among three categories. *Above normal* refers to precipitation amounts falling within the wettest one-third (or tercile) of the years in a 30 yr database of precipitation observations, *normal* indicates precipitation within the center tercile, and *below normal* indicates precipitation within the driest tercile. Within the above- and below-normal categories, color gradations are used to indicate the probability of the prediction. White areas indicate equal chances (33.3%) that the precipitation will be above, near, or below normal (i.e., climatological probabilities). Small bar graphs superimposed upon the colored map indicate the probabilities of precipitation falling within each of the three categories.

The other map (Fig. 1B) shows the observed precipitation during that 3 mo period. The observed precipitation is shown as "precipitation anomaly," which is defined as the ratio of the

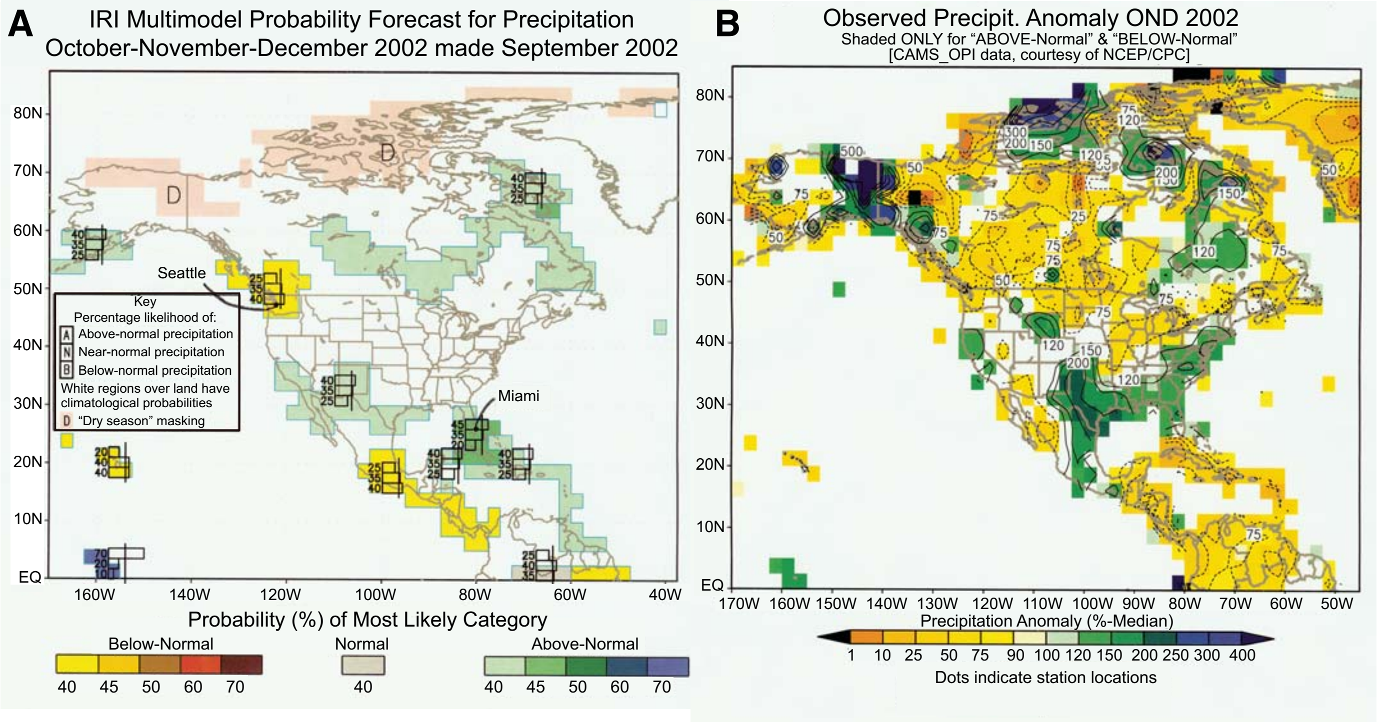


Figure 1. (A) Precipitation forecast map issued by the International Research Institute for Climate and Society. Colors indicate the probability for the most likely category among above-, near-, and below-normal. Small bar graphs superimposed on the colored areas show probabilities for the three categories. For example, the yellow color over Seattle, Washington, shows that below-normal has the highest forecast probability, and the bar graphs indicate that there is a 25%–35%–40% chance of above-, near-, and below-normal precipitation. (B) Map showing the observed precipitation in terms of precipitation anomaly for the 3 mo forecast period in part A. For example, blue regions received three times as much, and yellow regions half as much, as the median precipitation for those regions in the 30 yr database.

observed precipitation at each location during that specific 3 mo period to the median precipitation in the 30 yr database multiplied by 100 to give percentage of normal.

We varied the levels of accuracy of the forecasts shown to participants, so that we could examine the level of quality that people demand for climate forecasts. We chose a pair of real forecast and observation maps (Fig. 1), and then we constructed "artificial" forecast maps having varying spatial correlations with the observed patterns of precipitation, while preserving realism in the scales of spatial coherence of the predicted patterns (Figs. 2A–2D). In other words, the predicted patterns on the constructed forecast maps were designed to conform to their typical spatial scales, based on the large-scale atmospheric circulation patterns that affect departures from the climatological amounts of precipitation, but the phasing of these patterns with those observed was allowed to vary, ranging progressively from those yielding noticeable negative correlations to those yielding noticeable positive correlations. The ranked probability skill scores (RPSS) for the five forecast maps were −4.6, −1.7, 1.7, 5.3, and 9.7, and we refer to them as poor (Fig. 2A), real (Fig. 1A), typical (Fig. 2B), very good (Fig. 2C), and excellent (Fig. 2D) forecasts, respectively. The RPSS is a measure of forecast accuracy with its probabilistic nature taken into account; for details, see Goddard et al. (2003).

Test Questions and Procedure

The study activities took place during a regularly scheduled class session for the environmental-sciences module of the master's degree program (Louchouarn was the course instructor). Participants answered all questions individually.

Question 1 assessed participants' understanding of the standard IRI precipitation forecast map (Fig. 1A). It asked participants to give the probabilities that Seattle, Washington, and Miami, Florida, will receive more precipitation and less precipitation than normal. Question 2 assessed participants' understanding of the observation map of precipitation for a specific forecast period (Fig. 1B). It asked participants to identify a region that may have suffered drought conditions and a region that may have suffered flood damage during this forecast period. These questions replicate Ishikawa et al.'s study (2005), where more details about the questions may be found.

Question 3 examined participants' evaluation and use of precipitation forecasts by showing five different forecast maps. Participants viewed the real forecast map (Fig. 1A) first, and

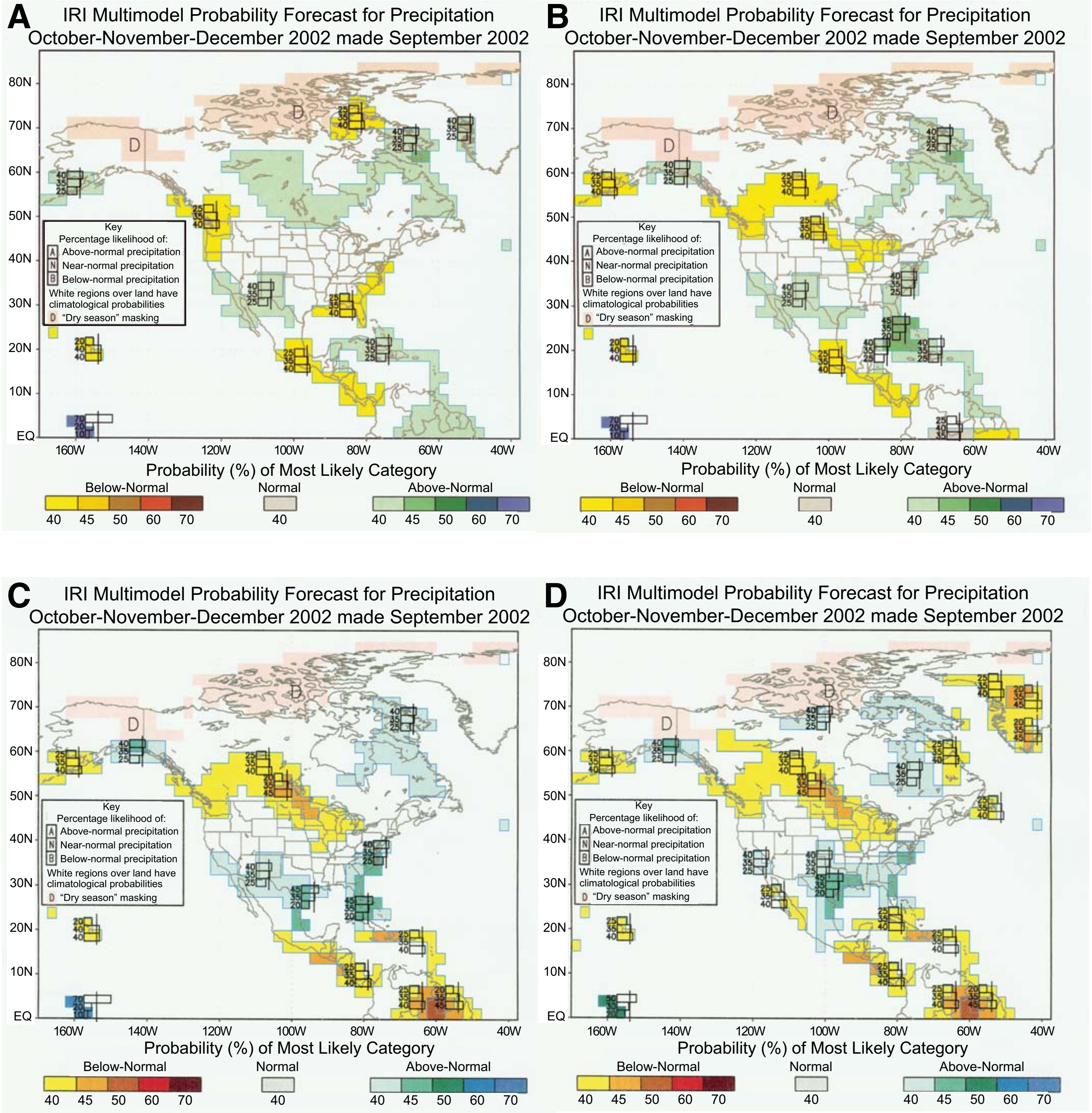


Figure 2. Forecast maps artificially constructed with varying degrees of agreement with the observed precipitation map in Figure 1B, respectively called (A) poor, (B) typical, (C) very good, and (D) excellent forecasts. The forecast map in Figure 1A was the real forecast issued for this 3 mo period.

then the remaining four forecast maps (Figs. 2A–2D) in random order, without being allowed to turn back to look at previous maps. For each of the five forecast maps, participants were asked (1) to evaluate the degree of its correspondence to the observation map on a 5 point scale (1—they tend to be opposite; 2—they are unrelated to each other; 3—they agree only slightly; 4—they agree somewhat; 5—they agree quite closely); and (2) to indicate whether they recommend using these forecasts to make decisions about which crops to plant on a 5 point scale (1—strong recommendation for using the forecasts; 2—weak recommendation for using the forecasts; 3—no recommendation in either way; 4—weak recommendation for not using the forecasts; 5—strong recommendation for not using the forecasts). We refer to the former as an evaluation question and the latter as an action question.

Results

Questions 1 and 2: Understanding of Forecasts

When asked to interpret the observation map, the majority of participants correctly identified regions for drought condition (94%) and flood damage (93%). When asked to interpret the standard forecast map, most participants (87%) correctly specified the probability that Seattle, Washington, would receive less precipitation than normal, whereas only half the participants (51%) correctly specified the probability that the city would receive more precipitation than normal. This pattern was reversed for Miami, Florida: 85% of the participants correctly answered the above-normal probability, whereas 59% correctly answered the below-normal probability (Fig. 3A). The difference in participants' performance on the four forecast-map questions was statistically significant (Cochran test, $p < 0.001$).

In summary, although almost all participants understood the observation map correctly, many had difficulty understanding the forecast map, particularly the above-normal probability for a city for which the most likely forecast category was below normal, and the below-normal probability for a city for which the most likely forecast category was above normal. These findings are consistent with Ishikawa et al.'s (2005) findings (shown in Fig. 3B). When participants can read the probability directly from the color scale, over 80% answer correctly, but when they must refer to the superimposed bar graphs, performance drops below 60%.

Question 3: Evaluation and Use of Forecasts

Participants' responses to the evaluation and action questions for the five maps are shown in Figure 4. For both questions, participants' ratings increased on average as the objective measure of agreement between the forecast and observation maps increased, $F(4, 45) = 53.85$ and $F(4, 46) = 41.35$, respectively ($p < 0.001$), with a significant cubic trend, $F(1, 48) = 9.60$, $p < 0.01$, and $F(1, 49) = 17.64$, $p < 0.001$.

For both evaluation and action, there were significant gaps in mean ratings between the poor and real forecasts, between the typical and very good forecasts, and between the very good and excellent forecasts (Bonferroni, $\alpha = 0.05/4$). For the real and typical forecasts, most participants chose "agree only slightly" or "agree somewhat" for evaluation (mean ratings of 3.4 and 3.6, respectively), and chose either "no recommendation in either way" or "weak recommendation for using the forecasts" for action (mean ratings of 2.5). Even for the very good forecast, many participants chose "agree somewhat" for evaluation (mean rating of 4.2), and "weak recommendation for using the forecasts" for action (mean rating of 2.0). Only when the forecast was excellent did participants evaluate it as "agree quite closely" (mean rating of 4.7) and become inclined to use it by choosing "strong recommendation for using the forecasts" (mean rating of 1.4).

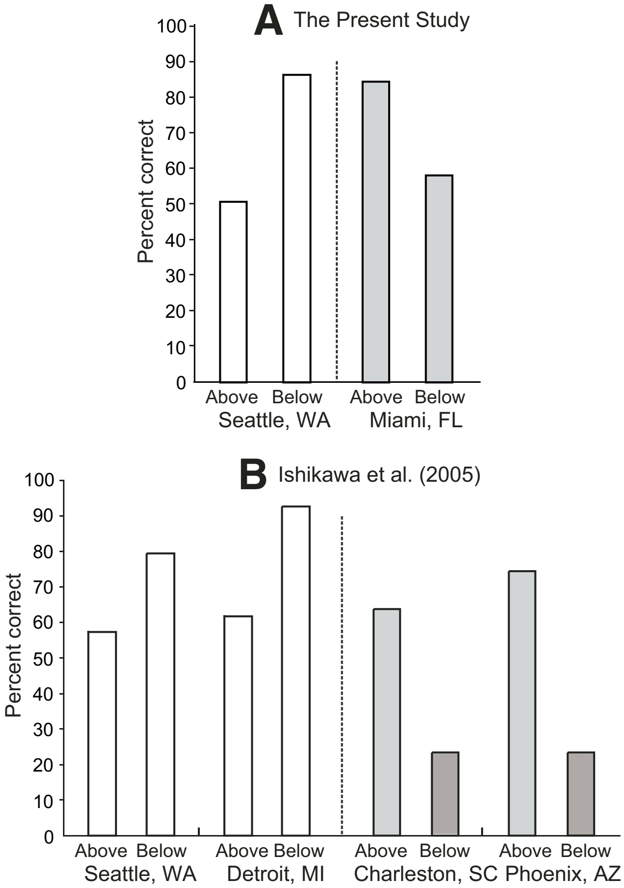


Figure 3. (A) Participants' performance on the questions about the above- and below-normal probabilities for Seattle, Washington, and Miami, Florida. (B) Results from Ishikawa et al.'s (2005) study, in which Charleston, South Carolina, and Phoenix, Arizona, had the highest likelihood for the above-normal category, and Seattle, Washington, and Detroit, Michigan, had the highest likelihood for the below-normal category. In both panels, data for below-normal cities are shown by white bars, and data for above-normal cities are shown by gray bars. Note that in both studies, performance was good when participants could read values directly from the color scale, but performance was poorer when they had to refer to the superimposed horizontal bar graphs.

At the individual level, there was large variability in evaluation and action ratings. For each map, participants' responses were scattered widely along the scales, especially for the action scale (Fig. 4). When each participant's evaluation and action ratings of the five maps and the RPSS measure of the forecasts were compared in terms of rank-order correlation, the values showed a wide distribution ranging from 0 to 1 (Fig. 5). However, at the same time, participants' evaluation and action ratings were significantly correlated ($r = -0.79$, $p < 0.001$; Fig. 6), showing that for specific maps, participants tended to be consistent in their evaluation and action ratings. That is, some people had difficulty visually estimating spatial correlation between the forecast and observation maps, and their rank ordering of the various forecast maps was close to random (at the left end of Fig. 5A), but once they had made a commitment about the quality of a specific forecast, they carried that commitment over to the action question. Figure 6 also depicts the extent of individual differences in participants' risk-taking or risk-averse tendencies: For a specific evaluation rating, their action ratings were distributed widely, especially for the middle three evaluation ratings.

We also note that many participants' rank orderings of the maps had high correlations with the objective measure of the maps' forecast quality (see Fig. 5). For example, correlations higher than 0.89 mean that the ratings for the five maps were monotonically increased as the objective measure increased. This is especially

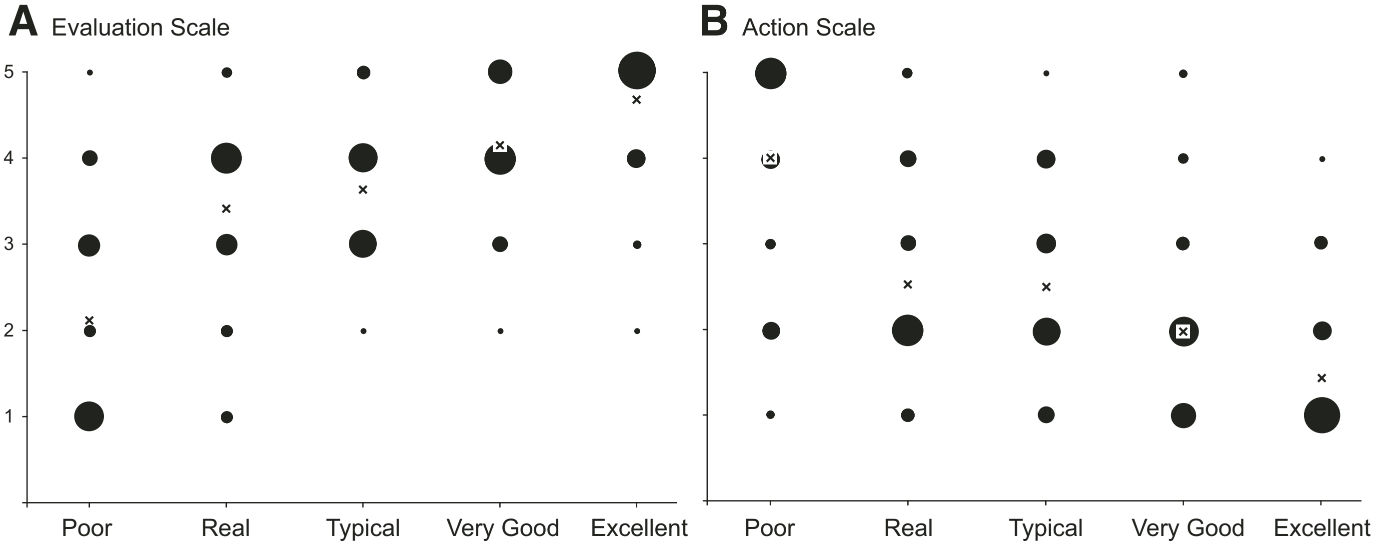


Figure 4. Participants' responses to the (A) evaluation and (B) action questions for each of the five forecast maps. The size of a circle is proportional to the number of participants. X's show the mean rating for each map. Note the high quality that participants demanded for climate forecasts, and the existence of large individual differences in ratings.

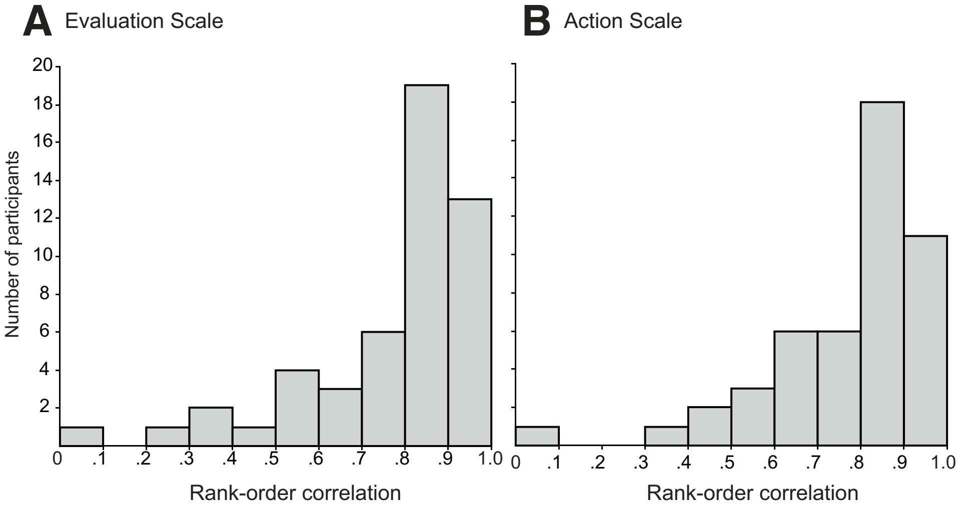


Figure 5. Distributions of rank-order correlations between the objective measure of the quality of forecasts and each participant's responses to the (A) evaluation and (B) action questions. Higher correlations mean better correspondence with the objective measure of forecast accuracy. Note that there were many high correlations but that the values ranged widely from 0 to 1.

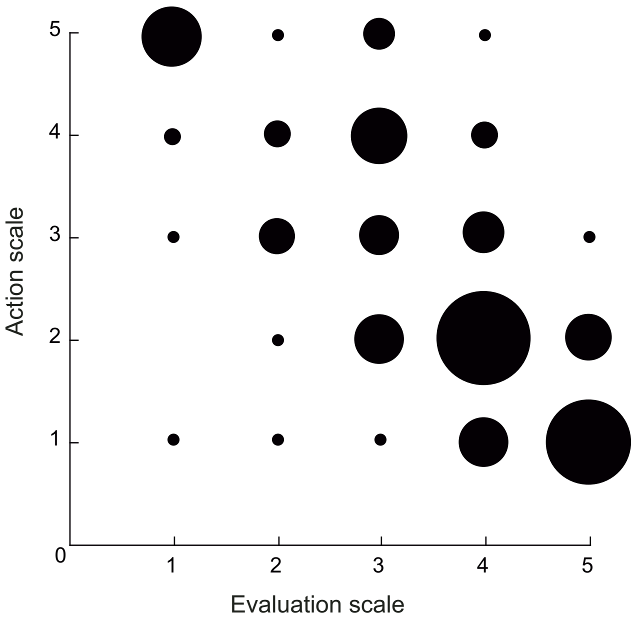


Figure 6. Relationship between participants' evaluation and action ratings for all the five maps combined. The size of a circle is proportional to the number of participants. Note that participants tended to be consistent in their evaluation and action ratings for specific maps. The vertical scatter of the graph shows the extent of individual differences in participants' risk-taking or risk-averse tendencies.

notable if we consider the fact that participants viewed the maps in random order and were not allowed to look back at previously viewed maps for comparisons.

QUALITATIVE STUDY: GROUP DECISION MAKING

In the first study, we found that people tend to have difficulties in understanding probabilistic three-category forecast information and feel reluctant to use it in environmental decision making. In addition, large variability among individuals' evaluations of forecasts was observed. If carried forward into these students' adult careers, such misunderstanding and reluctance could be detrimental to realizing the potential benefits of seasonal climate forecasts, and could result in suboptimal decisions by environmental policy makers. Our goals for the second study were to probe more deeply into the reasoning underlying the choice to use or not use climate forecasts in decision making, and to examine individual and group reasoning in a setting more typical of authentic environmental decision making.

Thus, in the second study, we placed participants in a situation where they had to make decisions taking into consideration climate information or historical water-availability data or both. The decision structure and "characters" played by each participant were fictitious; however, the data used to support these decisions and the structure of the scenario (water management for agricultural production) were authentic. Therefore, participants were provided with the specific context and background for their roles and decision making.

We sought to see in what way people evaluate and use forecast and other information in realistic situations, to elaborate on and extend the findings from the first study. This aim reflects the discussions in the literature about the importance of context for human thinking; in particular, the value of carrying out meaningful tasks in authentic contexts (e.g., Sawyer and Greeno, 2009). Research in situated cognition discusses learning environments as consisting of interactions with other learners and with material, informational, and conceptual resources (Greeno, 2006). In our study, other group members, forecast and historical data, water-allocation decision structures, and concepts of climate forecasting constituted these components. With that aim in mind, we conducted semistructured group interviews and qualitatively analyzed participants' verbal reports on their thought processes in decision making.

Our use of a qualitative approach was also based on the concern raised in the literature about Likert-scale questionnaires, such as acquiescence bias or central-tendency bias (e.g., Baker and Leary, 1995). In the first study, the validity (particularly content validity) of the scales used in the evaluation and action questions was not explicitly examined, except for replications. Thus, this second study examined whether the dissatisfaction with climate forecasts and reluctance to use them would be seen in the situated condition, as observed in the first study; and if so, why they come to their decisions as individuals or a group.

Method

Participants

A new group of 55 students (15 men and 40 women) in the MPA program in Environmental Science and Policy participated in this study. They were new students to the program in the following year, and their backgrounds were very similar to those described in the first study.

Materials

The situation that we devised for the second study involved agricultural water allocation in the Yaqui Valley, Mexico. Participants viewed a map showing the location of the Yaqui Valley, and a diagram of the Mexican water-management institutions shown in Figure 7 (Addams, 2004). The decision structure consists of three segments, at the farmer, local, and federal levels. First, the general assembly of farmers controls the 42 privatized Irrigation Modules of the Yaqui Valley Irrigation District, each ranging in area from 845 to 12,000 ha and being composed of 115–1640 farmers. It approves decisions made by the directors of the modules and is responsible for operation and maintenance of irrigation infrastructures. Second, the local representative of the National Water Commission (CNA) retains responsibility for operating, maintaining, and planning the surface-water reservoir systems of the Irrigation District. The Irrigation District and the

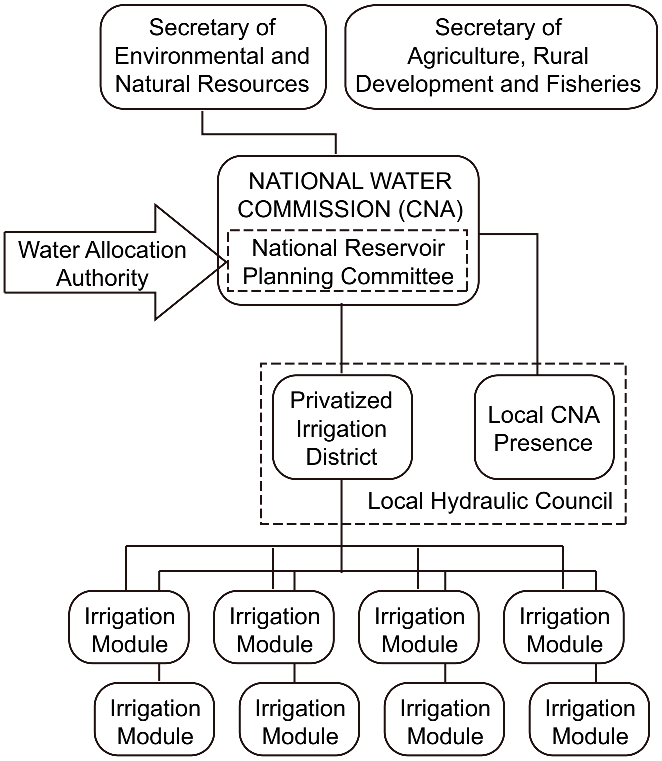


Figure 7. Decision structure that sets final water allocations for the Yaqui Valley Irrigation District. Participants acted as the representative of the general assembly of farmers controlling the District Irrigation Modules, as the local representative of the National Water Commission (CNA), or as the federal representative of the CNA.

local CNA representative make up the Local Hydraulic Council, which identifies irrigation needs for the upcoming water year and presents a request to the centralized CNA committee. Third, the federal representative of the CNA, heading the National Reservoir Planning Committee, reviews water-allocation requests made by the Local Hydraulic Council.

About the agriculture and water management in the area, participants were given the following scenario:

1. In the Yaqui Valley, there are 22,600 farms similar in surface area, and more than 99% of agriculture is supported by irrigation. Each farmer operates a 10 ha farm, and needs 10×10^4 m^3 of water to maintain crops. Water allocation for the past 20 yr has averaged $(2526 \pm 883) \times 10^6$ m^3. The economic crops that they plan to plant are water sensitive.
2. The Local Hydraulic Council is strongly requesting that the CNA approve its water allocation of 2750×10^6 m^3 to irrigate the 284,800 ha of farms that are planned to be planted with water-sensitive crops this year.
3. In the Yaqui Valley, the monsoon rains occur immediately before the major crop season. More than two thirds of the rain falls in the region between July and September and translates into surface runoff that fills a series of reservoirs. An early to mid-November wheat-planting date requires land preparation by late October of each year. Since water allocation depends greatly on the July to September rains, this water decision is necessarily delayed until early October, so that a planned water allocation can utilize the maximum possible initial storage volumes. Up until now, climate forecasts have not been used in water management.
4. The water level of the reservoir system has been fluctuating annually according to the graph in Figure 8. The "dead storage" corresponds to the volume of water that cannot flow above the lowest sill of the dam. The "committed storage" corresponds to the volume of water already allocated to sustain the municipal, industrial, and indigenous needs of the region. The "net storage" is the volume of water that is entirely available for irrigation.
5. Because the country is under pressure to bring more fiscal responsibility to its water-resource management, the emergency-relief funds often used as indirect agricultural subsidies in drought years have been slashed by more than 50%. Moreover, specific reforms to the relief-fund framework limit even further the potential subsidies for lost crops due to mismanagement of water resources.

In addition to this scenario, participants were given the IRI precipitation forecast maps and observation map for August-September-October 2004 (Figs. 9A and 9B) as a baseline for decision making and were told that this pair of forecasts and observations showed the representative degree of correspondence for North America for this 3 mo period. They were also given forecasts for three consecutive overlapping 3 mo seasons that covered October, the month when water-allocation decisions are made (Aug-Sept-Oct, Sept-Oct-Nov, and Oct-Nov-Dec; Fig. 9C).

Test Questions

Question 1 asked about participants' evaluation of the degree of correspondence between a forecast map (Fig. 9A) and an observation map (Fig. 9B), on the same 5 point scale used for the evaluation question in the first study (question 3-1).

Questions 2 and 3 were aimed at examining the way in which students evaluated and used climate forecasts in group decision making. In question 2, participants were asked to indicate how much they were inclined to use these forecasts to make decisions about water allocation, considering the representative pair of forecast and observation maps shown in question 1 (Figs. 9A and 9B). They answered this question on a 4 point scale (1—I would be strongly inclined to use the forecasts; 2—I would be weakly inclined to use the forecasts; 3—I would not be inclined to use the forecasts because I do not know if they are reliable; 4—I would be inclined not to use the forecasts because I think they are not reliable). Then, in question 3, participants were asked what would be their decision concerning the water allocation, in terms of how much they would increase or decrease the amount of 2750×10^6 m^3 requested by the Local Hydraulic Council.

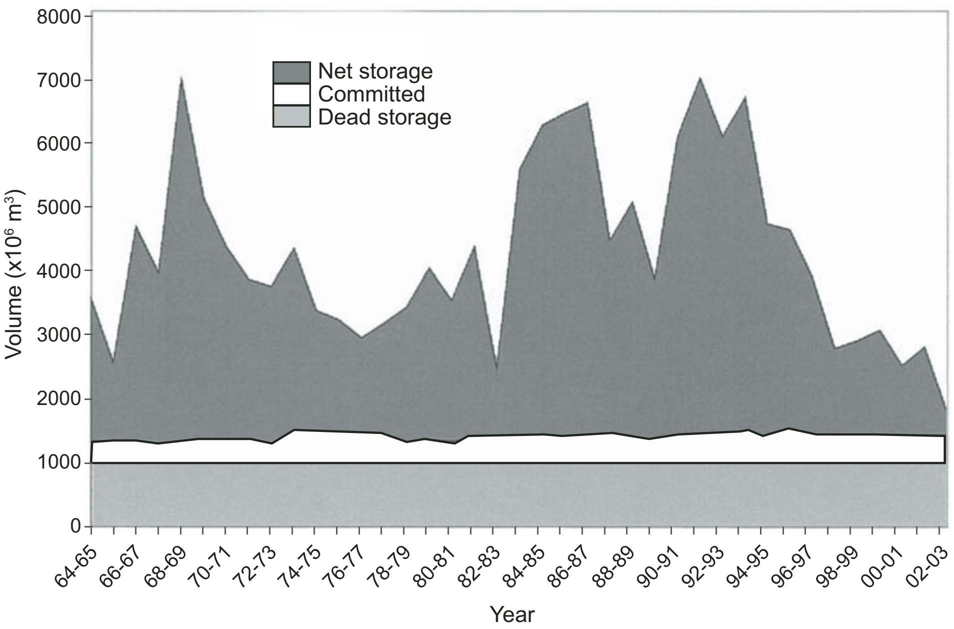


Figure 8. Graph showing annual changes in water level of the reservoir system in the Yaqui Valley since 1964.

Procedure

We assigned the 55 students randomly into 17 groups of three members and one group of four members, and assigned each member randomly to one of the three segments that set final water allocations for the Yaqui Valley Irrigation District. That is, in each group, one member acted as the representative of the general assembly of farmers, a second member as the local representative of the CNA, and a third member as the federal representative of the CNA (for the group of four members, two members acted as the farmer representatives).

Participants first answered questions 1–3 individually (i.e., as a farmer, a local representative, or a federal representative), and then they discussed the questions as groups for 15 min. After that, we interviewed 10 of the 17 groups, which were randomly chosen because of experimental time constraints, and we asked about their decision-making processes. The interviews were conducted in the form of a semistructured group interview, for 10 min on average for each group (Ishikawa, Kastens, and Louchouarn were the interviewers). We prepared basic questions beforehand on the basis of the first, quantitative study but adjusted when and how to ask them depending on participants' responses, so that we could prompt them to talk about their thought processes. The prepared questions were (1) whether they came to consensus about a decision as a group, (2) why they came to the group decision or individual decisions, and (3) what they thought about the provided or other desired information.

We videotaped the group interviews for subsequent transcription and analysis. The reason for videotaping rather than audiotaping was to make sure that we could transcribe or interpret their verbal reports accurately by observing their possible gestures (e.g., participants pointing to a specific forecast map as describing what they thought about it). Parenthetically, Kastens et al. (2008) discussed the role of gestures in geoscience teaching and learning.

Some comments on the backgrounds and roles of the authors in this qualitative study are in order, because they can possibly influence the way in which the qualitative data are analyzed and interpreted. The first through fourth authors' backgrounds are cognitive science, climate prediction and diagnostics science, geoscience learning and education, and environmental science, respectively. All of them were involved in the study processes and respectively provided expertise for designing the experiment and analyzing the data, preparing climate data and designing questions about forecast interpretation, setting contexts for geoscience education, and designing exercises about Yaqui Valley water allocation and running the class section.

Results

We conducted a content analysis (Hsieh and Shannon, 2005; Libarkin and Kurdziel, 2002) of participants' verbal reports from group interviews (see Table 1), and observed that five thematic contents emerged for the coding of their verbal protocols: (1) interpretation of provided data, (2) evaluation of the data, (3) roles or goals as individuals, (4) degrees of willingness to use the data, and (5) decisions as individuals and as a group. The first author was mainly responsible for creating these coding categories by looking into transcribed protocols in detail and then discussing the ideas with coauthors. Below, we describe results based on these coding schemes, assuming certain degrees of adequacy from the consensus among us. We still acknowledge that their validity needs to be further established, for example,

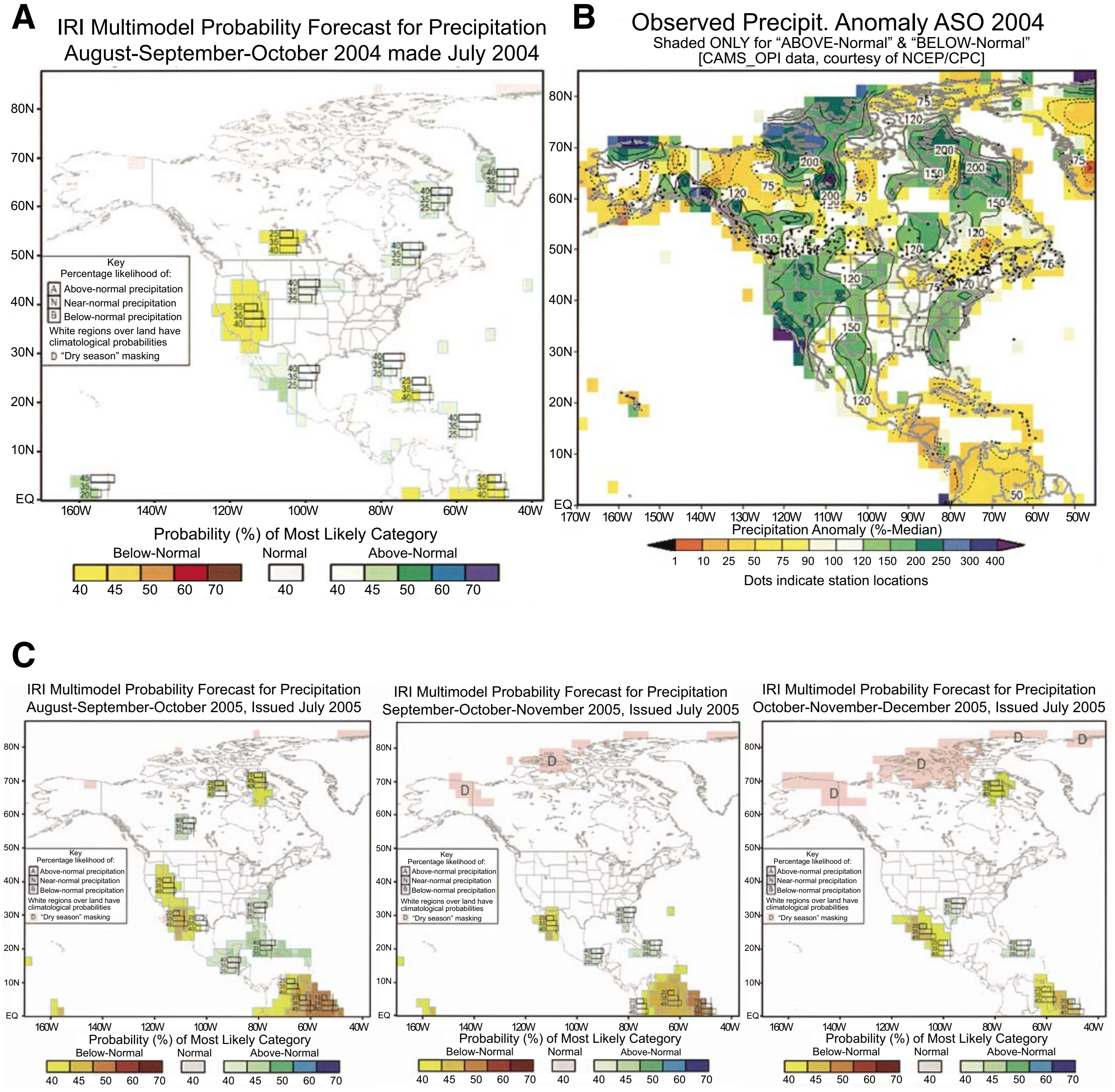


Figure 9. (A) Forecast map used for the evaluation question. (B) Map showing the observed amount of precipitation during the 3 mo forecast period for part A. (C) Forecast maps for three consecutive 3 mo seasons, presented to participants during group decision making about water allocation in the Yaqui Valley. The Yaqui Valley is located in northwestern Mexico, in the area where yellow and orange colors indicate that below-normal precipitation is most likely for the forecast months.

TABLE 1. COMMENTS FROM THE 10 GROUPS WE INTERVIEWED ABOUT DECISIONS ON WATER ALLOCATION

Group	Farmer	Local representative	Federal representative
1	Looks like there is going to be less rain than normal. (Req.)	I would be more responsible for managing the reservoir. It seems pretty clear that the projection is there is going to be quite a bit less precipitation. (Avg.)	I looked at the forecasts as far as how much water is expected and it is below average, and I looked at the past and it fluctuated a lot, and the average was less than what they are requesting. (Avg.)
2	We should not rely on predictions, because they were off ... and not reliable. However, the representatives are concerned, so I compromised by saying I will not take all the water requested. (Avg.)	Figures [forecast maps] are particularly unreliable. We should wait until October. We do not really have enough information at this stage to make a decision. (Avg.)	There is not definitely an increase of allocation, which was based on mostly Fig. 8, where the storage is decreasing since 1993. And the forecasts say they could probably have below-normal precipitation. I used Fig. 8 more than the forecasts given the fact that the forecast was incorrect in the year 2004. (Avg.)
4	I was less inclined to use the forecast because that would be against my position. (Req.)	Looking at Fig. 8, the amount of water we have is basically low, so I kind of thought about using the forecasts. We should keep the allocation at the historical average because it has got us through 20 years, and it also seems fair for my position representing the local farmers and the higher powers. (Avg.)	I looked first at Fig. 8, which showed the water in the reservoir has been dropping steadily, and I was nervous about our use of water. So I was strongly inclined to use the forecast. Even though the forecast was not entirely accurate for the region, I would still use it because it supports my idea that the region is kind of drying out. (Less)
5	Mostly we were thinking about the committed and dead storage [shown in Fig. 8]. Although there was some correspondence between the forecast and observations, we were afraid we did not want to trust it too much. (Avg.)	In the map as a whole, it was not necessarily so accurate, but in the region of Mexico, it was pretty close, so we decided to trust the forecast more than saying we are going to do this blindly. We did not want to decrease the allocation, which would have significant effects on our region economically, but we do not want to potentially run the risk of disaster by increasing it. (Avg.)	On behalf of the government, we would encourage them to become more efficient with their water use. (Avg.)
6	The forecast is not a good indicator of what actually happens. Also, how do you make a decision with 40% confidence? It should be at least 60% or 70%. We really do not have any facts, so we would just do as before because we were doing fine. (Req.)	We do not have very much confidence from the prediction last year, and also we need to manage their water and crops. “Same as last year” was an adequate solution. (Avg.)	We do not want to over-allocate or under-allocate, so we are more looking at a bigger picture. The maps [forecasts] are not so accurate, based on actual precipitation and predictions. (Avg.)
7	I was a little skeptical about the forecasts, just because it would be better to be on the safe side looking at it historically. But I got a little bit nervous when I saw how low the reservoir was in 2003. We definitely talked about the forecasts, but my decision is based more on Fig. 8 than the forecasts. (Less)	I thought my role was fiscally conservative and conservative with water allocation. But at the same time, I did not want to appear I was anticrops. Based on the historical data [Fig. 8], we should reduce it below the average, and I also considered the forecasts just because I did want to err on the side of caution. We would look bad five years from now if we had these data [forecasts of below-normal precipitation] and totally ignored them. (Less)	I was not 100% sure about the data [forecasts], but I did not want to totally ignore them, because I am representing the federal agency. (Less)

(*Continued*)

TABLE 1. COMMENTS FROM THE 10 GROUPS WE INTERVIEWED ABOUT DECISIONS ON WATER ALLOCATION (*Continued*)

Group	Farmer	Local representative	Federal representative
9	We need water, maximize the profits. (Req.)	We persuade the federal representative to give the amount of water the farmers want to get. The forecasts do not necessarily match up the observations, so we could not rely on them. (Req.)	Based on the forecasts, it is supposed to be drier. And due to the low reservoir, we wanted to use the forecasts. (Req.)
10	I am weakly inclined to use the forecasts because they were not accurate all the time, and based on the forecast for this area, there was a 50% or 60% chance that it could be wrong. I looked at Fig. 8, but I do not think I really need that in thinking about the decision either. (Avg.)	To represent the entire use of the reservoir, even though the farmers may get upset, I have to get based on very strongly the predicted amount of rainfall. And the water level that they were requesting has not been representative since 1997. (Less)	The reliability of the forecasts was like 50%, so I did not want to rely on them. And if we look at Fig. 8, we do not really have enough water left in the reservoir, so we cannot really allocate that much. (Less)
11	We want as much water as possible, even taking some risk. There is going to be an opportunity to adjust land preparation and crop choice based on the amount of precipitation that is actually received in early October. (Avg.)	We make the farmer understand that the plan needs to be made today and stick to it, the decision being based on the forecasts. We did not want any crop failure to cause any economic harm. Because the trends [Fig. 8] and forecasts agree and the three forecasts are consistent, we trust the forecasts and decrease the allocation. (Less)	We make the farmer understand that the plan needs to be made today and stick to it, the decision being based on the forecasts. We did not want any crop failure to cause any economic harm. The maps present the probability of lower than normal, but we do not know the amount of anomaly. The local representative just looks at the local area, which correlates well with observations. We federal representatives take an overall look at the map, which is not 100% accurate, so I am less confident with this prediction. (Less)
12	I would not be inclined to use the forecast, because it would ultimately say that we would have less water availability and therefore I would be getting less water. We do not really have enough data to know if the forecasts are reliable or not. (Avg.)	I would be strongly inclined to trust the forecast because that is the only forecast presented to me, and I am responsible for the management of water allocation. I am receiving pressure to be more fiscally responsible. (Avg.)	I am not inclined to use the forecast, because I do not think it is reliable. The data in front of me are not enough to justify. But I have to make a decision and I do not have other data. The implication of the data is very significant. (Avg.)

Notes: The group numbers correlate with those in Figure 10. Because of time constraints, not all groups were interviewed. Letters in parentheses indicate their decision: Req.—as much as requested; Avg. —the same as the 20 yr average; Less—less than the 20 yr average.

through replications or discussions with people who do not know our research questions.[1]

Concerning the interpretation of forecast data, the group interviews showed that our participants understood the provided precipitation forecasts, in a sense, correctly; that the forecasts predict that the Yaqui Valley area is likely to receive less rain than normal. However, this understanding tended to be coupled with a deterministic misinterpretation; that is, they thought that the area "will" receive less rain than normal and did not explicitly mention in the interview the possibility of observing the other two categories (normal or above-normal). This observation is consistent with our findings from the quantitative analysis in the first study. For example, the local representative in group 1 said, "It seems pretty clear that the projection is there is going to be quite a bit less precipitation."

Concerning the evaluation of forecasts, many participants expressed dissatisfaction with accuracy, on the basis of a "one-time" judgment about the degree of correspondence between forecast and observation maps for a specific past time. This observation coincides with their tendency to neglect the probabilistic nature of forecasts and implies the potential danger of people's losing trust in forecasts once the most likely category is not subsequently observed. For example, the local representative in group 6 said, "We do not have very much confidence from the prediction last year," and the farmers in groups 2 and

---

[1] We also assessed participants' understanding of forecast and observation maps with essentially the same questions as questions 1 and 2 in the first study and found that they had difficulty with the three-category probabilistic forecasts. In particular, we observed the reversed pattern of performance on the above-normal versus below-normal items that had been revealed in the first study. Thus, Ishikawa et al.'s (2005) findings were replicated again in the second study, with different participants and different maps.

6 said, "We should not reply on predictions, because they were off," and "the forecast is not a good indicator of what actually happens." This finding was reflected in our quantitative data. For the evaluation question (question 3), the majority of participants (79%) answered that the forecast and observation maps agreed only slightly or somewhat, a similar percentage to that observed in the first study (78%).

In explaining decision-making processes, participants tended to state explicitly what their goals were, in relation to their roles as a farmer, local, or federal representative. The farmer representative wanted to secure enough water to grow intended crops. In contrast, the federal representative felt it imperative to examine the requests from a larger viewpoint to keep the entire water system in good order. Thus, their willingness or reluctance to use the forecasts was related to their perception of their roles and expected courses of action.

Specifically, we observed from the protocol analyses four major types of decision-making processes: (1) adhering to roles, (2) comparing forecasts and goals, (3) comparing forecasts and historical data, and (4) considering societal connotations of predictions.

The first type of participant adhered to their roles and did not refer to any external information. For example, the farmers in groups 9 and 11 simply said, "We need water, maximize the profits" and "we want as much water as possible, even taking some risk," and the federal representative in group 5 said, "On behalf of the government, we would encourage them to become more efficient with their water use."

The second type of participant looked at the provided forecast data and expressed willingness or reluctance to use them depending on whether the forecasts would imply a course of action in support of or in opposition to their goals. For the farmers, the forecast of a dry season went against their desired outcome (the decision of a large water allocation), and so they did not use it. The farmers in groups 4 and 12 stated this explicitly: "I was less inclined to use the forecast because that would be against my position" and "I would not be inclined to use the forecast, because it would ultimately say that we would have less water availability and therefore I would be getting less water."

In contrast, some federal representatives expressed inclinations toward using the forecasts because the below-normal predictions would support their decision to constrain water allocation. In group 4, the federal representative said, "I would still use it because it supports my idea that the region is kind of drying out." Likewise, in group 9, the federal representative said, "Based on the forecasts, it is supposed to be drier, and due to the low reservoir, we wanted to use the forecast." These comments imply that both the farmers and the federal representative are choosing whether to use the forecasts based on whether the forecasts support their desired outcome or self-interest vis-à-vis their roles.

These findings about the effects of users' roles on forecast use were also observed in a quantitative analysis. For the question about the degree to which they were inclined to use the forecasts, there was a marginal difference in their responses among the three groups: local and federal representatives were likely to be more inclined to use the forecasts than were farmers, $\chi^2(6) = 10.84$, $p < 0.10$ (Table 2).

The third type of participant leaned toward using the forecasts because the below-normal predictions were in line with the connotation of a decreasing trend in the historical water-level data. The local representatives in groups 4 and 11 said, "The amount of water we have is basically low, so I kind of thought about using the forecasts" and "because the trends and forecasts agree…we trust the forecasts." The farmer in group 7 and the federal representative in group 2 explicitly stated that they compared the trustworthiness of forecasts and historical data: "My decision is based more on [the historical data] than the forecasts" and "I used [the historical data] more than the forecasts given the fact that the forecast was incorrect in the year 2004." These participants used the forecasts in a somewhat post hoc manner, placing more weight on past trends than on future predictions.

The fourth type of participant expressed that they felt the forecasts should be used because the connotations of below-normal predictions were societally significant, and subsequent disaster owing to nonuse of those predictions could harm their positions as representatives. It suggests an effect of social desirability. The local representative in group 7 said, "I also considered the forecasts just because I did want to err on the side of caution. We would look bad five years from now if we had these data [forecasts of below-normal precipitation] and totally ignored them." The federal representative in the group also said, "I was not 100% sure about the data [forecasts], but I did not want to totally ignore them, because I am representing the federal agency." The federal

TABLE 2. RESPONSES TO THE ACTION QUESTION BY FARMERS, LOCAL REPRESENTATIVES, AND FEDERAL REPRESENTATIVES

Scale	Farmer	Local representative	Federal representative
1	4	8	7
2	10	7	6
3	1	1	5
4	2	0	0

Notes: Scale 1—strongly inclined to use the forecasts; 4—inclined not to use the forecasts. Responses from two farmers and two local representatives were missing.

representative in group 12 explicitly stated that "The implication of the data [forecasts] is very significant." These people used the forecasts as a "hedge," not as credible evidence—the significance of the below-normal forecasts was too large to ignore, irrespective of their accuracy.

Our data seemed to suggest a relationship between the types of decision making and participants' roles. Comparable numbers of farmers (four) and local and federal representatives (three) were observed in the first two types of decision making (adhering to roles, and comparing forecasts and goals); while more local and federal representatives (eight, versus zero farmers) were observed in the latter two types (comparing forecasts and historical data, and considering societal connotations of predictions). This observation about possible effects of roles on decision making deserves further investigation.

Relating to the issue of evaluation, two other observations are noteworthy. First, some people expressed unease with uncertainty in forecasts. The farmer in group 6 said, "How do you make a decision with 40% confidence?" Similarly, the federal representative in group 10 said, "The reliability of the forecasts was like 50%, so I did not want to rely on them."

Second, some people commented on spatial scale, that is, about the accuracy of forecasts for a local area versus for the entire North America. The local representative in group 5 said, "In the map as a whole, it was not necessarily so accurate, but in the region of Mexico, it was pretty close, so we decided to trust the forecast." This points to subtlety and flexibility in using available information, and to the issue of relevance of information to the decision-making context.

Concerning the decisions of water allocation, some groups reached consensus and others did not. Figure 10 shows their decisions, which were classified into three categories: "as much as requested (2750×10^6 m^3)," "the same as the 20 yr average (2526×10^6 m^3)," and "less than the 20 yr average." It appears that farmers were likely to want more water, and federal representatives wanted to cut water allocation, with local representatives in the middle, but more than half (52%) of participants decided to reduce the requested amount to the 20 yr average. As shown in the content analysis, they mainly based this decision on the decreasing trend of the reservoir water; they had less confidence in the accuracy of the forecasts and were not inclined to use them.

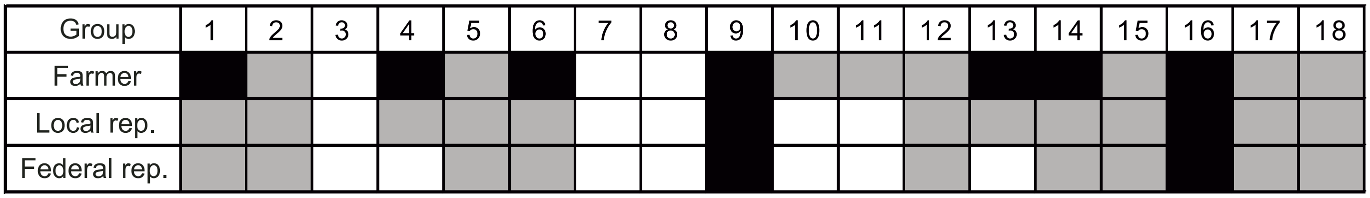
DISCUSSION AND CONCLUSIONS

This study empirically examined understanding, evaluation, and use of climate forecasts by prospective policy makers through a quantitative analysis of their interpretation of forecast information and a qualitative analysis of their decision making in a situated condition. Our results reveal some important characteristics of communication and decision making with climate forecasts.

Concerning the understanding of forecast data, difficulty in understanding the probabilistic nature of three-category forecasts was shown. In particular, the misunderstanding relating to the category having the highest likelihood versus the other two categories (Ishikawa et al., 2005) was replicated in the first, quantitative study. Our participants, although qualified and motivated students training in a professional master's program for prospective policy makers, did not interpret important aspects of the information as the forecaster intended. This suggests that the efficacy of the currently issued climate forecast maps as a communication tool is not optimal and leaves room for improvement through modification of data presentation formats, or pre- and in-service professional development for target users, or both.

Concerning evaluation, our results show that forecast users demand high accuracy, in fact, accuracy higher than is possible at the current state of the science in climate forecasting. Many participants did not positively evaluate the quality of forecasts or were reluctant to use the forecasts until the level of agreement between the forecasts and observations was better than is typically possible. Although the skill of climate forecasts is improving, it is not likely to soon reach the level desired by the participants who required excellent or even very good levels (Fig. 4B) before they would use this information in decision making. Educators working with such students need to help students develop methodologies for incorporating probabilistic, uncertain information into their decision making.

Although participants were generally accurate in evaluating the quality of forecasts (i.e., their evaluations became higher as the objective measure of forecast accuracy increased), they showed large individual differences both in the ability to visually inspect the degree of correspondence between forecast and observation maps and in their risk-taking or risk-averse tendencies. On the other hand, they were consistent in their evaluation and action



Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Farmer	■	▨	□	■	▨	■	□	□	■	▨	▨	▨	■	■	▨	■	▨	▨
Local rep.	▨	▨	□	▨	▨	▨	□	□	■	□	□	▨	▨	▨	▨	■	▨	▨
Federal rep.	▨	▨	□	□	▨	▨	□	□	■	□	□	▨	□	▨	▨	■	▨	▨

Figure 10. Participants' decisions about the amount of water to be allocated, classified into "as much as requested" (black), "the same as the 20 yr average" (gray), and "less than the 20 yr average" (white). Note that when the groups did not reach consensus, the farmer recommended the largest water allocation and the federal representative recommended the smallest.

ratings for specific maps, that is, they carried a commitment about the quality of a specific forecast over to the action question.

Potential conflicts arose during decision making by a group of people having different attitudes toward forecasts, which is in fact typical in real-world situations of policy making. In the group decision making, many participants disregarded the forecasts of precipitation and relied more on the time-series data showing changes in the reservoir levels for the past 20 yr. This means that the forecast information was not perceived to add additional value to the average historical data—participants put more weight on past or present information than predicted conditions in the future. Furthermore, people's decisions about whether to consider the forecasts in water allocation tended to be influenced by whether the forecasts supported or undercut their perceived self-interest (farmers) or predetermined positions (federal representatives). In this behavior, our role-playing graduate students resembled the middle-school students and adult nonscientists in the Hogan and Maglienti (2001) study, who were more likely than scientists to interpret ecological observations as supporting their prior personal values.

These results point out that, in the absence of specifically targeted training or instruction, it is not viable to assume that all users will understand the forecasts correctly. In other words, wide dissemination does not necessarily equate with effective communication. For the climate forecasts to be understood, evaluated, and used appropriately as input for decision making, improvements in design, training of the user, or preferably both, are necessary. Also, possible effects of users' attributes, such as gender or scientific backgrounds, were not specifically addressed in this study, but they are an important topic for future study.

Concerning the design of forecast maps, even though current presentation formats of climate forecasts are complex, perhaps too complex for untutored users, they are still a simplified representation of the forecaster's conception of the climate system. The forecaster needs to select a subset of information that is deemed important and necessary to the user. Methods for simplifying representations of complex scientific data, without sacrificing their richness and subtlety and without overwhelming the recipient, still represent a difficult enterprise that deserves further investigation.

The two key issues that probably require the most attention during the education of forecast users are the concept of probability and the existence of uncertainty in forecasts. As observed in our interviews, people tend to lose trust in forecasts once they "fail" (e.g., above-normal precipitation is observed when the below-normal probability had the highest likelihood in the issued forecast), despite the fact that the three-category forecast explicitly shows that all the three outcome categories are possible. This indicates that people interpret the probabilistic forecast in a rather deterministic way. A discussion of the role and extent of uncertainty in forecasts, and in science generally, is paramount during educational programs to avoid misconceptions in decision making (e.g., Louchouarn, 2008; Pollack, 2003). We also note that people without a solid understanding of uncertainty can err in both directions—either disregarding forecasts entirely and relying only on the "real data" from the past and present (the historical water-availability graph in our study), or completely believing the forecasts. The trick is to find an educational strategy that helps them find a middle path of weighting forecasts along with other factors, in an appropriate balance. For this purpose, a time-series plot showing all the outputs from the climate model over a long period of time seems to help broaden the user's understanding of climate forecasting (Hansen et al., 2004). The wide distribution of individual model outputs might help the user to appreciate the existence of uncertainty in forecasts, and hence to evaluate forecasts appropriately.

It may also be pointed out that from their prior education, most people have more experience in interpreting graphs than interpreting spatial data. Making a visual evaluation of the degree of correlation between two spatial representations is a fairly complicated kind of spatial thinking, which many or most participants would not have done previously. In fact, spatial thinking is undertaught and underpracticed in the current K–12 education (National Research Council, 2006). Furthermore, research has shown that people generally have difficulty comprehending spatial representations such as maps (e.g., Liben et al., 2002), graphs (e.g., McDermott et al., 1987), and diagrams (e.g., Larkin and Simon, 1987). As a possible educational strategy, training of environmental policy makers could include work with spatial data, especially GIS (geographic information systems) data, which are increasingly used in environmental decision making. Alternatively, some people might be more inclined to use a forecast that came in the form of a time-series graph for a specific location of interest, in which the forecast takes the form of a dotted line or a shaded range extending off into the future.

In this study, participants were asked to make decisions with limited information. Prior studies have proposed that decision making requires more detailed information than the common three-category probabilistic forecasts and should include the total amount of rainfall, the duration and distribution of rainfall over time and space, and the temporal and spatial resolution of forecasts (Ingram et al., 2002; O'Brien et al., 2000; Patt and Gwata, 2002). For example, our participants mentioned in group interviews that they would like information about the types of crops and their sensitivity to water supply, average yields, growing costs, market values, and possible subsidies. Also, participants in this study were given forecast products from only one institution. It would be interesting to see how they evaluate or weight multiple forecasts from different sources, including experience-based, traditional predictions (Phillips and Orlove, 2004).

Finally, this study hinted at the value of employing qualitative research methods to study students' understanding and evaluation of climate forecasts and use of these uncertain but potentially useful data in environmental decision making. In addition, this series of exercises pointed to the benefits of employing interviews or group-discussion activities to stimulate students' motivation and interest. Our student participants all conducted the group discussions with enthusiasm and responded to our interview questions with thoughtful ideas and interest in our research questions. We hope that they left the classroom with clearer and more solid understanding of the climate system and of humans' interactions with the climate system.

When we go back to the four characteristics of forecast products—accessibility, interpretability, credibility, and relevance (Pagano et al., 2002), we found that none is perfect, and interpretability and credibility probably suffer the most from limitations. We hope that this study serves to foster further attempts to achieve better communications and relationships between the forecaster and user, and to increase decision makers' capacity to make wise use of all environmental information, even when that information is incomplete or contradictory.

ACKNOWLEDGMENTS

We are grateful to James Hansen, David Krantz, and Chester Ropelewski for their valuable comments, and to the prospective policy makers who participated in the empirical studies. We thank Lee Addams for helping us understand the structure of the hydrological, economic, and agronomic context of the Yaqui Valley, and for providing the water data used in the group decision making. We also thank the three reviewers, especially Alison Stokes for her careful reading of the manuscript and constructive suggestions for revision. Part of the results of this study was presented at the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Applications Science Workshop, held in Palisades, New York, in March 2005, and at the annual meeting of the Association of American Geographers, held in San Francisco in April 2007. This is Lamont-Doherty Earth Observatory contribution 7358.

REFERENCES CITED

Addams, L., 2004, Water Resource Policy Evaluation Using a Combined Hydrological-Economic-Agronomic Modeling Framework: Yaqui Valley, Sonora, Mexico [Ph.D. thesis]: Stanford, California, Stanford University, 354 p.

Baker, D., and Leary, R., 1995, Letting girls speak out about science: Journal of Research in Science Teaching, v. 32, p. 3–27, doi:10.1002/tea.3660320104.

Barnston, A.G., Li, S., Mason, S.J., DeWitt, D.G., Goddard, L., and Gong, X., 2010, Verification of the first 11 years of IRI's seasonal climate forecasts: Journal of Applied Meteorology and Climatology, v. 49, p. 493–520, doi:10.1175/2009JAMC2325.1.

Cane, M.A., Eshel, G., and Buckland, R.W., 1994, Forecasting Zimbabwean maize yield using eastern equatorial Pacific sea surface temperature: Nature, v. 370, p. 204–205, doi:10.1038/370204a0.

Changnon, S.A., Changnon, J.M., and Changnon, D., 1995, Uses and applications of climate forecasts for power utilities: Bulletin of the American Meteorological Society, v. 76, p. 711–720, doi:10.1175/1520-0477(1995)076<0711:UAAOCF>2.0.CO;2.

Goddard, L., Mason, S.J., Zebiak, S.E., Ropelewski, C.F., Basher, R., and Cane, M.A., 2001, Current approaches to seasonal to interannual climate predictions: International Journal of Climatology, v. 21, p. 1111–1152, doi:10.1002/joc.636.

Goddard, L., Barnston, A.G., and Mason, S.J., 2003, Evaluation of the IRI's "net assessment" seasonal climate forecasts 1997–2001: Bulletin of the American Meteorological Society, v. 84, p. 1761–1781, doi:10.1175/BAMS-84-12-1761.

Greeno, J.G., 2006, Learning in activity, *in* Sawyer, R.K., ed., Cambridge Handbook of the Learning Sciences: New York, Cambridge University Press, p. 79–96.

Hammer, G.L., Hansen, J.W., Phillips, J.G., Mjelde, J.W., Hill, H., Love, A., and Potgieter, A., 2001, Advances in application of climate prediction in agriculture: Agricultural Systems, v. 70, p. 515–553, doi:10.1016/S0308-521X(01)00058-0.

Hansen, J.W., 2002, Realizing the potential benefits of climate prediction to agriculture: Issues, approaches, challenges: Agricultural Systems, v. 74, p. 309–330, doi:10.1016/S0308-521X(02)00043-4.

Hansen, J.W., Marx, S., and Weber, E., 2004, The role of climate perceptions, expectations, and forecasts in farmer decision making: International Research Institute for Climate and Society Technical Report 04-01, 55 p. (http://iri.columbia.edu/outreach/publication/report/04-01/index.html).

Hartmann, H.C., Pagano, T.C., Sorooshian, S., and Bales, R., 2002, Confidence builders: Evaluating seasonal climate forecasts from user perspectives: Bulletin of the American Meteorological Society, v. 83, p. 683–698, doi:10.1175/1520-0477(2002)083<0683:CBESCF>2.3.CO;2.

Hogan, K., and Maglienti, M., 2001, Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions: Journal of Research in Science Teaching, v. 38, p. 663–687, doi:10.1002/tea.1025.

Hsieh, H.-F., and Shannon, S.E., 2005, Three approaches to qualitative content analysis: Qualitative Health Research, v. 15, p. 1277–1288, doi:10.1177/1049732305276687.

Ingram, K.T., Roncoli, M.C., and Kirshen, P.H., 2002, Opportunities and constraints for farmers of West Africa to use seasonal precipitation forecasts with Burkina Faso as a case study: Agricultural Systems, v. 74, p. 331–349, doi:10.1016/S0308-521X(02)00044-6.

Ishikawa, T., Barnston, A.G., Kastens, K.A., Louchouarn, P., and Ropelewski, C.F., 2005, Climate forecast maps as a communication and decision-support tool: An empirical test with prospective policy makers: Cartography and Geographic Information Science, v. 32, p. 3–16, doi:10.1559/1523040053270747.

Jones, J.W., Hansen, J.W., Royce, F.S., and Messina, C.D., 2000, Potential benefits of climate forecasting to agriculture: Agriculture Ecosystems & Environment, v. 82, p. 169–184, doi:10.1016/S0167-8809(00)00225-5.

Kastens, K.A., Agrawal, S., and Liben, L.S., 2008, Research in science education: The role of gestures in geoscience teaching and learning: Journal of Geoscience Education, v. 56, p. 362–368.

Kirshen, P.H., 2002, Potential impacts of global warming on groundwater in eastern Massachusetts: Journal of Water Resources Planning and Management, v. 128, p. 216–226, doi:10.1061/(ASCE)0733-9496(2002)128:3(216).

Larkin, J.H., and Simon, H.A., 1987, Why a diagram is (sometimes) worth ten thousand words: Cognitive Science, v. 11, p. 65–100, doi:10.1111/j.1551-6708.1987.tb00863.x.

Libarkin, J.C., and Kurdziel, J.P., 2002, Research methodologies in science education: Qualitative data: Journal of Geoscience Education, v. 50, p. 195–200.

Liben, L.S., Kastens, K.A., and Stevenson, L.M., 2002, Real-world knowledge through real-world maps: A developmental guide for navigating the educational terrain: Developmental Review, v. 22, p. 267–322, doi:10.1006/drev.2002.0545.

Louchouarn, P., 2008, How authentic can scientific inquiry be in the classroom, and does it matter? A case study addressing graduate professional programs in environmental sciences: Geological Society of America Abstracts with Programs, v. 40, no. 6, p. 215.

McDermott, L.C., Rosenquist, M.L., and van Zee, E.H., 1987, Student difficulties in connecting graphs and physics: Examples from kinematics: American Journal of Physics, v. 55, p. 503–513, doi:10.1119/1.15104.

National Research Council, 2006, Learning to Think Spatially: Washington, D.C., National Academy Press, 313 p.

Nicholls, N., 1999, Cognitive illusions, heuristics, and climate prediction: Bulletin of the American Meteorological Society, v. 80, p. 1385–1397, doi:10.1175/1520-0477(1999)080<1385:CIHACP>2.0.CO;2.

O'Brien, K., Sygna, L., Næss, L.O., Kingamkono, R., and Hochobeb, B., 2000, Is Information Enough? User Responses to Seasonal Climate Forecasts in Southern Africa: Center for International Climate and Environmental Research (CICERO) Report 2000:03, 70 p. (http://www.cicero.uio.no/media/22.pdf).

Pagano, T.C., Hartmann, H.C., and Sorooshian, S., 2002, Factors affecting seasonal forecast use in Arizona water management: A case study of the

1997–98 El Niño: Climate Research, v. 21, p. 259–269, doi:10.3354/cr021259.

Patt, A., and Gwata, C., 2002, Effective seasonal climate forecast applications: Examining constraints for subsistence farmers in Zimbabwe: Global Environmental Change, v. 12, p. 185–195, doi:10.1016/S0959-3780(02)00013-4.

Patt, A.G., and Schröter, D., 2008, Perceptions of climate risk in Mozambique: Implications for the success of adaptation strategies: Global Environmental Change, v. 18, p. 458–467, doi:10.1016/j.gloenvcha.2008.04.002.

Pfaff, A., Broad, K., and Glantz, M., 1999, Who benefits from climate forecasts?: Nature, v. 397, p. 645–646, doi:10.1038/17676.

Phillips, J., and Orlove, B., 2004, Improving Climate Forecast Communications for Farm Management in Uganda: Silver Spring, Maryland, National Oceanic and Atmospheric Administration, Office of Global Programs, Final Report, 23 p. (http://www.climate.noaa.gov/cpo_pa/sarp/2000_phillips_2004.pdf).

Pollack, H.N., 2003, Uncertain Science … Uncertain World: Cambridge, UK, Cambridge University Press, 243 p.

Rayner, S., Lach, D., Ingram, H., and Houck, M., 2002, Weather Forecasts Are for WIMPS: Why Water Resource Managers Don't Use Climate Forecasts: Silver Spring, Maryland, National Oceanic and Atmospheric Administration, Office of Global Programs, Final Report, 80 p. (http://www.climate.noaa.gov/cpo_pa/sarp/1998_rayner_final.pdf).

Sawyer, R.K., and Greeno, J.G., 2009, Situativity and learning, *in* Robbins, P., and Aydede, M., eds., Cambridge Handbook of Situated Cognition: New York, Cambridge University Press, p. 347–367.

Thomson, M.C., Palmer, T., Morse, A.P., Cresswell, M., and Connor, S.J., 2000, Forecasting disease risk with seasonal climate predictions: Lancet, v. 355, p. 1559–1560, doi:10.1016/S0140-6736(05)74616-X.

Thomson, M.C., Mason, S.J., Phindela, T., and Connor, S.J., 2005, Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana: The American Journal of Tropical Medicine and Hygiene, v. 73, p. 214–221.

Manuscript Accepted by the Society 23 June 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

Counting tectonic plates: A mixed-methods study of college students' conceptions of plates and boundaries

Karen M. Kortz*
Physics Department, Community College of Rhode Island, Lincoln, Rhode Island 02865, USA

Scott K. Clark[†]
Department of Geological Sciences, Michigan State University, East Lansing, Michigan 48824, USA

Kyle Gray
Department of Earth Science, University of Northern Iowa, Cedar Falls, Iowa 50614, USA

Jessica J. Smay
Physical Science Department, San José City College, San José, California 95128, USA

Brendalee Viveiros
Physics Department, Community College of Rhode Island, Lincoln, Rhode Island 02865, USA

David Steer
Department of Geology and Environmental Science, University of Akron, Akron, Ohio 44325, USA

ABSTRACT

We explored students' conceptions of plate tectonics using a combined qualitative and quantitative approach consisting of multiple-choice ConcepTest questions, questionnaires, and interviews. When shown schematic images illustrating plate tectonics, half of the students were unable to determine the correct number of tectonic plates. These students appeared to have the most difficulty determining whether or not to count a divergent boundary as a plate boundary, but additional difficulties include confusion between continent-ocean boundaries (shorelines) and plate boundaries, and failure to see the larger picture as a result of focusing on individual boundaries. We propose that the underlying causes for these difficulties stem from the tendency for students to construct their understanding of plate tectonics based on inappropriately applied prior knowledge. For example, when viewing a divergent boundary, many students activate two lines of prior knowledge: (1) if entities are the same (such as ocean plates on both sides of a divergent boundary) then they are not considered separate; and (2) if there is no obvious break (which is not seen on diagrams of

---

*kkortz@ccri.edu
[†]Current address: Department of Geology, University of Wisconsin–Eau Claire, Eau Claire, Wisconsin 54703, USA.

Kortz, K.M., Clark, S.K., Gray, K., Smay, J.J., Viveiros, B., and Steer, D., 2011, Counting tectonic plates: A mixed-methods study of college students' conceptions of plates and boundaries, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 171–188, doi:10.1130/2011.2474(12). For permission to copy, contact editing@geosociety.org.

divergent boundaries), then they are also not considered separate. The application of both of these lines of prior knowledge results in students concluding the two sides of a divergent boundary are the same plate. Retention of these alternative concepts prevents conceptual change from occurring during the period of instruction and results in students not recognizing divergent boundaries as plate boundaries, leading them to incorrectly count the number of plates.

INTRODUCTION

The theory of plate tectonics describes how plates of the rigid, outer layer of Earth move relative to and interact with one another. Processes related to plate tectonics generate many geoscience phenomena, and the theory is therefore often called the unifying theory of geology (e.g., Monroe et al., 2007; Grotzinger and Jordan, 2010). It is necessary for students to have a fundamental understanding of plate tectonics in order to build accurate mental models of Earth.

Given its importance to geology, plate tectonics is considered one of the key theories in the geosciences that students should learn (Rutherford and Ahlgren, 1991; National Research Council, 1996) and scientifically literate citizens should know (Earth Science Literacy Initiative, 2009). It is a key concept taught in middle school and high school earth science courses as listed in the National Science Education Standards (National Research Council, 1996), and it is covered in nearly all undergraduate introductory geology courses (Kelso et al., 2000). With repeated and near universal coverage, one might expect students who have completed an introductory geoscience college course to have a basic understanding of fundamental aspects of plate tectonics.

Prior research on students' understandings of several aspects relating to plate tectonics has revealed a number of alternative conceptions held by students. For example, students are unsure about the location of tectonic plates within Earth. Two documented alternative conceptions observed in students are that tectonic plates are arranged like a stack of layers (Marques and Thompson, 1997; Gobert, 2000), and that plates are located as layers within Earth, at the core, or in the atmosphere (Libarkin and Anderson, 2005; Libarkin et al., 2005; Libarkin, 2006). The incorrect conceptions of the location of tectonic plates are likely due to spatially incorrect mental models of the interior of Earth (Gobert, 2000; Steer et al., 2005). Some students hold the alternative conception that continental boundaries are the same as plate boundaries (Marques and Thompson, 1997). These alternative conceptions were found in high school students (Marques and Thompson, 1997), college students (Libarkin and Anderson, 2005; Libarkin et al., 2005; Libarkin, 2006), and in-service teachers (Dahl et al., 2005).

In addition, Clark and Libarkin (2008) found that students often used terms relating to plate tectonics, such as mantle, hotspot, melting, and convergent, without having a clear understanding of the concepts behind those terms. Clark and Libarkin (2008) also documented that one quarter of students who indicated motion at divergent boundaries interpreted the plates to be moving toward each other instead of apart. Students appear to perceive the ridge at the boundary as being analogous to two pieces of hard rubber being pushed together on a flat surface. A similar conceptualization has been documented by Sibley (2005) in regard to mountain formation.

PROBLEM STATEMENT

Because plate tectonic processes are responsible for so many Earth surface features, students need to possess an understanding of the basic principles pertaining to plate-tectonic theory in order to accurately understand geological concepts. One of the most fundamental aspects of plate tectonics is the ways in which plates interact with each other along their boundaries. Although the details can admittedly get very complicated, how plates interact and what constitutes a boundary are important concepts within the plate-tectonic theory. Therefore, this study examined college-level geoscience students' conceptions of tectonic plate boundaries. Guiding our study were the following questions:

1. When presented a plate-tectonic schematic diagram, how many tectonic plates do students count?
2. Why do students correctly or incorrectly count the number of plates?

THEORETICAL FRAMEWORK

When learning a new concept, students connect the new information to previously organized knowledge structures within their memory (e.g., Bartlett, 1932; Bransford et al., 2000, and references within; Pellegrino et al., 2001; McVee et al., 2005; Eshach and Schwartz, 2006). That is, their understanding of a given concept changes to accommodate the new information. Initially, these conceptions stem from everyday observations about the natural world but eventually are shaped by academic learning in the classroom (Vosniadou and Brewer, 1992; Vosniadou, 1994). Therefore, one goal of teachers is to replace naïve or inaccurate understandings with more complex and scientifically accurate concepts. The term conceptual change is used to denote this process.

Conceptual change is not a uniform theory but rather encompasses several research models (see Ozdemir and Clark, 2007; Scott et al., 2007; Vosniadou, 2007). An early model of conceptual change was articulated by Posner et al. (1982), whose seminal theory stated that people change their conceptions only when they are dissatisfied with their current schemata and are presented with a new conception that is more intelligible, plausible, and fruitful at explaining the target phenomenon. This theory drew heavily from three sources, including Piaget's (1952) view that knowledge is organized into cognitive schemes, as well as Ausubel's (1968) findings that a person's prior knowledge influences their current understanding, and Kuhn's (1970) idea that paradigm shifts radically alter one's view of reality (diSessa, 2002). In this framework, an earth science student adopts a more scientific understanding of plate tectonics only when she cannot explain the information presented in class using her prior knowledge and conceptual schema, or categories for information. If the older conceptual understanding adequately explains the new information, conceptual change will not occur regardless of the accuracy of the new information.

More recently, two cognitive models of conceptual change have focused on the changes that occur within the learner. Vosniadou (1994) described conceptual change in terms of modifications to the learner's explanatory theories, whereas diSessa (1993) and diSessa and Sherin (1998) described changes in the connections between rudimentary observations held by the learner. For example, Vosniadou (1994) argued that students begin with a naïve explanatory theory, but as the students acquire new and more sophisticated knowledge, they progressively modify their personal theories to accommodate the new data. Furthermore, Vosniadou (1994, 2007) found that students often refuse to discard their older, less accurate explanations when presented with a more scientifically accurate model. By contrast, diSessa (1993) argued that students collect small knowledge structures based on observations of the natural world and that conceptual change consists of students adjusting or modifying their connections between these primitive chunks of knowledge. In either case, the cognitive approach to conceptual change suggests that when the earth science student is confronted with new information that cannot be explained by his or her current schema, he or she modifies or adjusts his/her knowledge structures to accommodate these new data.

Among the different models of conceptual change, there are three common perspectives. First, a student's prior knowledge impacts the student's ability to formally learn a new concept (Hewson and Hewson, 1983; Scott et al., 2007). Second, students resist change to their preconceived knowledge structures (Pintrich et al., 1993; Chi, 2005; Steer et al., 2005; Vosniadou, 2007). Third, the process of conceptual change occurs over an extended period of time (Mikkilä-Erdmann, 2001).

This paper focuses on student understanding of plate tectonics, specifically, the rationales used to determine the number of plates in a given illustration. To analyze these data, we used a cognitive approach to conceptual change theory as described by Scott et al. (2007) and Vosniadou (2007). In this approach, we are most interested in the explanatory descriptions used by students to identify tectonic plates and interpret our data from this perspective.

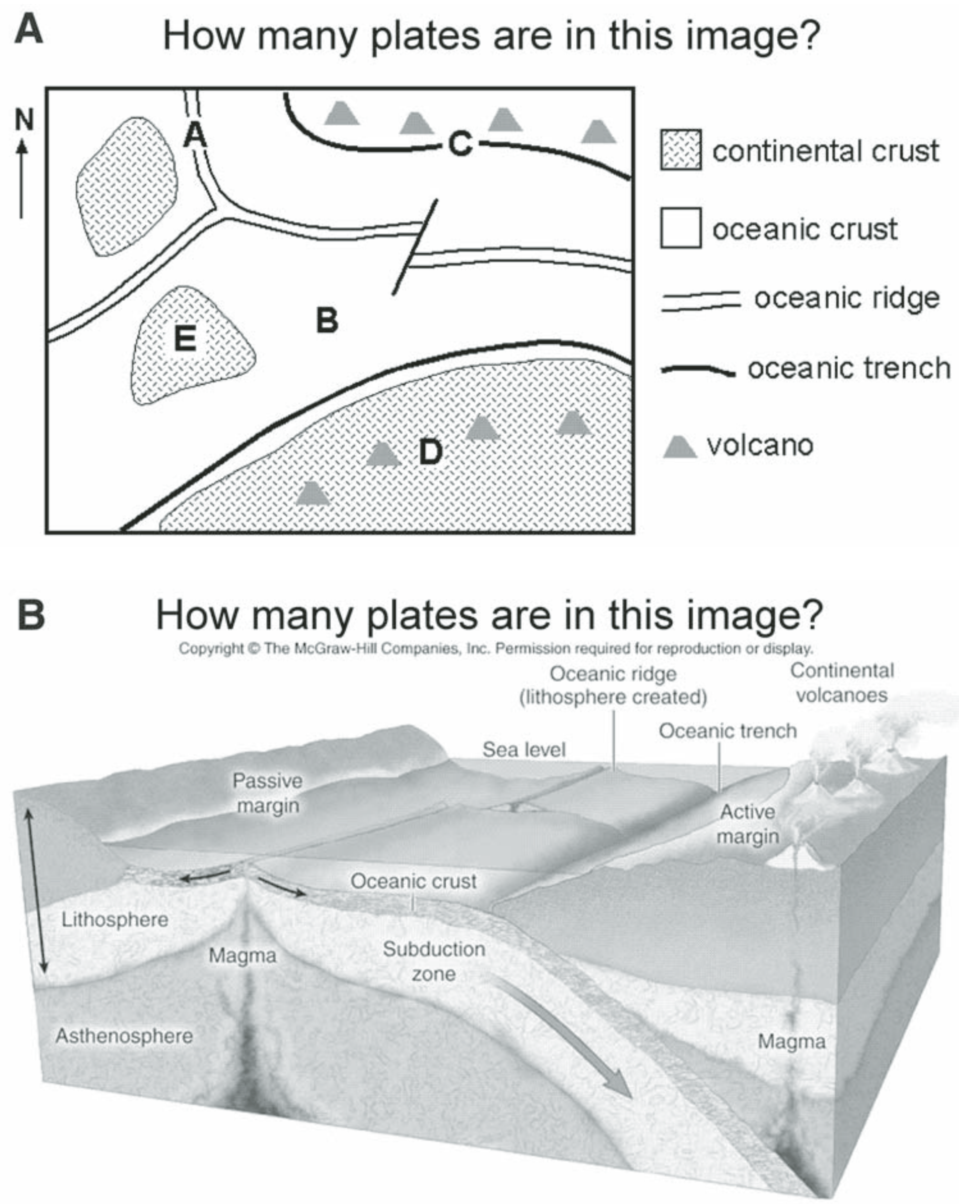
THREE STUDIES

Methodological Approaches

Teddlie and Tashakkori (2003) noted that studies that combine quantitative and qualitative methods are more likely to identify diverse points of view and generate stronger inferences than studies that use a single methodology. In addition, such mixed-methods studies allow researchers to simultaneously conduct exploratory and confirmatory research, which leads to a deeper understanding of the target phenomenon than what could be produced by a study from either the qualitative or quantitative traditions. This paper combines qualitative and quantitative data from three separate studies of student understanding of plate tectonics. In study 1, two multiple-choice ConcepTest questions (see study 1 for an explanation of ConcepTests) exposed the difficulty students have when counting tectonic plates on diagrams and the resiliency of alternative conceptions (Fig. 1). Study 2 probed students' conceptions of tectonic plate boundaries via a six-question questionnaire (Fig. 2) that included open-ended questions and a diagram showing plate-tectonic boundaries. Student responses to many of the open-ended questionnaire questions were transformed from qualitative data to quantitative, or numerical, data by placing student responses into intrinsically generated categories, resulting in a defined set of values. This transformation was done to facilitate the analysis of and comparisons between large numbers of student responses.

The multiple-choice and open-ended questions of studies 1 and 2, respectively, allowed us to identify where students perceived plate boundaries and what they identified as a tectonic plate. However, these studies did not give us much insight into the student thought process underlying those difficulties. In study 3, interviews probed the "why" behind student answers. Unlike responses on questionnaires, interviews allowed us to probe the students' thought process behind their answers. Students were asked similar questions as the questionnaire in study 2, allowing us to hypothesize why the broader population of students had difficulties with the questions in the first two studies. Institutional Review Board (IRB) approval was received for all three studies.

We report our findings from the three studies here, applying both qualitative and quantitative approaches in order to provide a robust answer to our research questions. We chose to use a mixed-method approach because quantitative data allowed us to answer our first research question about how many plates students counted, and qualitative data were best suited to answer the second research question about why they counted a certain number of plates. Previous studies of student misconceptions have used both qualitative and quantitative methods, although the emphasis has tended to be on qualitative research methods (e.g., Vosniadou and Brewer, 1992; Gobert, 2000; Trend, 2000; Kusnick, 2002; Teddlie and Tashakkori, 2003; Libarkin and Anderson, 2005; Libarkin et al., 2005). The quantitative data collected during our study (i.e., responses to multiple-choice ConcepTest

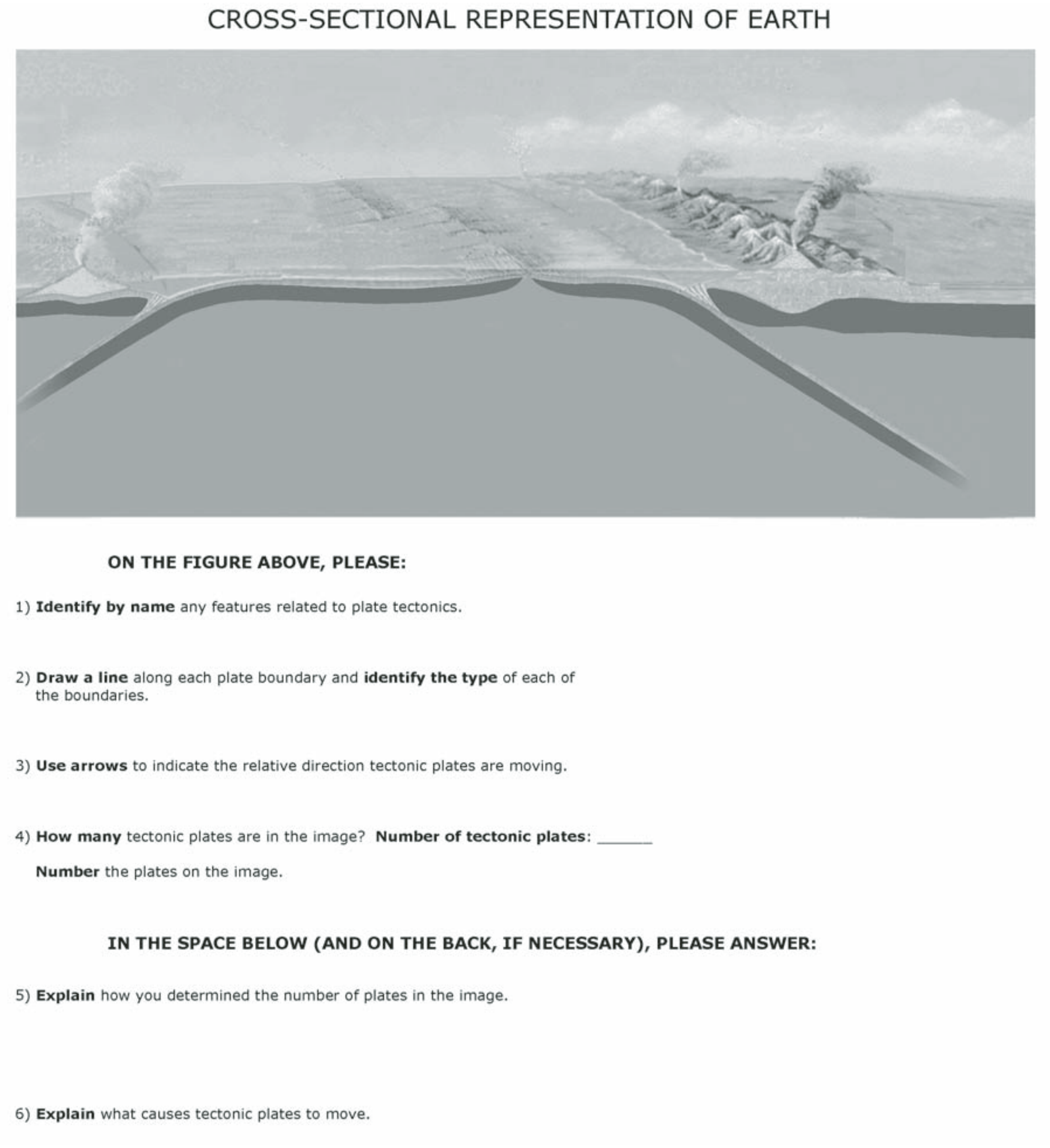


Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.

Figure 1. ConcepTest questions used in study 1. Figure 1B is from McConnell et al. (2008; their Fig. 4.17) and is used with permission of the McGraw-Hill Companies.

questions and categorized answers to the six-question questionnaire) illuminate the widespread struggle by students to correctly identify plate boundaries, as well as the students' apparent resistance to discard an inappropriate application of prior knowledge in favor of adopting a more accurate mental model. Our qualitative data (i.e., interviews and open-ended questions in the questionnaire) probed why the students had difficulties and provide insights into student reasoning and thought processes. Because part of our study was exploratory in nature, the rich details from the qualitative data allowed us to gather new information and generate new explanations concerning the ways in which students apply their understanding of plate tectonics to novel situations. Our interpretations of the data were directly tied to the students because the students expressed their own views. The combination of qualitative and quantitative methods lends confidence to the study's conclusions because each study's individual biases (such as method or inquirer bias) are offset by biases from other studies. This approach, known as triangulation, improves the validity of the study (Greene et al., 1989). In addition, the mixed-methods approach allows for elaboration and clarification of the results of one method to the results from the other methods, resulting in an increased meaningfulness and validity (Greene et al., 1989).

The researchers were "located" differently within each of the three studies (Feig, this volume). This location is described for each study in order to make transparent how the researchers fit into the study and to provide a context for potential bias.



CROSS-SECTIONAL REPRESENTATION OF EARTH

ON THE FIGURE ABOVE, PLEASE:

1) **Identify by name** any features related to plate tectonics.

2) **Draw a line** along each plate boundary and **identify the type** of each of the boundaries.

3) **Use arrows** to indicate the relative direction tectonic plates are moving.

4) **How many** tectonic plates are in the image? **Number of tectonic plates**: ______

Number the plates on the image.

IN THE SPACE BELOW (AND ON THE BACK, IF NECESSARY), PLEASE ANSWER:

5) **Explain** how you determined the number of plates in the image.

6) **Explain** what causes tectonic plates to move.

Figure 2. The questionnaire used in study 2. The original version of this questionnaire was developed by Clark and Libarkin (2008), and its development is described by Clark and Libarkin (this volume).

In study 1, two of the authors (Kortz, Steer) were researcher-participants (Feig, this volume; Patton, 2002), since they generated data from multiple-choice ConcepTest questions within their own courses. ConcepTest questions were also collected by additional instructors in their courses not involved in this study. In studies 2 and 3, the researchers were, for the most part, researcher-observers (Feig, this volume), since data were collected from students outside of normal classroom instruction, using passive questionnaires (study 2) and active engagement in interviews (study 3). Study 2 questionnaires were administered by instructors in their own courses but analyzed by the authors, although in one course, the author and instructor were the same (Kortz). In that case, the instructor mailed the responses to Clark for de-identification prior to analyzing those data. Study 3 interviews were all conducted by an author (Clark) who was not involved in the instruction of the students.

Building Trustworthiness

Trustworthiness in qualitative research is similar to validity and reliability in quantitative research. Based on the paradigm established by Lincoln and Guba (1985), Erlandson et al. (1993, p. 29) wrote that trustworthiness "demonstrate[s] its truth value, provide[s] basis for applying it, and allow[s] for external judgments to be made about the consistency of its procedures and the neutrality of its findings or decisions." Four aspects of trustworthiness are credibility, transferability, dependability, and confirmability (Lincoln and Guba, 1985). Each of these, as it applies to this study, is described in Table 1. The strongest support of trustworthiness in these findings comes from the multistudy aspect. Collecting information from different points of view with a variety of questions within and between three different studies (triangulation) provided us with support for credibility, dependability, and confirmability of our findings. Although study 3 had a slightly wider study population relative to studies 1 and 2, there is overlap between the levels of geoscience expertise of the students, allowing comparisons to be made between the studies. To counteract their inherent biases in order to enhance validity, each method needs to address the same conceptual phenomena but be implemented independently from other methods, as done in our three separate studies (Greene et al., 1989). In addition, the independent agreement between authors of the data (inter-rater reliability) and interpretations made from them helps to augment the trustworthiness.

Study 1

Study 1 used quantitative, multiple-choice ConcepTest questions (described in the following section) to determine how many plates introductory-level students ($n = 479$) counted on two separate plate-tectonic schematic diagrams.

Participants

The ConcepTest questions of study 1 were given between fall 2006 and spring 2008 to students in nine different introductory geology courses in which plate tectonics was taught. These courses were taught at two large public universities in the western United States, a large community college in the Northeast, and a large private university in the Midwest. The number of students answering the ConcepTest questions in each course ranged from 15 to 88, for a total of 479 students answering ConcepTest questions. The students who answered ConcepTest questions were 44% female and 56% male and had an average age of 21. They were 78% White, 10% African American, 4% Hispanic, 3% Asian/Pacific Islander, 0.4% Native American, and 5% not reported.

Data Collection

ConcepTest questions are conceptual multiple-choice questions that focus on one key concept (Mazur, 1997). They were used in class as a method of formative assessment of student understanding during lecture as part of peer instruction (Mazur, 1997;

TABLE 1. TRUSTWORTHINESS AND HOW IT IS APPROACHED IN THIS STUDY

Aspect of trustworthiness	What it measures	Comparison to quantitative research	How it is approached in this study
Credibility (whether the research conclusions match what the participants thought)	Truth value	Internal validity	Triangulation (collection of information from different points of view) with ConcepTest questions, questionnaires, and interviews Triangulation with questions on different topics to collect student views from different perspectives More than one author verified the credibility of the interpretations from the questionnaires and interviews Use of student quotes to demonstrate link between students' words and interpretations
Transferability (the extent to which the findings can be applied outside of the study)	Applicability	External validity	Description of classes and students from which data were collected Sampling of a large range of institutions and students to maximize the range of information collected
Dependability (whether the findings would be repeated under similar conditions)	Consistency	Reliability	Triangulation (described above) Code-recode procedure of analysis of questionnaire and interviews Use of rubric to analyze questionnaire responses Establishment of inter-rater reliability and discussion of coding Review by coauthors
Confirmability (whether conclusions can be tracked to the source)	Neutrality	Objectivity	Triangulation (described above) Student quotes used to illustrate link between source and interpretations Paper trail recorded

Note: Information adapted from Lincoln and Guba (1985) and Erlandson et al. (1993).

Crouch and Mazur, 2001; McConnell et al., 2003, 2006). During instruction on plate tectonics, instructors posed ConcepTest questions to their class using PowerPoint slides. Initially, students responded individually with personal response systems, or clickers; students then divided into small groups, wherein they discussed the question for a few minutes prior to revoting. The use of personal response systems facilitated data collection by allowing student answers to be automatically recorded and exported into a spreadsheet for analysis.

The two ConcepTest questions and accompanying images used in this study are shown in Figure 1. Figure 1A shows a map view of tectonic features on five plates. Divergent boundaries are indicated by the ridges, and subduction-type convergent boundaries are indicated by trenches. Figure 1B is a block diagram showing three plates separated by a subduction-type convergent boundary and a divergent boundary. These ConcepTest questions were written by faculty and evaluated by peers to verify content validity, which ensures the question measures an important aspect of geoscience understanding. Communication validity, or whether or not the students understand the question as intended, was ensured by writing the questions using simple vocabulary and choices. In addition, when testing the ConcepTest questions before this study, faculty listened to student discussions and talked to the students directly to determine if the students had any trouble in understanding the ConcepTest questions.

The map-view ConcepTest question (Fig. 1A) was given in all classes that participated in study 1; however, in some classes, this question was asked after the block diagram question (Fig. 1B). To avoid the potential influence of the first ConcepTest question on the student responses to the second, we analyzed only the ConcepTest question asked first. Responses given after peer discussion likely represent the collective knowledge of the group rather than the individuals (Steer et al., 2009), so they were not included in the study except to document how entrenched student conceptions were.

Results

The distributions of the number of plates students counted for each ConcepTest question (Fig. 1) are given in Figures 3 and 4. On the ConcepTest question that asked students about the number of plates on the map view (Fig. 1A), 40% of the 141 students answered the question correctly. After discussions with their peers, this percentage improved to 58% (Fig. 3). On the block diagram ConcepTest question (Fig. 1B), slightly over half of the 338 students correctly answered the question the first time, and the question was not revoted on during class.

Study 2

Study 2 used a six-question questionnaire that asked introductory geology students ($n = 35$) basic questions about a schematic diagram showing plate-tectonic boundaries, with the goal of having a better understanding of how many plates students identified and where the plates and boundaries were located. These questions were analyzed using both qualitative and quantitative methods.

Participants

Questionnaires for study 2 were completed by 35 students in six different courses taught by two professors at the same large community college in the Northeast where a subset of both ConcepTests (study 1) and interviews (study 3) were collected and conducted, respectively. The courses include introductory earth science courses that teach plate tectonics, such as physical geology, historical geology, natural disasters, and oceanography. The students who filled out the questionnaires were 51% female and 49% male and had an average age of 24. They were 74% White, 9% Hispanic, 6% Asian, 3% Black, and 9% other or not reported. Twenty-nine percent of the students had taken another introductory-level geoscience course in college, and 46% reported they had taken geoscience in high school. Fourteen

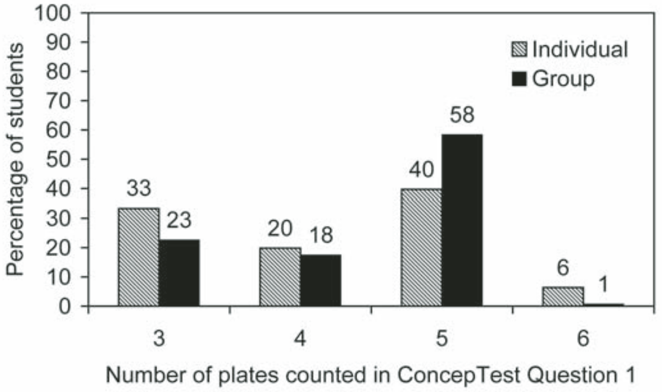


Figure 3. Percentage of students counting a specific number of plates on the ConcepTest question in Figure 1A (map) from individual responses after lecture and responses after group discussion (classes 1–4). The correct answer is 5.

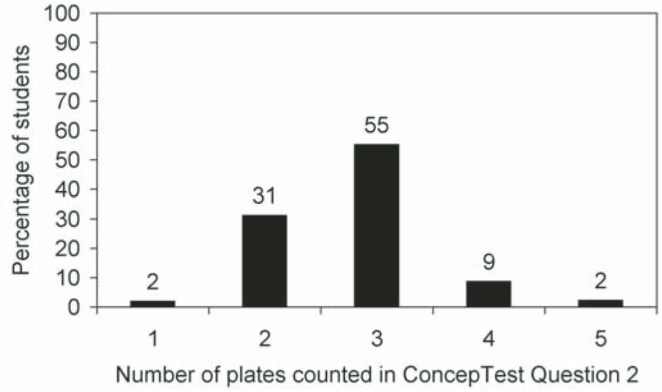


Figure 4. Percentage of students counting a specific number of plates on the ConcepTest question in Figure 1B (box diagram) from individual responses (classes 5–12). The correct answer is 3. Students were not asked to revote after peer discussion.

percent of the students intended to major in a science other than geology or engineering, 14% intended to major in education, and the rest indicated a different major. The questionnaires were distributed in class, and students participated on a voluntary basis. As a result, not all students answered all questions on the questionnaires, and some of the more difficult questions, such as numbering the plates for question 4, were left blank by over half the students (Table 2).

Data Collection

The questionnaire used in study 2 (Fig. 2) was a modified version of Clark and Libarkin's (2008) questionnaire, which was designed to assess people's conceptions of plate tectonics. Questions containing both qualitative and quantitative elements were directed at a figure that combines surface and cross-section views of four plates separated by one divergent and two convergent boundaries. The modified version used in this study asked only questions directly related to the issue of identifying tectonic plates and plate boundaries. The development of the original questionnaire, including a description of the ways in which validity and reliability were determined for the instrument, is described by Clark and Libarkin (this volume).

Several different analyses were performed on student responses to the questionnaires. The analysis of question 1 is described next. Question 2 on the questionnaire asked students to draw a line along the plate boundaries. We examined whether or not students drew lines, correctly or incorrectly, at each of the plate boundaries. For example, at the convergent boundaries, students drew lines along the trench, along the coast, crossing the subducting slab, and parallel to the subducting slab, and we considered all these lines as indicating the identification of the convergent boundary.

Question 3 asked the students to use arrows to indicate the direction of plate movement. We used a rubric created by Clark and Libarkin (2008) to provide a way to locate the arrow, relative to each of the three plate boundaries, and identify the direction of each arrow or pair of arrows drawn by students.

Question 4 asked the students how many plates were in the image and to number them on the image. We used a modified version of constant comparative analysis (Lincoln and Guba, 1985; Erlandson et al., 1993) to analyze student responses to numbering plates on the image. As we examined student answers, we looked for trends in the locations of the numbers, and if students located numbers in similar locations (e.g., numbering the two ocean plates as one or numbering the plate boundaries), we grouped those into categories. As we progressed through student responses, we compared the categories we created with each new student response and modified the categories as necessary. One author (Kortz) created the categories, and a second author (Viveiros) verified that they represented the student responses. Both authors were in full agreement on the assignment of student responses to each category. As a result, we took originally qualitative data and placed student responses into quantitative categories derived from the data.

Question 5 asked students how they determined the number of plates in the image. The responses between the questionnaire and interviews contained similar explanations (see Table 3), but the students more fully explained their reasoning during the interviews. Because student responses on the questionnaires could not be probed, we analyzed this question by using the insight provided by the interviews in study 3. Question 6 was not analyzed for this study.

We compared individual student answers to questions 1, 2, and 3 to examine whether or not the students identified each of the three plate boundaries. The students used a variety of means to identify individual boundaries, including: drawing lines along the boundary (question 2), drawing arrows at the boundary indicating relative plate motion (question 3), writing the boundary name (question 1), and/or identifying features or actions at the boundary associated with plate tectonics (ridge, trench, or subduction) (question 1).

Results

The distribution of the number of plates counted by students (30 responses) on the questionnaire (Fig. 2, question 4) is given in Figure 5. Nearly one third of students correctly counted four plates, and an equal number counted three plates. One student responded with a different number of plates to two questions (questions 4 and 5). Because question 5 required an explanation of how the number of plates was counted, we used this number in our analysis.

Of the 35 students who were asked to write a number on each of the plates in the image (Fig. 2, question 4), 12 students provided numbers. Five students correctly determined that the image contained four plates. Four students incorrectly numbered three plates because they combined the two oceanic plates into a

TABLE 2. NUMBER OF STUDENTS IDENTIFYING EACH BOUNDARY ON THE QUESTIONNAIRE (*N* = 35)

	Left convergent	Right convergent	Divergent	*n**
Line near boundary	17	19	10	20
Arrows at boundary	20	21	13	24
Type of boundary labeled	13	11	8	13
Features labeled	6	7	6	10
Boundary identified by any of the above indicators	26	28	19	29

Note: The responses did not need to be correct to be counted.
**n* is the number of students who answered each question with a response at a plate boundary.

TABLE 3. CATEGORIES, FREQUENCIES, AND INTERPRETATIONS OF STUDENTS' EXPLANATIONS FOR COUNTING TECTONIC PLATES

Category	Description	Frequency in study 3 (*N* = 20)	Frequency in study 2 (*N* = 21)[*]	Interpretation[†]
Divergent boundaries are on one plate				
Grow in middle	New material is being added to the middle of the plate, making it bigger.	5	0	Different means separate
Continuous	Both sides are the same and continuous, although there may be a small break.	5	0	Different means separate
No clear edge	There is no distinct edge. The two sides are not separate.	3	1	Obvious break means separate
No subduction	It is not a boundary because there is no subduction occurring.	5	5	Obvious break means separate
Definition	Student uses a definition to describe plate boundaries.	3	0	Plate-tectonic knowledge
Example	Student uses a real-world example to reason that it is one plate.	4	0	Plate-tectonic knowledge
Divergent boundaries divide two plates				
Clear edge	There is a distinct edge. The two sides are separate.	8	3	Obvious break means separate
Move apart	The two sides are moving in opposite directions, so there are two plates.	15	0	Plate-tectonic knowledge
Example	Student uses a real-world example to reason that it is two plates.	4	0	Plate-tectonic knowledge
"Boundary"	Because the term "boundary" is used, there are two plates by definition of a boundary.	3	1	Plate-tectonic knowledge
Use feature	Features are used to identify the boundary, explaining how it is known the boundary is there.	11	0	Plate-tectonic knowledge
Convergent boundaries divide two plates				
Different	The two sides are different from each other so they are separate plates.	3	0	Different means separate
Subduction	There are two plates because there is subduction happening.	15	7	Obvious break means separate
Clear edge	There is a distinct edge. The two sides are separate.	8	5	Obvious break means separate
"Boundary"	Because there is a boundary between the plates, there are two plates.	3	1	Plate-tectonic knowledge
Move separately	The two sides are moving in different directions, so there are two plates.	9	4	Plate-tectonic knowledge
Use features	Features are used to identify the boundary, explaining how it is known the boundary is there.	7	1	Plate-tectonic knowledge
Continents and oceans can be on one plate				
No activity	Activity defines a plate boundary.	2	0	Plate-tectonic knowledge
Continents and oceans are two plates				
Different	The two sides are different from each other so they are separate plates.	3	0	Different means separate
Land = plate	Continent or land and plate are used interchangeably.	3	0	Different means separate
Mix plates and boundaries				
Mix plates and boundaries	When describing plates, boundaries are instead described.	3	3	Plate-tectonic knowledge
Do not see continuous picture				
Arrows	Arrows are drawn in contradicting directions on a single plate.	5	7[§]	Plate-tectonic knowledge
Two plates per boundary	Because there are two plates at each boundary, the number of plates equals the number of boundaries times two.	0	3	Plate-tectonic knowledge

Note: This table shows the categories created from students' explanations for counting tectonic plates, the frequency with which students express statements within each category, and the interpretation of underlying factors causing each category.

[*] 21 of the 35 students who filled out the questionnaire gave explanations for counting tectonic plates.

[†] Different means separate is the deeply ingrained mental model that is used wherein "if the two sides are different, then they are separate entities (and the opposite)." Obvious break means separate is the deeply ingrained mental model that connotes "if there is an obvious break, then there are two separate entities (and the opposite)." Plate-tectonic knowledge is information students learned about plate tectonics that cannot be directly observed in the diagram.

[§] Twelve students overall drew contradictory arrows on the questionnaires, but only 7 students who wrote explanations drew them.

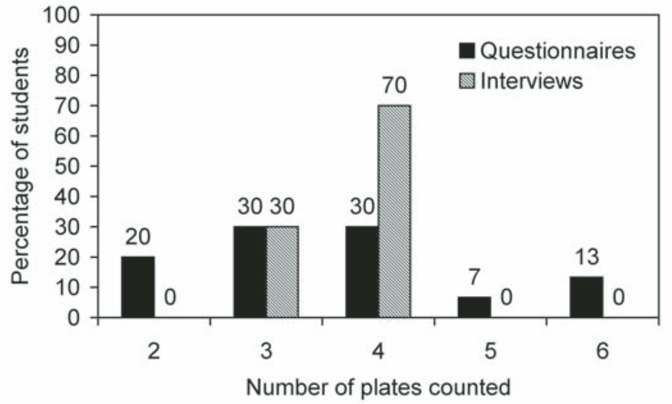


Figure 5. Percentage of students counting a specific number of plates on the questionnaires shown in Figure 2 and during interviews with the same figure. The correct answer is 4.

single plate. Two students numbered the three plate boundaries instead of the plates, and one student placed a number on either side of each boundary resulting in six total plates, two for each boundary.

For the question that asked students to draw lines along the plate boundaries (Fig. 2, question 2), 20 students attempted to delineate the boundaries, but none correctly drew lines along all three boundaries. More students drew lines (correctly or incorrectly) near the convergent boundaries than the divergent boundary (Table 2). No student drew lines along the transform boundary.

Students were asked to draw arrows indicating the direction of motion of the plates (Fig. 2, question 3). Of the 24 students who answered this question, more drew arrows at the convergent boundaries than the divergent boundary (20 and 21 compared to 13; Table 2). No student drew arrows at the transform boundary. Three of the 13 students who drew arrows at the divergent boundary indicated convergence, similar to the findings of Clark and Libarkin (2008), who reported 25% of students drawing converging arrows at the divergent boundary. Twelve students drew arrows within what they delineated as a single plate that indicted either compression or stretching within that plate.

On the questionnaire, 26 of the 29 students who indicated something at any plate boundary through labeling, drawing lines, or drawing arrows, did so for the left convergent boundary. Twenty-eight students did so for the right convergent boundary, and fewer students (n = 19) did so for the divergent boundary (Table 2).

Study 3

In study 3, students ($n = 20$) participating in one-to-one interviews were asked to determine the number of plates and locations of the plate boundaries on the plate-tectonic diagram used on the study 2 questionnaire. The ability to probe responses during interviews allowed us to examine students' explanations of their answers, giving us insight into their thought processes. Responses were evaluated from a qualitative methods perspective.

Participants

As part of a separate, larger study, the interviews for study 3 were conducted with 20 undergraduate students at four different institutions in the northeastern United States: two private universities, a medium-sized public university, and a large community college. The self-reported demographic information for the interviewed students is given in Table 4 (all names used are pseudonyms but do represent correct gender). The average age of the students was 22, 45% were female, and the race or ethnicity of the students was predominantly White.

Compared to the students in studies 1 and 2, the students interviewed in study 3 had, on average, taken more geoscience courses. However, 29% of study 2 participants had taken more than one introductory-level geoscience course. Only two of the 20 students interviewed in study 3 had taken a single geoscience course, although at least four others had taken only introductory-level courses. Interviewed students were recruited on a voluntary basis and were reimbursed $20 for their participation.

Data Collection

Semistructured interviews were conducted by one of the coauthors (Clark). A semistructured interview format was chosen because we wanted to both ask specific questions and have the flexibility to probe student responses, as warranted. The follow-up, probing questions varied from student to student, depending on their responses. A questionnaire very similar to that shown in Figure 2 was used to guide the interviews (for the actual questionnaire used in the surveys, see Clark and Libarkin, this volume). Interviewees were questioned about plate boundaries, plate motion, terminology, and other tectonic features and processes. The structured questions relevant to this study were:

1. Can you draw a line along the surface of any plate boundaries that you see?
2. How many plates do you see in this image?
3. How do you know that this [point to a specific location] was a plate boundary?

The first two questions were chosen because they are similar to questions on the questionnaire. The wording of the questions was sometimes slightly rephrased, and not all 20 interviewees were asked all of these questions. The first two questions were each asked of 19 students. The third question, which was asked of 11 students, was added to the protocol after a discussion between the interviewer and another coauthor (Kortz). The semistructured nature, wherein interviewees were encouraged to elaborate on their thinking throughout the interview, and the 30 min time constraint meant that not all anticipated questions were asked during every interview. The 30 min interviews were video- and audiotaped and transcribed verbatim.

The interviews were analyzed using thematic content analysis, a form of constant comparative analysis in which the researcher identifies patterns, or common themes, in the qualitative data

TABLE 4. DEMOGRAPHIC INFORMATION FOR INTERVIEWED STUDENTS (*N* = 20)

Student	Age	Gender	Race/ethnicity	High school geoscience course?	Number of geoscience courses	Major
Large community college in the Northeast						
Derrick	20	Male	White	Yes	2 to 5	Other
Eric	18	Male	White	Yes	2 to 5	Other
Frederick	21	Male	White	No	1	Other
Gary	20	Male	White	No	2 to 5	Other
Hazel	20	Female	White	Yes	2 to 5	Geology, STEM*
Medium-sized public university in the Northeast						
Ian	32	Male	White	No	2 to 5	Geology
Julia	31	Female	White	Yes	>5	Geology
Karla	20	Female	Hispanic	No	>5	Geology
Laura	20	Female	White	No	2 to 5	Geology
Mike	20	Male	White	No	2 to 5	Geology
Nicholas	23	Male	White	Yes	2 to 5	Geology
Oliver	26	Male	White	Yes	>5	Geology
Pamela	37	Female	White	No	>5	Geology
Private university no. 1 in the Northeast						
Ashley	21	Female	White	No	>5	STEM
Bailey	20	Female	White	No	2 to 5	STEM
Carl	19	Male	White	Yes	2 to 5	Geology, STEM
Private university no. 2 in the Northeast						
Rose	18	Female	White	Yes	1	Geology
Shawn	19	Male	White	No	2 to 5	Geology, STEM
Tyler	21	Male	White	No	>5	Geology
Victor	20	Male	Mixed	No	>5	Geology, other

Note: Self-reported demographic information for the interviewed students. Names are pseudonyms.
*STEM consists of science, technology, engineering, and math majors other than geology.

(Glaser and Strauss, 1967; Lincoln and Guba, 1985; Erlandson et al., 1993; Denzin and Lincoln, 1998; Patton, 2002). With the goal of maintaining the intent of the student responses and facilitating analysis of multiple student perspectives, this method of analysis groups student comments into common categories that reflect their views during the interviews. After categories are identified, they are grouped and organized into more abstract themes, permitting interpretation of deeper meaning.

The interviews were initially coded for themes that explained why students counted a certain number of plates. Student answers were divided into segments representing distinct ideas relevant to our research questions. The segments ranged in size from phrases to multiple sentences. We grouped related segments into categories, and as we analyzed additional interviews, we compared categories and modified them to incorporate the new information. After groups of five interviews were analyzed, we revisited the categories to make any necessary major modifications, such as splitting categories that were too heterogeneous or combining categories that conveyed the same idea. Once the categories captured the variations in the students' ideas, a final list of themes was generated. In this way, common themes representative of students' answers emerged from the data. After the categories were generated, we read through the interviews again, and recoded them using the categories. Two authors (Kortz and Smay) coded all interviews, and their inter-rater reliability was initially 79%. After discussion, full agreement was achieved.

Once the themes explaining the number of plates were established, we analyzed them to determine if there were any underlying explanations for the students' perspectives. To do so, we grouped the themes, looking for similarities that may indicate that previous knowledge was activated.

Results

Eight of the 20 students (40%) initially counted three plates in the image, rather than the correct number of four (Fig. 5). However, two of the students who initially answered three plates changed their answers to four plates upon discussion with the interviewer, and this is the answer recorded in Figure 5. The following is an example of one of those discussions:

> Interviewer: Let's go back to where you counted three plates. Where are the three plates at?
>
> Bailey: So, there's the one on the left side, here. The middle one [the ocean], that's two, and this third one right here [pointing to continent]….
>
> Interviewer: You call that [ridge] a plate boundary, but you're saying it [the ocean] is one plate. [The interviewer was revisiting a definition provided by Bailey earlier in the interview.]
>
> Bailey: Yeah, I am saying that. (Laughs.) So, then that would make four [plates].

Six of the eight students who initially counted three plates debated with themselves as to whether the image showed three or four plates; three additional students out of the 20 interviewed also debated between seeing three or four plates before settling on four plates. Victor is an example of one of the students that had trouble deciding between three and four plates:

> Victor: I mean I could be convinced that it [the ocean] was one plate. I could be convinced that it is two plates. I'm not 100% certain that it's two plates. But there is definitely a divide in there. And it's definitely like going into two different directions.

The categories derived from student explanations for counting plates and the overarching themes derived from them (Table 3) are described and discussed in the following section. Students' intended major and the type of school they attended did not correlate with the number of plates they counted and the difficulties they had.

DISCUSSION

Our discussion of why students counted a certain number of plates is derived primarily from students' responses during the interviews in study 3, although it also draws from student responses on the questionnaire in study 2. We are using findings from the qualitative aspects of our research to provide explanations for the quantitative findings. Because our three studies overlap in topic and scope, we arrange our discussion by topic instead of by study, pulling together common threads of all three.

Although, based on the number of geoscience courses taken, the overall average geoscience experience level of students in study 3 was higher than that for students in studies 1 and 2, four study 3 students who were interviewed at the large community college in the northeast United States had enrolled in the same course during a different semester as students from the same college who participated in study 2. We did not collect data on which geoscience courses students in study 3 had taken, and it is possible that the additional courses did not emphasize plate-tectonic concepts. In addition, student responses on the interviews were very similar to responses on the questionnaires from study 2. Students in both studies were making similar mistakes when counting the number of plates. This similarity in responses leads us to interpret a similarity in reasoning. Since the study 3 students who had taken more geoscience courses were continuing to make similar mistakes to students in their first geoscience course, they cannot be considered experts, and this study should not be considered an expert-novice study. However, we do acknowledge that our findings could potentially be biased by making interpretations of introductory student thought processes based on the reasoning of some students who have taken more geoscience courses.

We begin the discussion with the ways in which students' conceptions of divergent boundaries affected their determination of the number of plates present in each illustration. We continue with their conceptions of convergent boundaries, followed by additional difficulties observed that affected the counting of plates. We use these conceptions to hypothesize why students counted a certain number of plates in studies 1 and 2. Finally, we put forward possible underlying factors for these difficulties.

Divergent Boundaries

The most common reason students miscounted the number of plates is that they did not consider the ocean ridge to be a plate boundary. Students who counted three plates instead of four plates incorrectly claimed that the ocean floor consisted of just one plate, while concurrently and correctly identifying the ocean ridge. Students provided a range of explanations for this perspective that centered on misinterpreting the ocean divergent plate boundary.

When describing the ocean ridge, students cited the lack of subduction as the most prevalent reason why the ridge was not a plate boundary. The following conversation exemplifies this perspective:

> Interviewer: How did you know there was a boundary between two plates there [pointing to the convergent boundary]?
>
> Mike: Because of the subduction zone….
>
> Interviewer: And how do you know this middle one [ocean] is only one plate?
>
> Mike: Because there is no subduction zone dividing it.

Students apparently view subduction as being an easily recognized division between two plates, so if a subduction zone is not present, such as at an ocean ridge, then a plate boundary is not present. Other students cited the lack of a distinct edge at the ocean ridge as a reason why the ocean floor did not contain a plate boundary, as summarized by Karla, "There is no separate, like definite boundary that separates it from the next plate."

Those students who only identified three plates seem to have categorized divergence as an intraplate process rather than a plate boundary. It appears that they thought that the ocean ridge represented a plate that was growing in the middle instead of two plates moving apart. For example, Julia used the analogy of the growth of bones to explain how a single plate could grow in the middle:

> Julia: New material is coming up and it's growing but from the inside… I don't know too much about bones, but I've had the growth of bones explained to me… I think bones might grow from areas that are here and here, but it's all the same bone even though the ends are getting further apart. So I think that's how I think of the spreading center there.

Students with this perspective correctly understood that the two sides of a divergent boundary were separating, but, as Karla further explained, "It doesn't move relatively to other plates by itself. It moves within one plate." Because the two plates look the same on both sides of the ridge, our students apparently viewed the ridge as a feature within one plate (such as a "puncture" or "cut") instead of something that divides two separate plates.

Four students also included interpretations of real-world examples to explain why ridges are not considered plate boundaries. For example, Julia did not think Iceland is defined as a plate boundary, so she did not consider a spreading center a divergent boundary between two plates.

Student responses on the questionnaires in study 2 further supported our interpretation of the interview data. Four of the 12 students who numbered the plates counted the entire ocean floor as one combined plate. In addition, more students identified convergent boundaries compared to the divergent boundary (Table 2).

The failure to recognize the ocean ridge as a plate boundary cannot be attributed to the students failing to see it on the diagram. Of the students who labeled features on the questionnaire, approximately the same number labeled a physical feature at the divergent boundary as at either of the convergent boundaries (6 at the divergent boundary versus 6 and 7 for the left and right convergent boundaries, respectively; Table 2), suggesting that students saw the ocean ridge as frequently as they saw subduction-related features. In addition, when initially asked during the interviews what they saw in the image, 17 of the 20 students said they saw a ridge or divergent boundary. The other three students discussed the ridge or divergent boundary later during the interview without being prompted by the interviewer.

In contrast, students who correctly identified four tectonic plates correctly explained that divergent boundaries divide the ocean basin into two separate plates. The most common explanation students gave for a boundary existing between two plates at the ridge was that the two plates moved apart ($n = 15$). These students also viewed the distinct edges to the two plates as a boundary. Mike summarized this perspective:

> Mike: Well, you can say it's two plates because it's got this big divider in the middle, and one plate's going this way and one plate's going this way. So you have two sections going opposite directions so you'd say, "Ah, it's obviously two different things."

In addition, many students ($n = 11$) also used physical features at the boundary (such as the ridge) to explain how they knew there was a boundary dividing two plates. As Rose explained, "We've come to understand that when you see a ridge of that nature, one plate is moving in one direction, the other is moving away from it. These two plates are divergent."

Convergent Boundaries

Students overwhelmingly recognized subduction zones as boundaries between two plates. As previously mentioned, they cited the presence of a subduction zone to explain why there were separate plates on either side. It appears that they saw the distinct edge of the plate and inferred that the plates were moving in different directions. The students also explained that subduction zones separate two plates because either side is different in appearance from the other side, such as oceanic lithosphere on one side and continental lithosphere or volcanic islands on the other side. Tyler explained this rationale:

> Tyler: From what I've learned [about] the difference between oceanic lithosphere and continental lithosphere.… As [the oceanic lithosphere] is being subducted, that implies that there is no real connection between these. So they are two distinct bodies.

Additional Difficulties Counting Plates

On the study 2 questionnaire, seven students did not identify the divergent boundary through labeling, identifying features, or indicating motion at the divergent boundary, and they answered the question asking them to count the plates. Of these seven students, four counted two plates and three counted three plates. In contrast, 16 students labeled, identified features, or indicated motion at the divergent boundary in addition to counting the plates. These students counted between two and six plates. This large range of values suggests that other factors are involved in addition to not counting the divergent boundary as a plate boundary. These factors are discussed next.

Many students appeared to view the continents as being a separate plate from oceans, whether or not subduction is occurring. For example, Shawn explained, "They're different plates in that one is oceanic and one is continental. So, I'd say they're different plates." Victor further explained using a real-world example why continents are different plates than the oceans, "I guess, I would say they are still different plates. The North American plate would have, like, distinct chemical and physical properties from the ocean basalt." Students may also view continents as being the same thing as plates, ignoring the oceans. When asked why she hesitated in answering that she counted four plates, Ashley stated, "I was hesitating because at first I was, this is really bad. Ok, I am going to say this. Because usually I think that the common misconception is that continents are different plates." This perspective was investigated only in the last few interviews, so we are unsure how many of the interviewed students held this view, but it is probably more widespread than we report in Table 3. Of the 20 students who drew lines along plate boundaries on the questionnaire, three of the students drew lines along the coast, which suggests that they think the plate boundary is the same as the coastline between the land and the ocean. These findings are in agreement with Marques and Thompson (1997).

Some of our students also used the terms plates and plate boundaries interchangeably. When asked about plates in study 3, two interviewees responded by initially describing plate boundaries in addition to plates. In study 2, two of the 12 students numbered plate boundaries instead of plates. We interpret this to suggest that students are not comfortable with the terminology, but this apparent confusion may simply be a result of carelessness when reading or hearing the question.

Lastly, some students seemed to consider plate boundaries to be separate, isolated features rather than interconnected boundaries that define larger, continuous plates. When asked to place numbers on each of the plates on the study 2 questionnaire, one of the 12 students who numbered plates placed a number on either side of each of the boundaries and stated that the illustration showed six plates—"two for each boundary." This student appears to have not considered that two of these plates were continuous with each other. Of the 12 students in study 2 who indicated multiple plate directions within a single plate, none counted the plates correctly.

This fragmented understanding of plate-tectonic processes was also observed when the transform boundary between the two divergent sections of the ocean ridge (Fig. 2) was discussed during study 3 interviews. When 12 of the interviewed students were asked to draw arrows showing plate motion relative to the transform boundary, four of them settled on motion along the transform boundary that was opposite to the direction of motion for the divergent boundary overall. Additional students corrected this initial internally inconsistent conceptualization after discussion with the interviewer, as Nicholas explained, "[Drawing the arrows the] other way would make a lot more sense since they are in congruency with the piece of it as a whole instead of just that one boundary."

Revisiting the Number of Plates Counted

Data from all three studies suggest that a large percentage of students have difficulties counting the correct number of tectonic plates. Fifty percent of the 529 students from all three studies correctly counted the number of plates. The use of three different diagrams shown to hundreds of students at eight different schools signifies that it is not a problem isolated to a particular group of students or a specific instructor. Neither is it a matter of students not understanding any one particular diagram. The high number of incorrect responses after peer discussion of the map-view ConcepTest question in study 2 can be interpreted to suggest that student concepts on this topic are deeply ingrained, or that no one in many of the groups was confident enough in the correct answer to try to sway the other students.

Based on insights from interview data, we can postulate on causes of the alternative conceptions documented in studies 1 and 2. On the ConcepTest map-view question (Fig. 1A), 33% of students incorrectly chose three plates, and the following explanations may elucidate the students' reasoning: if the divergent boundary is not counted there are three plates; there are three land masses; and there are three plate boundaries. Additional research could sort out these interpretations to determine which of these strategies, if any, students use.

On the ConcepTest question showing the block diagram of a convergent and divergent boundary (Fig. 1B), 55% of students chose the correct answer of three plates, but nearly one third of students chose two plates. As with the previous question, this response corresponds with the views that a divergent boundary is not a plate boundary, continents are plates, or plate boundaries are counted instead of plates.

On the study 2 questionnaire, 30% of students (as many as those who correctly counted four plates) counted three plates for the following possible reasons: they did not count the divergent boundary as a plate boundary; the diagram contained two landmasses and one ocean; and there are three plate boundaries. Another 20% of the students in study 2 chose two plates. This count may have been because they saw two subduction zones, they counted the continent and ocean as two separate plates, or they counted the island arc and the continent each as a plate. No interviewees gave this answer, so we do not have data that distinguishes which, if any, of these three interpretations is more likely.

A wide range of students had difficulties counting the number of tectonic plates in all three studies. Whether geology majors or nonscience majors, and regardless of the institution, a significant percentage of students were not able to correctly count the number of tectonic plates. Although, for the most part, the interviewed students have taken more geoscience courses than the students in the other two studies, more than half of them had difficulties in determining the number of plates in the image. This result further justifies our application of the interview findings to explain the reasoning of introductory-level students in studies 1 and 2. These questions were not trivial to the students. For example, when Pamela, who has taken more than five geoscience courses, was asked, "What is it that you are keying in on that would make you say, 'Oh, this is one plate and this is another plate'?" she responded, "You ask very good questions, by the way. This is really challenging."

Possible Reasons for All Difficulties Counting Plates

Once we determined why students had difficulties counting the numbers of plates, we examined underlying causes for those difficulties. Because students use their prior knowledge to process new information, we examined the explanations for counting the number of plates for preconceived knowledge structures used by the students to make sense of information taught about plate tectonics.

Plate tectonics is at a scale that is both spatially and temporally unobservable to students, (Jacobi et al., 1996; Gobert, 2000). Previous knowledge of specific plate-tectonic concepts is most likely based on what students have been taught in a classroom, experienced themselves, or seen in books or the media (Libarkin, 2006). However, students may also apply more tangential, deeply ingrained prior knowledge (e.g., Vosniadou and Brewer, 1992; diSessa, 1993; Vosniadou, 2002) to make sense of the information presented about plate tectonics. It appears that our students' understanding of plate tectonics did not match the scientifically accepted explanation, which would indicate that our students did not achieve complete conceptual change from their prior knowledge when learning about plate tectonics. We determined that, in addition to knowledge of plate tectonics, there were two deeply ingrained preexisting mental models that students applied when they examined an image showing plate tectonics and were asked to count the tectonic plates (Table 3, last column).

The first deeply ingrained preexisting mental model is that an obvious break indicates separate entities. In other words, for things to be separate entities, such as separate plates, there needs to be an obvious break between them. The opposite is also true—if there is no obvious break, then it is not a separate entity. This application of prior knowledge may help explain why students have difficulties conceptualizing that ocean ridges are plate boundaries, because although the lithosphere becomes very thin at the divergent boundary, it may not be perceived by students as a distinct break between the two sides. Subduction zones, on the other hand, provide an obvious separation between the two sides, so students have no problems visualizing them as dividing two separate plates. Conceptual change is not required.

The other deeply ingrained preexisting mental model is that if things are different, they are separate entities. Again, the reverse is true—if something is the same, it is not two separate entities. This mental model also helps explain why students have difficulties viewing ocean ridges as plate boundaries. The plates on either side of an ocean ridge are the same, which makes it difficult for some students to view them as separate entities. On the other hand, the plates on either side of a subduction zone are different, in that one is a continent or has volcanic islands, so the students view either side as a different plate. Again, conceptual change is not required to learn that subduction zones divide two plates.

The "different means separate" preexisting mental model also helps to explain why some students view land and ocean as being separate plates. For example, a student could observe differences between land and ocean and conclude they are very different. Consequently, when this hypothetical student learns that land is made of continental lithosphere and the oceans are underlain by oceanic lithosphere, he or she might continue to cling to a belief that there is a fundamental difference between the land and the sea and therefore inaccurately conclude that pieces of continental and oceanic lithosphere cannot be found on the same tectonic plate.

In addition to deeply ingrained mental constructs, students use information they were taught while learning plate tectonics. This taught information would include the direction of plate motion, where new plates are created, or the application of real-world examples to unknown plate boundaries. This information cannot be directly gathered from the image without some interpretation in light of plate tectonics.

Derrick is a student in study 3 who initially said there were three plates in the image, although he appeared to waiver as he applied his preexisting mental models to the question, as shown in the following discussion:

> Interviewer: So give me an argument that you can think of for [the ocean] being one plate and another argument why it might be two plates.
>
> Derrick: Well, it could be one plate just kind of growing. That's what I personally, if I had to say something, that's what I would agree with. That it's just one plate just kind of separating and—but then when you think about it then I feel that would be two plates then.
>
> Interviewer: Because?
>
> Derrick: Because they're moving apart even though it could still be the same rock it's just stretching and breaking apart. But I feel once it stretches and breaks apart it's no longer the same one, so it would make it two even though it's still conjoined by the new rock that forms there.

Derrick reasoned through whether or not a divergent boundary separates one or two plates using both examples of deeply ingrained mental models described previously. One plate growing results in a plate that is the same on both sides, which would trigger the different means separate entities mental model (or rather the inverse, that same means not separate). However, he also utilizes the knowledge that an obvious break means separate entities when he explains that if a plate breaks apart, then it would be two separate plates. For this student, the different means separate entities explanation eventually "wins out" and the student decides the ocean is made up of one plate.

Nine interviewees in study 3 did not express comments that could be explained by the deeply held mental model that different means separate entities, and all of them counted the plates correctly. Of the 11 students who either incorrectly counted plates or debated their answer before settling on the correct answer, most of them (n = 8) used reasoning that could be explained by the deeply ingrained knowledge that different means separate entities. Therefore, in some students, this preexisting knowledge appears to be inappropriately relied upon when counting plates, resulting in incorrect answers.

IMPLICATIONS FOR GEOSCIENCE EDUCATION AND GEOCOGNITION

The strength of alternative conceptions may be unintentionally reinforced by commonly used terms and instruction practices. For example, in the experiences of the authors, textbooks (e.g., Monroe et al., 2007) and instructors tend to talk about oceanic plates and continental plates at plate boundaries. Although instructors are referring to the type of plate specifically at that boundary and not overall, many students may view the entire plate as being identified by whether it is land or water. Also, if the types of plate boundaries are discussed as unique entities, (e.g., represented in separate block diagrams) without explicit instruction showing how the boundaries are interconnected, then some students are unlikely to grasp the big picture of plate tectonics. One exercise that does address this issue is, "Discovering Plate Boundaries" developed by Sawyer et al. (2005). An extensive list of other interactive instructional methods, including Lecture Tutorials (Kortz et al., 2008; Kortz and Smay, 2010) and ConcepTests (McConnell et al., 2003, 2006) that are designed specifically for geoscience topics such as plate tectonics can be found on the Pedagogy in Action (2010) Web site hosted by the Science Education Resource Center (SERC) at Carleton College, Minnesota.

Although most of the interviewed students had taken multiple geoscience courses, many still had difficulties counting the number of plates. Based on these results, use of additional active-learning instruction techniques regarding plate tectonics in core geology courses is recommended as a means to further enhance student understanding of this fundamental concept. A factor that we found to be interesting is that most students gave explanations based on a reasoned approach derived from the processes operating at divergent boundaries, even if they did not consider the boundary as separating two plates. In this regard, we see those students as understanding the process but forgetting the definition. If they can explain the processes that occur at divergent boundaries, then they have made an important stride in understanding plate tectonics. For nonscience majors, this raises the question of how important is it that students are able to define plate boundaries?

CONCLUSIONS

Not only did this study identify some of the difficulties that students have in understanding plate tectonics, but it also illustrates the strength of a mixed-methods research design. Our study both confirms that students struggle to successfully complete a seemingly simple task (identify tectonic plates) and also explores the reasons behind these struggles. By blending the qualitative and quantitative research traditions, our findings are stronger and more diverse than if we had conducted just one type of study. The quantitative data allowed the rich detail provided by the interviews to be applied to a larger data set, whereas the explanatory power of qualitative interview data gave meaning to the frequently chosen misconceptions observed with questionnaires and ConcepTest questions. In addition, by conducting our student interviews, we were able to identify some novel (and unforeseen) explanations that the ConcepTest questions or questionnaire would have missed. Such findings support Teddlie and Tashakkori's (2003) assertion that mixed-method studies can blend the best of both traditions and produce a self-consistent set of results.

We found that half of the students from a variety of institutions were not able to correctly identify the number of plates on an image. This difficulty was present across three different diagrams, at a variety of institutions, and with introductory-level through upper-division geoscience students. These difficulties suggest that some students are retaining alternative conceptions of basic plate tectonics well into their upper-level courses, which points toward the difficulty of conceptual change on this topic. Additional research on a novice-expert continuum would help address when and why geology students begin to think of plate tectonics more like expert geoscientists.

Students did not have difficulty identifying subduction as a process occurring between two plates. The visual image of a subducting slab plunging into the asthenosphere with two different sides allowed the students to overwhelmingly view convergent boundaries as separating two plates. Divergent boundaries, on the other hand, are less obvious. Although students correctly described processes that occur at the ocean ridge, some did not count the ridge as a plate boundary. Because both sides of the ocean ridge look identical and are not divided by a subduction zone, some students viewed the divergent boundary as one plate growing in the middle. By seeing the two ocean plates as one, students undercounted the number of plates present. Other difficulties students had when counting plates were that some confused the boundary between continents and oceans as the boundary between plates, and some focused their attention on individual boundaries while failing to see a larger picture. Future research on the ways in which students interpret a plate boundary and how that affects their counting of plates would help to distinguish between these explanations.

An underlying cause of student difficulties when counting plates is that they incorrectly apply preexisting mental models when counting plates. The two deeply ingrained mental models identified in this study are that if entities are different, then they are considered separate, and if there is an obvious break, then they are considered separate entities. Therefore, to identify that a divergent boundary separates two plates, students need to utilize knowledge they were specifically taught about plate tectonics rather than utilize a mental model based on inappropriately applied prior knowledge. Future research can help identify causes of the inappropriate applications of preexisting mental models and distinguish if students consistently apply these mental models in different situations. In regard to education, we recommend for instructors to be made aware of these preexisting mental models. Therefore, they can design activities to directly confront them so conceptual change can occur, and students will be able to have a fuller understanding of what a tectonic plate is and how plates interact with one another.

ACKNOWLEDGMENTS

We thank all students for their participation. This manuscript benefited greatly from comments by Alison Stokes and two anonymous reviewers. Partial support for this work was provided by the National Science Foundation's DUE Course, Curriculum, and Laboratory Improvement programs under awards 0716397 and 0837185. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES CITED

Ausubel, D.P., 1968, Educational Psychology: A Cognitive View: New York, Holt, Rinehart, and Winston, 685 p.

Bartlett, F.C., 1932, Remembering: An Experimental and Social Study: Cambridge, UK, Cambridge University Press, 344 p.

Bransford, J.D., Brown, A.L., and Cocking, R.R., eds., 2000, How People Learn: Brain, Mind, Experience, and School (expanded edition): Washington, D.C., National Academy Press, 374 p.

Chi, M.T.H., 2005, Common sense conceptions of emergent processes: Why some misconceptions are robust: Journal of the Learning Sciences, v. 14, p. 161–199, doi:10.1207/s15327809jls1402_1.

Clark, S.K., and Libarkin, J.C., 2008, Post-instruction alternative conceptions about plate tectonics held by non-science majors: Geological Society of America Abstracts with Programs, v. 40, no. 6, p. 366.

Clark, S.K., and Libarkin, J.C., 2011, this volume, Designing a mixed-methods research instrument and scoring rubric to investigate individuals' conceptions of plate tectonics, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(07).

Crouch, C.H., and Mazur, E., 2001, Peer instruction: Ten years of experience and results: American Journal of Physics, v. 69, p. 970–977, doi:10.1119/1.1374249.

Dahl, J., Anderson, S.W., and Libarkin, J.C., 2005, Digging into earth science: Alternative conceptions held by K–12 teachers: Journal of Science Education, v. 12, p. 65–68.

Denzin, N., and Lincoln, Y.S., eds., 1998, Collecting and Interpreting Qualitative Materials: Thousand Oaks, California, Sage Publications, 462 p.

diSessa, A., 1993, Towards an epistemology of physics: Cognition and Instruction, v. 10, p. 105–225.

diSessa, A., 2002, Why "conceptual ecology" is a good idea, *in* Limón, M., and Mason, L., eds., Reconsidering Conceptual Change: Issues in Theory and Practice: Dordrecht, the Netherlands, Kluwer Academic Publishers, p. 29–60.

diSessa, A., and Sherin, B., 1998, What changes in conceptual change?: International Journal of Science Education, v. 20, p. 1155–1191, doi:10.1080/0950069980201002.

Earth Science Literacy Initiative, 2009, Earth Science Literacy Initiative: http://www.earthscienceliteracy.org/index.html (accessed 13 September 2009).

Erlandson, D.A., Harris, E.L., Skipper, B.L., and Allen, S.D., 1993, Doing Naturalistic Inquiry: A Guide to Methods: Newbury Park, California, Sage Publications, 199 p.

Eshach, H., and Schwartz, J.L., 2006, Sound stuff? Naïve materialism in middle-school students' conceptions of sound: International Journal of Science Education, v. 28, p. 733–764, doi:10.1080/09500690500277938.

Feig, A.D., 2011, this volume, Methodology and location in the context of qualitative data and theoretical frameworks in geoscience education research, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(01).

Glaser, B.G., and Strauss, A.L., 1967, The Discovery of Grounded Theory: Strategies for Qualitative Research: New York, Aldine Publishing Company, 271 p.

Gobert, J.D., 2000, A typology of causal models for plate tectonics: Inferential power and barriers to understanding: International Journal of Science Education, v. 22, p. 937–977, doi:10.1080/095006900416857.

Greene, J.C., Caracelli, V.J., and Graham, W.F., 1989, Toward a conceptual framework for mixed-method evaluation designs: Educational Evaluation and Policy Analysis, Autumn, v. 11, p. 255–274.

Grotzinger, J., and Jordan, T.H., 2010, Understanding Earth (6th ed.): New York, W.H. Freeman and Company, 654 p.

Hewson, M.G., and Hewson, P.W., 1983, Effect of instruction using students' prior knowledge and conceptual change strategies on science learning: Journal of Research in Science Teaching, v. 20, p. 731–743, doi:10.1002/tea.3660200804.

Jacobi, D., Bergeron, A., and Malvesy, T., 1996, The popularization of plate tectonics: Presenting the concepts of dynamics and time: Public Understanding of Science (Bristol, England), v. 5, p. 75–100, doi:10.1088/0963-6625/5/2/001.

Kelso, P.R., Brown, L.M., Mintzes, J.J., and Englebrecht, A.C., 2000, Key concepts in geoscience subdisciplines, teaching strategies, and core course requirements: Results of a national survey of geoscience faculty [abs.]: Eos (Transactions, American Geophysical Union), v. 41, p. 304, updated online http://geology.lssu.edu/survey.html (accessed November 2005).

Kortz, K.M., and Smay, J.J., 2010, Lecture Tutorials for Introductory Geoscience: New York, W.H. Freeman and Company, 88 p.

Kortz, K.M., Smay, J.J., and Murray, D.P., 2008, Increasing student learning in introductory geoscience courses using Lecture Tutorials: Journal of Geoscience Education, v. 56, p. 280–290.

Kuhn, T., 1970, The Structure of Scientific Revolutions (2nd ed.): Chicago, University of Chicago Press, 210 p.

Kusnick, J., 2002, Growing pebbles and conceptual prisms—Understanding the source of student misconceptions about rock formation: Journal of Geoscience Education, v. 50, p. 31–39.

Libarkin, J.C., 2006, College student conceptions of geological phenomena and their importance in classroom instruction: Planet, v. 17, p. 1–9.

Libarkin, J.C., and Anderson, S., 2005, Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory: Journal of Geoscience Education, v. 53, p. 394–401.

Libarkin, J.C., Anderson, S., Dahl, J., Beilfuss, M., Boone, W., and Kurdziel, J., 2005, College students' ideas about geologic time, Earth's interior, and Earth's crust: Journal of Geoscience Education, v. 53, p. 17–26.

Lincoln, Y.S., and Guba, E.G., 1985, Naturalistic Inquiry: Newbury Park, California, Sage Publications, 416 p.

Marques, L., and Thompson, D., 1997, Misconceptions and conceptual changes concerning continental drift and plate tectonics among Portuguese students aged 16–17: Research in Science & Technological Education, v. 15, p. 195–222, doi:10.1080/0263514970150206.

Mazur, E., 1997, Peer Instruction: A User's Manual: Upper Saddle River, New Jersey, Prentice Hall, 253 p.

McConnell, D.A., Steer, D.N., and Owens, K.D., 2003, Assessment and active learning strategies for introductory geology courses: Journal of Geoscience Education, v. 51, p. 205–216.

McConnell, D.A., Steer, D.N., Owens, K.D., Knott, J.R., Van Horn, S., Borowski, W., Dick, J., Foos, A., Malone, M., McGrew, H., Greer, L., and Heaney, P.J., 2006, Using ConcepTests to assess and improve student conceptual understanding in introductory geoscience courses: Journal of Geoscience Education, v. 54, p. 61–68.

McConnell, D.A., Steer, D.N., Knight, C., Owens, K.D., and Park, L., 2008, The Good Earth: An Introduction to Earth Science: Dubuque, Iowa, McGraw-Hill, 526 p.

McVee, M.B., Dunsmore, K., and Gavelek, J.R., 2005, Schema theory revisited: Review of Educational Research, v. 75, p. 531–566, doi:10.3102/00346543075004531.

Mikkilä-Erdmann, M., 2001, Improving conceptual change concerning photosynthesis through text design: Learning and Instruction, v. 11, p. 241–257, doi:10.1016/S0959-4752(00)00041-4.

Monroe, J.S., Wicander, R., and Hazlett, R., 2007, Physical Geology: Exploring the Earth (6th ed.): Belmont, California, Thomson Brooks/Cole, 690 p.

National Research Council, 1996, National Science Education Standards: Washington, D.C.: National Academy Press, 272 p. (http://www.nap.edu/catalog.php?record_id=4962; accessed 13 September 2009).

Özdemir, G., and Clark, D.B., 2007, An overview of conceptual change theories: Eurasia Journal of Mathematics, Science, and Technology Education, v. 3, p. 351–361.

Patton, M.Q., 2002, Qualitative Evaluation and Research Methods (3rd ed.): Thousand Oaks, California, Sage Publications, 598 p.

Pedagogy in Action, 2010, The Science Education Research Center Portal for Educators: http://serc.carleton.edu/sp/index.html (accessed 30 March 2010).

Pellegrino, J.W., Chudowsky, N., and Glaser, R., eds., 2001, Knowing What Students Know: The Science and Design of Educational Assessment: Washington, D.C., National Academy Press, 382 p. (http://www.nap.edu/catalog.php?record_id=10019; accessed 28 August 2009).

Piaget, J., 1952, The Origins of Intelligence in Children: New York, International University Press, 419 p.

Pintrich, P.R., Marx, R.W., and Boyle, R.A., 1993, Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change: Review of Educational Research, v. 63, p. 167–199.

Posner, G.J., Strike, K.A., Hewson, P.W., and Gerzog, W.A., 1982, Accommodation of a scientific conception: Toward a theory of conceptual change: Science Education, v. 66, p. 211–227, doi:10.1002/sce.3730660207.

Rutherford, F.J., and Ahlgren, A., 1991, Science for All Americans: New York, Oxford University Press, 272 p. (http://www.project2061.org/publications/sfaa/online/sfaatoc.htm; accessed 13 September 2009).

Sawyer, D.S., Henning, A.T., Shipp, S., and Dunbar, R.W., 2005, A data-rich exercise for discovering plate-boundary processes: Journal of Geoscience Education, v. 53, p. 65–74.

Scott, P., Asoko, H., and Leach, J., 2007, Student conceptions and conceptual learning in science, *in* Abell, S., and Lederman, N.G., eds.,, Handbook of Research on Science Education: Mahwah, New Jersey, Lawrence Erlbaum Associates, p. 31–56.

Sibley, D.F., 2005, Visual abilities and misconceptions about plate tectonics: Journal of Geoscience Education, v. 53, p. 471–477.

Steer, D.N., Knight, C.C., Owens, K.D., and McConnell, D.A., 2005, Challenging students' ideas about Earth's interior structure using a model-based, conceptual change approach in a large class setting: Journal of Geoscience Education, v. 53, p. 415–423.

Steer, D.N., McConnell, D., Gray, K., Kortz, K.M., and Liang, X., 2009, Analysis of student responses to peer-instruction conceptual questions answered using an electronic response system: Trends by gender and ethnicity: Science Educator, v. 18, p. 30–38.

Teddlie, C., and Tashakkori, A., 2003, Major issues and controversies in the use of mixed methods in the social and behavioral sciences, *in* Tashakkori, A., and Teddlie, C., eds., Handbook of Mixed Methods in Social and Behavioral Research: Thousand Oaks, California, Sage Publications, p. 3–50.

Trend, R., 2000, Conceptions of geological time among primary teacher trainees, with reference to their engagement with geoscience, history, and science: International Journal of Science Education, v. 22, p. 539–555, doi:10.1080/095006900289778.

Vosniadou, S., 1994, Capturing and modeling the process of conceptual change: Learning and Instruction, v. 4, p. 45–69, doi:10.1016/0959-4752(94)90018-3.

Vosniadou, S., 2002, On the nature of naïve physics, *in* Limón, M., and Mason, L., eds., Reconsidering Conceptual Change: Issues in Theory and Practice: Dordrecht, the Netherlands, Kluwer Academic Publishers, p. 61–76.

Vosniadou, S., 2007, Conceptual change and education: Human Development, v. 50, p. 47–54, doi:10.1159/000097684.

Vosniadou, S., and Brewer, W.F., 1992, Mental models of the Earth: A study of conceptual change in childhood: Cognitive Psychology, v. 24, p. 535–585, doi:10.1016/0010-0285(92)90018-W.

Manuscript Accepted by the Society 23 June 2010

Printed in the USA

The Geological Society of America
Special Paper 474
2011

Student interpretation of a global elevation map: What it is, how it was made, and what it is useful for

Sandra Swenson*
Columbia University Teachers College, 525 West 120th St., New York, New York 10027, USA

Kim Kastens[†]
Lamont-Doherty Earth Observatory, Columbia University, 61 Route 9W, Palisades, New York 10964, USA

> *Looking at a contour map, the student sees lines on a paper,*
> *the cartographer a picture of a terrain.*
>
> —Thomas S. Kuhn, *The Structure of Scientific Revolutions* (1962)

ABSTRACT

Visual representations of scientific data make these data accessible and enable students to examine the evidence used to build scientific arguments and test theories, even when the underlying data set is large or complicated. It is becoming more common in science education to use data visualizations based on data that students did not collect themselves. Teachers and instructional designers need to understand how students perceive and interpret such visualizations. This research examined the nature of students' interpretations about a colored, shaded-relief global digital elevation map useful for reasoning about a wide range of Earth processes. One hundred and ninety-six middle and high school students wrote answers to three open-ended questions while viewing the map projected on an overhead screen: "What do you think this is?" "How do you think this was made?" and "What do you think this is useful for?" Nearly half the students surveyed made no mention of topography/bathymetry or an equivalent concept. Twenty percent of the students misinterpreted the map to contain information other than elevation, including inappropriate interpretations such as water, temperature, and weather. Over half of the students did not describe any aspect of data acquisition as a component of the data map creation. In describing the utility of the map, students focused on information-retrieval tasks rather than on making inferences about Earth processes. Based on our findings about geoscience data visualization, we suggest strategies that may be beneficial in designing curriculum for teaching and learning with data maps.

---

*Current address: Department of Sciences, John Jay College, City University of New York (CUNY), 445 West 59th Street, New York, New York 10019, USA; sswenson@jjay.cuny.edu.
[†]kastens@ldeo.columbia.edu.

Swenson, S., and Kastens, K., 2011, Student interpretation of a global elevation map: What it is, how it was made, and what it is useful for, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, p. 189–211, doi:10.1130/2011.2474(13).

INTRODUCTION

Role of Data in Science Education

Using data is an important process of science that involves understanding how data are collected, manipulated, and represented in order to make informed interpretations. The National Academy of Sciences has defined science as "The use of evidence to construct testable explanations and predictions of natural phenomena, as well as the knowledge generated through this process" (NAS, 2008, p. 10). "Evidence" in science is grounded in data. If the science educators accept the National Academy of Sciences definition of "science," then they must accept responsibility for helping students understand data/evidence and not merely the knowledge that scientists have generated from data.

Most science education research on students' understanding of data has dealt with data that students collected themselves. Both qualitative and quantitative methods of education research have proven fruitful. Students' actions and thought processes while recording, analyzing, and interpreting data have been researched as students engaged in traditional data-collecting experiments such as determining the variation in a pendulum's periodicity (e.g., Germann and Aram, 1996; Kanari and Millar, 2004; Hug and McNeill, 2008) or computer-mediated activities such as using a microcomputer-based laboratory to measure distance, velocity, or thermodynamics (Brasell, 1987; Linn and Songer, 1991; Mokros and Tinker, 1987).

However, for many topics in geosciences curricula, it is not feasible to have students collect their own data. Earth phenomena are often too large (e.g., global atmospheric circulation), too far away (e.g., diminishing summer ice in the Arctic), too slow (e.g., rise in atmospheric carbon dioxide concentration), too dangerous (e.g., tornadoes), or require instrumentation that is too expensive (e.g., seafloor hydrothermal vents) for students to examine directly. Fortunately, government agencies and academic institutions have collected vast amounts of data about Earth processes that have been calibrated, quality controlled, archived, and are freely available via the Internet to the public, including schools.

Teaching and learning with data sets that students did not collect differs from working with student-collected data. Researchers are just beginning to explore the differences in teaching and learning when students do and do not collect the data they analyze. Hug and McNeill (2008) found considerable overlap in the classroom discourse stimulated by the two data types, but less discussion of error sources, more reliance on personal experiences, and different approaches to drawing conclusions from data when students had not personally collected the data used in their inquiry.

Hug and McNeill's (2008) study dealt with data that students would have been capable of collecting themselves, insofar as the methods, materials, and equipment were suitable for students of their age and experience. Other combinations of data-acquirer and data-interpreter are possible, as detailed in Table 1. All of these combinations have a potential role in science education and are ripe for educational research. The data set used to create the representation used in our study was acquired and provided by professionals (scientists, technologists, information specialists) and interpreted by students. We will refer to this configuration as "professionally collected data."

Role of Data-Based Visualizations in Science Education

The practical realities of student laboratory work means that student-collected data sets tend to be small, and thus amenable to relatively simple forms of representation, such as data tables or graphs with a few to a few hundred data points. However, when the flood gates of professionally collected data are thrown open, the volume and intricacy of the incoming data require an expanded repertoire of data-handling techniques. Scientists themselves extract insights from large data sets by rendering the data into "data visualizations," using computers to craft images that convey aspects of the data through position, color, texture, shading, and other perceptual devices that tap into human's powerful perceptual, spatial, and pattern recognition abilities (Edelson et al., 1999; Ware, 2004). Data visualizations also provide perspectives of phenomena that cannot be seen with the unaided human eye, for example, the morphology of the seafloor.

In science education, data visualizations have the potential to allow students to examine the evidence used to build scientific arguments, and to develop and test theories, even when the underlying data set is large or complicated. Use of sophisticated data visualization is growing rapidly in the applied sciences, business, and government (e.g., International Research for Climate

TABLE 1. TAXONOMY OF RELATIONSHIPS BETWEEN DATA COLLECTOR AND DATA INTERPRETER

Who collects the data?	Who interprets the data?	Literature	Terminology
Student	Same student	e.g., Germann and Aram (1996), Kanari and Millar (2004)	"First-hand data"*
Student	Another student, similar experience and ability	Hug and McNeill (2008)	"Second-hand data"*
Scientists, technologists, information specialists	Student	This study	"Professionally collected data"
Students	Professionals	Trumbull et al. (2000)	"Citizen science"

*Usage of Hug and McNeill (2008).

& Society/LDEO Climate Data Library, Tableau Software, and National Oceanic and Atmospheric Administration [NOAA] Data Explorer). Learning to extract insights from complex data is a skill that students will find useful far beyond geosciences; yet in our public school systems, this content knowledge area and skill development are sorely lacking (MacKay, 2006).

It is tempting to presume that because modern visualizations are so appealing to the eye, that their message is self-evident to all students. Evidence is beginning to accumulate that this is not true, that there are substantial differences in how people perceive and interpret data visualizations, both among students and between students and experts (Gilbert, 2005; Ishikawa and Kastens, 2005; Roth et al., 2007). Teachers and curriculum developers need to understand how students perceive and interpret data visualizations in order to craft effective instruction.

The scientific visualization of interest for the present paper is a specific type of data visualization in which the two spatial dimensions of the paper or computer screen are used to depict the two spatial dimensions of Earth's surface—in other words, a data map (Tufte, 2001). Maps, including data maps, are pervasive in geosciences, and fairly common in other sciences. Recent science education curricula that incorporate data maps have been developed by a variety of institutes and universities to help educators integrate global data sets into their instruction (e.g., Edelson et al., 1999; Hays et al., 2000; Prothero, 2006; Roushias and Anderson, 2001; Sawyer, 2005). An early pioneer in such curriculum development asked: "Do students know that data maps represent 'real, quantitative measurements about the Earth'?" (Sambrotto and Anderson, 2001, p. 57). Almost a decade later, this question remains unanswered.

Theoretical Framework

This research follows a grounded theory approach to understanding students' interpretations of a global elevation map, whereby qualitative data were collected by open-ended survey questions, but the analysis of the data was quantified. In quantifying the qualitative data, the researcher is examining the data for patterns and trends that emerge from the data and then categorizes these according to codes or concept indicators (Chi, 1997; Feig, this volume). The data are then quantified to determine frequency of responses. This approach provides a "middle ground" between traditional quantitative analysis and newer models of qualitative analysis.

Our work follows in the research tradition of probing students' conceptions (also referred to as preconceptions, prior conceptions, and misconceptions) as a necessary starting point for designing effective instruction (Driver et al., 1985, 1996; Libarkin and Kurdziel, 2002). However, we extend this line of research by examining students' understanding of data and data visualization rather than their understanding of a science concept. In a sense, we are probing their understanding of earth science—the physical and intellectual tools and techniques by which scientists learn about Earth—as opposed to probing their understanding of Earth.

Rationale and Context for Present Study

The specific intellectual and physical tools of interest in this study are bathymetric and topographic data, and the means by which such data are gathered and used. Of all data sets used in geosciences, bathymetry/topography is one of the richest in interpretive power. Solid Earth geoscientists invoke such data in identifying tectonic plate boundaries, and hydrologists use such data in defining watersheds. Oceanographers and paleoceanographers use bathymetry for identifying the gateways and boundaries that steer ocean currents. Atmospheric scientists view topography as a critical boundary constraint in explaining phenomena as varied as the location of tornadoes and the onset of monsoons. In addition, land-use planners, military officers, and civil engineers use topography, and fishermen and ships' officers use bathymetry, for making practical decisions every day.

The data visualization used in this study is a global map of the world's topography and bathymetry, created by the late William F. Haxby (Fig. 1). Haxby combined ship multibeam bathymetry data and satellite altimetry data for the oceans, plus radar interferometric data for the continents, to create a digital elevation model (DEM) that can be used to generate seamless topographic/bathymetric representations of the entire globe or portions thereof. For one commentator on the history of cartography, the Haxby map "has thematically reversed centuries of terrestrial bias" (Hall, 1992, p. 83) by displaying a detailed view of the seafloor. To an experienced geoscientist, this map can tell stories about Earth and Earth's processes, stories about plate tectonics, erosion, and deposition, and even about the placement of cities and the boundaries of nations, but what do students see when they look at the same map?

This study explored the nature of students' perception and understanding by asking them to write answers to three open-ended questions, as they viewed the Haxby map:

1. What do you think this is?
2. How do you think this was made?
3. What do you think this is useful for?

Context for Survey Question 1: "What Do You Think This Is?"

Several decades of research on children's understanding of maps have shown that mastery of what a map is and what it represents develops only gradually across childhood and even into adulthood. Liben and Downs (1989, p. 193) framed the question well: "Underpinning our discussion of maps is a fundamental question: How do children know what they are looking at? When and how do children understand that a pattern of lines and colors or gray tones on a sheet of paper stands for a particular place in the real world?"

Liben and Downs (1989) studied children's understanding of maps by asking them to identify whether or not various images were maps. Children correctly categorized as "maps" those representations that show places on a small to medium scale, have

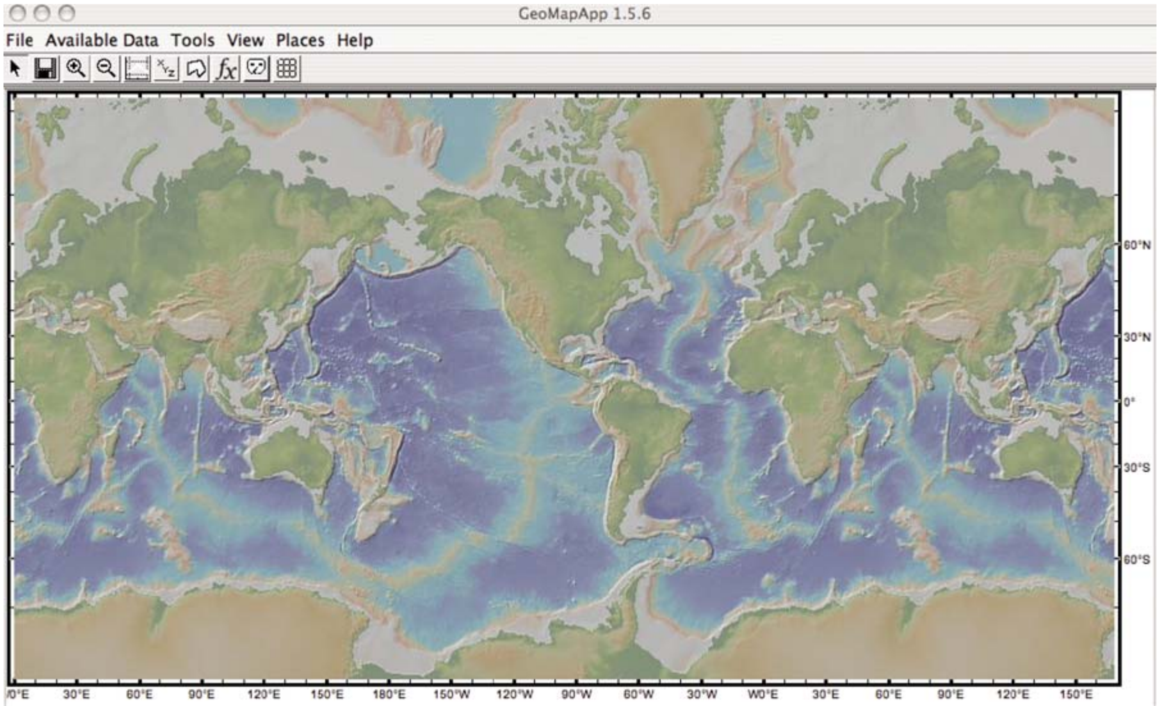


Figure 1. Digital elevation map produced by GeoMapApp as viewed by student participants. Students viewed map in color on a classroom screen, projected by a computer projector. The color version can be accessed from http://www.geomapapp.org/.

color, are seen directly from overhead, and have conventional cartographic symbols, for example, a standard road map. Any deviation from this kind of map increased the likelihood that children would not categorize the representations as a map.

All external representations, including maps, have a dual existence, in that they *are* something and at the same time they *stand for* something (DeLoache, 2000; Liben, 2003). A map exists as an entity on paper or screen, characterized by observable attributes such as color and size. At the same time, the map *stands for* something—Earth or a portion of Earth. To distinguish between these concepts, we will refer to the map as the "representation," and Earth, or more specifically the depicted aspects of Earth, as the "referent" of the map (MacEachren, 1995). Understanding the nature of a map representation does not imply that a person necessarily understands the referent or the connection between referent and representation. For example, in Liben and Down's (1989) study, students who could successfully identify a representation as "a map" did not "reasonably understand" (p. 181) the particular places the maps were intending to represent.

Context for Survey Question 2: "How Do You Think This Was Made?"

The second question examined students' epistemological model for the information displayed in the data map. Prior research (e.g., Brasell, 1987; Mokros and Tinker, 1987; Nachmias and Linn, 1987) has shown that collecting and displaying data in a microcomputer-based laboratory improves students' interpretations of graphs. When students experience the connections among the actions of the person collecting the data, the instruments collecting the data, the referent (i.e., the phenomena being measured), and the resulting representation (the graph), their ability to make insightful and accurate interpretations of the representation improves.

For many geoscience data representations, including global bathymetry/topography maps, such direct experience is lacking. Bathymetry and topography have been collected over long periods of time, using sophisticated tools that students do not have access to, and include submarine areas of Earth that people cannot view directly. Although a few exemplary geography curriculum materials involve students in making maps by direct observation of authentic environments (e.g., Sobel, 1998), most students do not have experience with making even the simplest of maps by making observations of the referent.

Based on the research on student-collected data (Germann and Aram, 1996; Kanari and Millar, 2004; Hug and McNeill, 2008), we consider it plausible, but unproven, that students' understanding of data that they did not collect would be stronger if they understood the basics of how the data were acquired and processed. There is no perfect substitute for being there oneself, making decisions as the experiment unfolds and gaining an embodied sense of the scale and scope of the phenomena of study. However, an intellectual, nonexperiential understanding of the origin of the data could enable students to detect flaws in the data (Nachmias and Linn, 1987) and may help to protect them against other forms of ignorance-based misinterpretation. An understanding of where data come from is also part of the larger science education agenda of helping students develop an epistemological model for science, i.e., an understanding of how scientists know what they know (Bransford et al., 2000).

We could find no prior research on K–12 students' understandings or misunderstandings about the ways in which bathymetric and topographic data are collected or processed into data visualizations. As a starting point for comparison, we developed an epistemological model of how an expert envisions the sequence of knowledge-generation processes that underlie the global elevation map used in our study (Fig. 2).[1] In our experts' epistemological model, data are first acquired from Earth, using a variety of sensors mounted on ships or satellites, using different technologies for the subsea and terrestrial parts of the globe. Second, the individual data streams are processed, using various assumptions and calibrations, to turn the raw sensor data plus navigation into depths and heights as a function of latitude and longitude. Next, the data from different sources are merged, coping with gaps and overlaps in data such that every point on the globe is associated with a single elevation value, to form a global digital elevation model. Finally, a representation is generated from the DEM according to choices made by the user as to color palette, vertical exaggeration, etc. Every step along this information chain is mediated by human decision makers and by software.

Context for Survey Question 3: "What Do You Think This Is Useful For?"

Maps have a purpose (Liben, 2003). Maps are useful for recording, conveying, organizing, and finding out information about the location, shape, and configuration of features on Earth's surface. Because Earth processes cause location, shape, and configuration of natural features, maps are also useful for making causal inferences about events in Earth history that shaped Earth's surface. Location, shape, and configuration of surficial features also impact human society in terms of land use, watercourses, transportation pathways, and so forth; thus, maps can be useful for explaining and planning human/environment interactions.

Another way of probing students' preconceptions is to ask them about the purpose of the map. In order to comment on purpose, students already need to have some kind of interpretation of what the symbols and colors mean as explored in question 1. However, their interpretation need not be detailed or correct in order to formulate ideas about purpose. For example, when people see the classic "Rand McNally" cartographic conventions, they know this is a road map and that road maps are for finding one's way (Downs and Liben, 1987), even though they cannot yet interpret all the symbols.

Some maps advertise their purpose, as for example, road maps and navigational charts. The map used in this study does not state its purpose, and so the students must make inferences from their prior knowledge and life experience, plus evidence within the map itself to answer these questions.

One way to formulate ideas about utility would be to consider the map itself. MacEachren (1995) suggested that individuals use feature identification and feature comparison to make sense of scientific visualizations. For example, the inclusion of latitude and longitude may be interpreted as an indication that the map was intended for navigation.

Another possible approach would be to think in terms of potential users (e.g., for students, scientists, fishermen, or the creator of the map). Thinking about the map-creator's intention may be important in light of Myers and Liben's (2008) recent finding that children's interpretations of maps depend on whether or not they were aware of the map-creator's symbolic (semiotic) intent. Since the students in our study did not collect the data or create the map itself, they do not have the insight about the creator's mindset that they would have had if they had collected the data and created the map themselves.

Contributions of This Study

The purpose of this study was to investigate conceptions that students have about a data map of a type that is widely used by the geoscience community in order to design curriculum and pedagogical methods suitable for using such a map in the K–12 science classroom.

After analyzing 196 student responses, we were able to make observations and inferences about student awareness of:

1. the nature of the representation;
2. the scope of the referent;
3. the aspect of the referent that is depicted;
4. the fact that some kind of information/data/observation had to be acquired from Earth to make the map;

---

[1] Development of the experts' epistemological model: Prior work that informed the epistemological model included Robinson and Petchenik's (1975) classic depiction of cartographic information flow from the represented space, through the mapmaker, to the map, and thence to the map reader, and Chayes' (1999) diagram of information flow from sensors to geoscience data products. Kastens extended Robinson and Petchenik's concept to environmental policy and Earth system education in Ishikawa et al. (2005) and Kastens and Turrin (2006). The initial draft of the current model was developed by Kastens, drawing from her training and experience in marine geology. Kastens' initial model was then refined through iteration with two additional experts: Dale Chayes and Andrew Goodwille. An oceanographic engineer, Chayes is the codeveloper of a widely used software system for processing and analyzing swath bathymetric data (Caress and Chayes, 2009), has installed and supported seafloor mapping hardware and software on numerous oceanographic ships, and has collected geoscience data on over 100 research expeditions on land, sea, and ice. Trained as a geophysicist, Goodwille is the data manager and education coordinator for the Marine Geoscience Data Center, the facility that developed, maintains, and serves the database and visualization tool used to make the visualization used in this study. Kastens, Swenson, Chayes, and Goodwille cycled through multiple versions of the epistemological model seeking a balance among the following criteria: accuracy, simplicity for communication with a nonspecialist audience, and extensibility to other Earth data types. The prototype version of the epistemological model motivated our decision to ask question 2. The final version of the epistemological model was informed by the student responses to question 2.

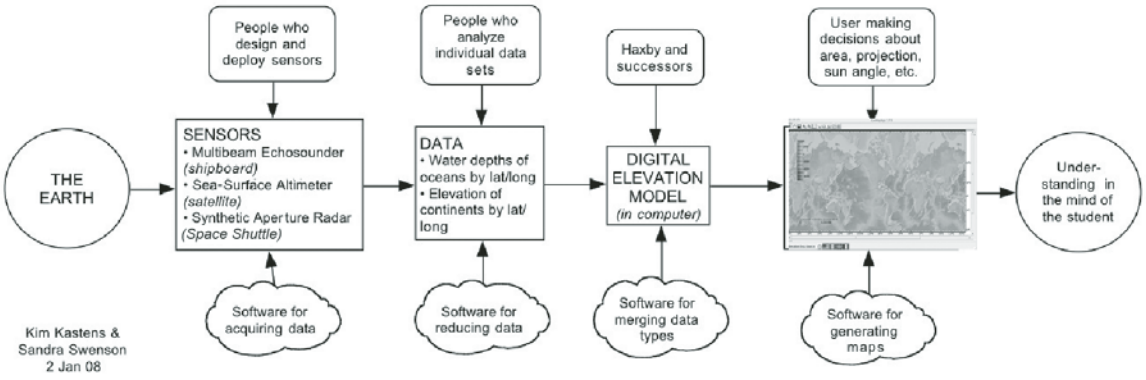


Figure 2. Diagram summarizes an expert's epistemological model for the knowledge encapsulated in the GeoMapApp representation. How much of this do teachers and students need to know to teach and learn effectively from the representation?

5. the roles, if any, of people, instrumentation and equipment, and computers and computer software in making the map; and
6. the usefulness of the map and for whom the map would be useful.

METHODOLOGY

Our methodology builds on prior research on learner's conceptions and alternative conceptions (e.g., Driver et al., 1985, 1996); however, we extend this line of research by examining students' understanding of data and data visualization rather than their understanding of a science concept. This research is grounded in the views of the participants of this study (Creswell, 2003, p. 14) who were surveyed in the natural setting of their classroom with their teacher and a researcher (Swenson) present.

Participants

In total, 196 science students participated in the study. The students were studying various science courses in grades 8 through 12 in suburban New Jersey and New York. A breakdown of the participants is as follows: 105 eighth-grade earth science students, 26 ninth-grade biology students, 43 twelfth-grade marine science students, and 22 twelfth-grade Advanced Placement (AP) biology students. Except for a few students who moved in from out of state, all participants had studied earth science for at least a half-year in the current year or an earlier grade. All had studied landforms and topographic maps in their class.

The objective of this sample was to cast a broad net to gather a wide range of conceptions on a previously under-researched topic from a relatively large and varied population of students. Grades 8–12 were targeted because it is in those years where students' "knowledge and use of representations should expand in scope and complexity" (National Council of Teachers of Mathematics, 2000, p. 361) and they should be developing the skill of "mak[ing] inferences and drawing conclusions from maps and other geographic representations" (Geography Education Standards Project, 1994, p. 55). It was not our goal to make comparisons between schools or across grades.

Materials

The digital topographic and bathymetric map that students viewed (Fig. 1) was the default global map created by GeoMapApp (Carbotte et al., 2004, 2005). GeoMapApp is a scientists' tool, and at the time of our study, the map was not incorporated in any educational or outreach materials. This representation is a Mercator projection with latitude and longitude tick marks along the margin. Latitude spans from 60°S to 80°N, while longitude includes 360° plus a repeat of another 170°. This projection allows an uninterrupted view of all the world's oceans and seas except for the central Arctic and far southern oceans. The topography and bathymetry are shown as color-coded shaded relief, with a vertical exaggeration of 2×. Topography is represented by shades of green and brown, while deepening shades of blue represent bathymetry. The color choices and shaded relief combine to create a representation that resembles Earth as seen with the human eye. The extent to which a representation is similar to its referent (rather than relying on arbitrary or culturally specific symbols) is called "iconicity" (MacEachren, 1995); the map used in this study has a high degree of iconicity. No map key is included on the GeoMapApp default map.

Instrument

A survey design was chosen in order to sample the greatest number of students. The survey was conducted using three open-ended questions: (1) What do you think this is? (2) How do you think this was made? (3) What do you think this is useful for? Respondents wrote their ideas on one sheet of paper.

Open-ended questions were used because there was little applicable prior research, and we did not want to prejudge or constrain the nature of the understandings and misunderstandings that could emerge from the study. This instrument was not pilot-tested because the researchers did not want to be bound to pre-established constructs but rather to utilize an inductive approach to analyze the collected data. Miles and Huberman (1994) explained that predesigned and structured instruments may preclude the researcher from context-rich phenomena that emerge from the data. The questions asked of students in this study were intended to be general enough so as not to be leading, allowing key themes to emerge naturalistically from the data rather than fitting the data into a predetermined framework.

Researchers' Location Relative to this Study

The first author's background is in science education, and she came to this area of research through an interest in understanding the conceptions (and alternative conceptions) students have about scientific data and data visualizations. The first author has been a science teacher of middle school, high school, and college. She completed a doctorate in earth science education based on dissertation research (Swenson, 2010) that examined another type of topography/bathymetry representation as well as the one used in this study. Her role in this study was to recruit the volunteer teachers, acquire the data in the classroom, develop the initial coding scheme, and draft the manuscript.

The second author is a marine geologist by training who has extensive experience collecting and analyzing data of the type found in the Haxby map, including ~22 mo at sea and publication of original bathymetric maps (e.g., Kastens et al., 2000). She came to this study through an interest in students' understanding of maps (Kastens et al., 2001; Kastens and Liben, 2007) and as the education and outreach coordinator for the Ridge 2000 Open Data Exchange System (RODES). She identified the data set as important and deserving of educational research, brought expertise on spatial thinking to the project, developed the epistemological model, served as second coder on all the data, and assisted in writing the manuscript.

As an earth science educator and an earth scientist, we are deeply familiar with the data visualization used in this study, and we realize that this may have interfered with our ability to "see though the eyes of" a student viewing this data set for the first time.

Procedures

The location of this research was the students' normal classroom environment. The researcher was invited into the classroom by the instructor to administer the survey and to provide follow-up discussion about the three questions the next day. At the beginning of a regular class session, students were able to observe the teacher and experimenter using the classroom computer to retrieve from the Internet the topography/bathymetry representation and then project it on a screen.

Students were then handed the questionnaire and were requested to respond as best they could while viewing the overhead display. They were told that there were no right or wrong responses because the researcher was seeking intuitive responses from the students, that is, what came naturally to them rather than a response the participants thought the researcher or instructor would want to hear. What students "choose to talk about is an indication of what they think is important, even if they don't talk about everything they know" (Chi, 1997, p. 305). Students were given the class period to complete the questionnaire, but all finished within 35 min of class time. In a class meeting the following day, the experimenter debriefed the students by having a discussion about the three survey questions. She also led a hands-on activity in the computer laboratory, making use of some of the more advanced capabilities of the GeoMapApp tool. After participating in these activities, the students could download the data set onto their home computer if they chose to do so, because the data set is freely accessible via the Internet.

We also had an opportunity to survey 33 geoscientists with these three questions as they viewed this same data map. The experts were given the survey prior to a research seminar on students' understanding of maps and were therefore self-selected for an interest in this topic. These expert responses informed our suggestions about how to move students toward greater expertise in the use of data maps.

Coding

In thematic content analysis, the themes are extracted from the text of the participants' responses rather than established a priori, so that themes emerge naturally from the data and can be linked or reorganized to develop a dominant structure (Miles and Huberman, 1994; Chi, 1997; Libarkin and Kurdziel, 2002). In the analysis, the experimenters examined the keywords and phrases that students used to describe what they were seeing, how they thought the representation was made, and what they thought the representation was useful for.

After a discussion of a sample of student responses, the lead author created an initial coding scheme, criteria, and examples for each question. Coding categories were iterated until both authors felt the categories captured the range of responses. Each researcher then read each student response separately and tallied the responses under the code that they thought best matched the student response. Inter-rater reliability was determined to be 91% for question 1, 96% for question 2, and 93% for question 3. Disagreements in analysis were resolved through discussion until a consensus analysis could be agreed upon.

Examination of the initial broad-scale coding generated follow-up questions, which we pursued by further dividing or combining some initial categories. This second-order coding was treated the same as the first-order coding with respect to iterating coding categories, independent tallying by both researchers, and resolution of discrepancies through discussion.

Coding for question 1 was completed before beginning coding for question 2, and coding for question 2 was completed before beginning question 3; however, the researchers were free to use the other two responses by the same student in order to clarify ambiguities in a response.

In interpreting the results, readers should keep constantly in mind that whereas presence of an element in a student's response surely means that they had awareness of that element, absence does not necessarily mean that they were *un*aware of this element.

Establishment of Validity and Reliability

As discussed already, we intentionally designed the instrument with broad, open-ended, simply worded questions to allow whatever was in the forefront of the students' minds to emerge in their own words. The use of open-ended questions is in the tradition of research on students' prior and alternative conceptions (Driver et al., 1985). Although such questions bring forth the students' ideas cast in the students' own words, a limitation of this technique is that respondents may not say everything that they know.

The reliability of this instrument was corroborated in two subsequent contexts. Swenson (2010) used questions 1 and 2 along with follow-up interviews with a population of college, non–science majors. Swenson (personal observ.) used all three questions with a population of geoscience experts. In both cases, the same broad themes emerged from the responses.

Within the current study, inter-rater consistency of the coding categories and the assignment of student responses to coding categories were evaluated through dual coding by both authors of every response to every question. Inter-rater reliability was calculated for each question, as described previously, and all were above 90%.

RESULTS

Question 1: "What Do You Think This Is?"

Primary Coding

In 88% of the responses, students indicated (1) that the displayed image was a map and (2) that it was about Earth. In other words, a high percentage appeared to grasp both the nature of the representation, and the basic representation-referent relationship.

Within the near-universal understanding that students were viewing a map of Earth, the most common theme (Table 2) that emerged included responses that provided only very basic geographic information, such as the existence of continents and oceans and a latitude/longitude grid. Illustrative are "1 and a half map[s] of the world," and "map of the world with all of the coordinates." A geoscientist would view such a map as a basemap onto which additional data types could be layered, so we coded such responses as "Basemap." A visualization of the "Basemap" construct might look something like the diagram in Figure 3, a map showing just basic geographical information. Forty-four percent of the total student responses (87/196) fell into the "Basemap" category (Table 3; Fig. 4). For all classes except eighth-grade earth science, "Basemap" was the modal response.

The second theme was student descriptions about topography and/or bathymetry ("Topo"). This is the accepted interpretation of the representation that was intended by the data map creator and would be offered by most geoscientists. Responses in this category may include "Basemap" information, but they stated or implied something about height (elevation) or depth below sea level or the shape of Earth's surface or the existence of specific landforms. Most students are not familiar with the word "bathymetry," so we relied upon descriptions about physical features of the seafloor, such as "showing all land and water mass on Earth, including undersea mountain ranges" (Table 2). Thirty percent of the total student population (59/196) described the map as representing topography/bathymetry (Table 3; Fig. 4).

The third theme was student descriptions of aspects of the Earth *other than* topography/bathymetry. "NonTopo" responses were usually alternative observations or interpretations of the representation referring to attributes of the Earth that were not shown on the provided map. Examples include: weather patterns, clouds, ocean currents, tides, or even the level of sodium (Table 2). Some alternative interpretations seem to have been triggered by the map's colors, for example, "… It looks like the different colors in the water especially are showing different currents," and "… the type of terrain found on certain regions of the Earth. While green represents a lush and treeful environment, dark browns symbolize a barren and desert-like terrain." Between 11% and 14% of each class stated that the map showed aspects of Earth that were not in fact on the map (Table 3; Fig. 4).

Some responses referred to topography/bathymetry but also included NonTopo (Table 2). Such responses were coded as "Topo&NonTopo," for example, "I think this is an elevational topics chart. Showing elevation by ridges and tan color. Shows temperature by different blues also white for cold water." Between 5% (twelfth grade AP) and 11% (eighth grade) responded in the "Topo&NonTopo" category (Table 3).

Finally, we included an "Ambiguous" and a "No response" category for responses that were not clear or when a student did not respond.

Secondary Coding

To better understand students' conceptions about what the map represented, we further subdivided the nontopographic responses in the "NonTopo" and "Topo&NonTopo" categories (Table 4). There were 24 "NonTopo" responses plus 17 "Topo&NonTopo" responses, giving a total of 41 responses analyzed; however, some descriptions included multiple elements and so were tallied in multiple categories. Seventeen out of 41 students (41%) described the digital elevation map as displaying something about the fluid Earth, where most of these responses were about tides and currents. Twenty-two out of 41 (54%) responded with a description about the solid Earth, including plate tectonics. A small percentage (7%) discussed the global elevation map within the context of biology. Information about the fluid Earth, biomes, and plate tectonics might be inferred from the data map, but the map itself does not represent any of these phenomena directly.

TABLE 2. CODING OF MAJOR THEMES THAT EMERGED FROM QUESTION 1: “WHAT DO YOU THINK THIS IS?”

Category/criteria	Keywords	Examples
Basemap: • States or implies an area of Earth’s surface, may be a region or the entire planet. • Focus on basic geography, including continents, lat/long grid • No mention of any earth science data types	Map, globe, Earth, World, latitude and longitude, continent, coastlines	• I think this is a map showing the latitudes and longitudes of a part of the world. 8th gr. • A map creator that allows you to focus on certain coordinates. 8th gr. • A world map. 8th gr. • Map of the world with two Africas, Europes, and Australias. 9th gr. • 1 and a half map[s] of the world. 9th gr. • A map of the world. 12th gr. OS • A map of the globe, but flattened. 12th gr. AP
Topo: • Must mention either height/depth or shape of Earth’s surface • May also include “basemap” information • Need not include terms “bathymetry” or “topography”	Height, depth, ocean depth, elevation, topography, physical features Also names for specific landforms such as mountain	• I think this is a topographic map, and it shows different level[s] of continent and ocean. 8th gr. • This is a map of the world showing elevations of different areas of the Earth (altitudes). 9th gr. • A map showing a physical landscape of the Earth. I think it shows elevation/depression indicated by color. 9th gr. • World map (Mercator map) showing all land and water mass on Earth, including undersea mountain ranges. 12th gr. AP
NonTopo: • May include “Basemap” information • Must not include “Topo” information • Must also include one or more other aspect of Earth, neither “Topo” nor “Basemap”	Terms for natural phenomena, such as weather, climate, biomes, or geological structures Terms for Earth processes such as tides, plate movement, or continental drift	• A map of the world’s boundaries along with the climates of the Earth (ex. white near the pole is snow/glaciers, tundra). 8th gr. • High tide and low tide. 8th gr. • A map showing the different kinds of lands around the world, example: swamp, dessert *[sic].* 8th gr. • Fault lines on the map of the world. 9th gr. • I think this is a map of the world that shows weather conditions and things like that. It looks like there are clouds. 12 gr. OS
Topo&NonTopo: • Must include “Topo” information • Must include “Other” information • May include “Basemap” information	See above	• I think this is an elevational topics chart. Showing elevation by ridges and tan color. Shows temperature by different blues also white for cold water. 8th gr. • I think this is a topographic map of the world which also depicts fault lines of major tectonic plates. 9th gr. • It is a map of the Earth, showing not only topography but sediment deposits as well. 12th gr. OS • This is a map of the world after much of the ocean was evaporated away or underwater volcanoes. 12th gr. AP
Ambiguous • Response was not clear		• I think this is a map of everything and telling you where you can find it. 8th gr. • I think this is what the world used to look like. 9th gr. • Earth’s shifting over time. 9th gr.
No Response • Student did not answer question 1		

Because GeoMapApp was developed for marine geology and geophysics, the map design was optimized for examinations of the ocean basins. To examine to what extent students attended to the ocean basins, we further divided question 1 responses in the "Topo," "Basemap," and "NonTopo" categories into "Continent," "Ocean," "Both oceans and continents," or "Neither" (Table 5). As a group, the Basemap responders attended to neither the oceans nor continents—77% of "Basemap" responses fell in the Neither category. Instead, most "Basemap" descriptions were about "a map of the world," with some references to "latitude and longitude" or "coordinates." The "Topo" group paid more attention to continents than oceans (44% vs. 5%), whereas the "NonTopo" group paid more attention to oceans than continents (38% vs. 25%).

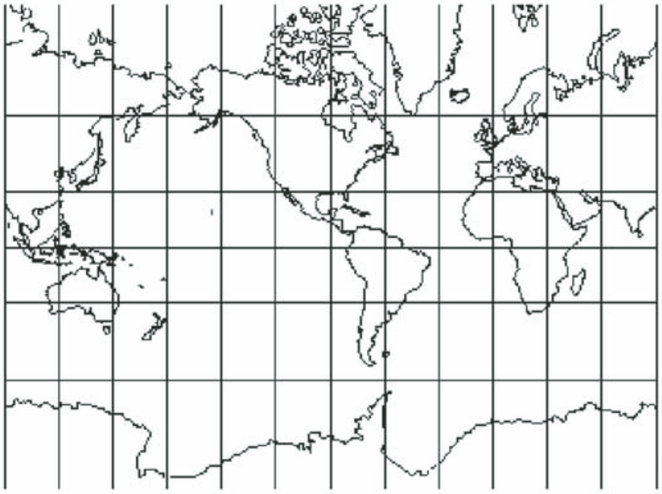


Figure 3. A visual representation of the "Basemap" coding category of question 1. This type of response mentioning only area, latitude, longitude, and perhaps continents or coastlines was the most common response across the population studied (from G. Evenden, U.S. Geological Survey, 1995, public domain; accessed June 2008).

Question 2: "How Do You Think This Was Made?

Primary Coding

Five themes emerged for question 2 (Table 6). Many of the responses included multiple themes; therefore, the researchers allowed more than one coding category for each response.

The first theme was that students stated or implied that the representation was made by a person or people. Such answers might mention a specialist such as a "scientist" or a "cartographer," or a more generic "person," or "someone." Diverse roles were described for these people, related to both collecting the data and generating the representation. Illustrative responses are: "I think this map was made by people who discover and research the features of the land," and "A cartographer most likely started out by mapping the land masses and water regions of the Earth as he would normally do for a spherical globe, but then spread his reproduction into the shape of a rectangle." Eleven percent (21/196) of the total student population responses was coded as mentioning that a person or people were involved in making the map (Table 7; Fig. 5).

The second theme encompassed various aspects of data acquisition, subdivided according to whether the focus was on the type of data acquired or on the tool used for data acquisition. The essential element of this theme is that the response stated or implied engagement with the referent, Earth. Data types mentioned by students included data height and depth (coded as category 2A-1), or a physical property other than topography/bathymetry (2A-2), for example, salinity. Data acquisition tools ranged from "satellite," spaceship, or "space station" (2B-1) to "pictures" (2B-2) to "ship" or "submarine" (2B-3). We also included categories for other tools (2B-4) and data acquisition without a specified tool or other data type (2C).

Data acquired from a satellite or spaceship (2B-1) was the highest percentage in all of the categories for data acquisition,

TABLE 3. DISTRIBUTION OF STUDENT RESPONSES TO QUESTION 1: "WHAT DO YOU THINK THIS IS?"

	Eighth-grade earth science (%)	Ninth-grade honors (%)	Twelfth-grade ocean science (%)	Twelfth-grade Advanced Placement (%)	Total (%)
	n = 105	*n* = 26	*n* = 43	*n* = 22	*n* = 196
Basemap	32	50	65	55	44
Topo	39	23	14	27	30
NonTopo	11	12	14	14	12
Topo&NonTopo	11	8	7	5	9
Ambiguous	3	8	0	0	3
No response	4	0	0	0	2

Note: Table shows percentage of student responses in each class that were coded in each category. For question 1, primary coding, each response was coded in only one category. Columns may not sum exactly to 100% because of rounding errors.

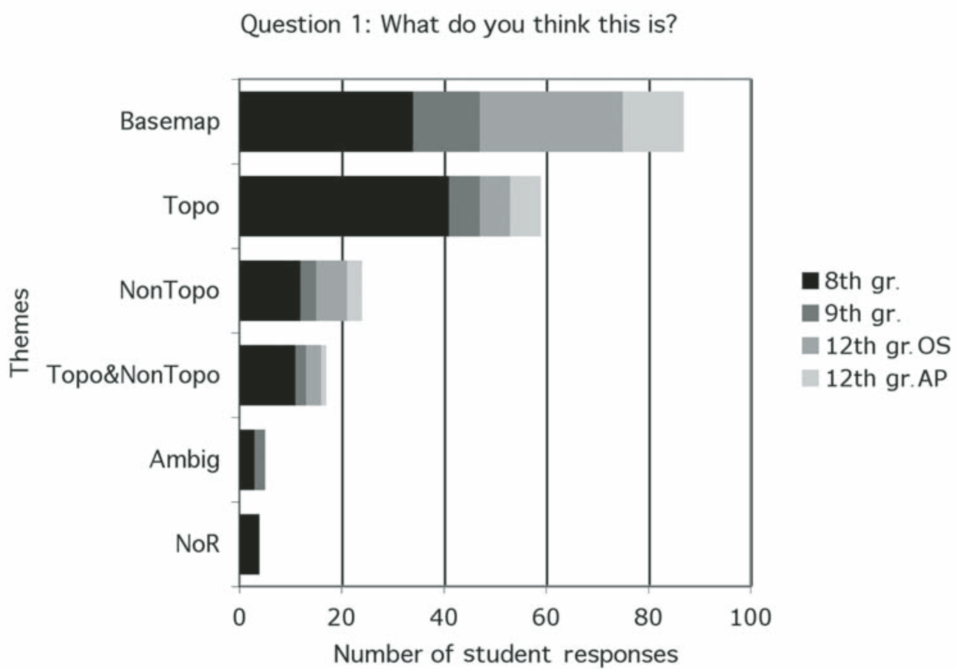


Figure 4. Student response to question 1: "What do you think this is?" The graph shows the number of student responses per coding category, as defined in Table 2. Close to half of the students (87/196 or 44%) described a map with only basic geographic information ("Basemap" category). Less than a third (59/196, or 30%) of total students interpreted the data elevation map to represent elevations/depths or landforms ("Topo" category), which is the professionally accepted interpretation. Eleven percent of the students described other aspects of Earth ("NonTopo" category) that were not in fact represented on the map.

TABLE 4. SECONDARY ANALYSIS OF NONTOPOGRAPHIC RESPONSES TO QUESTION 1

Key concepts	Keywords	Examples	Number of occurrences	Total
Fluid Earth: Student states or implies attributes about the ocean or atmosphere.	Tides, currents, ice	• A map of the world showing tides • I think this is a picture of a map with longitude and latitude degrees on it. It looks like the different colors in the water especially are showing different currents.	9	
	Weather, winds, clouds	• In addition to demonstrating measurements of longitude and latitude, the map seems to exhibit weather patterns.	3	
	Sodium, salinity, saltiness	• I think this is a map of where there are high levels of sodium.	1	
	Air temp., water temperature	• A temperature/thermo map • This is an image of the water temperature for all of Earth's oceans.	4	**17**
Solid Earth: Student states or implies attributes about the solid Earth.	Kinds of land/types of terrain	• A map of the world[']s boundaries along with the climates of the earth (ex. white near the pole is snow/glaciers, tundra). • A map showing the different kinds of lands around the world, example swamp, dessert [*sic*].	6	
	Plate tectonics/plate boundaries, crust, fault lines, volcanoes	• A map of the world's plate boundaries and countries. • It's a map of the world, however, Australia is noticed twice instead of once. Also plates are noticed in the background while countries are in the foreground.	11	
	Rock types; sediment	• It is a map of the Earth, showing not only topography but sediment deposits as well. • Some sort of geological map w/coordinates so you can pinpoint certain locations.	5	**22**
Biology: Student states or implies attributes about living things.	Animal/ migration patterns; plants/vegetation; biomes	• A map of the world that shows vegetation with green and either desert or tundra with off white. • This is a map of the world showing tidal water flow or migration patterns.	3	

Note: This table encompasses 24 responses coded "NonTopo" plus the erroneous portion of 17 responses coded "Topo&NonTopo." A response could be counted in more than one category of this tally.

TABLE 5. TALLIES OF RESPONSES ABOUT THE CONTINENTS, OCEANS, BOTH, OR NEITHER

Response type	"Topography"	"Basemap"	"Other"
	n = 59	*n* = 87	*n* = 24
Continent	44%	14%	25%
Ocean	5%	0	38%
Both ocean and continents	36%	9%	8%
Neither	15%	77%	29%

Note: Underlined values are the most abundant in both their row and column.

TABLE 6. CODING OF MAJOR THEMES THAT EMERGED FROM QUESTION 2: "HOW DO YOU THINK THIS WAS MADE?"

Category/criteria	Subcategories and keywords	Examples
1. *People/scientist or man-made:* The student response refers to a person involved. Focus on description about a person.	"Scientists, "human," "someone."	• I think it was made by scientists who explored the ocean, and somehow predicted what the layout of the ocean would be if the world continues its process. 8th gr. • A combination of satellite and human examination and surveying. 9th gr.
2A. *Data acquisition type:* Student states or implies that some kind of data of a specific type were acquired.	**2A-1.** Student response refers to acquisition of height or depth data: "depth," "height."	• By measuring the land depth and sea depth. 8th gr. • This was probably made by collecting measurement of each region's height in relation to sea level, and then shown in this image comparatively. 9th gr.
	2A-2. Student response refers to acquisition of specified data type, neither topo/bathy.	• I think it was made by scanning a map into the computer & highlighting the places of high *sodium*. 8th gr. • I think this was made by scientists who study *precipitation*. 12th gr. AP
2B. *Data acquisition tools:* Student states or implies that a data acquisition tool was used to make the representation.	**2B-1.** Student response refers to a tool in space: "satellite," "space ship."	• Satellite pictures. 8th gr. • This was made from satellite images. 8th gr.
	2B-2. Student response refers to "photos" or "pictures" to acquire data.	• I think it was made by taking a picture of the Earth, and then making a flat, square kind of Earth. 8th gr. • Picture from space. 8th gr.
	2B-3. Student response refers to tool in the ocean: "ship," "submarine," "sonar."	• I think that this was made from information taken by a satellite and possibly deep water submarines or submersibles such as "Alvin." 12th gr. OS
	2B-4. Student response refers to other tools: "heat sensors," "infrared heat imaging."	• Most likely using a compilation of satellite imaging and geological surveying as well as infrared heat imaging. 12th gr. AP • Heat sensors. 12th gr. OS
2C. *Data acquisition:* Unspecified; student does not specify data type or tool.	"Measurement," "observations," "accurate data collection," "studying physical features."	• This map was probably made from accurate data collection and carefully planned pointing out of some of the Earth's physical features. 8th gr.
3. *Representational technique and technology*	**3A.** High-tech: Computer or computer software	• This was probably made with spiffy editing equipment that allows the image to be copied and pasted. 9th gr. • A computer and other technologically advanced equipment. 12th gr. AP • This was made by computer graphics. 12th gr. OS
	3B. Low-tech: sketch, clay	• Plastics/ paper/gluing. 8th gr. • By taking two pictures of Earth and sticking them together. 9th gr.
4. *Prior map:* Student states or implies a map was the starting point used to create the representation.		• This was made with a map then adding different color[s] where they wanted areas to stand out. 8th gr. • I think this was made by a computer flattening out a globe and connecting the two resulting pictures.12th gr. AP
5. *Referent:* Response describes how the referent or an aspect of the referent was made rather than the representation.		• From the moon pulling from the Earth. 8th gr. • Continental drift. 9th gr. • By God. 12th gr. OS

TABLE 7. TALLIES OF RESPONSES FOR QUESTION 2: "HOW DO YOU THINK THIS WAS MADE?"

Coded category	Primary coding (%)	Secondary coding (%)
1. People	**11**	
2A-1. Data acquisition: Data types: height, depth	**9**	**44**
2A-2. Data acquisition: Data types: other	**2**	
2B-1. Data acquisition: Tools: satellite, or from space in general	**27**	
2B-2. Data acquisition: Tools: photography/taking pictures	**12**	
2B-3. Data acquisition: Ship, submarine, sonar	**2**	
2B-4. Data acquisition: Other	**5**	
2C. Data acquisition: Unspecified tool and technique	**8**	
3A. High-tech: Computer: computer software or program	**39**	
3B. Low-tech: representation methods (sketch, etc.)	**14**	
4. Prior map	**13**	
5. Referent	**8**	
6. No idea, not sure	**4**	

Note: Primary coding equals number of students providing this response divided by 196 students; column does not sum to 100 because some responses contained multiple elements. Secondary coding equals number of students providing *any* data acquisition (2A-1 through 2C) divided by 196 students.

with 27% (53/196) of the student responses coded in this category. The second most frequent response was in the category for data acquisition tools such as photography, or taking pictures 12% (24/196). Only 2% of the students responded that data from a ship, submarine, or sonar were used to create this representation (Table 7; Fig. 5).

The third theme focused on making the representation rather than acquiring the data. The means of making the representation might be high-tech (coding category 3A, distinguished by keywords such as "computer," "computer software," "computer applications," "scanner," or "printer"), or low-tech (category 3B, such as "made by clay," or "sketching"). Out of the total student responses, 39% (76/196) tallied in the "computer" category. This was by far the most frequent response type for question 2, making up more than a third of the students surveyed. Fourteen percent (27/196) of the students described a "low-tech" method of map-making (Table 7; Fig. 5).

The fourth theme was found when a student stated or implied that the starting point for making the representation was an existing map, picture, or globe, rather than Earth (Table 6). For example, "This was made with a map then adding different color[s] where they wanted areas to stand out," or "I think this was made by adding texture to a map of the world to show the physical features." We coded such responses as "Prior map." Thirteen percent (26/196) responded that the starting point for making the map was a map itself (Table 7; Fig. 5).

Finally, for the fifth category (Table 6), students answered the question by describing how they thought the referent was made rather than how the representation was made. For example, some students suggested "the moon pulling from the Earth," "continental drift," and "by God." We coded such examples as "Referent." Eight percent (16/196) of the students responded in terms of making the referent rather than the representation (Table 7; Fig. 5).

Secondary Coding

A key distinction among the responses to question 2 was whether or not the student indicated any kind of data acquisition (Table 7, secondary coding category). Some students provided a rich description of the data acquisition process, including multiple category 2 subcodes, but 56% (110/196) gave no indication at all that some kind of observation or measurement or data collection or engagement with Earth itself was required to make the representation.

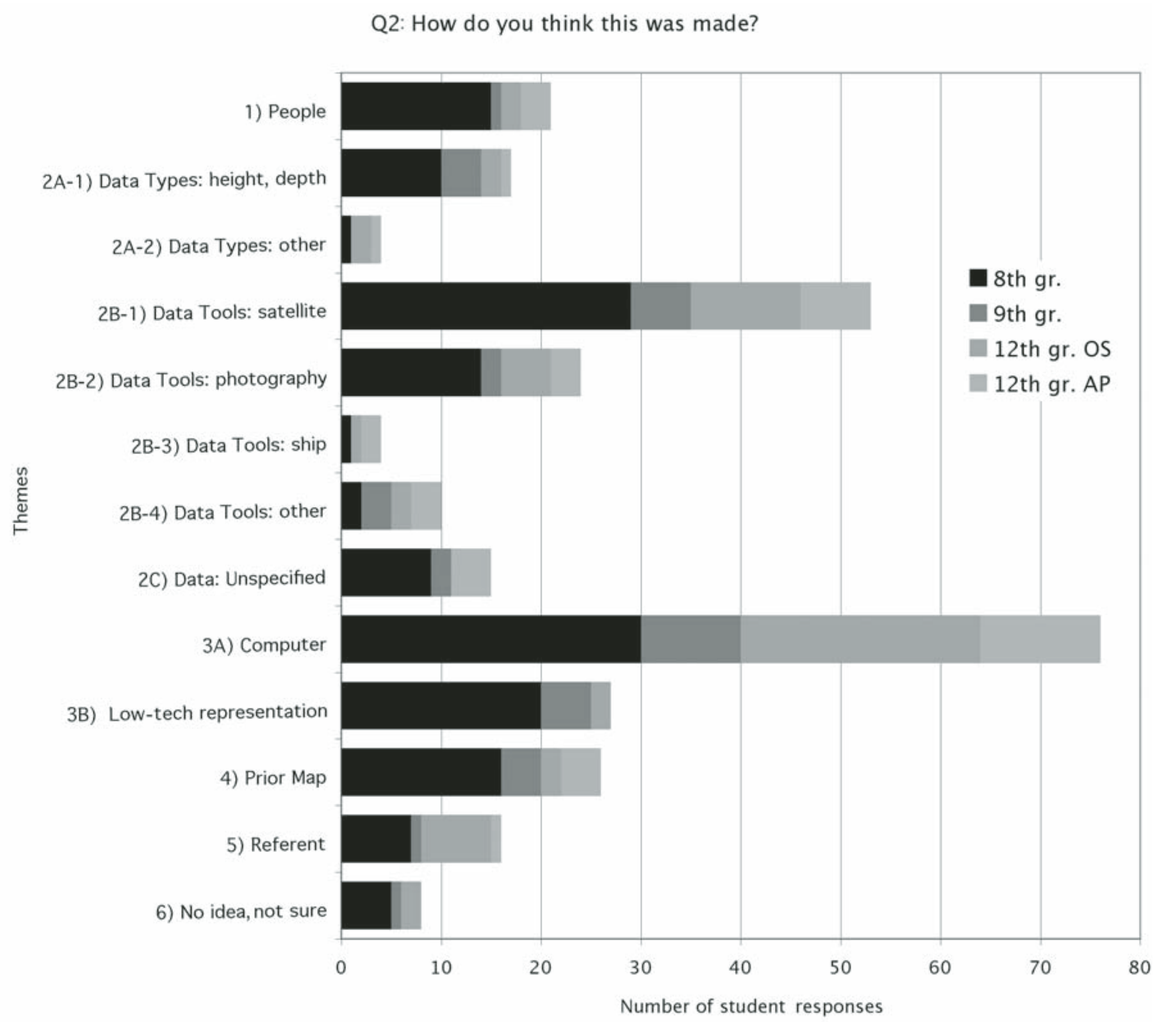


Figure 5. Student response to question 2: "How do you think this was made?" Bars represent the number of student responses exhibiting each coded theme or subtheme; some responses contained multiple themes. Among the responses involving data acquisition (theme 2), data acquired from a satellite or spaceship (2B-1) was the most common, with 27% (53/196) of the students providing responses that were coded in this category. Almost nobody mentioned use of a ship or other seagoing survey platform. Across all of the question 2 subthemes, use of a computer (hardware and/or software) was the most frequently mentioned, with 39% (76/196) of the student responses coded into this category.

Question 3: "What Do You Think This Is Useful For?"

Three major themes emerged from the third research question (Table 8). As in question 2, many of the responses included multiple elements; therefore, the researchers tallied the data, allowing more than one coding category per student response (Table 9; Fig. 6).

The first coding category was "Navigation," where the student described finding one's way or directing the course of a vehicle or vessel. A key element of this category was a sense of movement through space within or across the referent. Illustrative are: "I think this is useful for navigating the ocean floor," or "To show sailors' boats where land masses are on our planet" (Table 8). As a result, 7% (14/196) of the student responses were about navigation (Table 9).

The second theme, "Observations," concerned structures or features of Earth that could be observed directly from the provided map. One subcategory was "Observation about location of Earth's structures or features" (Table 8, 2A), for example, "I think this is useful for finding underwater mountain ranges, deep parts in the ocean and high mountains." The second subcategory was observations about "the shape or configuration [geomorphology] of Earth's structures or features" (Table 8, 2B). For example, "This is useful for discriminating elevations and depths, and finding landmasses and bodies of water." The total number of student responses for observations about location was 39%

TABLE 8. CODING OF MAJOR THEMES THAT EMERGED FROM QUESTION 3: "WHAT DO YOU THINK THIS IS USEFUL FOR?"

Category	Criteria	Examples
1. *Navigation*	Student response includes directing the course of something. May be oneself or an external object. Includes answers that are about travel, but without a specific mention of figuring out where you are while traveling.	• I think this is useful for **navigating the ocean floor**. 8th gr. • To show sailors['] boats where land masses are on our planet **for navigation purposes.** 8th gr. • To **predict the next islands and/or to avoid those places when in a ship.** 9th gr. • Visualizing the continents; **navigating the wide open waters** of our beautiful oceans. 12th gr. AP
2A. *Observation* about location of Earth's structures or features.	The response includes learning about the Earth by observing the map. For example, finding another location, physical feature, or structure.	• I think that this is useful for **finding underwater mountain ranges,** deep parts in the ocean and high mountains. 8th gr. • This is useful for trying to **figure out where everything is.** 8th gr.
2B. *Observation* about the shape or configuration of Earth's structures.	The response includes information about height and/or depth, shape, or physical attributes. Student is making an observation about geomorphology.	• This is useful because it shows us the fault lines and the **different elevations of physical features**. 8th gr. • This is useful for **discriminating elevations and depths,** and finding landmasses and bodies of water. 9th gr. • See different elevations such as mountains. Underwater volcanoes, plates. 9th gr.
3A. *Inferences* about solid Earth, including solid Earth processes.	Student response is about identifying, understanding, or predicting Earth processes. May include identifying patterns or making predictions.	• I think this is useful for **predicting possibilities of volcanic activity or earthquakes/other seismic activity** on the Earth. 9th gr. • Understanding the dynamics of the Earth's crust and the **movements of the tectonic plates.** 12th gr. AP • **Predicting earthquake threats** (as impossible as it might seem); studying marine geography/topography; oil drilling rigs. 12th gr. AP
3B. *Inferences other than* solid Earth.	Student response is an inference about anything other than solid Earth, including atmosphere, ocean (water), plants or animals, people and human activities.	• I think this lets people know **where a lot of sodium is so that they can fish in certain places.** 8th gr. • I think this is useful for **the directions of the currents & how the water moves around the earth.** 8th gr. • If people wish to **predict natural disasters or observe climate patterns,** this map would be a simple way to do it. 9th gr. • Predicting weather patterns and effects. 12th gr. AP
4. *Too general*	Response is too general to convey how this map is useful.	• Learning about the world. 8th gr. • Estimating how Earth will change in the future, and how it has in the past. 8th gr.
5. *No response*	Students did not respond, stated they did not know, or apparently misunderstood the question.	• I have no clue. 8th gr. • It is useful for people to live on and survive. 12th gr. OS *[Here, the student appears to be describing the referent.]*

Note: Only the portion of each response shown in boldface fits the coding category for the row. Response could be coded in more than one category.

(77/196) and for observations about geomorphology was 37% (72/196) (Table 9). These two were by far the most frequently coded response types for question 3.

The third theme, "Inferences," built on the idea that the global elevation map can be useful for making inferences about Earth, even of phenomena that are not directly shown on the map. Making inferences from data is an important skill at the heart of what scientists do. The two subcategories were inferences about the solid Earth (3A) and inferences about anything other than solid Earth (3B). The first subcategory included student responses such as, "I think this is useful for predicting possibilities of volcanic activity or earthquakes/other seismic activity on the Earth," and "Understanding the dynamics of the Earth's crust and the movements of the tectonic plates." The second category included inferences such as, "I think this is useful for the directions of the currents & how the water moves around the earth," or to "predict natural disasters or observe climate patterns" (Table 8). Most inferences focused on making predictions (e.g., of earthquakes) or making inferences from patterns (e.g., location of plate boundaries). No student explained how their suggested inferences could be made from the provided map. Total student responses in these two coding categories were 17% (34/196) for interpretations about the solid Earth and 21% (41/196) for inferences other than solid Earth (Table 9).

TABLE 9. DISTRIBUTION OF RESPONSES BY THEME TO QUESTION 3: "WHAT DO YOU THINK THIS IS USEFUL FOR?"

N = 196	Total (%)
1. Navigation	7
2A. Observations: locations	39
2B. Observations: geomorphology	37
3A. Inferences: solid Earth	17
3B. Inferences: other	21
4. Too general	8
5. No response	9

Note: Column does not sum to 100% because some responses included more than one information type.

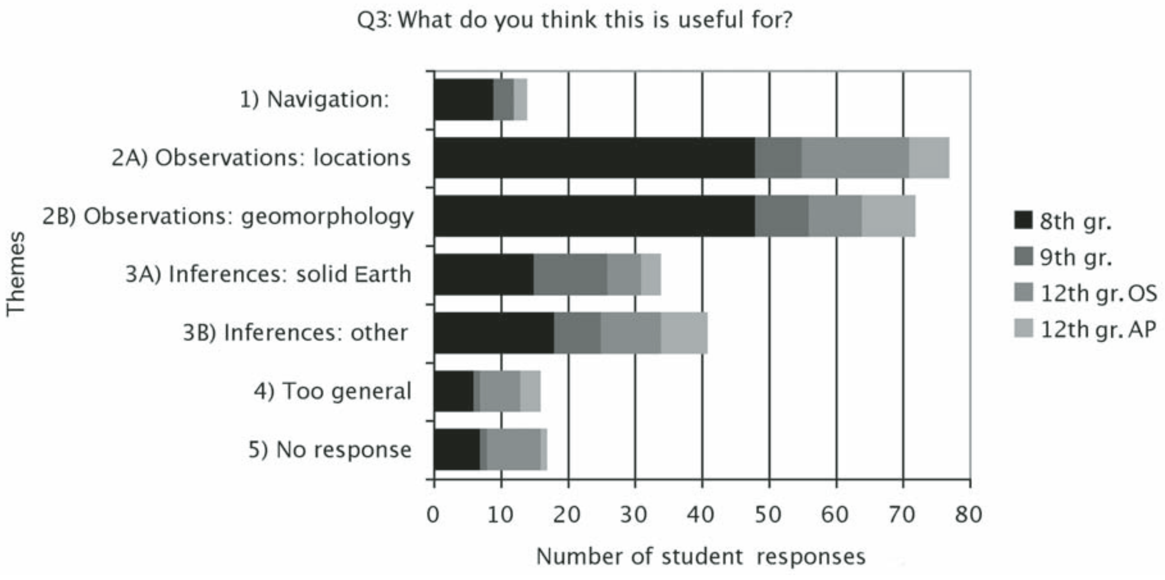


Figure 6. Student response to question 3: "What do you think this is useful for?" Bars represent the number of student responses exhibiting each coded theme or subtheme; some responses contained multiple themes. Most responses focused on using the map to obtain information that could be observed directly on the map, including location and shape of features.

The last three coding categories were about student responses that were too general to convey how the map could be useful (8%) or did not answer the question (9%).

DISCUSSION

Students' Understandings about What the Representation Is

Almost without exception, student responses to question 1 showed that they recognized that they were viewing a map, and that this map represents Earth. This was not a foregone conclusion; in a similar study (Swenson, 2010), involving a less iconic representation of global bathymetry/topography, a nontrivial minority (6%) of college non–science majors described only the colors of the representation, as though describing a work of abstract art, without mentioning the referent. The fact that our study participants were able to recognize the referent, Earth, on a representation they had probably never seen before, suggests that the distinctive visual pattern of the shapes and configurations of the continents are widespread in the visual recognition vocabulary of the population represented by these students.

Forty-four percent of the student responses (those coded as "Basemap") went only as far as identifying the scope of the referent (Earth with its continents and oceans), but they failed to mention any specific data type or aspect of the referent depicted by the representation. Our interpretation of what these students extracted from the viewed map resembles Figure 3 rather than Figure 1.

Only 30% of the responses (those coded "Topo") described a map of topography and bathymetry or allied concepts such as landforms, physical features, mountains, height of the land, or depth of the ocean. We find this result surprisingly low for several reasons. The participants were either current earth science students or had studied earth science within the past 4 yr, and the global bathymetry/topography data set is a fundamental constraint on solid Earth, ocean, and atmospheric processes. Moreover, the map portrays an aspect of Earth (relief) that students have had an opportunity to experience through direct perception and is an iconic representation that resembles the referent.

Students' Alternative Conceptions of the Representation

Responses to question 1 that were coded as "NonTopo" or "Topo/NonTopo" stated or implied that the data map contained information that simply is not there, for example, "I think that this is a world map of different ocean currents of the world. I think that the white and lighter blue lines on the map are the currents and the darker blue is the water," and "…green represents a lush and treeful environment, dark browns symbolize a barren and desert-like terrain."

Why might this be? We infer several possible sources of students' alternative conceptions from the nature of their responses.

First, some students seem to have made assumptions about iconicity that do not match the intentions of the map creator, for example, in interpreting green as vegetation (rather than low elevation) and white as ice or clouds (rather than shallow water depths). They may have assumed that the color scheme they saw on another map (for example, a biome map or weather map) carried over to this map. This would be an instance of "negative transfer," a situation in which a student's "experience with one set of events could hurt performance on related tasks" (Bransford et al., 2000, p. 53). Students may not have sufficient knowledge of visual representational strategies in general or cartographic conventions in particular to realize that the same colors can have different meanings on different maps.

Some students may not have had the life experiences that would have allowed them to take advantage of the iconicity put there by the map-maker. If they have never examined a terrain from an airplane or scenic overlook, the shaded relief aspect of the provided map may not be communicative. If they have not seen how water shades darker as it deepens going offshore across a beach or harbor, the significance of the varied shades of blue on the provided map may have escaped them.

Alternative conceptions were more common for the oceanic parts of the map than for the continents. Among responses coded as "Topo," a plurality (44%) mentioned only the continents; they constructed their answers by ignoring the oceans (Table 5). Conversely, among responses coded as "NonTopo," a plurality (38%) mentioned *only* the oceans. Among the hybrid "Topo&NonTopo" responses, a common pattern is that the alternative conception (NonTopo) refers to the oceans, for example: "I think this is an elevational topics chart. Showing elevation by ridges and tan color. Shows temperature by different blues also white for cold water." Although we do not have data to this effect, we strongly suspect that study participants had greater prior knowledge of the land surface than of the seafloor, because earth science and geography curricula focus on continents and because students' life experience is on land. If this is true, then the pattern of responses to question 1 suggests that students are more likely to misinterpret aspects of the representation where they come to the map with weak prior knowledge of corresponding aspects of the referent.

Finally, some students seem not to have adequately distinguished between "what this is a map of" and "what might be interpreted from the map." Strictly speaking, this is not "a map of fault lines," or "volcanoes," or "plate boundaries." Although the shape of landforms can be suggestive of causal processes such as faulting or volcanism, those are interpretative assertions rather than attributes of the map, and thus are not a correct response to the question "what is this?" Such answers to question 1 actually require a high level of knowledge of earth sciences, and can be considered as *overinterpretations* of the map, the opposite problem from the underinterpretations seen in the "Basemap" responses.

Students' Individual and Collective Knowledge of How the Map Was Made

Data maps are made by acquiring data in the field, processing the data into a manageable form with the intention of communicating something about the referent, and then generating a representation, with human decision making and software mediating each step along the way. Figure 2 illustrates experts' understanding of how the provided representation was created. We had no expectation that students would know how this data map was created, and, not surprisingly, no individual student articulated the entirety of Figure 2. However, encouragingly, most ideas expressed by students fell within the experts' model, and, collectively, the group of students is aware of all the major elements of the model. Students' existing knowledge provides a starting point to inform more purposeful curriculum design going forward.

With respect to the 11% of students who mentioned that a person or people were involved in making the data map, awareness that people are involved in making the representation humanizes the process of science, including the many different contributions made by technicians, engineers, cartographers, and scientists. Moreover, scholars who study symbol systems, of which maps are an example, emphasize that it is a developmental accomplishment for a young symbol-interpreter to understand the "intentionality" of the symbol-creator (Callaghan and Rochat, 2003), in other words, that "symbols mean what they are intended to mean by a creator (not what they happen to resemble)" (Myers and Liben, 2008, p. 682). This is particularly true for arbitrary symbols that have little or no physical resemblance to their referents, such as the colors on our study map. "Intentionality" is a human trait, and thus understanding that a person designed the map is prerequisite to understanding intentionality.

Among the 39% of students who mentioned that computer hardware and/or software was involved in creating the representation, some responses reflect only the awareness that the experimenter used a computer projector to show the image, but others indicate more varied roles for hardware and software in collecting and processing the data, for example, "computer design, drawing software'" "with a satellite through a computer," and "This was made using different programming languages using images probably Flash."

Within the secondary coding category, 44% of all students surveyed said something about data being acquired to make this representation (Table 7). Data acquired from a satellite or spaceship (2B-1) was the highest percentage among the subcategories of data acquisition, followed by photography (2B-2). These two elements were often combined (i.e., taking pictures from space), perhaps because students reason that the field of view of a photograph gets larger as the photographer moves farther away from the subject (Liben, 2008), and so to see the whole globe requires a vantage point in outer space. The actual data type (height and depth) and one of the actual tool types (ship, submarine, or sonar) were rarely mentioned.

On the other hand, 56% of the students did not mention that data were acquired from Earth. Of this group, nearly half indicated that the representation was made by a computer or some computer method with no mention of data acquisition (i.e., their response included 3A but none of the data acquisition codes). In terms of the epistemological model of Figure 2, these students expressed an understanding that reaches only slightly upstream from the representation itself, to the process of making the representation, and not all the way upstream to the processes of collecting and analyzing the underlying data.

Only a handful of responses indicated some process or element that is completely outside of the experts' epistemological model. The 14% (27/196) of students who answered in category 3B low-tech (e.g., "clay," "sketching") described a concept-driven visualization made with artistic techniques rather than a data-driven visualization (Clark and Wiebe, 2000; Kastens, 2009a).

In summary, individual students in our study possessed a partial understanding of where the data map came from and how it was made. Although no student articulated the entire big picture, *collectively*, the group was aware of all of the major elements in the experts' model: people, data acquisition of various types with various instruments, computer hardware and software, the image itself, and cartographic strategies (Fig. 2).

Students' Understanding of Utility of the Map

Students' responses to question 3 ("What do you think this is useful for?") spanned both observations and inferences. Most responses described using the map as a source of information that is actually shown on the map and can be directly observed on the map. Such responses encompassed the concepts of location (response category 2A) and shape (geomorphology, category 2B). As described previously ("Context for Survey Question 3"), maps can be used to depict location, shape, and configuration. Configuration is easy to depict on a map but difficult to express in words, so we could not tell whether students understood that maps are useful as a source of information about the configuration of Earth features.

A substantial fraction of students' responses to question 3 conveyed that maps can be used as the basis for inferences about features or phenomena that are not actually shown on the maps (Table 9; Fig. 6, categories 3A and 3B, "Inferences"), such as location of tectonic plates or volcanoes. In order to use a map to make inferences, the user needs to bring to the table additional information that is not in the map. In the case of the provided data map, the user needs to have and make use of the insight that Earth is dynamic and undergoes processes that cause it to differ from place to place and time to time. The bumps and wiggles of Earth's surface carry meaning or significance, in terms of (1) causative processes (e.g., plate tectonics) and (2) societal and human consequences (e.g., constraints on land use) (Kastens, 2009b).

As is apparent from this discussion, most responses concerned the purpose or purposes for which a map could be used. Some respondents also mentioned the types of people who might find the map useful. A map can be useful to the map creator, as a means of recording and organizing information. A map can also be useful to a recipient of the completed map for finding out or thinking about information. Among such responses to question 3, the apparent beneficiary was always a map recipient; no response suggested the insight that a data map is also useful to the creator of the map.

Liben and Downs (1989) noted that young children can recognize a road map as something useful for finding places before they can interpret the details of the map, let alone use it themselves for personal navigation. Similarly, students in our study were able to recognize and describe uses for the topography/bathymetry map even without full mastery of either the representational strategies or how to use the map themselves. This is a promising finding from an instructional perspective, because it suggests that it should be possible to sequence instruction by beginning with a motivational discussion of what the data map is useful for, without having to first slog through the details of the map's representational strategies and symbol system.

Implications for Instructional Design

The overwhelming finding from our study is that many students who are currently studying or have recently studied earth science do not demonstrate a robust understanding of one of the most fundamental data sets in geosciences, the shape of the solid Earth's surface. Substantial fractions of the study population misinterpreted an iconic representation of global bathymetry/topography and displayed at best fragmentary knowledge of how such a representation could have been made. Curriculum designers and instructors need to guard against the assumption that students will find data visualizations easy or obvious just because they appear more intuitively accessible than, for example, graphs or tables of numbers.

Participants in our study were more likely to misread those parts of the provided map where they had less prior knowledge of the referent, i.e., the seafloor. Student descriptions of the oceanic parts of the map deviated far from the normative answer, encompassing currents, tides, water temperature, and level of sodium, suggesting that although they knew conceptually that the oceans had currents, tides, sodium, etc., they knew very little about the spatial distribution of those phenomena, which bear no resemblance to the provided map. Learning about Earth through maps and other representations would seem to be an iterative or spiral process, in which one needs to know something about the referent to understand the representation (Dutrow, 2007), at which point one can use the representation to deepen one's knowledge of the referent, after which one may be able to appreciate more subtle nuances of the representation, and so on. Uttal (2000, p. 247–248) documented a reciprocal relationship: "As children acquire new and more sophisticated ways of mentally representing and using spatial information, their understanding of maps improves. Likewise, children's developing conception of maps affects how they understand and conceive of spatial information." This reciprocal relationship suggests that students will benefit from repeated exposure to rich data sets such as bathymetry/topography, which continue to yield new insights as students' knowledge of both Earth processes and representational strategies grows from elementary school through graduate school.

At present, the burden of providing frequent exposure to data maps and other data visualizations lies with the teacher. The illustrations for middle school and high school earth science textbooks are overwhelmingly concept-driven visualizations and photographs, with data-driven visualizations comprising only a few percent of the figures (Kastens, personal observ.). We would encourage teachers to hang data maps on their classroom walls and use the rich assortment of data maps available from the Internet as visual aids in explanations and as the focal point for class discussions, modeling how the data can be used as evidence to support inferences about natural processes and human-Earth interactions. Textbook authors should move toward incorporating more data-driven visualizations alongside photographs and concept-driven visualizations; the college textbook by Reynolds et al. (2007) is a good model.

Students need to do more than look at data visualizations passively; they need to engage with them actively (Dutrow, 2007). Wiggins and McTighe (2006) suggested that instructional design should be guided by a vision of what learning performance students should be able to do after instruction. For this design tradition, a useful roadmap is provided by Liben's (1997) research-based taxonomy of four ways in which children can demonstrate map understanding. Liben's first two methods take place in a field setting, where the representation and the referent can be directly compared and contrasted. The realization that children need to have firsthand experience with a terrain rather than just learning from a map goes back at least to John Dewey (1902, p. 26), who wrote, "The map is not a substitute for a personal experience. The map does not take the place of an actual journey." In Liben's "production methods" (1997), the learner produces a map based on observations of the referent or adds information to an existing map based on direct observation of the referent, as in geological mapping. In "comprehension methods," the child interprets a map in the field and demonstrates understanding by performing an action within the real world guided by information on the map, for example, by moving to a series of sampling stations. Liben's third and fourth demonstrations of map mastery are suitable for classroom use, where a map or maps are present but the referent is not. In "representational correspondence methods," the child transfers information from one form of representation to another, as from a relief map to a profile. A variant of this method would be to compare and contrast information from two or more representations, for example, a geological map and a relief map, or a population distribution map and a relief map, and draw inferences about the referent based on this comparison. Liben's final category of map mastery is "metarepresentional methods," in which the child reflects on the relationship between the representation and the referent, for example, by explaining the meaning of the colors in Figure 1 or by describing how the data were collected to make a specific data map. In our opinion, all four methods have a place in a thorough earth science education.

Although individual students had only fragmentary understanding of how the data map was made, the group of students collectively had knowledge of all of the elements contained in the experts' epistemological model of Figure 2: data acquisition by sensors in the field, involvement of people who collect data and make representations, use of computer hardware and software, and decisions about how to represent the data. This suggests that students might benefit from a group activity in which students combine their fragmentary knowledge to assemble a more nearly complete group understanding of how the provided representation was created, for example, by collaborating to fill in a partially incomplete version of Figure 2. To guide such an activity, the teachers themselves will need a good understanding of the epistemological model.

One final suggestion emerges from our finding that some of the students who had the strongest apparent knowledge base about the provided map responded to question 1 by making interpretive assertions (e.g., this is a map of faults, volcanoes, or plate boundaries). In fact, the map shows physiography; faults, volcanoes, and plate boundaries can be inferred but are not part of this map. As in all other aspects of science education, teaching with data maps requires constant attention to the distinction between what is observation (e.g., this is a bathymetric trench) and what is interpretation (e.g., this is a subduction zone). Students need multiple opportunities to examine experts' interpretations from data maps, teasing out the data-based evidence from the line of reasoning that leads from the data to the interpretation. Next, they need opportunities to make their own interpretations and defend those interpretations with evidence derived from the data map.

Directions for Future Research

In order to keep the research design tractable, the map used in this study was static (Libarkin and Brick, 2002). However, GeoMapApp, like other modern data visualization tools, provides a rich suite of interactivity, including the capability to zoom into areas of interest, to create profiles at any desired azimuth and position, to create "3-D" terrain-like representations, and to adjust color, sun angle, and vertical exaggeration at will. In what ways would access to any or all of these functions improve or change students' understanding of what this data map is or is useful for? A fruitful line of research would be observational and think-aloud studies of how individuals navigate through the rich set of functions provided by modern data visualization tools in pursuit of answers to authentic geoscience inquiries; such observations would help researchers understand how students are conceptualizing and prioritizing the information in the database.

The data map used in this study depicts a data type that quantifies an aspect of the referent Earth that students have experienced directly, by walking across nonhorizontal terrain and by viewing landscapes with their own stereoscopic visual system. For this data type, students are relatively rich in direct knowledge of the referent (in terms of Liben, 1999, 2006). Similar research should be conducted on students' understanding of representations of geoscience data types where students come equipped with less direct knowledge of the referent. Such a research agenda should encompass aspects of the referent that students can sense but not see as a spatial array in nature (e.g., sea-surface temperature) as well as aspects that are not sensible at all through human senses (e.g., magnetic field).

We infer that difficulties for our study participants arose from *both* inadequate grasp of representational strategies (as when they assumed that green symbolized vegetation) and incomplete knowledge of the referent (as when the same students interpreted the continents correctly and the oceans incorrectly). As Edelson (1998) pointed out, when a scientist interprets a scientific visualization, he or she draws on a rich knowledge of scientific phenomena. Many of the specialized representations used by geoscientists present a chicken-and-egg situation, in which learners must understand something about Earth to interpret the representation and yet the representation is the means by which we teach about Earth (Kastens and Manduca, 2009). Following the lead of Dutrow (2007), we suggest a spiraling instructional progression in which gradually deepening knowledge of Earth and gradually more sophisticated mastery of representational strategies are built up in parallel. The burgeoning field of research on learning progressions (e.g., Mohan et al., 2009; Duschl et al., 2007) may be able to provide insight into the situation where the learner needs to have some understanding of A to understand B, and yet needs to understand B in order to understand A. Blades (2000) stressed how little research has been done on the ways in which learners integrate information learned from spatial representations with information gained through direct experience with the environment.

The relationship between students' ability to extract insights from data they did not collect and their knowledge of how the data were collected remains an area of active research. This question could be addressed through intervention studies: Does the experience of collecting and interpreting a small data set in one's own locality transfer into increased ability to extract insights from professionally collected large-scale data sets, perhaps by providing needed context (Winn et al., 2006)? Does learning about how scientists collect and process data (for example, through videos of field research) transfer into deeper insights about Earth when students later work with data maps and other data visualizations? Geoscientists, especially geoscientists who do field-based research, would tend to say "yes, obviously." However, there is little educational research to test this assertion or to elucidate the nature of the transfer; this topic is ripe for a combination of qualitative and quantitative research.

Liben (1999, 2006) makes the case that if learners do not grasp the representational strategies that have been used by a map-maker, they are vulnerable to "mis-mediated knowledge" of the referent when the representation is used as the means to study the referent. We consider that "representational strategies" are not limited to merely the last step of generating an external representation from data, but rather constitute the entire "chain of inscriptions" (Latour, 1986, 1987) from the referent to the representation shown in Figure 2. It is unclear how much students at different levels need to know about the processes shown in Figure 2 in order to avoid mis-mediated knowledge of Earth. Surely an eighth grader does not need a complete understanding of Figure 2 in order to avoid mis-mediated knowledge at an educationally appropriate level, but a doctoral student in marine geology certainly does. What about students in between? What is the nature of the mis-mediation caused by various forms of missing or mistaken knowledge of representational strategies?

CONCLUSIONS

Almost all study participants recognized the most basic elements of the provided visualization: the nature of the representation (a map), the scope of the referent (Earth), and a familiar pattern (the outlines of the continents). A substantial minority of the students recognized that the land (and less often the ocean) portions of the map show elevation/relief/landforms. Among the group of students taken as a whole, there was some awareness of each of the major processes that had contributed to making the data map, and some awareness of the utility of the map for both practical purposes (such as navigation) and scientific research (such as interpreting plate boundaries).

On the other hand, many students described the map in terms that depart wildly from the normative answer, as a map of vegetation, climatic zones, tides, migration patterns, sodium level, currents, weather, clouds, etc. No individual student presented a coherent explanation of how the data map was made, linking something about data acquisition, something about data processing, and something about representational techniques. Most students' ideas about what the map might be useful for were confined to low-level information-retrieval tasks, such as finding out where something is located or what its shape is.

Although the documented level of understanding provides a good foundation for further instruction, it seems to be a low level for students who are currently enrolled in or have completed an earth science course. Elevation is not esoteric: it is one of the most fundamental global geoscience data sets for explaining solid Earth, ocean, and atmospheric processes, and one best grounded in everyday experience. We recognize that different probes or follow-up questions might have revealed broader knowledge or deeper insight, but the pattern of responses taken as a whole suggests knowledge that is rather fragmentary, in which the elevation data do not connect back to Earth through a series of data acquisition and processing steps, nor forward to interpretation though lines of logical reasoning. Certainly, most of these students do not seem ready to use global elevation data in the way envisioned by the National Academy of Sciences in the opening quote in our introduction, as "evidence to construct testable explanations and predictions of natural phenomena."

The pattern of responses, interpreted in light of prior research on spatial thinking and student learning, suggests several factors that may have contributed to the observed difficulties. Students may have inappropriately interpreted the symbol system of the provided map because they expected that colors on the map would directly correspond to colors in the referent, because they negatively transferred symbol systems from other maps, or because they lacked relevant personal experience such as viewing terrain from an airplane or a mountaintop. The alternative conceptions that emerged in response to question 1 suggest that knowledge of the referent and understanding of the representation intertwine in a complicated way: To produce an answer of "currents" or "temperature" or "sodium level," it seems that the student must simultaneously *possess* the conceptual knowledge that oceans have currents, temperature variation, etc., but *lack* the spatial knowledge of how those attributes are distributed. To use the information on the data map as evidence in support of inference, students need to understand that the map records the bumps and wiggles in the referent, *and* that the bumps and wiggles in the referent record Earth processes.

To build on the documented level of student understanding through instruction, we offer the following suggestions. Data visualizations, including data maps, should feature prominently in all aspects of earth science instruction, including teachers' presentations, class discussions, inquiry activities, and textbooks (where the ratio of concept-driven visualizations to data-driven visualizations is currently overbalanced toward concept-driven). The abundance of geoscience data visualizations available through the Internet makes this suggestion viable as never before in educational history. However, teachers need support in developing the pedagogical content knowledge that will enable them to choose data wisely and use it effectively (Edelson, 1998). Students should work with local data maps in the field, in production and comprehension activities that require them to translate back and forth between the representation and the referent when both are within view. In preparation for this form of teaching, preservice and in-service teacher professional development for earth science teachers should include instruction and practice in field-based education. Finally, there is a need for additional middle- and high-school level inquiry activities in which students use evidence from data maps to construct explanations and predictions.

ACKNOWLEDGMENTS

The authors gratefully acknowledge advice and encouragement from O. Roger Anderson, Dale Chayes, Andrew Goodwille, the late William Haxby, Ann Rivet, William B.F. Ryan, and the Advisory Committee of the Marine Geoscience Data System. We appreciate the insights provided by the participating students and their teachers and by two anonymous reviewers and Special Paper 474 co-editors Stokes and Feig. Partial funding was provided by the National Science Foundation RIDGE2000 program through grant OCE-03-28117. This is a Lamont-Doherty Earth Observatory contribution.

REFERENCES CITED

Blades, M., 2000, Commentary: Young children's understanding of indirect sources of spatial information: Developmental Science, v. 3, no. 3, p. 265–266.

Bransford, J.D., Brown, A.L., and Cocking, R.R., eds., 2000, National Research Council: How People Learn: Brain, Mind, Experience, and School: Washington, D.C., National Academy Press, 319 p.

Brasell, H., 1987, The effect of real-time laboratory graphing on learning graphic representations of distance and velocity: Journal of Research in Science Teaching, v. 24, no. 4, p. 385–395, doi:10.1002/tea.3660240409.

Callaghan, T.C., and Rochat, P., 2003, Traces of the artist: Sensitivity to the role of the artist in children's pictorial reasoning: The British Journal of Developmental Psychology, v. 21, p. 415–445, doi:10.1348/026151003322277784.

Carbotte, S.M., Arko, R., Chayes, D.N., Haxby, W., Lehnert, K., O'Hara, S., Ryan, W.B.F., Weissel, R.A., Hipley, T.S., Gahagan, L., and Hank, T.S., 2004, New integrated data management system for Ridge2000 and margins research: Eos (Transactions, American Geophysical Union), v. 85, no. 51, p. 553, 559; available at http://www.marine-geo.org/about/docs/reports/Carbotte_etal_2004EO510002.pdf (accessed 4 November 2010).

Carbotte, S., Chayes, D., Ryan, W., Arko, R., Haxby, W., Lehnert, K., O'Hara, S., Shank, T., and Tolstoy, M., 2005, Status Report on the Ridge 2000 Data Management System in Ridge 2000 Events: Newsletter of the Ridge 2000 Program: http://www.ridge2000.org/science/downloads/newsletters/R2KNews0705.pdf (accessed January 2010).

Caress, D., and Chayes, D., 2009, MB-System: Mapping the Seafloor: Software for the Processing and Display of Swath Sonar Data: http://www.ldeo.columbia.edu/res/pi/MB-System/ (accessed 3 February 2010).

Chayes, D., 1999, Bits to Data: http://www.ldeo.columbia.edu/~dale/dataflow/ (accessed 3 February 2010).

Chi, M.T.H., 1997, Quantifying qualitative analyses of verbal data: A practical guide: Journal of the Learning Sciences, v. 6, no. 3, p. 271–315, doi:10.1207/s15327809jls0603_1.

Clark, A.C., and Wiebe, E.N., 2000, Scientific visualization for secondary and post-secondary schools: Journal of Technology Studies, v. 26, no. 1, p. 24–32.

Creswell, J.W., 2003, Research Design: Qualitative, Quantitative, and Mixed Methods Approach: Thousand Oaks, California, Sage Publications, 246 p.

DeLoache, J.S., 2000, Dual representation and young children's use of scale models: Child Development, v. 71, p. 329–338, doi:10.1111/1467-8624.00148.

Dewey, J., 1902, The Child and the Curriculum: Contributions to Education Number V: Chicago, University of Chicago Press, 40 p.

Downs, R.M., and Liben, L., 1987, Children's understanding of maps, *in* Ellen, P., and Thinus-Blanc, C., eds., Cognitive Processes and Spatial Orientation in Animal and Man: Volume 2. Neurophysiology and Development Aspects: Dordrecht, Martinus Nijhoff, p. 202–219.

Driver, R., Guesne, E., and Tiberghien, A.E., 1985, Children's Ideas in Science: Philadelphia, Open University Press, 208 p.

Driver, R., Leach, J., Millar, R., and Scott, P., 1996, Young People's Images of Science: Philadelphia, Open University Press, 172 p.

Duschl, R.A., Schweingruber, H.A., and Shouse, A.W., 2007, Taking Science to School: Learning and Teaching Science in Grades K–8: National Academy Press, 404 p.

Dutrow, B.L., 2007, Teaching with visuals: Do you see what I see?: Elements (Special Issue on Teaching Mineralogy, Petrology & Geochemistry), v. 3, p. 119–126.

Edelson, D.C., 1998, Realising authentic science learning through the adaptation of scientific practice, *in* Fraser, B.J., and Tobin, K.G., eds., International Handbook of Science Education: Dordrecht, the Netherlands, Kluwer Academic Publishers, p. 317–331.

Edelson, D.C., Gordin, D.N., and Pea, R.D., 1999, Addressing the challenges of inquiry-based learning through technology and curriculum design: Journal of the Learning Sciences, v. 8, p. 391–450.

Feig, A.D., 2011, this volume, Methodology and location in the context of qualitative data and theoretical frameworks in geoscience education research, *in* Feig, A.D., and Stokes, A., eds., Qualitative Inquiry in Geoscience Education Research: Geological Society of America Special Paper 474, doi:10.1130/2011.2474(01).

Geography Education Standards Project, 1994, Geography for Life: National Geography Standards: Washington, D.C., National Geographic Research & Exploration, 272 p.

Germann, P.J., and Aram, R.J., 1996, student performances on the science processes of recording data, analyzing data, drawing conclusions, and providing evidence: Journal of Research in Science Teaching, v. 33, no. 7, p. 773–798, doi:10.1002/(SICI)1098-2736(199609)33:7<773::AID-TEA5>3.0.CO;2-K.

Gilbert, J.K., ed., 2005, Visualization in Science Education: Dordrecht, the Netherlands, Springer, 346 p.

Hall, S.S., 1992, Mapping the Next Millennium: The Discovery of New Geographies: New York, Random House, 477 p.

Hays, J.D., Pfirman, S., Blumenthal, B., Kastens, K., and Menke, W., 2000, Earth science instruction with digital data: Computers & Geosciences, v. 26, p. 657–668, doi:10.1016/S0098-3004(99)00101-6.

Hug, B., and McNeill, K.L., 2008, Use of first-hand and second-hand data in science: Does data type influence classroom conversations?: International Journal of Science Education, v. 30, no. 13, p. 1725–1751, doi:10.1080/09500690701506945.

Ishikawa, T., and Kastens, K.A., 2005, Why some students have trouble with maps and other spatial representations: Journal of Geoscience Education, v. 53, no. 2, p. 184–195.

Ishikawa, T., Kastens, K.A., Barnston, A.G., Louchouarn, P., and Ropelewski, C.F., 2005, Testing the efficacy of climate forecast maps as a means of communicating with policy makers: Cartography and Geographic Information Science, v. 32, no. 1, p. 3–16, doi:10.1559/1523040053270747.

Kanari, Z., and Millar, R., 2004, Reasoning from data: How students collect and interpret data in science investigations: Journal of Research in Science Teaching, v. 41, no. 7, p. 748–769, doi:10.1002/tea.20020.

Kastens, K., 2009a, Data-driven versus concept-driven visualizations: Earth & Mind: The Blog: http://serc.carleton.edu/earthandmind/posts/datadrivenvis.html (accessed 3 November 2009).

Kastens, K., 2009b, The meaning of "meaning": Causes and consequences: http://serc.carleton.edu/earthandmind/posts/meaning_meaning.html (accessed 3 November 2009).

Kastens, K.A., and Liben, L.S., 2007, Eliciting self-explanations improves children's performance on a field-based map skills task: Cognition and Instruction, v. 25, no. 1, p. 45–74.

Kastens, K.A., and Manduca, C.A., 2009, Synthesis of research on thinking & learning in the geosciences: Developing representational competence: Geological Society of America Abstracts with Programs, v. 41, no. 7, p. 249.

Kastens, K., and Turrin, M., 2006, To what extent should human/environment interactions be included in science education?: Journal of Geoscience Education, Special Issue on Earth System Science Education, v. 54, p. 422–436.

Kastens, K.A., Bonatti, E., Caress, D., Carrara, G., Dauteuil, O., Frueh-Green, G., Ligi, M., and Tartarotti, P., 2000, The Vema Transverse Ridge (central Atlantic): Marine Geophysical Researches, v. 20, p. 533–556, doi:10.1023/A:1004745127999.

Kastens, K.A., Kaplan, D., and Christie-Blick, K., 2001, Development and evaluation of Where Are We? Map-skills software and curriculum: Journal of Geoscience Education, v. 49, p. 249–266.

Kuhn, T.S., 1962, The Structure of Scientific Revolutions: Chicago, University of Chicago Press, 172 p.

Latour, B., 1986, Science in Action: How to Follow Scientists and Engineers through Society: Cambridge, Massachusetts, Harvard University Press, 196 p.

Latour, B., 1987, Science in Action: How to Follow Scientists and Engineers through Society: Milton, Keynes, UK, Open University Press, 271 p.

Libarkin, J.C., and Brick, C., 2002, Research methodologies in science education; visualization and the geosciences: Journal of Geoscience Education, v. 50, p. 449–455.

Libarkin, L., and Kurdziel, J.P., 2002, Research methodologies in science education: Qualitative data: Journal of Geoscience Education, v. 50, p. 195–200.

Liben, L.S., 1997, Children's understanding of spatial representations of place: Mapping the methodological landscape, *in* Foreman, N., and Gillett, R., eds., Handbook of Spatial Research Paradigms and Methodologies: Volume 1. Spatial Cognition in the Child and Adult: East Sussex, UK, The Psychology Press (Taylor & Francis Group), p. 41–82.

Liben, L., 1999, Developing an understanding of external spatial representations, *in* Sigel, I.E., ed., Development of Mental Representation: Theories and Applications: Hillsdale, New Jersey, Lawrence Erlbaum Associates, p. 297–321.

Liben, L.S., 2003, Thinking through maps, *in* Gattis, M., ed., Spatial Schema and Abstract Thought: Cambridge, Massachusetts, Massachusetts Institute of Technology Press, p. 45–78.

Liben, L., 2006, Education for spatial thinking, *in* Renninger, K.A., and Sigel, I.E., eds., Handbook of Child Psychology (6th edition): Volume 4. Child Psychology in Practice: Hoboken, New Jersey, Wiley, p. 197–247.

Liben, L.S., 2008, Embodiment and children's understanding of the real and represented world, *in* Overton, W.F., Muller, U., and Newman, J.L., eds., Developmental Perspectives on Embodiment and Consciousness: New York, Lawrence Erlbaum Associates, p. 191–224.

Liben, L., and Downs, R.M., 1989, Understanding maps as symbols: The development of map concepts in children, *in* Reese, H.W., ed., Advances in Child Development and Behavior, v. 22: New York, Academic Press, p. 145–201.

Linn, M., and Songer, N.B., 1991, Teaching thermodynamics to middle school students: What are appropriate cognitive demands?: Journal of Research in Science Teaching, v. 28, no. 10, p. 885–918.

MacEachren, A.M., 1995, How Maps Work: Representation, Visualization, and Design: New York, Gilford Press, 523 p.

MacKay, B., 2006, Teaching with Visualizations. Starting Point: Teaching Entry Level Geoscience: Science Education Research Center, Carleton College: http://serc.carleton.edu/introgeo/visualizations/index.html (accessed 17 July 2007).

Miles, M.B., and Huberman, A.M., 1994, Qualitative Data Analysis (2nd ed.): Thousand Oaks, California, Sage Publications, 338 p.

Mohan, L., Chen, J., and Anderson, C.W., 2009, Developing a multi-year learning progression for carbon cycling: Socio-ecological systems: Journal of Research in Science Teaching, v. 46, no. 6, p. 675–698, doi:10.1002/tea.20314.

Mokros, J.R., and Tinker, R.F., 1987, The impact of microcomputer-based labs on children's ability to interpret graphs: Journal of Research in Science Teaching, v. 24, no. 4, p. 369–383, doi:10.1002/tea.3660240408.

Myers, L.J., and Liben, L.S., 2008, The role of intentionality and iconicity in children's developing comprehension and production of cartographic symbols: Child Development, v. 79, no. 3, p. 668–684, doi:10.1111/j.1467-8624.2008.01150.x.

Nachmias, R., and Linn, M.C., 1987, Evaluations of science laboratory data: The role of computer-presented information: Journal of Research in Science Teaching, v. 24, no. 5, p. 491–506, doi:10.1002/tea.3660240509.

National Academy of Sciences, 2008, Science, Evolution, and Creationism: Washington, D.C., The National Academies Press, 70 p.

National Council of Teachers in Mathematics, 2000, Principles and Standards for School Mathematics: Reston, Virginia, National Council of Teachers in Mathematics, 402 p.

Prothero, W., 2006, Earth Education Online: http://earthednet.org/ (accessed 15 January 2010).

Reynolds, S., Johnson, J., Kelly, M., Morin, P., and Carter, C., 2007, Exploring Geology: Dubuque, Iowa, McGraw-Hill Science/Engineering/Math, 640 p.

Robinson, A.H., and Petchenik, B.B., 1975, The map as a communication system: Cartographic Journal, v. 12, p. 7–15.

Roth, W.-M., Pozzer-Ardenghi, L., and Han, J.Y., 2007, Critical Graphicacy: Understanding Visual Representation Practices in School Science: Dordrecht, the Netherlands, Springer, 285 p.

Roushias, C.R., and Anderson, O.R., 2001, Earth View Explorer: A novel, student-centered computer-based learning environment, *in* Proceedings of the Second Mediterranean Computer Using Educators Conference, 4–5 November 2000: Nicosia, Cyprus, p. 172–187.

Sambrotto, R., and Anderson, O.R., 2001, Earth View Explorer: Earth View Teacher Guide: New York, The Learning Team, 198 p.

Sawyer, D., 2005, Discovering Plate Boundaries: Rice University: http://www.geophysics.rice.edu/plateboundary/home.html (accessed 15 December 2009).

Sobel, D., 1998, Mapmaking with Children: Sense of Place Education for the Elementary Years: Portsmouth, New Hampshire, Heinemann, 164 p.

Swenson, S., 2010, Undergraduate Non-Science Majors' Descriptions and Interpretations of Scientific Data Visualizations [Ed.D. thesis]: New York, Columbia University Teachers College, 204 p.

Trumbull, D.J., Bonney, R., Bascom, D., and Cabral, A., 2000, Thinking scientifically during participation in a citizen-science project: Science Education, v. 84, p. 265–275, doi:10.1002/(SICI)1098-237X(200003)84:2<265::AID-SCE7>3.0.CO;2-5.

Tufte, E.R., 2001, The Visual Display of Quantitative Information: Cheshire, Connecticut, Graphics Press, 197 p.

Uttal, D.H., 2000, Seeing the big picture: Map use and the development of spatial cognition: Developmental Science, v. 3, no. 3, p. 247–264, doi:10.1111/1467-7687.00119.

Ware, C., 2004, Information Visualization: Perception for Design: Amsterdam, Elsevier, 438 p.

Wiggins, G., and McTighe, J., 2006, Understanding by Design (expanded 2nd ed.): Columbus, Ohio, Merrill Prentice Hall, 371 p.

Winn, W., Stahr, F., Sarason, C., Fruuland, R., Oppenheimer, P., and Lee, Y.-L., 2006, Learning oceanography from a computer simulation compared with direct experience at sea: Journal of Research in Science Teaching, v. 43, no. 1, p. 25–42, doi:10.1002/tea.20097.

Manuscript Accepted by the Society 23 June 2010

CONTENTS

THE
GEOLOGICAL
SOCIETY
OF AMERICA®

3300 Penrose Place • P.O. Box 9140
Boulder, CO 80301-9140, USA

ISBN 978-0-8137-2474-4