



A typology of causal models for plate tectonics: Inferential power and barriers to understandingⁱ

Janice D. Gobert Department of Learning and Teaching, Harvard University, Cambridge, MA 02138 and Department of Science Studies, Western Michigan University, Kalamazoo, MI 49008, USA

Forty-seven fifth grade students (40 group-tested and 7 individually interviewed) read a text describing plate tectonics. At four points they drew diagrams of the spatial, causal, and dynamic processes inside the earth. These diagrams along with students' corresponding explanations, think-aloud protocols (for those individually interviewed), and answers to inference questions were analysed in order to characterize students' models of the interior of the earth, and models of its causal and dynamic processes. Types and characteristics of models, and reasoning associated with them are presented. Additionally, data from two exemplary students are presented as case studies. One student has considerable misunderstandings regarding both her understanding of the spatial layout of the interior of the earth and its causal mechanisms. The second student is more typical in terms of his initial models, but makes large gains in revising his understanding about the causal and dynamic processes inside the earth. In both cases, data are used to infer how each student used their diagrams as artefacts for externalizing knowledge, inference making, and model-revision.

Introduction

In a previous study (Gobert and Clement 1999) it was found that having students generate diagrams during their reading of a text about plate tectonics was better at promoting students' post-text conceptual understanding of the spatial, causal, and dynamic features of the domain compared to generating summaries while reading the text or simply reading the text only (control). Although the diagram group outperformed the summary group on post-text performance, the summaries (generated by the summary group) during the reading of the text contained more semantic information than did the diagrams (generated by the diagram group). These data were interpreted in accordance with current literature on constructing mental models from textual information sources (Johnson-Laird 1983, Kintsch 1998, Schmalhofer 1998) as follows. For the summary group, because the media was the same (i.e. they were reading and generating text), they were able to rely on a rote memory of the textual material in order to generate their summaries. For the diagram group, on the other hand, the task of generating diagrams was a higher-level task which required them to do additional processing on the textual material they had read in order to generate their diagrams. Thus, the cognitive processing of the textual materials and interaction of the text with the learning tasks, i.e. either diagram-drawing or summary-writing, are reflected in each groups' performance on the post-text assessment and the intermittent tasks during students' reading of

International Journal of Science Education ISSN 0950-0693 print/ISSN 1464-5289 online © 2000 Taylor & Francis Ltd
<http://www.tandf.co.uk/journals>

the text. More specifically, the summary group's summaries generated during their reading of the text contained a great deal of content-related information which was processed on a rote level, as evidenced by their resulting mental models which were not as rich as those in the diagram group. The diagram groups' diagrams contained less content-related information due to the difficulty of generating diagrams from text, but their resulting mental models were richer and allowed greater inference-making as evidenced by their superior scores on the post-text (see Gobert and Clement 1999 for a more thorough explanation of these findings).

In extending the findings of Gobert and Clement (1999), the purposes of the present study were to: identify and characterize the different types of models held by middle schools students about the inside of the earth and of the causal and dynamic mechanisms involved in plate tectonics, and to characterize the nature of the reasoning associated with these models. Additionally, since it was found that drawing diagrams promoted a richer understanding of the domain compared to generating summaries (Gobert and Clement 1999), this research was conducted in order to identify the learning gains and inferences afforded when students construct diagrams for learning and use these diagrams to support model revision.

Domain studied

The present research addresses students' models and model construction processes central to learning a middle school science domain, namely plate tectonics. The research draws on current findings from research on causal models (White 1993, Schauble *et al.* 1991, Raghavan and Glaser 1995), model-based teaching and learning (Gilbert, S. 1991, Gilbert, J. 1993); model revising (Clement 1989, 1993, Stewart and Hafner 1991); diagram generation and comprehension (Gobert 1994; Gobert and Frederiksen 1988; Kindfield 1993; Larkin and Simon 1987, Lowe 1989, 1993), the integration of text and diagrams (Hegarty and Just 1993), and text comprehension (van Dijk and Kintsch 1983, Kintsch 1998). Findings about both students' causal models and model-based learning should be applicable to other science domains involving convection (e. g., other earth science topics and weather systems), other science topics at the middle school level (e. g., photosynthesis, properties of matter, heat and temperature, day/night cycle, seasonal change, planetary motion, and density), as well as scientific reasoning in general (Clement 1993).

Plate tectonics, representative of a difficult middle school science topic, was chosen for the domain of study because of the important role that model building and causal reasoning play in understanding the hidden, explanatory mechanisms, i.e. convection, underlying continental drift, earthquakes, volcanoes, mountain formation, and sea floor spreading. Briefly, the theory of plate tectonics, which offers a unified explanation of the past, present, and future geographic distribution of the earth's landmasses and oceans (Bencloski and Heyl 1985) proposes that the outer layer of the earth (the crust) is broken up into slabs (the plates) which move on the partially molten layer of the earth (the mantle) due to the slow convective movement of hot magma in the mantle (Feather *et al.* 1995, Plummer and McGeary 1996). This topic is typically covered in the fifth or sixth grade and then again in the eighth or ninth grade (MEGOSE 1991, Massachusetts Department of Education 1996).

Plate tectonics is difficult to learn for many reasons: (1) the earth's internal layers and unobserved processes, e. g., convection, are outside our direct experience (Ault 1984, Gobert and Clement 1994, 1999); (2) the size scale is difficult for children to understand (Ault 1984); (3) the time scale of geological processes is difficult for people to conceptualize since it surpasses our reference of a human lifetime (Jacobi *et al.* 1996); and (4) it involves the comprehension and integration of several different types of information, namely, spatial, causal, and dynamic information (Gobert and Clement 1994, 1999).

It is important to note that the goal in this program of research is to facilitate students' understanding of simplified, qualitative models of plate tectonics. As such, issues like whether radioactive decay in the mantle acts, in part, as a heat source in addition to the earth's core (Feather *et al.* 1995) are not addressed. Also, the physics involved in convection is not addressed. It is assumed that the models that students develop through instructional interventions such as those in the present study will scaffold further model revision and knowledge integration in later years when plate tectonics is addressed again in high school (MEGOSE 1991, Massachusetts Department of Education 1996), at which point more conceptually-difficult aspects of the domain can be addressed.

Previous research on earth science

Plate tectonics falls under the larger domain of Earth Science. More generally, the topic of learning in Earth Science has not been well studied, particularly when compared to students' learning and conceptions in the physical sciences (Stofflett 1994). The lack of research on learning in the Earth Sciences is likely due to the fact that in the past, it has received much less emphasis than the Physical and Life Sciences. Now however, the National Science Education Standards (National Committee Science on Science Education Standards and Assessment 1996) are recognizing Earth Science as a necessary and important component of science training across elementary, middle, and high school levels and considered equivalent in importance to training in the Life and Physical sciences (AAAS 1989, 1993).

Of the studies that have been carried out in the domain of earth science, some of the sub-topics and concepts that have been addressed are: the earth as a cosmic body (Vosniadou and Brewer 1992, Nussbaum 1979, Nussbaum and Novak 1976, Sneider and Pulos 1983); knowledge of rock-cycle processes (Stofflett 1994); conceptions of earth and space as it relates to seasons and phases of the moon, (Schoon 1992, Bisard *et al.* 1994); sea floor dynamics (Bencloski and Heyl 1985); knowledge of the earth's gravitational field (Arnold *et al.* 1995); mountain formation (Muthukrishna, *et al.* 1993); modelling to promote understanding of subtopics of earth science including the geosphere, hydrosphere, atmosphere, and biosphere (Tallon and Audet 1999); and environmental problem-solving (Pin *et al.* 1995).

One study directly relevant to the present research is that by Ross and Shuell (1993) who investigated children from kindergarten through sixth grade regarding their beliefs about the characteristics and causes of earthquakes. Regarding their cause, the majority of children answered that they didn't know. Idiosyncratic responses included: that the core gets too hot and hits the surface of the earth; the earth is letting out air like a sneeze; and that earthquakes are caused by the wind, thunder and rain, or by mountains. Asked about what happens below the

surface when there is an earthquake, again, a large proportion of the children answered that they did not know. Incorrect responses included: that roots underground pop; the plants might get ‘screwed up’ because the seeds would jiggle around; and that the earth has too much energy just like children who need to get rid of it. The responses to these questions not only indicate a lack of knowledge regarding the underlying processes of plate tectonics responsible for earthquakes, but also indicate difficulties in understanding the size scale of the earth, as was also found by Ault (1994). It is also important to note that misconceptions regarding earthquakes are not just found in children; for example, Bezzi (1989) found that 1/3 of secondary students interviewed from an area with considerable seismic activity thought that the occurrence of earthquakes was related to the occurrence of volcanoes. Furthermore, Turner *et al.* (1986) found that of 1450 adults interviewed from southern California, many held the misconception that earthquakes could be predicted by ‘earthquake weather’. Thus, in the case of understanding plate tectonics, as with other science domains as well, it is not likely that children’s views become more scientifically accurate as they mature. This further necessitates the need to identify the nature of students’ pre-instruction models of plate tectonics, and design instructional strategies and tools to promote conceptual change towards more scientifically accurate models.

Method

Purpose

The purpose of this study was to identify and characterize the different types of models held by middle schools students about the inside of the earth and of the causal and dynamic processes involved in plate tectonics, and characterize the types of reasoning associated with these models. Additionally, this research was conducted in order to identify the learning gains and inferences afforded when students construct diagrams for learning and use these diagrams to support model revision.

Subjects

The data upon which this research is based was drawn from two classroom-based studies ($n = 40$) and additional students ($n = 7$) who were interviewed individually. All students were drawn from fifth grade classes in a rural town in Massachusetts; they ranged in age from 10–12 years. The students who were individually interviewed volunteered to participate in the study after an introduction to the research given by the interviewer; they were paid \$5.00 per hour for their participation.

Procedure

For the students who were interviewed on an individual basis; the interviews were approximately 45 minutes to 1 hour in length. Two video cameras and an audio-video mixer were used: one videocamera was embedded in the ceiling which recorded the student’s drawings as they were generated in real time; the second

video camera recorded all interaction between the interviewer and the student. For those who were group-tested, one class period was used (approximately 45 minutes), all data was in the form of paper and pencil, thus, no videotaping was done.

Students were given a short text describing plate tectonics (the full text is given in the appendix); at four intermittent points in the text, the students were given a prompt in the text that they would be requested to draw a diagram of what they had read *after* each of the four respective sections of text. These prompts provided an orienting task for the students while they were reading. Providing orienting tasks is a commonly used strategy in text comprehension research in order to focus learners' goals when reading (Schmalhofer and Glavanov 1986, Schmalhofer 1998). In the present study, informing students before each section of text that they would be subsequently asked to draw diagrams may have lead them to focus on specific features of the text that fostered their diagram-drawing activities and their mental model construction (Gobert 1997, Gobert and Clement 1999). It is important to note that students were not permitted to look back at the text in order to draw their diagrams; thus students' drawings are reflections of the mental models they formed on the basis of reading the text and remembering relevant information in order to draw diagrams. (For more detail on the interaction between text processing and diagramming, see Gobert 1997 or Gobert and Clement 1999).

The four diagram tasks requested during students' reading of the text were ordered as follows:

Thinking back to what you just read,...

- (1) ... draw a picture of the *different layers of the earth*.
- (2) ... draw a picture of the *movement in the different layers of the earth*.
- (3) ... draw a picture of the *movement in the different layers of the earth when mountains are being formed*.
- (4) ... draw a picture of the *movement in the different layers of the earth when volcanoes are erupting*.

Post-text assessment

After the students had finished reading the text and drawing their diagrams, they were asked questions about the domain to which they provided verbal responses; again, all verbalizations (for the students who were individually interviewed) were recorded via the videocameras (as described earlier). For the students who were group tested ($n = 40$) their responses to these items were done on paper. The questions were of several different formats including multiple choice, short answer, and explanation questions; additionally, diagrams were provided to the students for specific questions, and two diagrams to be drawn by the student again were requested. All items were designed to assess either knowledge of spatial/static aspects of the domain or causal/dynamic aspects of the domain. Examples of questions assessing spatial/static knowledge are: 'Where is the thinnest part of the crust?', 'If the continents were all together, would the rest of the earth be water?', as well as spatial features of diagrams depicting volcanic eruption and sea floor spreading. Examples of questions assessing causal/dynamic knowledge are, 'the movement in the crust of the earth is caused by...?', 'Rock from the floor of the Atlantic Ocean tests to be younger than rock from the middle of the North

American Continent because. . . ', and causal and dynamic features of diagrams depicting volcanic eruption and sea floor spreading (drawn during the post-text assessment).

All told, the data in this study on which students' mental models and reasoning (Gyselinck and Tardieu 1994) were examined are: their diagrams, think aloud protocols generated while drawing, verbal and/or written explanations accompanying their diagrams, and answers to the post-text assessment items. In the two case studies presented, the items of the post-text assessment were used to further investigate the types of inferences and reasoning each student was able to make on the basis of his or her models, as well as to test for consistency between each student's models and their respective answers to questions about the domain.

Tutoring to promote model revision

For the students who were individually interviewed ($n = 7$), each student was given tutoring in order to remediate their misconceptions and promote model revision. Since this was done on an individual basis, the type of tutoring depended on the nature of the student's models. In the two case studies presented, the tutoring given to each of these two students is outlined in the Results section.

Coding of students' diagrams

Propositional analysis (Frederiksen 1988) was conducted on the source text, allowing for the identification of all the semantic information given in the textual information source. From this analysis, four coding schemes were developed (one of each of the four diagramming tasks) to evaluate the spatial, causal and dynamic knowledge expressed in students' diagrams, text written on their diagrams, and corresponding think aloud protocols. Since the coding scheme is based upon semantic information as expressed via diagrams or textual/verbal descriptions, the coding schemes can be used to score verbal comments and or textual annotations made to the diagrams. An example of this coding for task 4 (volcanic eruption) is shown in table 1. The instructions for its corresponding diagram task were as follows: 'Thinking back to what you just read, draw a picture of the movement in the different layers of the earth when volcanoes are erupting. Include and label all the information about these layers that you can.'

Results

A. Types of models identified

Based on protocol analyses (Ericsson and Simon 1980) of students' interview data and detailed analyses of their diagrams as well as data from the classroom studies (Gobert 1997, Gobert and Clement 1999), two types of student models of the spatial layout of the inside of the earth were identified at this age level; five types of student models of the causal and dynamic mechanisms inside the earth have been identified at this age level.

Table 1. Coding protocol for diagrams of volcanic eruption.

<i>Spatial/Static Components</i>	<i>Score</i>
crust:	
LOCATION: on surface	1 point
PART: plates	2 points
mantle:	
PART: magma	2 points
LOCATION: below crust	1 point
magma:	
ATTRIBUTE: hot	1 point
ATTRIBUTE: liquid	1 point
core:	
LOCATION: center of earth	1 point
ATTRIBUTE: hot mass	1 point
	10 points
Causal/Dynamic Components	
heated core (label 'hot' is acceptable)	2 points
currents are shown	2 points
heat 'rises' from core to mantle	2 points
heat currents push on plates	2 points
plates move apart	2 points
magma rises from mantle (not from core)	2 points
magma rises above surface	2 points
	14 points

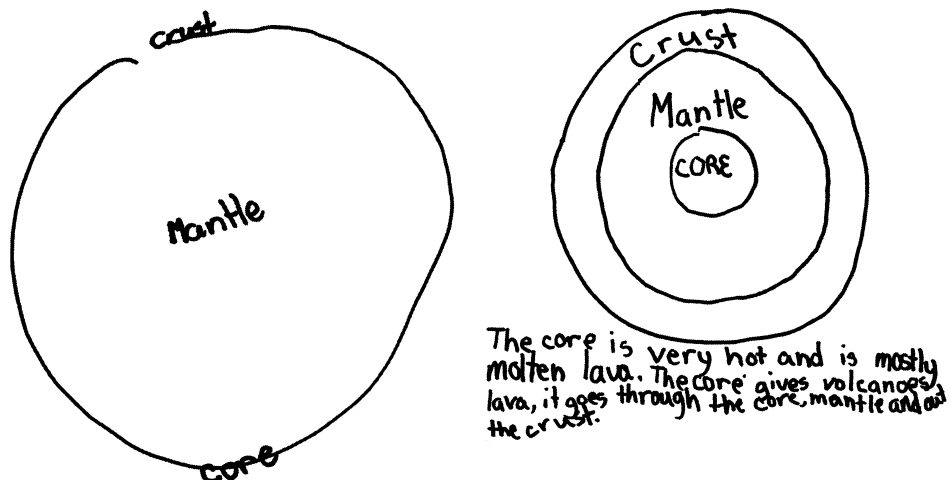
Models of the inside of the earth

There were two types of models identified regarding students' conceptions of the inside of the earth: spatially incorrect models and spatially correct models. These are referred to as Type 0 and 1, respectively. A diagrammatic example of each type of model is shown in table 2a. The percentage of each type of model observed in the data was 10.6% and 89.4%, respectively. A description of each of these types of models is shown in table 2b.

Models of the causal and dynamic processes inside the earth: the case of volcanic eruption

Models of the causal and dynamic processes inside the earth are those which were generated in response to the diagram task for volcanic eruption. The data from this task was chosen to examine students' models of the causal and dynamic processes inside the earth because students' models of volcanic eruption were, in most cases, the most detailed for the four diagramming tasks. This may be due to students having more prior knowledge about volcanic eruption than other plate tectonic-related phenomena.

Although the models are described as categories (1a, 1b, 2, or 3), they can be thought of as a continuum with Type 1a and 1b reflecting models with only heat-related or movement-related causal mechanisms, respectively, to Type 3 models reflecting the most sophisticated model (observed at this age level) which include multiple heat- and movement-related mechanisms. An integrated model of volcanic eruption refers to one in which students have integrated their spatial model



Type 0: Spatially Incorrect Model

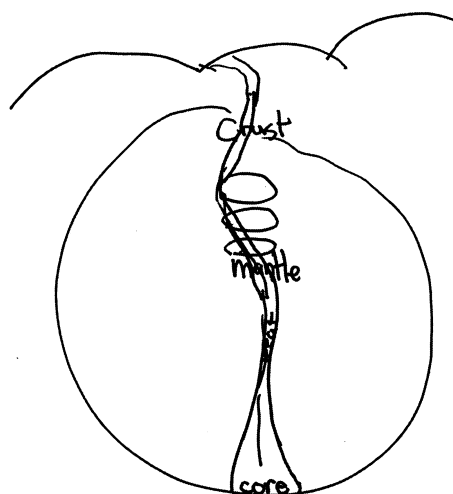
Type 1: Spatially Correct Model

Table 2b. Types of models of the inside of the earth and their characteristics ($n = 47$).

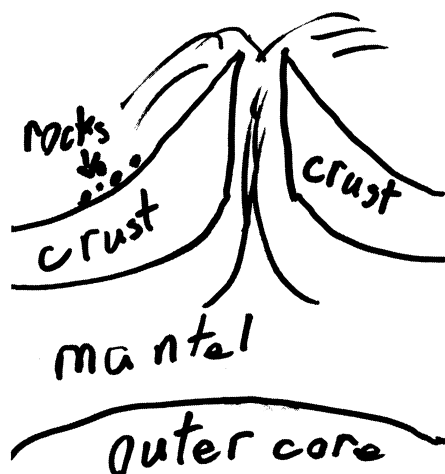
	Type of Model	Characteristics	Frequency	Percentage
TYPE 0	Spatially Incorrect Models	Spatial layout of interior is not correct; few inferences afforded on this type of model	5	10.6%
TYPE 1	Spatially Correct Models	Spatial layout of interior is correct	42	89.4%

of the earth with a number of heat-related mechanisms (i.e. heated core, convection currents, and currents pushing on plates) and movement-related mechanisms (plates moving apart, magma rising from mantle, and magma rising above the surface) into a rich causal model. It is assumed that from these rich causal models, inferencing can be done about the causal and dynamic processes in other plate tectonic phenomena. A canonical model (and coding protocol) of the components involved in volcanic eruption are shown in table 1. A diagrammatic example of each type of model of the causal and dynamic processes inside the earth when volcanoes are erupting is shown in table 3a. A summary table of their characteristics and the frequency with which these were observed in the data are provided in table 3b; a more detailed description of their characteristics can be found in table 4.

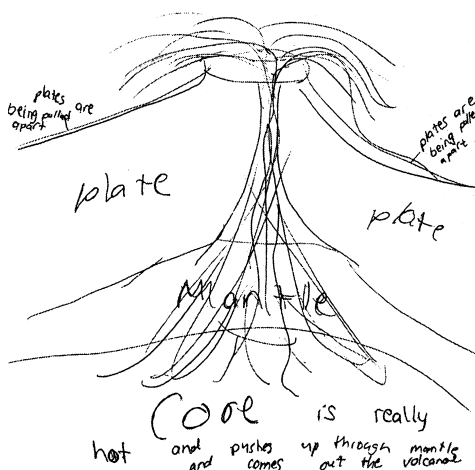
Table 3b summarizes the types of models of the causal and dynamic processes inside the earth during volcanic eruption, their characteristics, and the frequency with which each of these were observed in the data [this table is based on individual interviews with seven students as well as data from two classroom studies, $n = 40$, as described fully in Gobert 1997 and Gobert and Clement 1999]. The percentage of each type of model observed was 4.25%, 61.7%, 29.8%, and 4.25%, respectively. It can be seen that the local models which include movement-related mechanisms only (Type 1b) are the most frequently observed type of model at this age level.



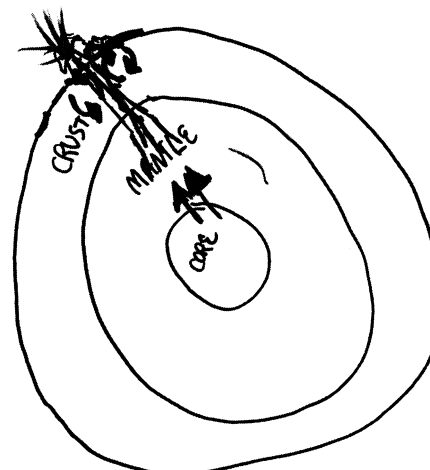
Type 1a: local 'Heat' model



Type 1b: Local 'Movement' model



Type 2: Mixed model



Type 3: Integrated model

B. Students' models: Barriers to deeper understanding or artefacts for model-revision and reasoning

In the next section two cases studies are described; each was selected as an exemplary case of model-based reasoning using diagrams as artefacts for inference and model-revision. From these data it is suggested that models, once constructed, can either serve as a barrier for further understanding, or alternatively, support and facilitate integration and inference-making. Each will be addressed in turn.

Student 1: Models as a barrier to deeper understanding

The first case presented is one in which the student does not have a correct 'concentric circle model' of the inside layers of the earth and she perseveres with this model in all her diagrams.¹ Furthermore, it is argued that until her model of the

Table 3b. Types of models of the causal and dynamic mechanisms in volcanic eruption ($n = 47$).

	<i>Type of Model</i>	<i>Characteristics</i>	<i>Frequency</i>	<i>Percentage</i>
TYPE 1a	Local 'Heat' Models	Heat-related mechanism(s) only; No movement-related mechanisms as causal	2	4.25%
TYPE 1b	Local 'Movement' Models	Movement-related mechanism(s) only; No heat as causal	29	61.7%
TYPE 2	Mixed Models	Few movement- and heat-related mechanisms; Notion(s) of heat and pressure	14	29.8%
TYPE 3	Integrated Models	Movement- and heat-related mechanisms; Includes heat as a causal agent	2	4.25%

Table 4. Characteristics of models of causal and dynamic mechanisms in volcanic eruption.

TYPE 1a: These models are simplistic, 'Local Models' involving heat as the only causal mechanism involved in volcanic eruption. The critical feature of Level 1a models is that they do *not* include any movement-related mechanisms involved in volcanic eruption (mantle or magma movement, plate movement, crust movement/breakage, etc.). These local, heat-only models are infrequently observed at this age group because most children by this age have prior knowledge about volcanoes which includes, at least, magma ('lava') rising above the surface (a movement-related mechanism). Thus, students' models are more likely to fall into the Type 1b classification, namely, Local, movement-only models.

TYPE 1b: These models are simplistic, 'Local Models' involving movement-related mechanisms as the only type of causal mechanisms used to describe and depict the causal and dynamic processes in volcanic eruption. Thus, heat is not seen as causal in these models. The inclusion of 'magma rising' as a movement-related causal mechanism is one which is frequently included in students' models of this type.

TYPE 2: Type 2 models are classified as 'Mixed Models' and are more sophisticated than Type 1a or 1b because they include both classes of causal mechanisms, i.e., both heat- and movement-related mechanisms, but these models are not very elaborate in that they do not include all of either type of causal mechanism. Students who hold these models lack an understanding of how and why convection currents form which then push on the plates, and as such, their models are not well integrated in terms of the heat-related and movement-related mechanisms included.

TYPE 3: Type 3 models are 'Integrated Models' which are well-integrated and include many heat-related and movement-related mechanisms. In these models there is some understanding of heat as a causal agent in causing the currents to form, some understanding of convection in the mantle, and the heat currents as a causal agent in pushing on the mantle and plates. Understanding reflected in models such as these have been achieved by some students at this age level in which students were given tutoring, particularly on the topic of convection (Gobert and Clement, 1994).

spatial arrangement of the inside layers of the earth is remediated to reflect the earth's layers as concentric circles, she will not be able to use her model of the inside of the earth to engage in model-building and inferencing about the causal and dynamic processes involved in plate tectonics.

Student 1's first diagram (figure 1) reflects a Type 0, spatially incorrect model in which she has depicted the core at the bottom rather than in the centre of the earth. In this case, the student has not formed a rich mental model of the internal layers of the earth, which reflects the correct spatial arrangement of the layers. Rather, it appears as though she has interpreted parts of the text literally because she has not depicted the mantle as a concentric circle around the core, rather, she has depicted the mantle 'under' the crust and the core 'under' the mantle. The section of the text which immediately preceded her drawing of figure 1 reads (the relevant sections of text which refer to the spatial layout of the earth's layers are italicized; the full text can be found in Appendix A):

The inside of the earth is made up of three different layers. If you could drill through the earth, *the first layer you would drill through is the crust*, which is 96 miles thick in some places. The continents we see and live on are only part of the crust. In other places, the crust dips down underwater to form the sea bed. The crust is divided into moving sections called plates. Some continents are made up of more than one plate. *Under the crust is the second layer called the mantle*. It is a layer made up of very thick liquid called magma. *At the centre of the earth's interior is the core* which is very hot.

It is hypothesized that spatially incorrect models such as this cannot support inferencing by means of perceptual cues such as spatial adjacency (Larkin and Simon 1987). Thus, her model of the inside of the earth as depicted in figure 1 cannot support knowledge integration and inferencing which is needed in order to understand the causal and dynamic processes inside the earth (figure 2).

As can be seen in her second diagram (figure 2) which requested that she depicts '... the movement and processes of the layers of the earth', she perseveres with her non-concentric model of the layers of the earth: she depicts the core at the bottom, rather than at the centre of the earth; the mantle layer is on top of the core, rather than surrounding it; and the plates are not embedded in the crust as they should be. Additionally, this figure does not include magma as liquid layer inside the mantle. Here again, it is likely that she has interpreted the text literally. In the section of text which immediately preceded her drawing (figure 2) the text reads (relevant sections are italicized):

Remember that the crust is divided into sections called plates. *Each plate can be thought of as a sheet of rock, riding on top of the mantle*. As mentioned before, *the core of the earth is very hot. This heat creates currents that rise up through the mantle. When these currents get near the top of the mantle, they push on the plates, and force the plates to move in many directions. As the mantle moves, the plates move with it*. Since the continents are part of the plates, the continents move too.

In figure 2 (and in figure 1) because her diagram does not depict the mantle layer surrounding the core and the crust surrounding the mantle as concentric circles, this figure cannot support reasoning by means of perceptual cues about how the heated core acts as a heat source for the magma which causes convection currents to form which then push on the plates. Thus, her diagrams at this point serve as a barrier for further understanding about the causal mechanisms responsible for plate tectonic-relate phenomena.

In her third diagram (figure 3) which requested that she draw a diagram to depict ‘... the different layers of the earth when mountains form’, she has replicated a portion of her two previous diagrams. This model, as depicted in figure 3, is a local, movement-only model (described in table 3b) which includes plate movement as the only causal mechanism. Here she has included the crust with the plates embedded in it, and a mountain ‘floating’ above the crust. Again, it appears that she has understood the text literally; for example, her diagram depicts the plates on top of each other. The text immediately preceding this figure reads as follows:

When two plates are forced together, mountains can form. As the plates are forced together, the edges may be arched like a deck of cards being squeezed from both sides. *Eventually one plate moves under the other plate.* As the plates continue to move together, the crust is slowly bent and crumpled, and mountains are formed. While the rock may rise only a quarter of an inch per year, over millions of years it can form very high mountains. The Himalayan Mountains are the best example; they were formed when the plate of India collided with the plate of Southern Asia.

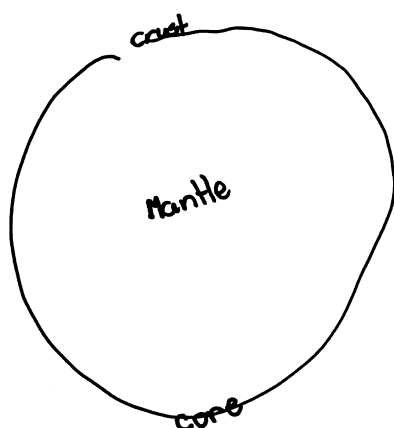
In her final diagram which depicts volcanic eruption (figure 4), she has again depicted the core at the bottom of the earth with the mantle on top of the core rather than surrounding it, and the crust on top of the mantle rather than surrounding it. As in her second diagram (depicting the movement and processes inside the earth), she has not included magma within the mantle, nor did she depict magma rising above the surface of the earth. She has drawn a vertical line, which divides the volcanic mountain in half. Perhaps this is intended to represent cracks in plates as taken literally from the text. The portion of the text, which immediately preceded her diagram, reads:

Volcanoes occur mostly along, or very near, the edges of plates. This is because it is at the edges where most of the stress and cracks occur. One way that volcanoes can form is when the plates move apart. As they move apart, hot liquid magma from the mantle rises up above the surface to form volcanoes.

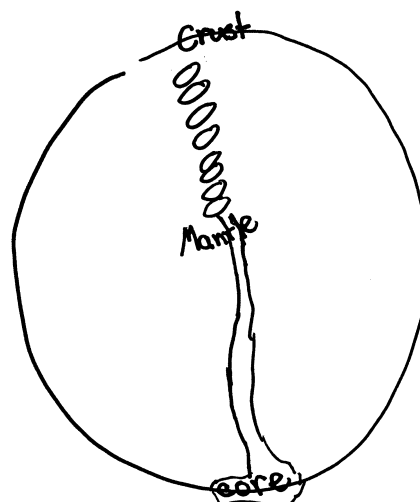
Again, her diagram, as depicted, can not support reasoning about how convection currents form in the mantle, nor how magma currents push on the plates and rise above the surface to cause volcanic eruption. Thus, her understanding of the causal mechanisms underlying volcanic eruption is limited to her model, which includes no causal mechanisms.

Post-text Assessment for student 1

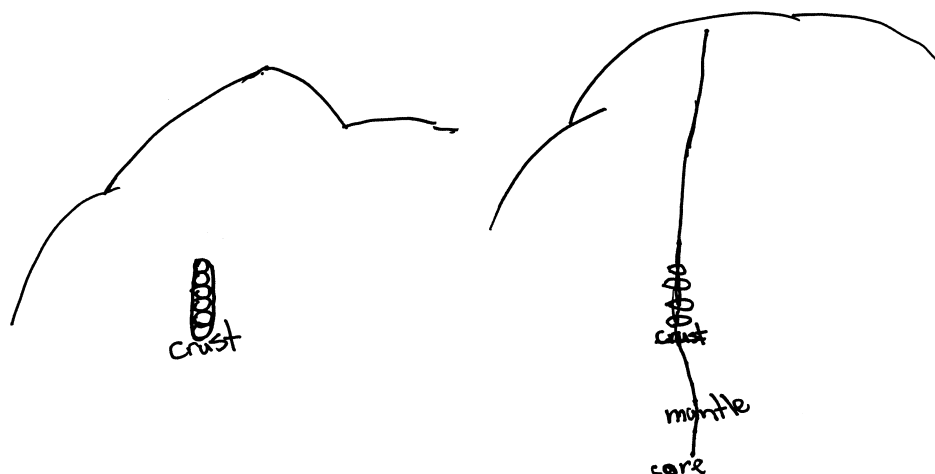
As previously stated in the method section, a post-text assessment was done after students finished reading and generating their diagrams. Specific answers from Student 1’s post-text assessment were selected to investigate the types of inferences the student was able to make on the basis of her models, as well as to test for consistency between her models and answers to questions about the domain. Here the data is discussed with respect to this student’s understanding of the spatial/static features of the inside of the earth, and the causal/dynamic processes of plate tectonics. Each will be addressed in turn.



Student 1: figure 1



Student 1: figure 2



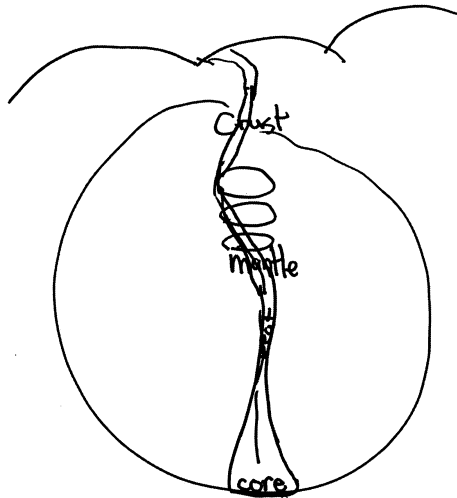
Student 1: figure 3

Student 1: figure 4

Understanding of the spatial arrangement of the inside of the earth

On the post-text assessment, when given a multiple-choice question asking 'Where is the thinnest part of the crust?', she circled (incorrectly) 'land at sea level'. From her answer to this question, it appears that she does not understand that the earth's crust dips under the ocean to form the seabed. From this response, as well as her poor, spatially incorrect diagram of the inside of earth (figure 1), it appears that this student has a poor understanding of the earth, both the spatial arrangement of the interior, as well as the three-dimensional nature of the exterior of the earth's crust.

When asked to draw a diagram depicting what happens to the layers of the earth when volcanoes erupt (this item was asked again in the post-text assessment), she drew figure 10. The spatial arrangement of the layers in this diagram are



Student 1: figure 10



Student 1: figure 11

incorrect: the core, mantle and crust are on top of each other, rather than depicted as concentric circles. This is consistent with her figure 4 drawn during her reading of the text; the only difference between figure 4 and figure 10 is that in the latter, she has included a 'channel' for the heat to move up through the earth. Again, figure 10 reflects her literal understanding of the text, as the mantle 'under' the crust and the core 'under' the mantle.

When asked to draw a diagram depicting what happens to the layers of the earth when the sea floor spreads, she drew figure 11. This figure is not at all informative because it does not include the interior layers of the earth.

Understanding of the causal and dynamic of the processes in plate tectonics

When asked 'Name three occurrences which suggest that the plates in the earth are still moving'. She replied that she did not know. Her response to this item again reifies her poor understanding of how the causal mechanisms inside the earth are responsible for geological phenomena on the earth's surface.

When asked 'movement in the crust of the earth is caused by ...', she circled 'Heat in the core.' This is only partly correct since it does not include any movement-related mechanisms as causal in plate tectonics. Her response to this item is consistent with her figure 2 (during reading of the text) which depicts a local heat-only model of the processes inside the earth. It is also consistent with her diagram depicting volcanic eruption drawn during post-text assessment (figure 10 shown above) which includes heat 'rising' from the core, but no plate movement as a causal mechanism responsible for volcanic eruption.

When asked 'The sea floor gets bigger over time because ...', she circled 'The tide washes up earth from the bottom of the ocean'. This misconception, which was also elicited in other students who participated in this study, indicates a lack of understanding of the causal mechanisms responsible for sea floor spreading. Her response to this question is also compatible with her diagram depicting sea floor

spreading (figure 11) since it includes no causality about how or why the sea floor spreads (in fact her figure 11 does not include the inside of the earth at all). Her lack of understanding about how the sea floor spreads is also reflected in her response to the question, 'Rock from the floor of the Atlantic Ocean tests to be younger than rock from the middle of the North American continent because . . .'; to this she responded that she did not know. The finding that her model of sea floor spreading does not include any correct causal mechanisms (not even heat-related mechanisms as do her other models) may indicate that she does not understand that sea floor spreading is another example of plate tectonic phenomena even though this sub topic was presented along with mountain formation and volcanic eruption in the text.

When asked 'How did India get to fit into Asia?', she replied that she did not know. This is compatible with her poor understanding of mountain formation as depicted in figure 3. It is interesting that she did not use her 'understanding' of plate movement as depicted in figure 3 to answer this question about mountain formation. It is likely that she did not have knowledge that the Himalayas were formed due to continental plate movement. It is also possible that she is very unsure of her conceptions. All told, these data reflect her poor understanding of the causal mechanisms involved in plate tectonics.

Tutoring used to promote model-revision

At this point the interviewer attempted to see if she could help the student revise her spatial model of the earth since she believed that if the student had a correct spatial model of the layers of the earth depicted as concentric circles, she would be better able to use this model to make inferences about the causal and dynamic processes involved in plate tectonics. The first strategy tried here was to allow her to re-read the first section of the text, 'The layers of the earth'. After she re-read the first section of the text, she drew figure 1, time 2 which again depicts the core at the bottom of the earth, the mantle on top of the core, and the crust on top of the mantle; the plates are not embedded in the crust, and there is no magma in the mantle layer. This diagram includes what appears to be a channel coming up from the core and ending on the earth's surface.

Since she did not revise her model substantially based on a second reading of the text, the interviewer then described the inside layers of the earth as an onion cut in half creating concentric circles and drew a cross-section of the earth with the core at the centre, the mantle layer surrounding the core, and the crustal layer surrounding the mantle. The interviewer asked if the student could see how her diagram and the interviewer's diagram differed; she replied, 'Yes'. The interviewer then asked her to draw another diagram to depict movement and processes inside the earth. She drew figure 2, time 2 and added circular lines of magma in the mantle layer. When the interviewer asked her how her original diagram (figure 2) and this one differed, she replied that she had '... forgotten the liquid' in her original diagram. This newer diagram (figure 2, time 2) is a significant advance over her previous drawing of the movement and processes in the earth (figure 2) because it depicts the core in the centre of the earth and includes magma in the mantle layer. The inclusion of both of these features provides a visual model from which she could reason and make inferences. As previously stated, it is argued that this type of model is a necessary condition for reasoning about the causal and

dynamic processes in plate tectonics. In particular, it is argued that a correct concentric model is a necessary condition for reasoning about how convection currents form and how currents of magma cause movement of crustal plates.

Next the interviewer requested that she draw another diagram to depict mountain formation. Her diagram, figure 3, time 2, appears to be a synthetic model (Vosniadou and Brewer 1992) of her original model (figure 3) and her newer model in that it depicts the core close to the bottom of the earth (as in her original model) but has the mantle layer surrounding the core. In her description of this, she said that 'the heat heats up the liquid'. Here she has revised her original mountain formation model to include the notion that the core acts as a heat source on the liquid magma. This is a significant advance over her original model because she is beginning to understand multiple causal mechanisms involved in plate tectonics.

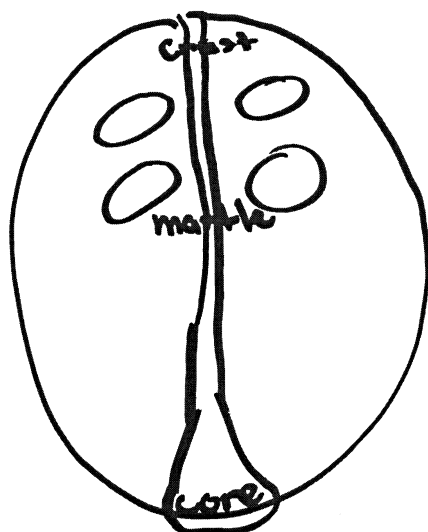
Lastly, the interviewer requested that the student draw another diagram of volcanic eruption. This model (figure 4, time 2) is a significant advance over her original volcano model (figure 4) as well as that generated during the post-text assessment (figure 10) for many reasons. First, it depicts the core at the centre of the earth, the mantle is surrounding the core, and magma is contained in the mantle layer. As such, her new model (figure 4, time 2) includes a spatially correct model of the inside of the earth. Additionally, she drew arrows to indicate plate movement, also an advance over her two previous volcano models (figure 4 and 10) in terms of including plate movement as a causal mechanism involved in volcanic eruption.

At this point, the student's understanding of the domain appears to be better than that which was reflected in her original diagrams (figure 1–4). In terms of the classification scheme in table 3b, this student now has a 'mixed model' of the causal mechanisms responsible for volcanic eruption since she has included both heat-related and movement-related mechanisms in her model (figure 4, time 2).

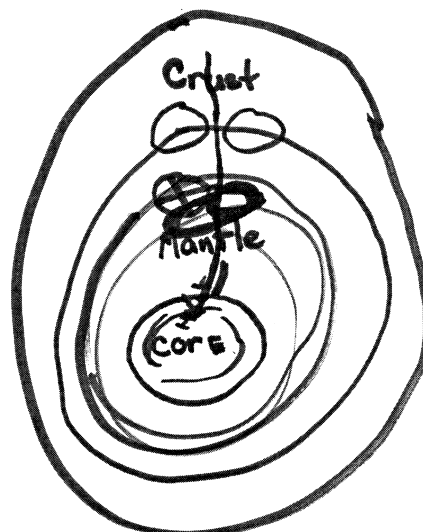
Student 2: Models as tools for model-revision and reasoning

In the second case study, it is shown how spatially correct models can serve as useful tools for reasoning (Kindfield 1993) about the causal and dynamic mechanisms inside the earth and how partially correct models can serve as useful tools for progressive model building.

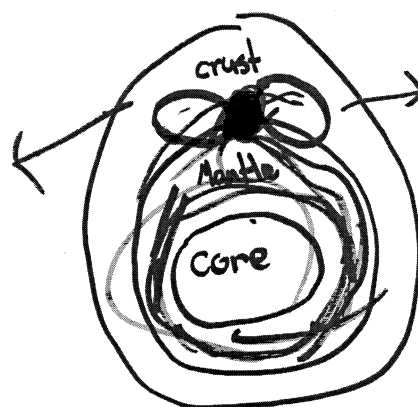
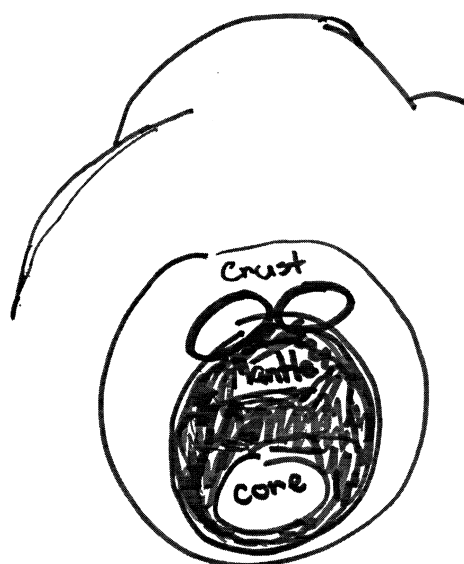
In Student 2's first diagram depicting the layers of the earth (figure 1), a spatially correct model of the interior layers of the earth is depicted. (In actuality, the crustal layer is proportionately much thinner than the student has depicted here but for the purposes of this research and grade level, this is not a grave error in that it does not limit the nature of the inferences afforded). He has annotated his diagram with text (unsolicited by the interviewer) which reads 'The core is very hot and is mostly molten lava, it goes through the core, mantle and out the crust'. This statement does not reflect a complete understanding of volcanic eruption but as the following analysis will demonstrate, he is able to use his spatially correct model to make inferences as he proceeds through the text and diagramming tasks. Since the textual information that he added to his diagram was not contained in the section of the text that he read before he generated his diagram, one can assume that this student has some prior knowledge of volcanoes.



Student 1: figure 1, time 2



Student 1: figure 2, time 2



During his reading of the next section of text 'Movement in the layers of the earth', he made a meta-level comment about how he might depict the plates as moving. He said, 'I was just wondering how you might draw the movement of the plates'. This is anecdotal evidence that the diagramming prompts in the text may have influenced some students to attend to specific features of the text that would be useful in producing their diagrams (see Gobert and Clement 1999 for more discussion on this aspect of the research). After he finished reading this section of text, he drew a diagram to depict the causal and dynamic processes inside the earth (figure 2). While he drew he said, 'The currents form here, push on the mantle so that the mantle moves and makes the crust move, and the plates will move if the mantle and crust moves, and if they overlap, they make an earthquake'.

The diagram and his corresponding protocol (above) indicate that he has a fairly good understanding of the movement-related mechanisms underlying plate tectonics, but he does not appear to understand the heat-related mechanisms involved in plate tectonics. More specifically, he does not appear to understand why or how the currents form (i.e. that the core acts as a heat source in causing convection currents to form). As such, his model as depicted in figure 2 is a mixed model (Type 2) as described in table 3b because although it includes multiple movement-related mechanisms, it only includes the presence of currents as a heat-related causal mechanism. At this point the interviewer asked if he knew what was causing the currents to form. He responded, 'I think its when the core gets too much pressure and if it didn't have earthquakes and volcanoes, it might explode'. This validates the assumption that the student did not understand how heat from the core was causal in heating the magma.

The student's third diagram depicts mountain formation (figure 3) and he has included 'force' in his diagram. His protocol (below) demonstrates that he has notions of pressure and force. It is important to note that although some of his assumptions are not correct (e.g., lava building up in the core and getting 'pressurized'), he appears to be trying to integrate what he is understanding from the text with his prior knowledge of heat and pressure. This is an example of how intuitive conceptions are rich, effective starting points for instruction (Clement *et al.* 1989). Below is an excerpt of our conversation:

Student: The plates on top of the crust, when they are forced together, they form a mountain. ... the mantle that is making the force so that mountains are formed when the two plates, I guess that those lines are force, and the force is coming from the mantle.

Interviewer: Why is there force coming from the mantle?

Student: Because maybe before it came from the core and then it went to the mantle and stayed there and then that got too pressurized. It has to let it go, it probably came from the core. ... I think that most of the movement of the earth comes from the core because the lava builds up that in there and there's too much pressure in it so it has to let out all the pressure and it will go through the core and go through the crust and let it out on the plates.

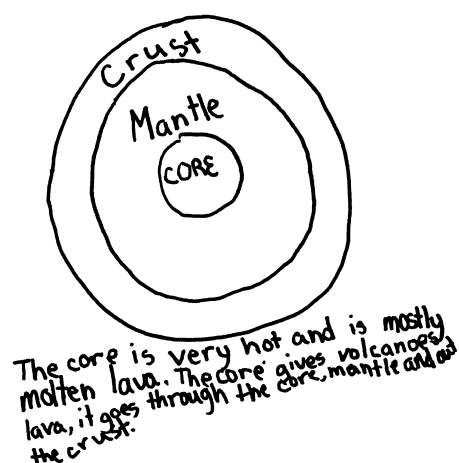
In his fourth diagram, which depicts volcanic eruption (figure 4), the student has drawn a local, movement only model of volcanic eruption. The only causal mechanisms that are included are magma rising from the mantle and magma rising above the surface. At this point, the interviewer attempted to see if the student could integrate some of his intuitions about pressure into his model in order to achieve a more causally sophisticated model of volcanic eruption. He generates the notion of heat 'rising' and develops a more causally sophisticated understanding of volcanic eruption. Important to note is the interesting inference he makes about the magma being hot based on a visual cue of spatial adjacency between the core and the mantle and his prior knowledge of heat 'rising'. Excerpts from our conversation are as follows:

Student: I'm drawing, this is coming from the mantle and then the magma is going up and then its coming through the volcano and its coming out of the volcano.

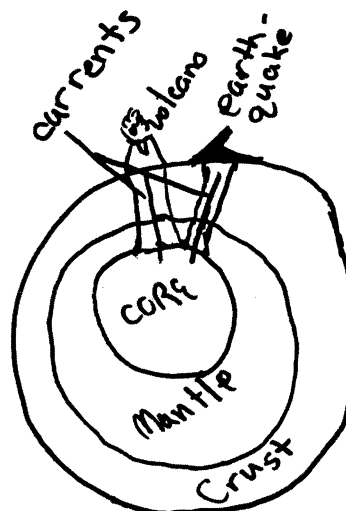
Interviewer: So why is the magma rising?

Student: I'm not sure.

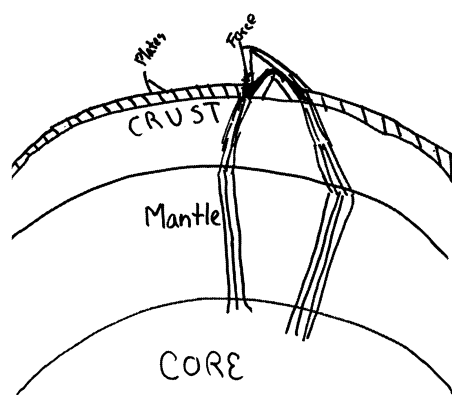
Interviewer: Any ideas?



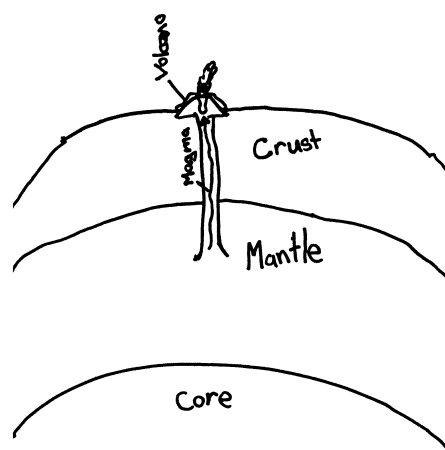
Student 2: figure 1, time 1



Students 2: figure 2, time 1



Student 2: figure 3, time 1



Student 2: figure 4, time 1

Student: Because *heat rises*, the magma is really hot and so it would go up, and basically any way that you face from the mantle is up and that's why it goes up (*good spatial reasoning*).

Interviewer: Let's pursue that idea a little bit – why is the magma hot?

Student: Because the magma is a hot liquid and the core is really hot and the mantle is right near the core... (*excellent inference using his diagram and prior knowledge*)

Post-text Assessment

For the next part of the analysis, specific items of Student 2's post-text assessment were selected to investigate the types of inferences the student was able to make on the basis of his models, as well as to test for consistency between his models and

answers to questions about the domain. Here, the data is discussed with respect to the student's understanding of the spatial/static features of the inside of the earth, and the causal and dynamic processes involved in plate tectonics. Each will be addressed in turn.

Understanding of the spatial arrangement of the inside of the earth

When given a multiple-choice question asking 'Where is the thinnest part of the crust?' he circled (correctly) 'at the ocean floor'. This response indicates that he has a good three-dimensional representation of the earth's crust. Additionally, his spatially correct models of the earth (figures 1–4, and 10), indicate that this student has a good understanding of the inside layers of the earth. His correct spatial understanding is also reflected in his answer to the following question, 'Plates moving apart causes...'; here he circled both b and c: 'causes volcanoes', and 'causes other plates to move together'. His answer to this question also reflects excellent spatial reasoning, i. e. that plates moving will cause others to move. Few students in this study were able to answer this question correctly.

Understanding of the causal and dynamic of the processes in plate tectonics

When asked 'Name three occurrences which suggest that the plates in the earth are still moving', he responded, '... the continents are still moving apart about 4 inches per year; there are earthquakes, and there are volcanoes.' All of these are correct, and as such, reflect a good understanding of the effects of plate tectonics on the earth's surface.

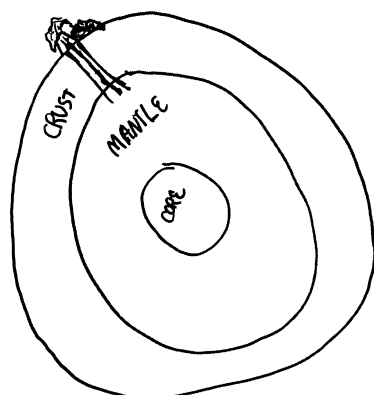
Other indicators of this student's understanding of the causal and dynamic mechanisms involved in plate tectonics are his figures 10 and 11 and their corresponding protocols. In both cases, the diagrams reflect local, movement-only models, compatible with his diagram (figure 4). More specifically, when asked to draw a diagram of volcanic eruption during the post-text assessment (figure 10), he said:

Well, right now I'm drawing the layers. . . . and this is the volcano and this is the currents coming up from the mantle up through the crust and out . . .

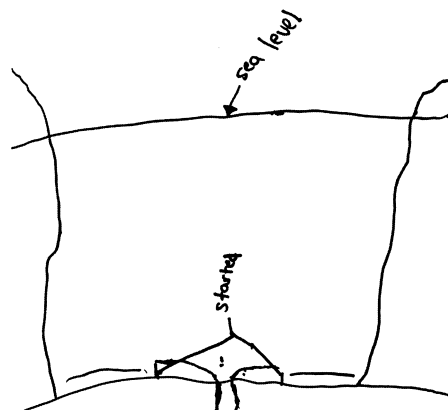
When asked to draw a diagram to depict sea floor spreading, he drew figure 11. While he drew he said:

Right now I'm drawing the sea level, and here's the bottom of the sea, say that this is one plate and this is another, and when the ocean floor gets bigger because the plates are moving. . . I'm going to say that this is where they started and then they move to here because the plates have moved because that's the section of the plates and then the currents fill up the gaps in the plates and then it makes the other plates move.

From this protocol, it is unclear whether this student has a full understanding of what causes the plates in the sea floor to move, i.e. there is no heat as causal in this model, but he appears to understand the relevant movement-related mechanisms, i.e. that magma rises into a gap and fills in the sea floor. There is some consistency between his model of the causal and dynamic processes inside the earth (figure 2,



Student 2: figure 10, time 1



Student 2: figure 11, time 1

above) and his model of sea floor spreading (figure 11) in that both include multiple movement-related mechanisms, which are well integrated into a causal chain.

Model Revision

At this point in the interview, we revisited the diagrams, which this student drew during his reading of the text. In each case, the interviewer asked him if there was anything that he wished to add to his diagrams. For figure 1, he added nothing but said, 'It isn't supposed to be the core. . . the mantle gives the volcanoes magma not lava, and it's not from the core.' The student here corrected the text which he had annotated to his first diagram, namely that the magma (not lava) comes from the mantle layer, not the core.

For his second diagram the interviewer asked if there was anything that he wanted to add to his diagram to show the movement in the different layers of the earth. He replied, 'I'm not really sure how they move besides the currents and how the crust and plates move apart or together.' At this point the interviewer gave him the relevant section of text to read again. Excerpts from our conversation after he had re-read the text are as follows:

Student: The heat rises up through the mantle from the core so that should be there, 'cuz it comes from the core and it goes to the mantle and creates currents from the mantle up to the crust so that would cause an earthquake or a volcano. When the current gets to the top of the plates it pushes on the plates, it forces the plates to move in many directions.

Interviewer: Do you understand that? Does that seem reasonable to you that the currents push on the plates?

Student: Yeah, it's sort of like currents on the water because they push on things in the water. They would push on the plates and they would either go together or apart or in any direction

Interviewer: So . . . can you explain to me clearly knowing what you do about how the plates move?

Student: When the currents get to the top of the plates, it will make the plates move in a lot of directions and that movement from the core, it gives heat to the mantle and the mantle would start off the currents in the mantle and then it would go to the crust, and would come out in the crust as an earthquake or a volcano . . . *It's sort of like a relay, the core to the mantle and the mantle to the crust!*

From this excerpt, it appears that the student has revised his model of movement and processes in the earth to include more heat-related mechanisms, (which was what was lacking in his original, figure 2). More specifically, he has now integrated into his model the notion that it is the heat from the core, which causes currents to form in the mantle, etc. As such, his model reflects a greater number of causal mechanisms. Furthermore, his description of the causal processes into a concise causal chain and his analogy of this process as a relay suggests that he understands the causal mechanisms as a causal system.

The interviewer then asked if there was anything that he wanted to add to his diagram depicting mountain formation (figure 3). He did not significantly revise his diagram. He was confused during this part of the interview about how the currents could be 'strong enough to push on the plates and cause them to move' (this is a common difficulty with this age level). Excerpts from our conversation are as follows:

Student: They're being squeezed from both sides, from the left and the right.

Interviewer: But why are they being pushed?

Student: the force is, but it doesn't say. . . .

Interviewer: Remember when we read about plate movement and what causes it, when the mantle heats up, the magma heats up, and it causes the mantle to move, it's like this layer that's wiggling and jiggling and the plates are on top of it and it causes the plates to move too, right? Do you think that that can cause the plates to move in the way that mountains are formed?

Student: If the mantle is wiggly like Jell-O think that it's not strong enough to push the plates together to make a mountain, it's the mantle that forms the mountain.

Interviewer: How?

Student: By pushing the plates together, but I don't know . . . if it's currents or if it's just the mantle moving around.

Tutoring to promote model revision

At this point the interviewer drew the student a visual analogical model of a boiling pot of water and macaroni on a stove. Together we compared the core to the element on the stove, the mantle to the pot, the magma to the water, and the plates to the macaroni. The interviewer described, in brief, how the water was hotter where it was closer to the element and less hot near the top of the surface of the water, and that the difference in the two temperatures and densities caused the hotter water to 'rise' and the less hot water to 'sink'. Further, the interviewer explained that the rising and sinking pattern was called convection and that this also happens within the mantle layer (Gobert and Clement 1994 presents additional data to demonstrate the efficacy of this visual analogy).

After this tutorial, we returned to his post-text diagram of volcanic eruption (figure 10), the interviewer asked if there was anything that he would like to add to it. He revised his diagram to include currents rising up through the mantle and

pushing on the plates of the crust. In his revised diagram (figure 10, time 2), the currents, as depicted, are not contained within the mantle layer, rather they are directly below the crustal plates. This could be due to the fact that he did not fully understand the relative thickness of the layers (excerpts from his transcript confirmed this) and if the interviewer had helped him to revise his model regarding the relative thickness of the layers, that he would not have depicted the currents in the crustal layer. Excerpts from our conversation are as follows:

Student: Yeah the heat would cause, from the core, since it was so hot, would push against the mantle, and then the mantle would have the magma go up and it would come up in a volcano.

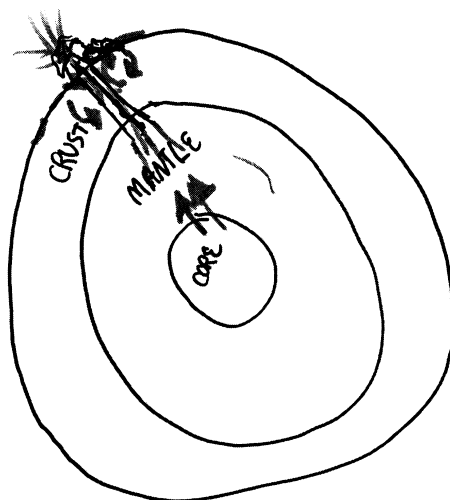
Interviewer: So what's happening with the plates?

Student: The plates are moving apart.

Interviewer: Why are they moving apart?

Interviewer: Because the magma is hitting right here (he points to section on the top of the mantle), it comes back down and keeps on hitting it, so that forms a gap and then if there is a gap, then the magma comes up and goes through the volcano.

From the student's transcript, it appears that he has a better understanding of what causes volcanic eruption and that he has some understanding that there is convection in the mantle, and that it is caused by heat from the core. As such, his revised model (figure 10, time 2) is categorized as an Integrated Model (as described in table 3b) because it includes both heat- and movement-related mechanisms, and includes heat as a causal agent in forming the convection currents. As previously said, the goal of this program of research is to promote the development of simplified, qualitative models of plate tectonics. This student's model, especially when taking into account that he is in fifth grade, is a good, qualitative model of the causal and dynamic mechanisms involved in plate tectonics.



Student 2: figure 10, time 2

Discussion

Focus on diagram-drawing rather than diagram presentation

The focus of the present research, diagram drawing, is in direct contrast to many studies, which simply present diagrams to students. In the latter type of studies, diagrams are given to students as adjunct sources of information to text, which is considered to carry the principal informational burden. The assumption is that the presence of the diagram should facilitate learning. More recently however, studies have demonstrated that there are problems associated with this. First, students often don't know how to search through diagrams in a systematic fashion in order to understand complex information (Gobert 1994, Lowe 1989), nor do they know what information is important (Anzai 1991). Additionally, scientific diagrams usually have domain-specific symbol systems which students are not skilled at interpreting (Gobert 1994, Hill 1988). Lastly, simply presenting diagrams to students puts them in a passive role as learners (Gobert and Clement 1999). From an applied perspective, this is deeply problematic because it does not promote the development of rich mental models from which inferences can be made (Lowe 1993), thereby defeating one of the main goals of science education. From a theoretical perspective, simply presenting diagrams to students is antithetical with both constructivist theories of learning (cf., von Glasersfeld 1987, 1995) and with current emphasis in science education on students' *active* model building (White 1993, Frederiksen *et al.* in press, Raghavan and Glaser 1995, Penner *et al.* 1997, Linn and Muilenberg 1996, Keys 1997). For these reasons, the present research employed active diagram construction as a means to promote model building and model revision.

Student-generated diagrams: Contributions to basic research and science education

These data on diagram generation are important in terms of what they can tell us about the cognitive processes used in constructing and making inferences from diagrams, and as such are important to basic research on cognition. Findings from these data are compatible with previous research in this area, namely, that drawing diagrams provides a means to externalize knowledge, thereby freeing up cognitive resources for inference-making (Kindfield 1993). In terms of contributions that these data make to science education, it is argued that detailed, systematic analyses of students' diagram drawing and their corresponding think aloud protocols provide fine-grained methodological tools which can be used to: evaluate students' pre-instructional models (Glynn 1997), trace conceptual change over time, particularly if multiple drawings are generated (as in the present study), and test for the robustness of students' models.

Types of models observed

Two types of models of the inside of the earth were identified: spatially incorrect models, such as in Student 1's case, in which the spatial arrangement of the layers is not correct; and spatially correct models, such as Student 2's case, in which the inside of the earth is depicted as concentric circles. The proportions of each type of

model observed in the data were 10.6% and 89.4%, respectively. At this age level, it is not surprising that the large proportion of students who participated in this research held spatially correct models of the inside of the earth. This is likely due to prior knowledge that students obtained from diagrams in books, the media, etc. It is possible that a greater proportion of younger students hold spatially incorrect models; however, this was not the focus of the present study. In terms of the spatially incorrect models observed, it is important to note that Student 1's model differed from the other models which were also classified as spatially incorrect; that is, of the 5 spatially incorrect models observed, each differed in terms of the spatial arrangement of the layers. Further research is needed in order to more fully understand the origin of these various types of spatially incorrect models.

Four types of models of the causal and dynamic processes involved in volcanic eruption were identified, ranging from local models including heat-related mechanisms only, local models including movement-related mechanisms only, mixed models including some heat- and movement-related mechanisms, and integrated models (the most sophisticated type of model observed at this age level) which include some understanding that heat acts as a causal agent in causing the formation of convection currents which then push on the plates and cause crustal movement/breakage. The proportions of each type of model in this category were 4.25%, 61.7%, 29.8%, and 4.25%, respectively. In terms of the relative proportions of these types of models, again, local heat-only models were observed very infrequently and these models tended to include the notion that heat only was responsible for plate tectonic phenomena but with no knowledge of *how* heat was causal. As such, these models are very simplistic but are effective points for instruction, as evidenced by the substantial gains made by Student 1. The highest proportion of models observed were Type 1b, local movement-only models, which included, in most cases, the idea that in volcanic eruption, magma rises above the surface. It is likely that the high proportion of models observed in this category is due to prior knowledge that students have from the media. This type of model is also very simplistic but partially correct and as data from Student 2's model building indicates, students' rich prior knowledge about magma rising presents an effective starting point for instruction. Probing questions such as 'Why does the magma rise?' may facilitate progressive model building by engaging the student to seek to integrate his or her prior knowledge. The third category, Mixed models, which include notions of heat and pressure as well as magma movement as causal mechanisms in volcanic eruption, were the second largest category observed in these data. These models are interesting in that they have many correct causal mechanisms included in them. Given the proportion of these in the data (29.8%), it is reasonable to assume that many students at this grade level could achieve at least this level of understanding if engaged in instructional tasks which promoted model-building such as the tasks in the present study. Research in which students generate their models and then pose questions to each other about their models in the planning stages (Gobert 1999), and it is hypothesized that if students engage in problem-posing in which they address questions such as 'Why is the magma rising?' and 'What causes the magma to be hot?', etc., that they will be able to progressively refine their models to include both heat-related and movement-related causal mechanisms. Previous research has shown that diagrams can serve as cognitive artefacts and that these can mediate conversations between individuals

(Pea *et al.* 1993). In terms of the students who were classified as having integrated models, it is important to note that they were classified so on the basis of individual interviews in which they were given a tutorial of a visual analogical model of a pot of boiling water on a stove. No integrated models were observed on the basis of the reading and diagramming tasks alone; however, the visual analogy which was used in these tutorials helped students to understand how convection currents were formed in the mantle and how these currents pushed on the plates. These data suggest that visual analogical models, such as the one used in the tutorials in the present study, might be used successfully in classroom instruction in order to promote deep learning.

Reasoning with spatially incorrect models: the case of Student 1

On the basis of Student 1's diagrams generated during her reading of the text (figures 1–4) as well as her responses to questions during the post-text assessment, student 1 had a flawed understanding of the spatial layout of the layers inside the earth, that is, a spatially incorrect model of the inside of the earth (a Type 0 model as described in table 2b). She also held a very rudimentary model, i.e. a local, heat-only model of the causal and dynamic mechanisms involved in plate tectonics (a Type 1b model as described in table 3b). These models appeared to reflect a very literal interpretation of the text.

It is hypothesized that her flawed model of the arrangement of the layers of the earth (figure 1 – core depicted at the bottom of the earth) could not support further model building about the causal and dynamic processes inside the earth because all of visual cues that are needed to understand the earth as a causal system were missing in her models. More specifically, her original, non-concentric earth model (figure 1) could not support inference-making by means of perceptual cues about the causal mechanisms and processes inside the earth (figure 2), and how these mechanisms cause mountain formation (figure 3) and volcanic eruption (figure 4). The lack of viability of her original models was further evidenced by her responses to post-text assessment items.

Regarding Student 1's model revision, it is important to note that once she was able to correctly depict the earth with the core at the centre and revised her diagrams to include magma in the mantle layer (figure 2, time 2), she was better able to use this model to reason about the causal mechanisms underlying mountain formation (figure 3, time 2) and volcanic eruption (figure 4, time 2). From this, it appears that having a correct model of the spatial layout of the earth is a necessary (but not sufficient) condition for understanding the causal and dynamic processes involved in plate tectonics because with a model which depicts the layers of the earth as concentric circles with the magma-filled layer surrounding the hot core it is easier to visualize and understand how the core acts as a heat source for the magma. Once students have understood the causal chain or attributive cluster (Brown 1993, 1995) that the core is a heat source for the mantle, it is then easier to address instructionally how convection currents form in the mantle, and how currents push on the plates, both of which are necessary to understand how plate tectonics accounts for crustal activity such as mountain formation, volcanic eruption, and sea floor spreading.

After her model revision, student 1's model of the causal and dynamic processes inside the earth can be described as a 'Mixed' model because it includes both heat- and movement-related mechanisms (figure 4, time 2). Thus, in terms of conceptual gains made, this student has made considerable progress in understanding the spatial arrangement of the layers of the earth in that she now understands that the layers are not depicted as piled on top of one another (as in her first diagram). Conceptual gains were also made in her understanding of the causal mechanisms involved in plate tectonics from her original model (figure 2) which included heat as the only causal mechanism to her final model of volcanic eruption (figure 4, time 2) which included both heat-related and movement-related causal mechanisms. It is likely that her understanding of heat as causal provided an important source of knowledge that was used in further developing her understanding of the causal mechanisms inside the earth. As such, the notion that there is heat in the core of the earth, which is relevant prior knowledge which many students have, is a usable anchor for conceptual change (Clement *et al.* 1989). It is also important to note that without feedback from the student's drawings, a teacher might not have detected what was blocking the student's understanding of the domain. Thus, this is an example of how students' diagrams can be used to diagnose misconceptions, and an example of how misconceptions, if they are not remediated, can adversely effect understanding.

Reasoning with spatially correct models: the case of Student 2

Student 2 was chosen as a case study to be described for two reasons. First, his pre-instruction model of the inside of the earth, i.e. a spatially correct model, is representative of the majority of students who participated in this research. Similarly, many of his pre-instruction models of the causal and dynamic processes inside the earth are local, movement-only models (figure 2, 4, 10, and 11), also representative of the majority of students who participated in this research.² Secondly, this student was chosen to describe in depth because of the ways in which he used his relevant prior knowledge of heat, pressure, and force as well as visual cues from his diagrams to revise his models.

After a brief tutorial in which the interviewer drew a visual analogy of a boiling pot of water on the stove, the student was able to better integrate his prior knowledge with his models in order to substantially revise his models, and thus, his understanding of the domain. The student constructed his understanding based on his prior knowledge of heat, pressure, and force, what he had learned from the text, inferences made on the basis of his own diagrams, and the visual analogy. In sum, he progressively revised his local, movement-only models to his final, Integrated model, which includes multiple heat- and movement-related mechanisms as well as the notion of the core acting as a causal agent in forming the convection currents. It is hypothesized that having a spatially correct model of the interior layers of the earth, as in the case of this student, facilitates inference-making by means of perceptual cues such as spatial adjacency (Larkin and Simon 1987) which then promotes model revision.

Models as coherent frameworks versus knowledge in pieces

Using a combination of qualitative and quantitative measures, these data demonstrate that there are a small number of well-defined models, which are held by students at this age level regarding the causal mechanisms involved in plate tectonics. Furthermore, there was a large degree of correspondence between students' diagrams, their corresponding explanations, and their answers to inference questions. Similar findings were reported by Vosniadou (Vosniadou and Brewer 1992). In Vosniadou's research, she has interpreted the commonality of the models observed and the correspondence between students' models and answers to inference questions as evidence that conceptual knowledge in the domain of astronomy is theory-like rather than fragmented and unconnected, as others have suggested (di Sessa 1985, 1993). Although data from the present study can be used to argue for a theory-like view of students' knowledge, a cautionary note is included. The present data suggest that students are using their models consistently in order to reason about plate tectonic-related phenomena, however, it is possible that they did so because they understood that the various phenomena addressed were all examples of plate tectonic-related phenomena. In fact, the text, which was used in the study, was written explicitly to promote this. It is possible, that if the text had not made these explicit connections, that the students may have reasoned in a manner more similar to a 'knowledge in pieces' fashion, that is, not attributing the same causal mechanisms to the various plate tectonic-related phenomena addressed. This study was not designed to address the knowledge in pieces versus knowledge in theories debate; however, empirical studies could be conducted in order to address the models as theories versus knowledge in pieces debate.

A comment on prior knowledge and epistemology

The prior knowledge and inference strategies which particular students bring to bear on the tasks in the present study can be thought of as part of a student's conceptual ecology (Toulmin 1976). Another aspect of a student's conceptual ecology which is likely important to their model construction and model revision is their epistemology of scientific models, i.e. their understanding of the nature and purpose of scientific models. Although students' epistemologies were not addressed in the present work, it has been argued elsewhere that this is an important component to be addressed in students' model construction and model revision (Grosslight *et al.* 1991, Gobert and Discenna 1997, Schwarz and White 1999). Moreover, it has been shown that students with a more sophisticated understanding of the nature and purpose of models are better able to make inferences from their models, once constructed (Gobert and Discenna 1997); findings consistent with studies of the effects of epistemology on the integration of textual material (Rukavina 1991, Rukavina and Daneman 1996) and studies on the effects of epistemology on the integration of scientific principles (Songer and Linn 1993). Further research is necessary and is planned in order to address the relationship between the nature of students' understanding of scientific models and how this influences their model building (Gobert 1999). It is hoped that this will provide further insight into both the nature of students' science learning as well as how this might be addressed pedagogically in science teaching.

Conclusions

In the existing literature, questions have been raised as to whether students can be taught to produce diagrams from which rich inferences can be drawn (Anzai 1991), and whether students can make inferences from their own diagrams, once constructed (Schwartz 1993). Results from these studies suggest that young students can construct rich mental models of complex causal and dynamic systems, which they can then use to make inferences. These data also suggest that developing rich integrated causal models may be facilitated when models are constructed by the learner beginning with the static components first followed by increasingly complex models involving causal and dynamic information.

This research utilizes student-constructed modelling tasks and progressive model revision in order to promote deep learning of subject-matter material, as well as promote the development of modelling skills required in science learning and in scientific reasoning in general. In doing so, it examines both the process and product of science learning, and focuses on model construction as a very important process skill and an integral part of science learning and science literacy (Linn and Muilenberg 1996, Penner *et al.* 1997). This approach to science education emphasizes diagrams as important tools for model-construction and reasoning as opposed to the more conventional use of diagrams as merely illustrations of science concepts.

Acknowledgements

This research was supported in part by a Spencer Post-Doctoral Fellowship from the National Academy of Education and in part by the National Science Foundation (#9806141). The author wishes to thank Dr. John Clement for many fruitful discussions regarding this research. The author wishes to thank Albert Wang of Harvard University's Technology Services Center for technical assistance with the diagrams.

Notes

- i. This research was supported in part by a Spencer Post-Doctoral Fellowship from the National Academy of Education and in part from a grant from the National Science Foundation (#9806141), both awarded to Janice Gobert.
1. Due to a technical problem with the recording equipment, this student's think aloud protocol was only partially audible; thus, no think aloud data for this student is included.
2. See tables 2b and 3b for percentages of students in each category

References

- AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE (1989) *Science for All Americans: Project 2061* (Washington, DC: American Association for the Advancement of Science).
- AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE (1993) *Benchmarks for Science Literacy* (New York: Oxford: Oxford University Press).

- ANZAI, Y. (1991) Learning and use of representations for physics expertise. In K. Anders Ericsson and J. Smith (eds) *Toward a General Theory of Expertise: Prospects and Limits* (New York: Cambridge University Press).
- ARNOLD, P., SARGE, A. and WORRALL, L. (1995) Children's knowledge of the earth's shape and its gravitational field. *International Journal of Science Education*, 17(5), 635–641.
- AULT, JR., C. R. (1984) The everyday perspective and exceedingly unobvious meaning. *Journal of Geological Education*, 32, 89–91.
- BENCLOSKI, J. W. and HEYL, R. J. (1985) Teaching plate tectonics with the aid of a model of sea-floor dynamics. *Journal of Geological Education*, 33, 274–276.
- BEZZI, A. (1989) Geology and Society: A survey on pupils' ideas as an instance of a broader prospect for educational research in earth science. Paper presented at the 28th International Geological Congress, Washington, DC.
- BISARD, W. J., ARON, R. H., FRANCEK, M. and NELSON, B. D. (1994) Assessing selected physical science and earth science misconceptions of middle school through university pre-service teachers. *Journal of College Science Teaching*, September/October, 38–42.
- BROWN, D. E. (1993) Refocusing core intuitions: a concretizing role for analogy in conceptual change. *Journal of Research in Science Teaching*, 30(10), 1273–1290.
- BROWN, D. E. (1995) Theories in pieces? The nature of students' conceptions and current issues in science education. Presented at the Annual Meeting of the National Association for Research in Science Teaching, April 22–25 (San Francisco, CA).
- CLEMENT, J. (1989) Learning via model construction and criticism: Protocol evidence on sources of creativity in science. In J. A. Glover, R. R. Ronning, and C. R. Reynolds (eds), *Handbook of Creativity: Assessment, Theory and Research* (New York: Plenum Press), 341–381.
- CLEMENT, J. (1993) Model construction and criticism cycles in expert reasoning. In the Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society (Hillsdale, NJ: Erlbaum).
- CLEMENT, J., BROWN, B. and ZIETSMAN, A. (1989) Not all preconceptions are misconceptions: finding 'anchoring conceptions' for grounding instruction on students' intuitions. *International Journal of Science Education*, 11, 554–565.
- DI SESSA, A. (1985) Knowledge in pieces. In G. Forman and P. Pufall (eds), *Constructivism in the Computer Age* (Hillsdale, NJ: Lawrence Erlbaum).
- DI SESSA, A. (1993) *Toward an epistemology of physics*. *Cognition and Instruction*, 10 (2 and 3), 105–226.
- ERICSSON, K. A. and SIMON, H. (1980) Verbal reports as data. *Psychological Review*, 87, 215–251.
- FEATHER, R., SNYDER, S. and HESSER, D. (1995) *Earth Science*. (Westerville, Ohio: Glencoe/McGraw-Hill).
- FREDERIKSEN, C. (1988) Text comprehension in functional task domains. In D. Bloom (ed.), *Learning to Use Literacy in Educational Settings* (Norwood, NJ: Ablex).
- FREDERIKSEN, J., WHITE, B. Y. and GUTWILL, J. (in press) Dynamic mental models in learning science: the importance of constructing derivational linkages among models. *Journal of Research in Science Teaching*.
- GILBERT, J. (1993) The role of models and modelling in science education. Presented at the Annual Conference of the National Association for Research in Science Teaching, April, 1993 (Atlanta, GA).
- GILBERT, S. (1991) Model building and a definition of science. *Journal of Research in Science Teaching*, 28(1), 73–79.
- GOBERT, J. (1994) Expertise in the comprehension of architectural plans: Contribution of representation and domain knowledge. Unpublished doctoral dissertation (Toronto, Ontario: University of Toronto).
- GOBERT, J. (1997) Summarizing, Explaining, and Diagramming: The differential effects on text-base representations and mental models. Presented at the Annual Meeting of the Cognitive Science Society, August, 7–10 (Stanford University, Palo Alto, CA).
- GOBERT, J. (1999) Making thinking visible: Promoting students' model-building and collaborative discourse in WISE. Proposal submitted to the National Science Foundation.

- GOBERT, J. and CLEMENT, J. (1994) Promoting causal model construction in science through student-generated diagrams. Presented at the Annual Meeting of the American Educational Research Association, April 4–8 (New Orleans, LA).
- GOBERT, J. and CLEMENT, J. (1999) The effects of student-generated diagrams on conceptual understanding of causal and dynamic knowledge in science, *Journal of Research in Science Teaching*, 36(1), 39–53.
- GOBERT, J. and DISCENNA, J. (1997) The relationship between students' epistemologies and model-based reasoning. Kalamazoo, MI: Western Michigan University, Department of Science Studies. (ERIC Document Reproduction Service No. ED409164).
- GOBERT, J. and FREDERIKSEN, C. (1988) The comprehension of architectural plans by expert and sub-expert architects. Proceedings of the Tenth Annual Meeting of the Cognitive Science Society (Montreal, Canada. Hillsdale, NJ: Lawrence Erlbaum).
- GROSSLIGHT, L., UNGER, C., JAY, E. and SMITH, C. (1991) Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28(9), 799–822.
- GLYNN, S. (1997) Drawing mental models. *The Science Teacher*, 30–32.
- GYSELINCK, V. and TARDIEU, H. (1994) Illustrations, mental models, and comprehension of instructional text. In W. Schnotz and R. Kulhavy (eds.), *Comprehension of Graphics*. Amsterdam (The Netherlands: North-Holland Publishers).
- HEGARTY, M. and JUST, M. (1993) Constructing mental models of machines from text and diagrams. In H. Mandl and J. R. Levin (eds), *Knowledge Acquisition from Text and Pictures* (North Holland: Elsevier Science Publishers B. V.).
- HILL, D. (1988) Difficulties with diagrams, *Journal of Science and Mathematics Education in S. Asia*, XI(2), 32–40.
- JACOBI, D., BERGERON, A. and MALVESY, T. (1996) The popularization of plate tectonics: presenting the concepts of dynamics and time. *Public Understanding in Science*, 5, 75–100.
- JOHNSON-LAIRD, P. N. (1983) *Mental Models: Toward a Cognitive Science of Language, Inference, and Consciousness* (Cambridge, MA: Harvard University Press).
- KEYS, C. W. (1997) An investigation of the relationship between scientific reasoning, conceptual knowledge and model formulation in a naturalistic setting. *International Journal of Science Education*, 19(8), 957–970.
- KINDFIELD, A. C. H. (1993) Biology diagrams: tools to think with. *Journal of the Learning Sciences*, 3(1), 1–36.
- KINTSCH, W. (1998) *Comprehension: A Paradigm for Comprehension* (Cambridge, UK: Cambridge University Press).
- LARKIN, J. and SIMON, H. (1987) Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–100.
- LINN, M. C. and MUILENBERG, L. (1996) Creating Lifelong Science Learners: What Models Form a Firm Foundation? *Educational Researcher*, 25(5), 18–24.
- LOWE, R. (1989) October, Scientific diagrams: How well can students read them? What research says to the science and mathematics teacher, 3. Available from Key Centre for School Science and Mathematics, Curtin University of Technology, GPO Box U1987, Perth, Western Australia 6001).
- LOWE, R. (1993) Constructing a mental representation from a abstract technical diagram. *Learning and Instruction*, 3, 157–179.
- MASSACHUSETTS DEPARTMENT OF EDUCATION (1996) Science and Technology Curriculum Framework: Owning the questions through science and technology.
- MEGOSE (1991) *Michigan Essential Goals and Objectives for Science Education, K-12* (Michigan State Board of Education, Michigan Department of Education).
- MUTHUKRISHNA, N., CARNINE, D., GROSSEN, B. and MILLER, S. (1993) Children's alternative frameworks: Should they be directly addressed in science instruction?. *Journal of Research in Science Teaching*, 30(3), 233–248.
- NATIONAL COMMITTEE ON SCIENCE EDUCATION STANDARDS AND ASSESSMENT (1996) *National Science Education Standards: 1996* (Washington, DC: National Academy Press).
- NUSSBAUM, J. (1979) Children's conceptions of the earth as a cosmic body: a cross age study. *Science Education*, 63, 83–93.

- NUSSBAUM, J. and NOVAK, J. D. (1976) An assessment of children's concepts of the earth utilising structured interviews. *Science Education*, 60, 535–550.
- PEA, R. D., ŠIPIŠIĆ, M. and ALLEN, S. (1993) Seeing the light on optics: classroom-based research and development of a learning environment for conceptual change. Draft manuscript.
- PENNER, D. E., GILES, N. D., LEHRER, R. and SCHAUBLE, L. (1997) Building functional models: designing an elbow. *Journal of Research in Science Teaching*, 34(2), 125–143.
- PINET, P. R. (2005) Rediscovering geological principles by collaborative learning, *Journal of Geological Education*, 43, 371–376.
- PLUMMER, C. and McGEARY, D. (1996) *Physical Geology* (Wm. C. Brown Publishers).
- RAGHAVAN, K. and GLASER, R. (1995), Model-based analysis and reasoning in science: The MARS curriculum. *Science Education*, 79(1), 37–61.
- ROSS, K. and SHUELL, T. (1993) Children's beliefs about earthquakes. *Science Education*, 77, 191–205.
- RUKAVINA, I. (1991) The effect of knowledge-building approach on advancing learning in science. Unpublished Doctoral Dissertation (Toronto, Ontario: University of Toronto).
- RUKAVINA, I. and DANEMAN, M. (1996) Integration and its effects on acquiring knowledge from competing scientific theories from text. *Journal of Educational Psychology*, 85, 406–411.
- SCHAUBLE, L., GLASER, M., RAGHAVAN, K. and REINER, M. (1991) Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences*, 1(2), 201–238.
- SCHMALHOFER, F. (1998) *Constructive Knowledge Acquisition: A Computational Model and Experimental Evaluation* (Hillsdale, NJ: Erlbaum).
- SCHMALHOFER, F. and GLAVANOV, D. (1986) Three components of understanding a programmer's manual: verbatim, propositional and situational representations. *Journal of Memory and Language*, 25, 279–294.
- SCHOON, K. J. (1992) Students' alternative conceptions of earth and space. *Journal of Geological Education*, 40, 209–214.
- SCHWARTZ, D. (1993) The construction and analogical transfer of symbolic visualizations. *Journal of Research in Science Teaching*, 30(10), 1309–1325.
- SCHWARZ, C. and WHITE, B. (1999) What do seventh grade students understand about scientific modeling from a model-oriented physics curriculum? Presented at the National Association for Research in Science Teaching, March 28–31, Boston, MA.
- SNEIDER, C. and PULOS, S. (1983) Children's cosmologies: understanding the earth's shape and gravity. *Science Education*, 67, 205–221.
- SONGER, N. and LINN, M. C. (1991) How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching*, 28(9), 761–784.
- STEWART, J. and HAFNER, R. (1991) Extending the conception of problem in problem-solving research. *Science Education*, 75(1), 105–120.
- STOFFLETT, R. (1994) Conceptual change in elementary school teacher candidate knowledge of rock-cycle processes. *Journal of Geological Education*, 42, 494–500.
- TALLON, R. and AUDET, R. (1999) Three-dimensional representation of student mental models: Exploring conceptual development in the Earthview classroom. Presented at the National Association for Research in Science Teaching, Boston, MA. March 28–31.
- TOULMIN, S. (1976) *Knowing and Acting: An Invitation to Philosophy* (New York: Macmillan).
- TURNER, R. H., NIGG, J. M. and DAZ, D. H. (1986) *Waiting for Disaster: Earthquake Watch in California* (Berkeley, CA: University of California).
- VAN DIJK, T. and KINTSCH, W. (1983) *Strategies of Discourse Comprehension* (New York: Academic Press).
- VON GLASERSFELD, E. (1987) *The Construction of Knowledge: Contributions of Conceptual Semantics* (Seaside, CA: Intersystems Publications, Inc.).
- VON GLASERSFELD, E. (1995) *Radical Constructivism: A Way of Knowing and Learning* (Washington, DC: The Falmer Press).

- VOSNIADOU, S. and BREWER, W. (1992) Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535–585.
- WHITE, B. (1993) ThinkerTools: causal models, conceptual change, and science education. *Cognition and Instruction*, 10(1), 1–100.

APPENDIX A: The text and diagramming tasks: What causes the continents to move?

Scientists have shown that the pattern of the continents has changed over time. They believe that at one time all of the continents were close together and formed one very large continent, called Pangaea. Scientists estimate that about 200 million years ago, Pangaea started to break into several pieces. Since that time the pieces have slowly drifted apart. These pieces have become the continents that we see today. The continents continue to move about three-quarters of an inch to 4 inches per year. These paragraphs explain how the change from one big piece of land to several continents happened.

The layers of the earth

NOTE: After this paragraph you will be asked to draw a picture (on the next page) of the different layers of the earth.

The inside of the earth is made up of three different layers. If you could drill through the earth, the first layer you would drill through is the **crust**, which is 96 miles thick in some places. The continents we see and live on are only part of the crust. In other places, the crust dips down underwater to form the seabed. The crust is divided into moving sections called plates. Some continents are made up of more than one plate.

Under the crust is the second layer called the **mantle**. It is a layer made up of very thick liquid called magma.

At the centre of the earth's interior is the **core**, which is very hot.

Please go on to the next page.

(D1) Draw a picture of the different layers of the earth. Include and label all the information about these layers that you can.

Movement in the layers of the earth

NOTE: After this paragraph you will be asked to draw a picture (on the next page) of the movement and processes in the layers of the earth.

Remember that the crust is divided into sections called plates. Each plate can be thought of as a sheet of rock, riding on top of the mantle.

As mentioned before, the core of the earth is very hot. This heat creates currents that rise up through the mantle. When these currents get near the top of the mantle, they push on the plates, and force the plates to move in many directions. As the mantle moves, the plates move with it. Since the continents are part of the plates, the continents move too.

Please go on to the next page.

(D2) Draw a picture of the movement and processes in the different layers of the earth. Include and label all the information about these layers that you can.

How movement causes: mountains, volcanoes, and sea floor spreading.

How mountains form

NOTE: After this paragraph you will be asked to draw a picture of mountain formation.

When two plates are forced together, mountains can form. As the plates are forced together, the edges may be arched like a deck of cards being squeezed from both sides. Eventually one plate moves under the other plate. As the plates continue to move together, the crust is slowly bent and crumpled, and mountains are formed. While the rock may rise only a quarter of an inch per year, over millions of years it can form very high mountains. The Himalayan Mountains are the best example; they were formed when the plate of India collided with the plate of Southern Asia.

Please go on to the next page.

(D3) Draw a picture of the movement and processes in the different layers of the earth when mountains are being formed. Include and label all the information about these layers that you can.

How volcanoes form

NOTE: After this paragraph you will be asked to draw a picture of volcanic eruption.

Volcanoes occur mostly along, or very near, the edges of plates. This is because the edges are where most of the stress and cracks occur.

One way that volcanoes can form is when the plates move apart. As the currents of hot magma push on the plates, they move apart. Hot liquid magma from the mantle then rises up above the surface to form volcanoes.

Please go on to the next page.

(D4) Draw a picture of the movement and processes in the different layers of the earth when volcanoes are erupting. Include and label all the information about these layers that you can.

Why the Atlantic Ocean is getting wider

Since North America is still moving away from Europe, the Atlantic Ocean is getting wider! This is how it happens.

There are two plates, which reach under the Atlantic Ocean. The place where these two plates meet runs right down the middle of the ocean floor, from north to south. Over time, the hot currents in the mantle causes these plates to move further apart; the magma then rises up through the crack and fills in the resulting gap in the ocean floor. As the magma cools down it forms rock, which becomes a new part of the ocean floor. This is how the ocean floor gets bigger over time.

**The effects of different learning tasks on model-building in Plate Tectonics:
diagramming versus explaining.**

Janice D. Gobert

The Concord Consortium

10 Concord Crossing, Suite 300

Concord MA 01742

jgobert@concord.org

Ph: (978) 318.6975

Fax: (978) 371.3354

Abstract

Geology is a complex, semantically rich domain involving the interpretation of geological maps as external visualizations. Geological maps are complex in particular because 3-dimensional features must be inferred from 2-dimensional representations depicted by differing line types and weights. Modeling building, as an internal mental activity, is also required in order to achieve deep understanding of textual materials in geology, of geological maps, as well as in understanding complex causal processes, e.g., convection, underlying geological phenomena. Using literature from Cognitive Psychology, a framework for teaching and learning with visualizations in Plate Tectonics is given as an example of one difficult topic in Geology which involves the understanding of visualizations. Based on previous work in students' conceptions in Geology, three studies

of students' conceptions and cognition in plate tectonics were designed. These studies highlight the importance of progressive model-building as a good pedagogical approach, as well as examine the efficacy of different learning tasks as strategies to promote model-building on the part of learners.

Keywords: learning with diagrams, model-building, learning in Plate Tectonics

Introduction

Geology is a complex domain which requires interpreting and reasoning with visualizations that are semantically-rich (Frederiksen & Breuleux, 1988). More specifically, the visualizations referred to herein are external visualizations, e.g., graphics, maps, diagrams, models, simulations, etc. These are distinguished from internal visualizations, i.e., internal mental constructs or mental models, used in reasoning (Johnson-Laird, 1985). (More on the role of mental models later in the paper). Furthermore, the visualizations of interest here are semantically-rich representations which involve complex, domain-specific symbol systems and as such are distinguished from iconic visual representations, e.g., a stop sign, which do not require a deep, conceptual knowledge base. Thus, the comprehension of and reasoning with semantically-rich visualizations is much more complex (Gobert, 1994). Because of the complexity involved in understanding geological maps, Geology is an excellent domain in which to think about the human cognition underlying visualizations.

In general, comprehending or interpreting complex visualizations is difficult because all the information is presented to the learner simultaneously in contrast to textual information sources in which the information follows the structure of the text (Larkin & Simon, 1987). For more details on the information-processing ramifications of these differences, see Gobert, 2005 (in press). In the case of graphics in geology, another level of complexity is added because 3-dimensional information is represented in 2-dimensional form. Thus, in order to understand a terrain from a geological map for example, learners must be able to

make inferences about 3-dimensional features from 2-dimensional information depicted by differing line types and hierarchies of pen weights. This is a complex and non-trivial task similar to understanding a building as a 3-dimensional entity from its plans which depict this information in 2-dimensions in architectural plans (Gobert, 1994, 1999).

In unpacking the learning processes from visualizations in Geology, the literature from Cognitive Science provides an excellent framework for both research and teaching with visualizations. The next sections of this paper are dedicated to this goal.

Cognitive science literature as a framework for research and teaching with visualizations in Geology

In thinking about learning processes for visualization, learning is viewed as an active and constructive process. This view of learning is largely due to a seminal paper entitled "Levels of processing: A framework for memory research" (Craik & Lockhart, 1972; Lockhart & Craik, 1990) which introduced the notion that the nature of the learner's processing of the stimulus material largely determines the learner's memory representations for that material. The levels of processing framework was originally developed for text materials, but the framework has been subsequently shown to be applicable with visual stimuli, including faces (Bower & Karlin, 1974) and cartoon figures (Bower, Karlin, & Dueck, 1975), as well as complex conceptual visual stimuli such as those found in chess (Lane & Robertson, 1979) and architecture (Akin, 1978, 1979; Gobert, 1989, 1994, 1999).

Expert-novice literature. A great deal of what is known about visual information processing has come from the expert-novice literature both in terms of how domain-related information is stored and chunked in human memory and the ways in which information processing is directed by prior domain knowledge. Differences between experts and novices have been studied in many, diverse domains including computer programming (Adelson, 1981, 1984; McKeithen, Reitman, Reuter & Hirtle, 1981), algebra (Lewis, 1981), physics (e.g., Chi, Feltovich & Glaser, 1981; Chi, Glaser & Rees, 1982; Larkin et al, 1980), and medicine (Frankel Tal, 1992; Groen & Patel, 1988; Patel & Groen, 1986; Patel et al, 1990; Patel et al, 1984). Bereiter and Scardamalia (1993) and Ericsson and Smith (1990) provide a good review of this literature.

In terms of previous research on expertise, relatively few of the studies deal with visual information sources (compared to the total number of expertise studies conducted). Some of these studies include research in the following domains: chess (Chase & Simon, 1973; deGroot, 1965, 1946/1978), Go (Reitman, J., 1976), gomoku (Eisenstadt & Kareev, 1975), bridge (Charness, 1979), radiology (Lesgold, et al., 1988; Myles-Worsley, Johnston, & Simons, 1988), geographical map reading (Ormrod et al., 1986; Gilhooly et al., 1988; Thorndyke & Stasz, 1980), topographical map reading (Chang et al., 1985); architecture (Akin, 1979; Chase & Chi, 1981; Gobert, 1989; 1994), electronics (Egan & Schwartz, 1979), and engineering (Vicente, 1991, 1992; Bedard, 1993).

In terms of expertise studies in semantically-rich domains like geology, a few studies have been conducted. Egan and Schwartz (1979) used a recall task to examine differences in chunking of information from electronic circuit diagrams between novices and experts in electronics. In addition to recalling larger chunks, skilled electricians related some of the chunks together and used their conceptual knowledge of the function of the various circuits in order to structure their recall. Furthermore, it was suggested that their knowledge organization was attributable to the functional units they had identified during their initial learning of the circuit diagram.

Ormrod, Ormrod, Wagner, and McCallin (1986) used faculty from geography, educational psychology, and sociology in order to examine their respective abilities to learn and recall two maps: a logical one (based on geographical principles), and an illogical one (the elements were randomly placed). Geographers, having a great deal of knowledge about map features, were hypothesized to use their domain knowledge to organize the map features in a meaningful way. Educational psychologists were chosen for their knowledge related to memory and learning principles. A control group of sociologists was also added. Results for the logical map showed that the best performance was attained by geographers, followed by educational psychologists; however, in the case of the illogical map, the recall of all three groups of subjects was equally low. Thus, the geographers, being "map experts", applied principles from their domain in order to learn the chosen map; educational psychologists, whose recall was greater than the sociologists, applied principles from their domain, e.g., memory and learning strategies in order to learn the map. Similarly, in the

case of topographic map reading (Chang et al., 1985; Gilhooly et al., 1988; Eastman, 1985) experts were found to have better comprehension of relative heights of the terrain depicted in the map. Search strategies identified by eye-tracking showed that they attended to the highest and lowest points depicted (implicitly) in the map in order to fully understand the terrain (Chang et al., 1985).

In studies conducted in architecture, experts were found to represent their knowledge in hierarchical structures made up of spatial chunks (Akin, 1979; 1986; Chase & Chi, 1981) and that the nature of the learning processes employed affected the resulting conceptual representations (Akin, 1979; 1986). In a study involving the understanding of a building from its plans, experts were found to better understand the building as a 3-dimensional entity compared to their less expert counterparts and that experts also employed more sophisticated search strategies in that they were both more systematic and 3-dimensional compared to sub-experts. Again, the resulting understanding of the building in both groups was found to reflect their initial knowledge acquisition strategies (Gobert, 1994; 1999).

Important in all of these studies is the finding that experts used knowledge acquisition strategies for learning from visualizations that are highly related to required task performance in their respective domain. Thus, in each case, skills for acquiring knowledge from visual information sources have evolved through experience and are especially adapted for performance in their respective domain. This domain-specific prior knowledge used in acquiring knowledge from visual information sources are referred to as schemata

(Brewer & Nakamura, 1984; Schank & Abelson, 1977; Rumelhart & Norman, 1975) which provide perceptual and cognitive structures that influence the amount and manner in which information is acquired and encoded in memory such that experts can process domain-specific material to a deeper level these prior knowledge schemata also account for the superior recall and inference-making evidenced by experts when they are working in their domains of expertise (Chang, Lenzen, & Antes, 1985; Gilhooly et al., 1988; Head, 1984).

Approaches to eliciting deep processing of visual information sources. Deep processing of information is a necessary requirement for conceptual understanding, and thus, much of the research which is carried out in cognitive science and education has higher level learning as its goal. One approach to eliciting deeper processing is providing students with orienting tasks (cf., Craik & Lockhart, 1972; Schmalhofer & Glavanov, 1986). Orienting tasks are instructions given to learners as part of the task in order to structure the learners' knowledge acquisition and processing. Orienting tasks for processing target material have significant effects on learning for both simple (Schulman, 1971) and complex textual material (cf., Schmalhofer & Glavanov, 1986). Results from studies using orienting tasks have shown that the beneficial effects on learning are greatest when the learner's attention is brought to features of the target material which would not be attended to otherwise, or when orienting tasks lead learners to engage in methods of learning which they would not use spontaneously, particularly learners lacking specific domain-knowledge (Mayer, 1989).

Implications for understanding visual information sources from text comprehension. As previously mentioned, information about objects or processes may be presented in either visual or textual form. Although the comprehension processes for textual information sources is fairly well understood (Frederiksen et al, 1988; Kintsch, 1988), very little is known about the comprehension processes for visual information sources. Briefly, models of text comprehension propose that understanding a text is a stratified process in which the semantic information presented in a text is represented by the learner in several levels. The comprehension process also inference-making from the information explicitly represented in the text by the learner (Frederiksen, Bracewell, Breuleux, & Renaud, 1989; Frederiksen & Breuleux, 1988; Kintsch, 1986, 1988; vanDijk & Kintsch, 1983). The three levels of representation hypothesized are (Frederiksen, Bracewell, Breuleux, & Renaud, 1989; Frederiksen & Breuleux, 1988; Kintsch, 1986, 1988; vanDijk & Kintsch, 1983):

- (a) **the linguistic/syntactic level** which reflects the syntactic structure of the text and word/morpheme sequences upon which the syntactic parsing is performed.
- (b) **the propositional level** which reflects the semantic information presented in the information source. The propositional level is regarded as an intermediate semantic level of representation, and
- (c) **the conceptual level** which refers to a higher -level of semantic representation also called situation models (Kintsch, 1988) or mental models (Johnson-Laird, 1985) and are postulated to be the way in which information is represented in long-term memory.

Using these theories of comprehension, methods for coding learners' understanding have been developed (Frederiksen, 1975, 1986; Frederiksen, Bracewell, Breuleux, & Renaud, 1989) and can be successfully used to code the conceptual information contained in a textual/linguistic information source as well as to code learners' understanding of various types of information sources, including visualizations which expressed in natural language, such as think aloud protocols from learners. For example, Frederiksen's propositional model has been used to represent the understanding of complex semantic information in: chemical equations (Kubes, 1988; Frederiksen & Renaud, 1989), algebraic expressions (Frederiksen & Renaud, 1989), a text describing plate tectonics (Gobert & Clement, 1999), think aloud protocols about architectural plans (Gobert & Frederiksen, 1988; Gobert, 1989), and think aloud protocols about electronics diagrams (Bedard, 1993). Thus, in terms of semantically-rich visualizations, the working hypothesis here is that similar cognitive processes used in the comprehension of textual material also should operate in the comprehension of graphic information sources (Gobert, 1994). (It is important to note that there are likely modality-specific levels of representation also required in the comprehension of visual information sources).

In two of the studies presented herein, text is used as a learning source, thus, the text comprehension model and methods of coding are appropriate for these data. Levels 2 and 3, the propositional level and resulting conceptual representation (both described above), are the levels of representation we are concerned with for the purposes of this research. As predicted by the comprehension model, simple recall and recognition tasks are best

supported by the representation of the propositional information contained in the text, i.e. level 2. Inference-making and reasoning tasks, reflecting higher-level understanding, are best supported by representations which reflect higher level, more integrated representations, i.e., situation models or mental models (Johnson-Laird, 1985), level 3 (described above).

Model-based teaching & learning as a framework for learning with visualizations in Science. Compatible with the text comprehension framework above (but at a more general level of description) is a framework called Model-based teaching and learning (Gobert & Buckley, 2000) which underlies much of the student conception work on model-based reasoning in Science Education.

Model-based learning and teaching is a theory about science learning based on a synthesis of research in Cognitive Psychology (including text comprehension) and Science Education (Gobert & Buckley, 2000). In model-based teaching and learning, it is assumed that learners construct mental models, i.e., internal visualizations, of phenomena in response to a particular learning task (assuming the task has engaged the learner); these are thought to be in the mind's eye and used in mental imagery and to solve problems whereby people read off their mental model (Johnson-Laird, 1985). In learning science, the model that is constructed integrates pieces of information about the spatial structure of the object, the causal mechanisms involved in the process under inquiry, and other relevant features of the process. Reasoning with the model may instantiate evaluation of the model, leading to its

revision or elaboration; model revision involves modifying parts of an existing model so that is better explains a given system. Model-based reasoning requires modeling skills to understand representations, generate predictions and explanations, transform knowledge from one representation to another, as well as analyze data and solve problems.

Types Knowledge and Models in Plate Tectonics.

In thinking about Geology from a pedagogical point of view, it is productive to identify the types of knowledge one needs in order to understand geological phenomena. A useful approach to thinking about Plate Tectonics was framed as a part of earlier work (Gobert, 2000) in which propositional analysis (Frederiksen, 1985) was conducted on an explanatory text about Plate Tectonics. (Propositional analysis is a method of semantic analysis by which the smallest units of meaning are identified and then a semantic network model is constructed which allows the experimenter to evaluate the learner's knowledge about the text, and in turn, assess the types and respective amounts of knowledge which the learner has acquired either from the text or on the basis of inferences on the text.) Here, three types of knowledge were derived (it is likely that these apply to other sub-domains of Geology as well): **spatial knowledge**, i.e., the spatial structure of a geological object; in the case of Plate Tectonics, the inside structure of the earth, **causal knowledge**, i.e., causal mechanisms underlying Plate tectonic phenomena, e.g., convection currents, and **temporal knowledge**, i.e., knowledge about the time scale of different geological phenomena (continental drift versus volcanic eruption). Thus, in teaching Plate Tectonics, it is

reasonable to assume that breaking down the conceptual knowledge into these types would elicit deep learning. Additionally, in deciding on the order of presentation of conceptual knowledge, we used a progressive model-building approach in which simpler conceptual knowledge provides conceptual leverage for more complex types of knowledge. The pedagogical strategy of progressive model-building has been shown to be successful for supporting students' learning in physics in which simpler models of density and force addition provided conceptual leverage for understanding buoyancy (Raghavan & Glaser, 1995). Additionally this approach has been successful for electricity (White & Frederiksen, 1990) and Newtonian Mechanics (White, 1993) in which students learn a series of causally more complex models. In the studies to be presented later in this paper, we used this progressive model-building approach in which we first had students think about the spatial structure of the earth, then we engaged them in thinking about causal and dynamic processes inside the earth, lastly, we engaged them in thinking about two plate tectonic-related phenomena, namely mountain formation and volcanic eruption, as two real-world examples of plate tectonic phenomena.

Science education work on student conceptions in Geology. The topic of learning in Earth Science has not been well studied, particularly when compared to students' learning and conceptions in the physical sciences (Stofflett, 1994). The lack of research on learning in the Earth Sciences is likely due to the fact that in the past, it has received much less emphasis than the Physical and Life Sciences in national and state curricular standards. Now however, the National Science Education Standards (1996) are recognizing Earth

Science as a necessary and important component of science training across elementary, middle, and high school levels and considered equivalent in importance to training in the Life and Physical sciences (AAAS, 1989, 1993).

The importance of learning in this domain is reflected in a number of more recent projects on Earth Science covering both teacher professional development projects (cf., Mayer, Fortner, & Hoyt, 1995) and student cognition projects including: knowledge of the causes for earthquakes (Ross & Shuell, 1993; Bezzi, 1989; Turner, Nigg, and Daz, 1986), mountain formation (Muthukrishna et al, 1993), knowledge of the earth as a cosmic body (Vosniadou & Brewer, 1992, Nussbaum, 1979, Nussbaum & Novak, 1976; Sneider & Pulos, 1983), knowledge of rock-cycle processes (Stofflett, 1994), conceptions of earth and space as it relates to seasons and phases of the moon, (Schoon, 1992; Bisard et al, 1994), conceptions of sea floor dynamics (Bencloski and Heyl, 1985), knowledge of the earth's gravitational field (Arnold, Sarge, and Worrall, 1995), and knowledge of the scale of the earth (Ault, 1994). There are also some recent programs of research that utilize visualizations in Plate Tectonics for student learning, including the Visual earth project (www.tercworks.terc.edu), the Science Odyssey project (www.pbs.org/wgbh/aso/tryit/tectonics), the Visualizing Earth project (www.visearth.ucsd.edu), and the Princeton Earth Physics Project, a high school and college-based project which uses an array of seismographs for the study of earthquakes (<http://lasker.princeton.edu>). However, none of these existing programs (to our knowledge) seeks to address the plate tectonics in an integrated fashion; that is, some

emphasize sea floor spreading, earthquakes, volcanoes, etc., but none integrate all types of plate tectonic phenomena. Additionally, none of these explicitly emphasize *active model building* on the part of the students.

Plate Tectonics as a sub-domain of study. As previously mentioned, Plate Tectonics is an excellent domain in which to investigate students' model-based learning because of the plethora of models, (i.e., external visualizations) used in Geology and the important role that model building, (as an internal mental activity) plays in understanding geological phenomena of hidden mechanisms, e.g., convection underlying continental drift, earthquakes, volcanoes, mountain formation, and sea floor spreading.

Plate tectonics, which is typically covered in fifth or sixth grade and then again in eighth or ninth grade is representative of a difficult school science topic. It is difficult to learn for many reasons: 1) the earth's internal layers are outside our direct experience, 2) the size scale and the unobserved processes, e.g., convection, are difficult to understand (Ault, 1984; Gobert & Clement, 1994; 1999), 3) the time scale of geological processes is difficult for people to conceptualize since it surpasses our reference of a human lifetime (Jacobi et al., 1996), and 4) it involves the comprehension and integration of several different types of information, such as, spatial, causal, and dynamic (Gobert & Clement, 1994; 1999).

Research on Fostering Students' Models and Reasoning in Plate Tectonics. Previous research addressing model-based learning in plate tectonics include: the effects of a

multimedia environment, CSILE (Scardamalia & Bereiter, 1991), on students' graphical and causal explanations of continental drift (Gobert & Coleman, 1993); learning difficulties encountered in this domain (Gobert & Clement, 1994); the nature of students' pre-instruction models and associated causal reasoning (Gobert, 2000); the benefits of student-generated diagrams versus summaries (Gobert & Clement, 1999); the influence of students' epistemologies of models on learning in this domain (Gobert & Discenna, 1997); and students' on-line collaboration about plate tectonics (mtv.concord.org; Gobert, 1998; Gobert & Pallant, 2004). Research most relevant to the topic of model-based learning in Earth Science are reviewed briefly in turn.

Previous Research, Study 1: Students' pre-instruction models and learning difficulties.

Gobert and Clement (1994) investigated fifth grade students' pre-instruction models of plate tectonics by conducting one-on-one interviews with children. Students' diagrams and think aloud protocols (Ericsson & Simon, 1980) were examined as a reflection of their mental models. Three main difficulties were identified in students' model construction processes: (1) problems with setting up a correct static model of the layers, (2) difficulty understanding causal and dynamic information (e.g., heat as causal in forming convection currents, or currents causing plate movement), and (3) difficulties with the integration of several different types of knowledge including causal and dynamic knowledge into a causal chain in order to build an integrated mental model of the system.

Based on protocol analyses of middle school students' diagrams and interview data (Ericsson & Simon, 1980) as well as data from classroom research (Gobert & Clement, 1994; 1999), two types of student models of the inside of the earth were identified at this age level (see figure 1 and table 1 below). These models (below) were drawn in response to the prompt, "Draw a diagram of the different layers of the earth".

Insert figure 1 and table 1 here

Based again on protocol analyses of middle school students' diagrams and interview data (Ericsson & Simon, 1980) as well as data from classroom research (Gobert & Clement, 1994; 1999), four types of student models of the inside of the earth were identified at this age level (see figure 2 and table 2 below). These models were drawn in response to the prompt, "Draw a diagram to depict what happens in the different layers of the earth when a volcano erupts". The models (below) are on a continuum from Type 1a and 1b reflecting models with only heat-related mechanisms and movement-related mechanisms, respectively, as the primary causal mechanisms responsible for volcanic eruption to Type 3 models which integrate multiple heat-related and movement-related causal mechanisms thus, reflecting the most sophisticated model observed at this age level. An integrated model in the case of volcanic eruption, for example, refers to one in which students have integrated their spatial model of the earth with a number of causal and dynamic mechanisms (i.e., core as a heat source, convection currents pushing on plates, plates moving apart, and magma rising above the surface). It is assumed that from these rich

causal models, inferences can be made about the causal mechanisms involved in other plate tectonic phenomena, e.g., sea floor spreading, etc.

Insert figure 2 and table 2 here

Model-Based Reasoning afforded by different types of models. Further analyses of students' models and think aloud protocols (Gobert, 2000) were used to demonstrate that if the student correctly depicts (i.e., understands) the layers of the earth in a spatial layout of concentric circles, then they are better able to revise this model to include (and understand) the causal and dynamic processes in the earth. If, alternatively, the student has a spatially incorrect model of the earth, this model will need to be revised before the model will support reasoning and inference-making by means of perceptual cues such as spatial adjacency (Larkin & Simon, 1987), e.g., one student (see Gobert, 2000) had a spatially incorrect model of the earth (such as Type 0 in Figure 1) which could not support the understanding of convection currents. By contrast, it was also shown that spatially correct models can serve as tools for reasoning (Kindfield, 1993) and model revision, e.g., another student who had a spatially correct model of the interior of the earth (such as Type 1 in Figure 1) made the correct inference that because the core was hot and the mantle was beside the core, the core acts as a heat source for the magma (see Gobert, 2000). (It is important to note that the goal in this program of research is to facilitate students' understanding of plate tectonics by means of qualitative, simplified models. As such, issues like whether radioactive decay in the mantle acts, in part, as a heat source in addition to the earth's core are not addressed in middle school but can be addressed in high school.)

Previous Research, Study 2. Promoting model-based learning: diagramming versus summarizing as an orienting task for deep science learning.

Based on the analyses from Study 1 it was hypothesized that understanding of the different types of information in this domain (i.e., spatial, causal/dynamic), as well as model construction is facilitated by diagram-based learning elicited in a progressive model-building order. This hypothesis was also based on previous research that has shown that diagrams both permit inferences based on perceptual cues such as spatial adjacency (Larkin & Simon, 1987) and explicitly indicate structural relationships (Schwartz, 1993) which are difficult from textual representations.

We tested empirically the efficacy of two different orienting tasks, namely, student-generated diagrams versus student-generated summaries as means to foster the development of rich, integrated models like the Type 3 models achieved in Study 1 described above. More specifically, here we investigated whether the task of constructing diagrams while reading would promote the development of richer causal models when compared to the task of generating summaries.

Two groups of students were asked to either construct diagrams or summaries at four specific points during their reading of a text describing plate tectonics; a control group who read the text only was also included. After students had read the text, they were given a

written post-test that assessed both spatial/static knowledge and causal/dynamic knowledge. There were two sets of data generated: the intermediate data (diagrams or summaries) which reflect students' understanding of the text, and a set of post-test data, which reflect students' higher-level conceptual understanding of the domain. In accordance with the text comprehension model underlying this research (van Dijk & Kintsch, 1983), simple recall and recognition tasks are best supported by a memory for the text itself, i.e., a text-base of the propositional knowledge contained in the text whereas, higher-level inference tasks are best supported by higher level, more integrated representations, i.e., situation models (Kintsch, 1988) made on the basis of the text plus inferences made on the text. In accordance with this theory, it is assumed that the understanding that the students exhibit on the post-test is due to an interaction of the processing induced by the orienting task (presented to the students before the relevant paragraphs) of either diagram-drawing or summarizing with the processing of the main passage itself.

Analysis of summaries and diagrams as intermediate representations. An overall manova on the semantic content comparing the summaries and diagrams on each of the four intermediate tasks revealed statistically significant differences favoring the summary group ($F = 5.718, p = .001$). (Since the coding scheme is based on semantic information regardless of medium, the coding scheme can be applied to either summaries or diagrams (Gobert, 2000). In terms of these findings, the intermediate representations, the **summary group outperformed the diagram group**, i.e., the summaries contained more semantic information than did the diagrams (see Gobert & Clement, 1999 for details on these data).

Analysis of post-test scores as mental model representations. A manova of the post-test revealed significant differences between the three groups (diagram, summary, control) for both the understanding of spatial information ($F = 4.38, p < .05$) as well as the understanding of causal/dynamic information ($F = 4.31, p < .05$). Thus, in terms of the students' resulting conceptual understanding, **the diagram group outperformed the summary group** and there were no significant differences found between the summary group and the control group (see Gobert & Clement, 1999 for details on these data).

This “discrepancy” between the findings for the intermediate tasks (summary group > diagram group) and the post-test (diagram group > summary group) was interpreted as follows. For the summary group, because the media was the same (textual information source and textual summarization task), they were able to rely on rote memory of what they had just read in order to produce their summaries, as evidenced by the inclusion of more semantic information than the diagram group on the intermediate tasks. However, the summarization task, because it only elicited only rote processing of the text, did not promote inferencing or mental model construction, as evidenced by poorer performance on the post-test than the diagram group. For the diagram group, these data suggest that constructing diagrams as part of the reading task required the students re-represent their knowledge into a diagrammatic format, and that they could not solely rely on rote memory of the text to do this, as evidenced by lower scores on the intermediate tasks. More specifically, diagramming required inferences in order to restructure what they read into

diagrammatic representations, and this processing lead to an advantage in terms of the resulting conceptual understanding. These findings are consistent with van Dijk and Kintsch's (1983) theory of text comprehension, as well as studies which have shown that learner's representations of material can be altered by changing their goals for learning (Schmalhofer & Glavanov, 1986).

However, an empirical question remained as to whether a different orienting task would elicit a deeper processing of the text compared to diagramming as an orienting task. A study which utilized a higher-level orienting task (higher level than summarizing) during reading would test the hypothesis whether it was the diagrammatic medium in particular or inferencing in general which was supporting mental model construction and higher-level reasoning yielded by those in the diagram condition in Study 2.

Study 3: The Effects of Diagramming versus Explaining on Text-based Representations and Mental Models.

Here, constructing diagrams as an orienting task during reading was compared to constructing explanations during reading. The choice of explanation as an orienting task was influenced by work which has shown that knowledge integration in science can be facilitated by providing an explanation to others (Coleman, 1992, 1995), as well as by providing self-explanations (Chi et al, 1994). Chi and her colleagues have suggested, although not empirically demonstrated, that explanation-based activities are likely to

promote the same type of inferences as diagram-drawing. Thus, in this study we sought to test out whether it was the “translation” of the textual information into diagrammatic representations which influenced students’ conceptual gains yielded on the post-test from Study 2 (above) or whether a higher level orienting task might elicit deep levels of processing and inference-making and thus higher conceptual understanding on the post-test.

Subjects. Two classes of grade five students participated. The students ranged in age from 10 to 12 years. Students were drawn from a small town in western Massachusetts, more specifically, from the same school and teacher, as in Study 2, thus it is reasonable to assume that they represent the same demographic.

Procedure. Students were given a short text (about 2 pages) about Plate Tectonics. One group was asked to draw diagrams at specific points during the text, and one group was asked to write explanations at the same points during the text. The prompts to draw or explain were given prior to each section of the text. For example, “After this paragraph, you will be asked to draw a diagram of the different layers of the earth”. Thus it is assumed that the students’ processing of the text interacts with the orienting task that the students were given.

For both groups the orienting tasks were requested in order of increasing difficulty, as in Study 2 (above), to promote progressive model-building. The instructions given to the subjects were as follows:

Explanation 3) "After this paragraph you will be asked to explain what happens in the different layers of the earth when mountains are formed. Include all the information about these layers that you can so that a friend who had never heard of this could learn about it."
OR

Diagram 3) " After this paragraph you will be asked to draw a picture of the different layers of the earth when mountains are formed. Include and label all the information about these layers that you can so that a friend who had never heard of this could learn about it."

The orienting tasks were requested of the groups were: 1) depict/explain the different layers of the earth; 2) depict/ explain the causal processes which are occurring in these layers; and 3) depict/explain what happens in the layers of the earth when mountains are formed.

Coding of data. Coding schemes were developed for each of the three orienting tasks; the scheme was used to code the diagrams and explanations for the inclusion of propositional information from the text source. More information about this type of coding scheme can be found in Gobert (2000). Using these data, the two groups were compared in terms of the semantic information contained in their explanations and diagrams during their reading of the text, as in Study 2; again, these data reflect their intermediate representations of the text.

The two groups were compared on their post-test, i.e., their resulting conceptual representations for the spatial as well as causal and dynamic aspects of the domain; again, this reflects their resulting conceptual understanding of the text plus inferences made on the text, i.e., their mental models.

Results

Analysis of explanations and diagrams as intermediate representations. For the comparison of the semantic information contained in explanations and diagrams which were generated during the students' reading of the text, a Manova yielded no statistically significant differences between the two groups ($F= 1.31$, $p= .283$ (Wilks); n.sig.). See the table below for the univariate Fs and the means.

Insert table 3 here

Analysis of post-test data. A Manova was performed with both spatial and causal dynamic understanding entered as variables. No statistically significant differences were obtained at either the multivariate [$F = 1.89$, $p=.162$; n. sig.] or univariate level for either the measure of the spatial layers of the earth [$F= 1.05$, $p=.310$; n. sig.], or the causal and dynamic processes involved in plate tectonics [$F = .075$, $p=.785$; n. sig.]. See Tables 4 and 5 for means and standard deviations.

Insert tables 4 & 5 here

Summary of results from Study 3. The explanations and diagrams that were constructed during students reading of the text contained approximately similar amounts of semantic information. In terms of the resulting conceptual representations, both groups also scored equally well in terms of their understanding of both the spatial layout of the layers of the earth as well as the causal and dynamic processes in the layers.

Discussion and Conclusions

In this paper relevant literature from Cognitive Science and Science Education are presented as a framework for thinking about learning and teaching with semantically rich-visualizations, such as those used in the Geology. Specifically, studies addressing expertise in learning with visually-complex representations are presented, as well as are findings about how to elicit deep processing of visually complex representations. Models of text comprehension are briefly described as framework for thinking about the comprehension of semantically-rich visualizations. Propositional analysis, derived from models of text comprehension, is briefly described in terms of how it can be used to systematically code learners' understanding on the basis of their think aloud protocols, diagrams, or written text (i.e., summaries or explanations). Lastly, model-based learning and teaching is described as a theoretical synthesis of cognitive psychology and science education; this framework, applied in the present studies, underlies (either explicitly or implicitly) much of the research on students' conceptions and conceptual change in science.

Regarding research on learning with visualizations, three studies are presented. Study 1 is an example of the types of mental models students hold; the reasoning associated with different types of mental models is described elsewhere (Gobert, 2000). This study makes a contribution to the literature since the types of models that students hold at this age level provide insight into why learning in this domain is difficult. Also, these models represent the pre-instruction conceptions that students bring to instruction in Plate Tectonics, thus, these findings have pedagogical implications for teaching Plate Tectonics.

Studies 2 and 3 employ methodologies from Cognitive Science i.e., the comprehension framework, and the semantic analysis that was applied to students' articulated models (i.e., their diagrams) and to their summaries and/or explanations. The data from Study 2 demonstrated the superior effects of diagramming over summarizing at intermittent points during reading as means to promote deep processing of textual material. These findings were interpreted as the diagramming orienting task as having a representational advantage over the summary task since the diagramming task provided affordances for both developing better mental models of the domain, and using these models, once constructed, as inference-making devices (Kindfield, 1993/1994). A follow-up study, Study3, was conducted in order to test whether it was the visual medium of diagramming or inferencing in general that was driving the superior learning exhibited by the diagram group in Study 2. In this study, explanation was chosen as an orienting task (versus diagramming) as means to promote deep processing of the text since explaining requires a higher-level of processing than does summarizing (Chi et al, 1994; Coleman, 1992). The data here yielded

interesting results, i.e., no differences were observed between the two groups on either their intermediate representations (as measured by the semantic information contained in their diagrams versus explanations) or their resulting understanding (as measured by the post-test). From these data, it is suggested that both types of orienting tasks, diagramming and generating explanations elicited deep processing on the part of the students. For example, it is possible that those in the explanation condition, i.e., who knew they were going to generate explanations at specific points in the text, were developing mental models in order to think deeply about the information needed in their explanation, and thus, the processing affordances are similar to those who were in the diagram condition. However, based on these data, we can only speculate about the modality-specific versus modality-general processing mechanisms. Research is currently underway in order to examine possible reading time differences for the different orienting task conditions, namely, summarization, explanation, and diagramming (Gobert, 2002) in order to try to empirically tease out possible differences. If processing differences are found across these varying orienting tasks, these data will be used to infer the interaction between the nature of the orienting task, the modality-specific as well as the modality-general processes employed in constructing and revising mental models, and the processing of the text itself. Lastly, these data will contribute to the cognitive science literature in terms of processing differences and affordances for learning in the different information modes; these data contribute to Science Education in terms of the implications for instruction with these information modes.

Acknowledgements.

Figures 1 and 2 and Tables 1 and 2 and appeared previously in Gobert, J. (2000). A typology of models for plate tectonics: Inferential power and barriers to understanding. *International Journal of Science Education*, 22(9), 937-977 and are reprinted with permission from Taylor and Francis, Inc.

Studies 1 & 2 were funded by a NSF grant awarded to John Clement. Study 3 was supported by a Spencer Post-Doctoral Award from the National Academy of Education awarded to Janice Gobert.

References

- American Association for the Advancement of Science. (1989). *Science For All Americans: Project 2061*. Washington, D.C.: American Association for the Advancement of Science.
- American Association for the Advancement of Science. (1993). *Benchmarks for Science Literacy*: 1993. New York Oxford: Oxford University Press.
- Adelson, B. (1981). Problem-solving and the development of abstract categories in programming languages. *Memory and Cognition*, 9, 422-433.
- Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 483-495.
- Akin, O. (1978). How do architects design? In J.-C. Latombe (Ed.), *Artificial intelligence and pattern recognition in computer aided design*. New York: North-Holland.
- Akin, O. (1979). Models of architectural knowledge: An information-processing view of design. *Dissertation Abstracts International*, 41(3), p. 833 A. (University Microfilms No. 72-8621.)
- Arnold, P., Sarge, A., & Worrall, L. (1995). Children's knowledge of the earth's shape and its gravitational field. *International Journal of Science Education*, 17(5), 635-641.
- Ault, Jr., C. R. (1984). The everyday perspective and exceedingly unobvious meaning. *Journal of Geological Education*, 32, 89-91.
- Bedard, D. (1993). *Utilisation des connaissances dans la résolution de problèmes en électronique*. Unpublished doctoral thesis. McGill University, Montreal.

- Bencloski, J. W., & Heyl, R. J. (1985). Teaching plate tectonics with the aid of a model of sea-floor dynamics. *Journal of Geological Education*, 33, 274-276.
- Bereiter, C., and Scardamalia, M. (1993). *Surpassing ourselves: An inquiry into the nature and implications of expertise*. Chicago: Open Court.
- Bezzi, A. (1989). Geology and society: A survey on pupil's ideas as an instance of a broader prospect for educational research in earth science. Paper presented at the 28th International Geological Congress, Washington, D.C.
- Bisard, W. J., Aron, R. H., Francek, M. & Nelson, B. D. (1994). Assessing selected physical science and earth science misconceptions of middle school through university pre-service teachers. *Journal of College Science Teaching*, September/October, 38-42.
- Board, C. (1984). Higher order map-using tasks: Geographical lessons in danger of being forgotten. *Cartographica*, 21, 85-97.
- Bower, G., and Karlin, M.B. (1974). Depth of processing of pictures of faces and recognition memory. *Journal of Experimental Psychology*, 103, 751-757.
- Bower, G., and Karlin, M.B., and Dueck, A. (1975). Comprehension and memory for pictures. *Memory & Cognition*, 3, 216-220.
- Brewer, W., and Nakamura, G. (1984). The nature and functions of schemas. In R.S. Wyer and T.K. Skell (Eds.), *Handbook of Social Cognition, Volume 1*. Hillsdale, NJ: Lawrence Erlbaum.
- Chang, K.-T., Lenzen, T., and Antes, J. (1985). The effect of experience on reading topographic relief information: Analyses of performance and eye movements. *The Cartographic Journal*, 22, 88-94.
- Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology*, 33, 1-50.
- Chase, W., & Chi, M. (1980). Cognitive skill: Implications for spatial skill in large-scale environments. In J. Harvey (Ed.), *Cognition, social behavior, and the environment*. Potomac, MD: Erlbaum.
- Chase, W., & Simon, H. (1973). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- Chase, W., and Chi, M. (1981). Cognitive skill: Implications for spatial skill in large-scale environments. In J. Harvey (Ed.), *Cognition, social behavior, and the environment*. Potomac, MD: Lawrence Erlbaum.
- Chase, W., and Simon, H. (1973). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- Chi, M.T., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Chi, M.T., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.
- Chi, M.T.H., de Leuw, N., Chiu, M. H., and LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439-478.
- Coleman E.B. (1992). *Facilitating conceptual understanding in science: a collaborative explanation-based approach*. University of Toronto, Unpublished doctoral dissertation.

- Coleman, E. (1995). Learning by explaining: Fostering collaborative progressive discourse in science. In R. J. Beun, M. J. Baker, and M. Reiner (Eds.), *Natural dialogue and interactive student modeling*, Berlin, Germany: Springer-Verlag.
- Craik, F.I.M., and Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- de Groot, A. (1965). *Thought and choice in chess*. The Hague: Mouton.
- de Groot, A. (1978). *Thought and choice in chess*. The Hague: Mouton. (Original work published in 1946).
- Eastman, J.R. (1985). Cognitive models and cartographic design research, *The Cartographic Journal*, 22, 95-101.
- Egan, D., and Schwartz, B. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition*, 7, 149-158.
- Eisenstadt, M., and Kareev, Y. (1975). Aspects of human problem-solving: The use of internal representations. In D.A. Norman and D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco, CA: Freeman.
- Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.
- Ericsson, K.A., and Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In K.A. Ericsson and J. Smith, (Eds.), *Toward a general theory of expertise: Prospects and limits*. Cambridge, MA: MIT Press.
- Frankel Tal, N. (1992). *Diagnostic reasoning of difficult internal medicine case: Expert, proto-expert, and non-expert approaches*. Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario.
- Frederiksen, C. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology*, 7, 471-458.
- Frederiksen, C. (1985). *Comprehension of different types of text structure*. Paper presented at the Annual Meeting of the Canadian Psychological Association, Halifax, Nova Scotia, Canada.
- Frederiksen, C. (1986). Cognitive models and discourse analysis. In C. Cooper & S. Greenbaum (Eds.), *Written communication annual volume 1: Linguistic approaches to the study of written discourse*, Beverly Hills, CA: Sage.
- Frederiksen, C. (1988). Text comprehension in functional task domains. In D. Bloom (Ed.), *Learning to use literacy in educational settings*. Norwood, NJ: Ablex.
- Frederiksen, C., & Renaud, A. (1989, March). *Representation, acquisition and use of procedural knowledge in chemical engineering*. Paper presented at the American Educational Research Association annual meeting, San Francisco, California.
- Frederiksen, C., & Breuleux, A. (1988). Monitoring cognitive processing in semantically complex domains. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition*. Hillsdale, NJ: Erlbaum.
- Frederiksen, C., and Renaud, A. (1989). *Representation, acquisition and use of procedural knowledge in chemical engineering*. Paper presented at the American Educational Research Association annual meeting, San Francisco, California.
- Frederiksen, C., Bracewell, R., Breuleux, A., & Renaud, A. (1989). The cognitive representation and processing of discourse: Function and dysfunction. In Y. Joannette

- and H. Brownell (Eds.), *Discourse ability and brain damage: Theoretical and empirical perspectives*. New York: Springer-Verlag.
- Frederiksen, C., Decary, M. & Hoover, M. (1988). The semantic representation and processing of natural language discourse: Development of a computational model. *Proceedings of the International Colloquium on "Informatique et langue naturelle."* Nantes, France.
- Gilhooly, K.J., Wood, M., Kinnear, P.R., and Green, C. (1988). Skill in map reading and memory for maps. *The Quarterly Journal of Experimental Psychology*, 40A (1), 87-107.
- Gobert, J. (1989). *The interpretation of architectural plans by expert and sub-expert architects*. Unpublished master's thesis. McGill University, Montreal, Canada.
- Gobert, J. (1994). *Expertise in the comprehension of architectural plans: Contribution of representation and domain knowledge*. Unpublished doctoral dissertation. University of Toronto, Toronto, Ontario.
- Gobert, J. (1999). Expertise in the comprehension of architectural plans: Contribution of representation and domain knowledge. In *Visual And Spatial Reasoning In Design '99*, John S. Gero and B. Tversky (Eds.), Key Centre of Design Computing and Cognition, University of Sydney, AU.
- Gobert, J. (2000). A typology of models for plate tectonics: Inferential power and barriers to understanding. *International Journal of Science Education*, 22(9), 937-977.
- Gobert, J. (2002). *Summarizing, Explaining, and Diagramming: The Differential Effects on Reading Times, Text-Base Representations, and Mental Models of Science Text*. Presented at the Winter Text Conference, Jackson Hole, WY, January 25-28.
- Gobert, J. & Buckley, B. (2000). Special issue editorial: Introduction to model-based teaching and learning. *International Journal of Science Education*, 22(9), 891-894.
- Gobert, J. & Clement, J. (1994). *Promoting causal model construction in science through student-generated diagrams*. Presented at the Annual Meeting of the American Educational Research Association, April 4-8. New Orleans, LA.
- Gobert, J. & Clement, J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching*, 36(1), 39-53.
- Gobert, J. & Coleman, E.B. (1993). Using diagrammatic representations and causal explanations to investigate children's models of continental drift. *Proceedings of the Society of Research in Child Development*. March 25-28. New Orleans, LA.
- Gobert, J. & Discenna, J. (1997). *The Relationship between Students' Epistemologies and Model-Based Reasoning*. Kalamazoo, MI: Western Michigan University, Department of Science Studies. (ERIC Document Reproduction Service No. ED409164).
- Gobert, J. & Frederiksen, C. (1988). The comprehension of architectural plans by expert and sub-expert architects. *Proceedings of the Tenth Annual Meeting of the Cognitive Science Society*. Montreal, Canada. Hillsdale, NJ.: Lawrence Erlbaum.
- Gobert, J., and Frederiksen, C. (1989). *Expert and novice semantic interpretation of architectural drawings*. Paper presented at the Annual Meeting of the American Educational Research Association, March,. San Francisco, California.

- Gobert, J.D., & Pallant, A., (2004). Fostering students' epistemologies of models via authentic model-based tasks. *Journal of Science Education and Technology*. Vol 13(1), 7-22.
- Groen, G., and Patel, V. (1988). The relationship between comprehension and reasoning in medical expertise. In Chi, M.T.H., Glaser, R., and Farr, M. (Eds.), *The Nature of Expertise*. Hillsdale, NJ: Lawrence Erlbaum.
- Head, C. (1984). The map as natural language: a paradigm for understanding. *Cartographica*, 31, 1-32.
- Jacobi, D., Bergeron, A., & Malvesy, T. (1996). The popularization of plate tectonics: presenting the concepts of dynamics and time, *Public Understanding in Science*, 5, 75-100.
- Johnson-Laird, P.N. (1985). Mental models. In A. Aitken, and J. Slack (Eds.), *Issues in cognitive modelling*. London: LEA Publishers.
- Kindfield, A.C.H. (1993/1994). Biology Diagrams: Tools to think with. *Journal of the Learning Sciences*, 3(1), 1-36.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Lawrence Erlbaum.
- Kintsch, W. (1986). Learning from text. *Cognition and Instruction*, 3(2), 87-108.
- Kubes, M. (1988). *The use of prior knowledge in integration of information from technical materials*. Unpublished doctoral dissertation. McGill University, Montreal.
- Lane, D., and Robertson, L. (1979). The generality of the levels of processing hypothesis: An application to memory for chess positions. *Memory & Cognition*, 7(4), 253-256.
- Larkin, J. & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65-100.
- Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.
- Lesgold, A., Robinson, H., Feltovich, P., Glaser, R., Klopfer, D., and Wang, Y. (1988). Expertise in diagnosing x-ray pictures. In Chi, M.T.H., Glaser, R., and Farr, M. (Eds.), *The Nature of Expertise*. Hillsdale, NJ: Lawrence Erlbaum.
- Lockhart, R., and Craik, F.I.M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology*, 44(1), 87-112.
- Mayer, V.J., Fortner, R.W., and Hoyt, W. H. (1995). Using cooperative learning as a structure for Earth Systems Education workshops. *Journal of Geological Education*, 43, 395-400.
- Mayer, R. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, Vol. 81(2), 240-246.
- Muthukrishna, N., Carnine, D., Gressen, B., & Miller, S. (1993). Children's alternative frameworks: Should they be directly addressed in science instruction? *Journal of Research in Science Teaching*, 30(3), 233-248.
- Myles-Worsley, M., Johnston, W., and Simons, M. (1988). The influence of expertise on x-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 553-557.

- National Committee on Science Education Standards and Assessment. (1996). *National Science Education Standards: 1996*. Washington, D.C.: National Academy Press.
- Nussbaum, J. (1979). Children's conceptions of the earth as a cosmic body: a cross age study. *Science Education*, 63, 83-93.
- Nussbaum, J. & Novak, J. D. (1976). An assessment of children's concepts of the earth utilizing structured interviews. *Science Education*, 60, 535-550.
- Ormrod, J., Ormrod, R., Wagner, E., & McCallin, R. (1986, April). Cognitive strategies in learning maps. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA.
- Patel, V., & Groen, G. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science*, 10, 91-116.
- Patel, V., Groen, G., and Arocha, J. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition*, 18(4), 394-406.
- Patel, V., HoPingKong, H., and Mark, V. (1984). Role of prior knowledge in comprehension of medical information by medical students and physicians. *Proceedings of Research in Medical Education*. Chicago: Association of American Medical Colleges.
- Raghavan, K. & Glaser, R. (1995). Model-based analysis and reasoning in science: The MARS curriculum. *Science Education*, 79, 37-61.
- Reitman, J. (1976). Skilled perception in GO. *Cognitive Psychology*, 8, 336-356.
- Ross, K. & Shuell, T. (1993). Children's beliefs about earthquakes. *Science Education*, 77, 191-205.
- Rumelhart, D., & Norman, D. (1975). The active structural network. In D. Norman, D. Rumelhart, and the LNR Research Group (Eds.), *Explorations in cognition*. San Francisco, CA: W.H. Freeman.
- Scardamalia, M & Bereiter, C (1991). High levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *Journal of the Learning Sciences*, 1(1), 37-68.
- Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum.
- Schmalhofer, F., and Glavanov, D. (1986). Three components of understanding a programmer's manual: verbatim, propositional and situational representations. *Journal of Memory and Language*, 25, 279-294.
- Schoon, K. J. (1992). Students' Alternative Conceptions of Earth and Space. *Journal of Geological Education*, 40, 209-214.
- Schulman, H. (1971). Similarity effects in short-term memory. *Psychological Bulletin*, 75, 399-415.
- Schwartz, D. (1993). The construction and analogical transfer of symbolic visualizations. *Journal of Research in Science Teaching*, 30, 1309-1325.
- Sneider, C. & Pulos, S. (1983). Children's cosmologies: understanding the earth's shape and gravity. *Science Education*, 67, 205-221.
- Spilich, G., Vesonder, G., Chiesi, H., & Voss, J. (1979). Text processing of domain related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18, 275-290.

- Spiro, R. (1977). Remembering information from text: The "state of schema" approach. In R. C. Anderson, R. J. Spiro, and W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum.
- Spiro, R. (1980). Constructive processes in prose comprehension and recall. In R. Spiro, B. Bruce, and W. Brewer (Eds.) *Theoretical issues in reading comprehension*. Hillsdale, NJ: Lawrence Erlbaum.
- Stofflett, R. (1994). Conceptual change in elementary school teacher candidate knowledge of rock-cycle processes. *Journal of Geological Education*, 42, 494-500.
- Thorndyke, P., & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. *Cognitive Psychology*, 12, 137-175.
- Turner, R. H., Nigg, J. M., & Daz, D. H. (1986). *Waiting for disaster: Earthquake watch in California*. Berkeley, CA: University of California.
- van Dijk, T. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Vicente, K. (1991). *Supporting knowledge-based behavior through ecological interface design*. Unpublished doctoral dissertation. University of Illinois, Chicago.
- Vicente, K. (1992). Memory recall in a process control system: A measure of expertise and display effectiveness. *Memory & Cognition*, 20(4), 356-373.
- Vosniadou, S. & Brewer, W. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535-585.
- White, B. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction*, 10 (1), 1-100.
- White, B. & Frederiksen, J. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, 24, 99-157.

ASSESSMENT AND ACTIVE LEARNING STRATEGIES FOR INTRODUCTORY GEOLOGY COURSES

David A. McConnell

David N. Steer

Kathie D. Owens

Department of Geology, University of Akron, Akron, OH 44325-4101, dam6@uakron.edu

Department of Geology, University of Akron, Akron, OH 44325-4101, steer@uakron.edu

Department of Curricular and Instructional Studies, College of Education, University of Akron, Akron, OH 44325-4205, kowens@uakron.edu

ABSTRACT

Educational research findings suggest that instructors can foster the growth of thinking skills and promote science literacy by incorporating active learning strategies into the classroom. Active learning occurs when instructors build learner participation into classes. Learning in large, general education Earth Science classes was evaluated using formative assessment exercises conducted by students in groups. Bloom's taxonomy of cognitive development was used as a guide to identify critical thinking skills (comprehension, application, analysis, synthesis, evaluation) that could be linked to specific assessment methods such as conceptests, Venn diagrams, image analysis, concept maps, open-ended questions, and evaluation rubrics. Two instructors conducted a series of analyses on sample classes taught with traditional lecture and inquiry-based learning methods. Qualitative and quantitative analyses show that such methods are preferred by students, improve student retention, produce no decrease in content knowledge, promote deeper understanding of course material, and increase logical thinking skills.

Keywords: active learning, inquiry-based learning, assessment, Bloom's taxonomy

INTRODUCTION

Several studies have emphasized the need to improve science literacy among non-science majors (American Geophysical Union, 1994; National Science Foundation, 1996; National Research Council, 1997) and college instructors have consistently ranked student intellectual development as a primary teaching goal (Angelo and Cross, 1993; Trice and Dey, 1997; Figure 1). Teachers can meet these complementary goals by, focusing on remedies that make content relevant to the intended audience, increasing student-student interaction in class, and encouraging conceptual understanding rather than rote memorization of facts (Chickering and Gamson, 1987; Tobias, 1990, 1992; Angelo, 1993; Astin, 1993). Such objectives can be realized by the combination of two teaching strategies, active learning and inquiry-based learning (Siebert and McIntosh, 2001). Active learning occurs when instructors build learner participation directly into classes using exercises that ask students to apply newly acquired knowledge to solve problems that may range from a single multiple-choice question to a class-length project (Silberman, 1996). Inquiry-based learning introduces elements of scientific inquiry into active learning exercises. Teaching strategies that promote inquiry-based learning (Allard and Barman, 1994; Mazur, 1997) emphasize higher-level thinking processes such as making observations, posing questions, analyzing data, making predictions, and

communicating ideas (Brunkhorst, 1996; National Research Council, 2000).

This paper describes a variety of learning strategies that may be adopted in introductory geology courses to encourage the development of higher-order thinking skills. We assume the reader has no prior experience in active learning methods and provide directions for implementing these techniques in the classroom. We discuss six hierarchical levels of student learning and link them to examples of appropriate assessment tools that were used successfully in several sections of a general education Earth Science course taught by two instructors at the University of Akron. These teaching strategies have been evaluated qualitatively using peer reviews, student written evaluations and semi-structured student interviews; and quantitatively by measuring improvements in student retention, exam scores, and scores on a logical thinking assessment instrument.

TEACHING, LEARNING AND ASSESSMENT

Teaching faculty consistently rank the development of higher-order thinking skills ahead of other teaching goals (Angelo and Cross, 1993). Unfortunately, large numbers of students in introductory courses frequently find themselves in an educational setting where learning is reduced to low level intellectual skills of listening and recording information that will be memorized for a multiple choice exam (Pinet, 1995; Prothero, 2000; McManus, 2002). Students familiar with high school experiential learning strategies allied with the national science standards will be unaccustomed to lecture delivery, especially in large-class settings (Collins, 1997). Content-driven coursework that can be efficiently graded by multiple-choice tests has proven ineffective in promoting deep student understanding of basic science concepts (Tobias, 1990). Furthermore, it can have a negative impact on student attitudes about science, even among majors (Allard and Barman, 1994; Gibbons, 1994; Sundberg et al., 1994; De Caprariis, 1997). As a result, such courses are usually poor recruiting and retention tools. In many institutions, pre-service teachers make up a significant proportion of introductory science courses. Teachers in K-12 schools not only learn what they will teach in these classes, but are also exposed to teaching models by their instructors (Collins, 1997). Finally, general education science courses represent an important opportunity for students to develop the critical thinking skills that are essential for success in college.

In recent years, college science instructors have attempted to encourage in-class learning by utilizing teaching methods that promote collaborative, active learning during lecture periods (Macdonald and Korinek, 1995; Ebert-May et al., 1997; Mazur, 1997; Reynolds and Peacock, 1998; Murck, 1999; Crouch and Mazur, 2001; Wyckoff, 2001). Student interaction through collaborative learning is a key determinant of student performance (Bykerk-Kauffman, 1995; Lord,

		Learning Tool (Assessment Method)					
Bloom's Taxonomy	Learning Skill	Concept-test	Venn Diagram	Image Analysis	Concept Map	Open-ended Question	Evaluation Rubric
Knowledge	memorization and recall	•	•	•	•	•	•
Comprehension	understanding	•	•	•	•	•	•
Application	using knowledge	•	•	•	•	•	•
Analysis	taking apart information		•	•	•	•	•
Synthesis	reorganizing information				•	•	•
Evaluation	making judgements					•	•

Table 1. Formative assessment methods and Bloom's taxonomy.

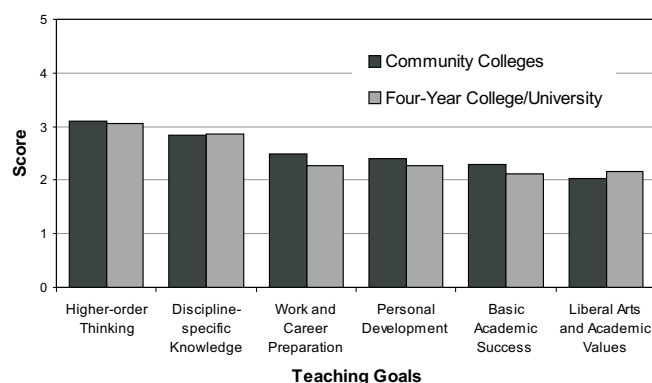


Figure 1. Relative scores on the Teaching Goals Inventory (Angelo and Cross, 1993). Instructors at both community colleges and four-year institutions ranked the development of higher-order thinking skills ahead of other teaching goals. n = 1873 for community colleges; n = 951 for four-year colleges and universities.

2001). The benefits of active learning and inquiry-based teaching methods can be seen in improvements in student attitudes about science (Gibbons, 1994; Ebert-May et al., 1997; Reynolds and Peacock, 1998) and increases in standardized test scores (Mazur, 1997; Hake, 1998).

Instructors typically assess learning by having students complete an exam following several weeks of lectures (McManus, 2002). This is a form of summative evaluation that comes at the end of a course of study, often too late to correct mistakes or identify gaps in comprehension. In contrast, formative assessment methods can be used to identify learning problems

during the presentation of information while there is an opportunity to recognize and correct misconceptions. The use of formative assessment can transform a traditional passive lecture into an active learning experience, as it requires that students provide feedback on their ongoing learning, thus giving the instructor an opportunity to highlight concepts that require additional explanation. For formative assessment to be effective, we must find questions to ask that will engage students and provide answers that can be used to signal understanding or confusion. Furthermore, we can link these assessment tools to different learning skills to nurture cognitive development.

A FRAMEWORK FOR LEARNING

Over forty years ago, Benjamin Bloom and several co-workers created a taxonomy of educational objectives that continues to provide a useful structure for organizing learning exercises and assessment experiences at all levels of education (Bloom et al., 1956; Anderson and Sosniak, 1994; Anderson and Krathwohl, 2001). Bloom's taxonomy divided cognitive learning into six levels (Table 1), from lower-level thinking skills such as memorization to higher order thinking that involves the evaluation of information. The taxonomy has been used by instructors in geology courses to guide the development of questions that address a full range of cognitive skills (Fuhrman, 1996; Nuhfer, 1996). Each taxonomy level is described briefly below and examples of specific questions linked to each level are presented.

Knowledge - Answers to knowledge questions indicate if a student knows and can recall specific information. Examples of questions that assess knowledge are some types of multiple choice questions, true/false questions, definitions, matching questions, or lists. Questions that ask students to define, identify, list, or name are often

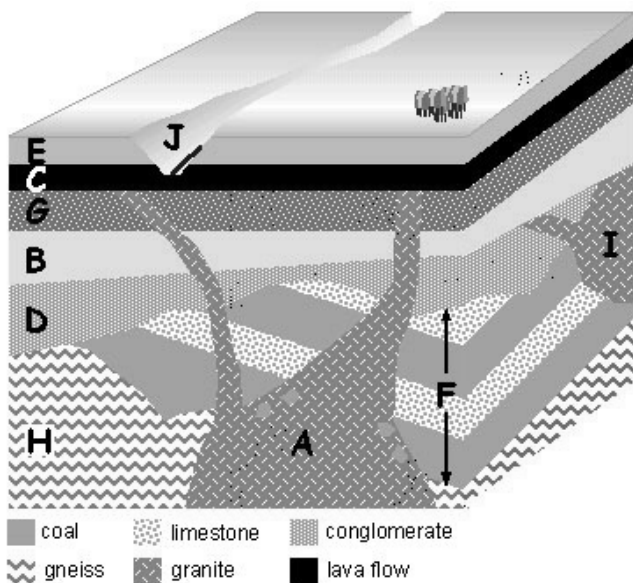


Figure 2. An application question would ask students to use the laws of superposition, cross cutting relationships, and original horizontality to determine the order of events for the labeled features in the idealized figure above.

“knowledge” questions. The following are knowledge questions.

K1. Which of the following is an igneous rock?

- a) limestone b) granite c) slate d) coal

K2. List the names of three major plates.

Multiple choice, true/false, and matching questions require that students recognize information stored in memory. Listing or fill-in-the-blank questions require that a student is able to remember specific information, as the questions themselves do not provide the answer choice.

Comprehension - Responses to comprehension questions report information or observations. Students must possess some basic knowledge of the subject to correctly answer these questions. Comprehension questions can fall into several categories and require students to convert, summarize, classify, infer, compare, or explain information. Examples of questions might include the following:

C1. Draw a diagram that shows the relationships between the principal components of the Earth system.

C2. View the motion picture Dante’s Peak and summarize the principal geological concepts presented in the movie.

C3. Take four pictures or samples of igneous rocks and sort them into volcanic and plutonic rock types.

C4. Fill in the blank to complete the analogy.

The yolk is to the egg as the _____ is to Earth.

C5. Contrast the floor of the Atlantic Ocean with the shape of a bathtub.

C6. Predict what would happen to sea level if it were to rain continuously worldwide.

Application - Application involves applying rules or principles to new situations, using known procedures to solve problems, or demonstrating how to do something. Questions that ask students to solve a problem using a known equation or to select a procedure to complete a new task would be considered application questions. An examples of an application question follows:

Ap1. Use the principles of superposition, cross cutting relationships, and original horizontality to determine the order of formation of labeled features in Figure 2.

Analysis - Answers to analysis questions may give directions, make commentaries, scrutinize data, explain how something works, or distinguish fact from opinion. Analysis requires that students break information into component parts to identify its organization. Students are expected to find links between data and interpretations and to discover which material is relevant to a task and which is extraneous. Questions that ask students to diagram, illustrate, outline or subdivide would be considered analysis questions, for example:

An1. Identify the hypothesis, observations, and conclusions in an assigned research report.

An2. Read a newspaper editorial and determine if it was written from a pro-environment or pro-development perspective.

Synthesis - Synthesis combines a series of parts into a greater whole. Good answers to synthesis questions may predict the outcome for a particular event and may involve making generalizations and developing a “big picture” view of a phenomenon or feature. Questions may ask students to create multiple hypotheses to explain a phenomenon, to develop a plan to solve a problem or to devise a procedure to accomplish a task. Examples of synthesis questions might include the following:

S1. How would you change building codes or zoning regulations in regions of volcanic activity to protect people and property?

S2. Plan an experiment to test if a landfill is polluting water from a nearby well.

Evaluation - Responses to evaluation questions use evidence and scientific reasoning to make judgments about facts, data, opinions or research results. Good answers require students to analyze and synthesize information and clarify ideas. Evaluation questions might ask a student to appraise, criticize, justify, or support an idea or concept. Examples of potential evaluation questions could include:

E1. Where is the greatest danger from an eruption of Mt. Shasta? Explain why.



Figure 3. Students' responses to conceptest questions are collated and tabulated using a classroom communication system. The system consists of transmitters (front left) that send signals to one or more receivers (center right) linked to a computer projection system.

E2. What is the most cost-efficient way to protect residents in a drainage basin from future flooding?

ASSESSMENT OF LEARNING

This section describes six methods of formative assessment aimed at recognizing and correcting misconceptions during lecture. Such learning tools can be assigned as in-class exercises or used by students outside of class in preparation for exams (Nuhfer, 1996). The assessment methods described below are keyed to Bloom's taxonomy in Table 1.

Concept Tests - Conceptests were developed as part of the peer instruction technique used to teach physics (Mazur, 1997). This teaching method has been widely used in physics courses at a range of institutions (Hake, 1998) and has been successfully adopted by faculty in a variety of other disciplines (e.g., chemistry, biology, astronomy; Crouch and Mazur, 2001). Peer instruction divides class time between short lectures and conceptual multiple-choice questions. Conceptest questions are designed to evaluate student understanding of the basic concepts behind the lecture material. Conceptests are not simple content-based multiple-choice (conceptest) questions that rely on the student re-reading their lecture notes or memorizing a fact or definition. Instead, these questions are designed to assess student understanding of the principal concepts underlying the lecture material. Conceptests generally correspond to the comprehension level of Bloom's taxonomy but may also be suitable for application questions (Table 1).

Students were given 30-60 seconds to consider a conceptest question and to choose an answer. We had previously provided large lettered answer cards that they would use to indicate their selection. This technique has been replaced by the use of an electronic personal

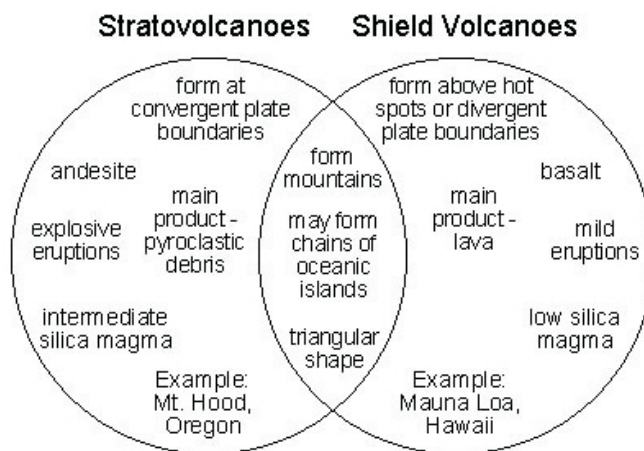


Figure 4. Venn diagrams can be used to compare and contrast the characteristics of related features. Students may be asked to complete a diagram in preparation for class, following a short reading assignment in class, or using lecture notes. Alternatively, students may be asked to locate a list of features in the correct place on the diagram.

response system (Figure 3) that registers student answers and generates a histogram of responses. This method has the advantage of providing students with a visual display of answers for the class while keeping individual answers anonymous. Furthermore, the technology provides students with the option of declaring their level of confidence in their answer choice. Following their initial answer, students are given 1-2 minutes to discuss the reasons for their choice with their neighbors (peer instruction) in pairs or small groups before voting again. This process usually results in an increase in the number of correct answers and an increase in student confidence in their answer choices (Mazur, 1997). Finally, a group spokesperson may be given an opportunity to provide a brief explanation of the group's answer and/or the instructor may clarify or expand on the correct response.

Venn Diagrams - Venn diagrams are a graphical method for comparing and contrasting features or phenomena. Such diagrams represent an opportunity for students to identify the characteristics of classification systems or to analyze the key components of complex sets of geological features. For example, students may be provided descriptions of the geological characteristics of two volcanoes and asked to compare and contrast their features using a Venn diagram (Figure 4). The use of Venn diagrams may involve knowledge, comprehension, application, and analysis levels of Bloom's taxonomy (Table 1). Examples of other possible comparisons are, igneous vs. sedimentary vs. metamorphic rocks, hurricanes vs. tornadoes, and convergent vs. divergent plate boundaries. Instructors may choose to provide a numbered list of characteristics that could then be placed in the correct locations on a labeled diagram. This assistance reduces the analysis aspect of the exercise, as students would not be identifying key components themselves. Such an

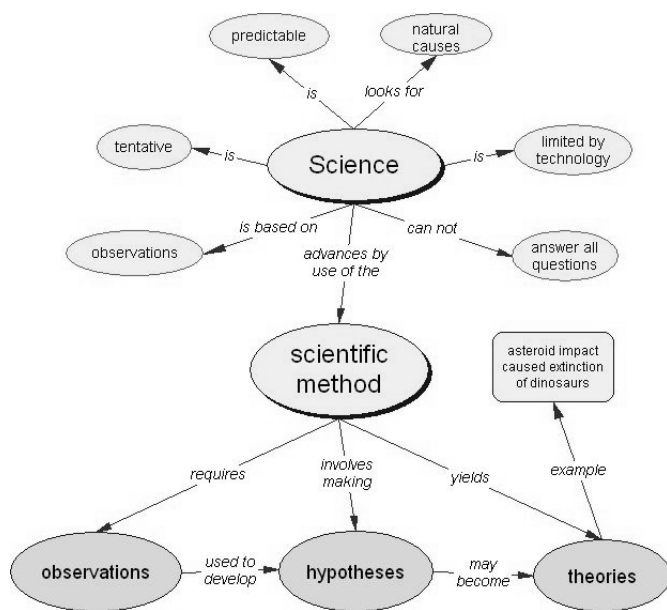


Figure 5. A simple concept map that illustrates the relationship between the elements of the scientific method. One potential scoring scheme would award 5 points per hierarchical level (5 levels present); 1 point for each reasonable linking phrase between adjacent points (12 links). Using this scheme the concept map would earn 37 points.

exercise would then be reclassified into the comprehension category.

IMAGE ANALYSIS

Image analysis is a form of slide observation (Reynolds and Peacock, 1998) where students are shown a photograph, map, or diagram and asked to make observations and interpretations. These types of exercises are an excellent way to begin a class as they immediately engage the student in the topic at hand. Image analysis involves knowledge, comprehension, application, and analysis levels of Bloom's taxonomy. Under certain circumstances exercises may also require students to synthesize and evaluate information. Readers are referred to Reynolds and Peacock (1998) for a thorough discussion of this technique.

CONCEPT MAPS

A concept map is a graphical representation of a student's knowledge about a topic (Zeilik et al., 1997). Concept maps are pictorial essays, a method of illustrating the principal concepts of a lesson, and include supporting information that indicates how a student has organized his/her ideas. Concept maps present a "big picture" view of a student's understanding of a topic. Good concept maps force their creators to challenge their own understanding and to build a strong foundation for information that follows. A

poorly constructed map allows a reviewer to quickly identify gaps in logic or comprehension. Concept maps will vary from person to person, no two are alike. They allow for creative thinking in their construction.

Concept maps have two principal components: 1. Terms or concepts - often presented in boxes; 2. Directional links (arrows) and linking phrases (prepositions) - that connect the terms (Figure 5). Concept maps identify the relationships between components and therefore correspond to synthesis in Bloom's taxonomy (Table 1). The number of levels in a concept map can be readily counted. The terms are joined by logical linking phrases appropriate for the topic. The maps can be readily evaluated as good, average, or poor to speed assessment. Alternatively, one can construct a formal scoring scheme (see caption, Figure 5).

Open-Ended Questions - Open-ended or divergent questions do not necessarily have a specific correct answer. Such questions can be written by the instructor to involve almost all levels of Bloom's taxonomy (Freedman, 1994). The creation of questions can serve as a method for promoting critical thinking among students. King (1995) used a series of generic question stems (Table 2) to prompt students to generate questions related to lecture and reading assignments. The question stems can be matched to specific levels of Bloom's taxonomy (Table 2). Student-generated questions could be used for self-examination, to assess comprehension of reading assignments, or in peer questioning exercises (King, 1995).

A form of open-ended question known as a minute paper is one of the most commonly utilized assessment methods in large classes. A minute paper is a short informal writing assignment that requires little time to complete and can be assessed easily (Angelo and Cross, 1993; Macdonald and Korinek, 1995; Murck, 1999). Students may be given literally one minute or a few minutes longer to complete the writing assignments. Minute papers can be used to determine whether students have grasped the key idea(s) presented during lecture. The papers may focus specifically on an important concept that students should understand but more commonly are the students' responses to the general question "What is the most important thing we discussed today?" This question challenges students to evaluate the lecture material and identifies whether they can discriminate between critical and routine information. Another technique known as a Muddiest Point exercise may start with the question, "What was the most confusing idea (muddiest point) presented in today's lecture?" (Angelo and Cross, 1993). Rather than asking students what they know, the focus may instead be on concepts they don't understand.

Assessment of answers following an in-class minute-paper or muddiest point exercise will indicate if student perceptions of lecture material corresponded to the instructor's lecture goals. Common misconceptions or gaps in comprehension should be addressed at the start of the next class period. Prompt feedback is a hallmark of good teaching (Chickering and Gamson, 1987; Angelo, 1993).

Evaluation Rubrics - Rubrics are used widely within society. When you complete a questionnaire that asks you to judge the quality of service in a restaurant you are

Bloom's Taxonomy	Question Stems
Knowledge	What is...?
Comprehension	What would happen if...? What does...illustrate about...? What is analogous to...?
Application	How could...be used to...? What is another example of...?
Analysis	How does...affect...? What are the differences (similarities) between...? What causes...?
Synthesis	What is a possible solution for the problem of...? How does...relate to what we learned before about...?
Evaluation	Why is...important? What is the best...and why? Do you agree/disagree that...?

Table 2. Critical thinking question stems, modified from King (1995).

Factors	Low Risk (1 pt)	Intermediate Risk (2pts)	High Risk (3 pts)
Proximity to fault	far (>200 km)	moderate (50-200 km)	close (<50 km)

Table 3. Template for earthquake risk evaluation rubric.

Factors	Low Risk (1 pt)	Intermediate Risk (2pts)	High Risk (3 pts)
Proximity to fault	far (>200 km)	moderate (50-200 km)	close (<50 km)
Time since last major earthquake	years	decades	centuries
Earthquake magnitude	small (<magnitude 4)	moderate (magnitude 4-5)	high (>magnitude 6)
Substrate	bedrock	rock and sediment mix	sediment
Utilization of building codes	all buildings built to code or retrofitted	building codes only partially enforced	no building codes

Table 4. Completed earthquake risk evaluation rubric. The values (distances, elevations, slopes) in the above example are arbitrary and are only intended to give an example of how quantitative data may be incorporated into a rubric.

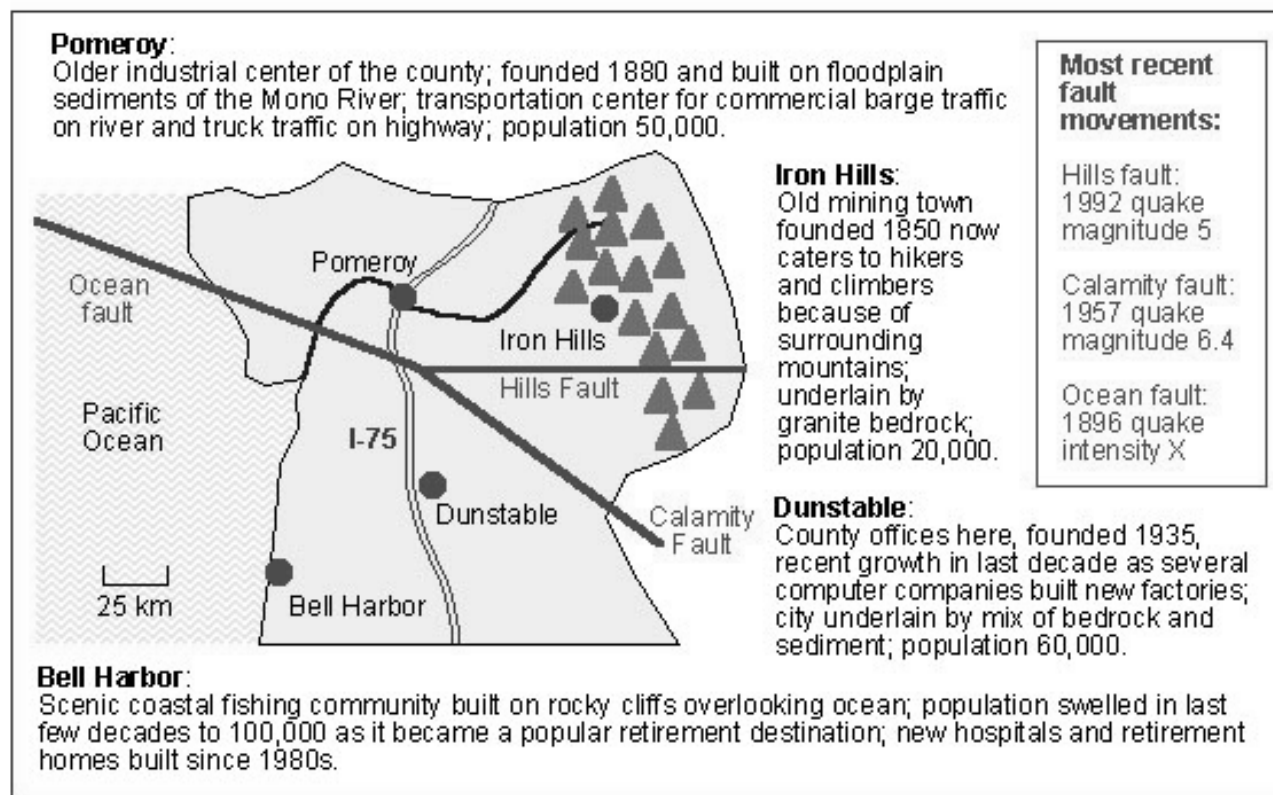


Figure 6. This idealized map of a county in California was used in the earthquake risk scoring rubric exercise. Students were asked to rank the four cities on this map in order of degree of risk of damage following a future earthquake.

using a rubric. When students judge the quality of teaching in a college class they often use a type of rubric. The relative scores on individual questions can be used to identify potential areas for improvement.

Scoring rubrics have traditionally been used by educators as assessment tools for student writing exercises. Rubrics provide a scoring scheme that can be keyed to specific performance goals and are especially useful for tasks where scoring could be subjective. The instructor compares each assignment with the standard of the rubric, ensuring a consistent scoring method. Rather than having students use an existing scoring rubric, we asked them to create their own rubrics for the purpose of evaluating specific geological situations. We termed these learning tools evaluation rubrics to differentiate them from the typical scoring rubric. Evaluation rubrics can involve all levels of Bloom's taxonomy (Table 1).

In this assessment method, students are required to generate their own rubrics. For example, students might create a rubric to assess the risk of an earthquake affecting a city (Tables 3, 4). We used two types of rubrics: 1. Rubrics with three scoring levels (Table 4); or, 2. Rubrics that required respondents to rank factors in order of significance.

Evaluation rubrics begin with a description of a specific situation. For example, the following instructions were given to students during a discussion of earthquakes. Students were asked to read the

instructions and complete a partially finished rubric (Table 3).

Following graduation you get a job working for a county planning task force in California. The task force must examine the setting of several different cities and identify which is at greatest risk for future earthquake damages from movements on known faults. You are given the assignment to create an evaluation rubric to assess factors that will influence the risk of potential damage from a future earthquake. The city that scores the highest using the rubric will receive additional county funds to protect key structures from earthquake damage.

Rubrics may be presented with one factor already identified to illustrate the scoring scheme. The quality of the student responses can be determined by the factors that are identified and the discrimination of the scoring methods (Table 4). A good rubric will identify several relevant factors and describe what constitutes a high or low score for each factor.

Students may be asked to distinguish which factor is the most important under the circumstances of the exercise. The score for this factor may be doubled. This requires making a judgment on the relative significance of the chosen factors. The final stage of a rubric exercise requires students to use their rubric in a hypothetical

situation. For example, students who had completed the earthquake risk rubric (Table 4) could be given information on the geology and characteristics of four cities (Figure 6) and asked to rank the cities in order of greatest to least risk of damage from a future earthquake.

ASSESSMENT OF TEACHING

The assessment methods described above have been used in several sections of an Earth Science course taught by two instructors (McConnell, Steer) over the last two years. Course enrollment ranged from 140-180 students per section. The majority (60-70%) of students were freshmen. The classes were taught in large auditorium-style classrooms with fixed seats facing a projection screen. Both instructors projected lecture materials using presentation programs such as Powerpoint, and had access to on-line materials through classroom internet connections.

The University of Akron is a large (22,000 students), open enrollment, state institution in northeast Ohio. The majority of students commute to class from surrounding communities. Students in an equivalent introductory geology course report that they work an average of 25 hours per week outside the University. Approximately a third of incoming freshmen do not return for the subsequent fall semester (UA Factbook, 2001). The student populations in Earth Science exhibit a broad range of skill levels. Students entering the University in Fall 2000 had an average ACT score of 20 and an average high-school GPA of 2.76. Fifty-eight percent of incoming students completed the college preparatory curriculum, in comparison to an average of 71% for the thirteen principal universities in Ohio (UA Factbook, 2001).

Most of the class sections discussed herein were offered in 50-minute blocks taught three-days a week at consistent times. Students were organized into informal groups made up of nearby students or permanent formal groups assigned by the instructor in all sections that employed active learning. Ideal group size was four students but groups varied from two to five students depending on attendance. For the purposes of this paper we will divide the classes into two types:

Traditional classes that followed a passive lecture format that did not involve groups and did not incorporate inquiry-based or active learning exercises during class;

Inquiry-based learning (IBL) sections that involved students working in groups, and the incorporation of active learning methods during lectures.

The contrast between these learning environments compares with teaching-centered and learning-centered classroom models (McManus, 2002). Steer gradually increased the degree of IBL material in his courses from only traditional lecture in early sections to incorporating daily exercises in the most recent versions of the course. During this investigation McConnell taught two consecutive sections using traditional methods in the first lecture and IBL techniques in the later class. One instructor (McConnell) has taught the course for twelve years whereas the other (Steer) was in his third year of

teaching. Material for both classes was divided into ten modules composed of three or four lectures each. The instructors shared many resources and identified common goals for the course, but did not necessarily use the same classroom materials or exercises.

Exams were either identical between sections or varied by a few questions due to slight differences in choice of material covered or pacing of the class sections. Exams were divided into three parts; knowledge questions, comprehension questions, and analysis questions. Equal numbers of knowledge and comprehension questions in a multiple-choice format were included on each exam and accounted for 80-90% of the total exam score. The remainder of the exam grade was from analysis questions that took several forms such as creating or completing concept maps, interpreting map data, open-ended questions, or drawing diagrams. Grading in the courses involved a combination of exam scores, homework assignments, and in-class exercises. Exams accounted for between 50-70% of the total course grade.

Peer Reviews of Teaching Methods - A colleague from the College of Education (Owens) visited the Earth Science classes on numerous occasions and collected field notes to record instructors' and students' behaviors. Table 5 compares and contrasts the classroom environments for the traditional and IBL sections of the course. In both the traditional and IBL classes students sat in groups of two to four people prior to the start of class. Students in the traditional class sat quietly or talked in hushed tones, in contrast, students in the inquiry class chatted among themselves and organized their group for the day's activities.

In the traditional class, the instructor did most of the talking about the topic of the day. Approximately 5% of the students participated when asked to give examples of phenomena or to respond to a question. Students paid attention but were passive receivers of the information. There was also some lecture in the IBL class, but it was often used to give instructions, to make transitions between one activity and the next, or to summarize the day's lesson. Students worked in groups to discuss and write responses to open-ended questions. During this time the noise level in the class increased dramatically, but a visitor listening to conversations would have discovered that students stayed on task. Conversations were peppered by technical terms but were characterized by less formal language and student idioms. After the group work ended, volunteer spokespersons were asked to report their groups' answers to the whole class. Frequently the students voted to decide a "best answer" as a way to come to an overall consensus. Sometimes the instructor presented a summary or asked the students to draw a concept map of the principal ideas of the lesson as a way to communicate their understanding.

The contrast in the methods was obvious to the College of Education observer. In the traditional class, only a handful of students participated in the discussion, whereas in the inquiry class, a substantial majority of students was actively engaged in class activities through group discussions. Essentially each class covered the same material, but the student involvement was much richer in the inquiry class.

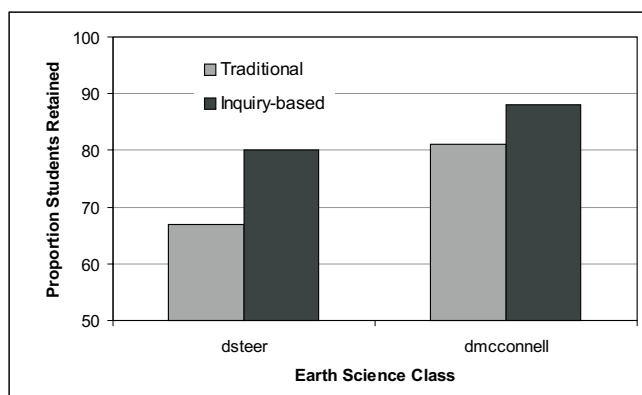


Figure 7. Student retention, the proportion of students present for the first exam who completed the last exam, was greater for the IBL classes than in traditional lecture classes. Gray bars for traditional lecture classes taught by Steer and McConnell black bars for inquiry-based learning classes.

Traditional	Inquiry-based Learning
Passive students	Active students
Quiet	Noisy
Instructor-focused	Student-focused
Information from instructor to student	Information from instructor to student and student to student
Students as individuals	Student collaboration
Competitive learning environment	Supportive learning environment
Limited assessment opportunities	Multiple assessment opportunities
Rigid setting (lack of mobility)	Mobile environment for instructor and student

Table 5. Traditional vs. IBL classroom characteristics.

STUDENT EVALUATIONS

The majority of student comments on the IBL teaching methods were positive. Written comments in student evaluations from McConnell's Fall 2000 IBL class were examined for references to in-class exercises. Forty-three students mentioned the exercises, 79% of the references were positive. Sample comments are included below.

... truthfully, this is the only class I made myself go to because it helped me learn in a more relaxed, interesting way.

The strongest part of the class is the group work. It helps you think about and understand the material.

The groups were an excellent learning tool. Students teaching students is the best way for them to learn!

Even the seemingly negative comments can sometimes be interpreted with a positive spin:

The in-class assignments were not clear...and we were expected to figure everything out for ourselves. The basic and overall outlook should be taught to understand Earth Science not to walk out as a scientist.

The answers should have been cut and dry and not up to our imagination.

STUDENT INTERVIEWS

Several students were randomly selected for semi-structured interviews that discussed course procedures in both traditional and IBL sections. All interviews were conducted by an assistant professor or graduate student from the College of Education. Students from all grade levels (A to F) were selected, but in at least one class, no "A" students were interviewed. Students reported that although they were initially skeptical, they preferred that the instructor chose to assign working groups and that they enjoyed getting to meet new people. They stated that the group arrangement took away the impersonal feeling of a large class, provided an opportunity to participate, gave students a peer to explain the material, and let them hear the opinions of others. Most students preferred the activities to a traditional lecture class.

Student comments on the use of the electronic personal response system (Figure 3) to answer concept test questions were universally positive. Using it gave the students an opportunity to test their understanding of course material, let them discuss their answers with others, and added vitality and interest. Students identified a strong link between class activities, the homework, and the tests. Responses to the question, "On a scale from 1 (take any other course but this) to 10 (don't miss this course), what number would you give this course?" ranged from 5 to 10 with 7 being the most common rating. Students enjoyed the participation resulting from the group work, admitted that they increased their knowledge, and would recommend the course to their peers.

Student Retention - We measured student retention by counting the number of students present for the first and last exams. Data from Steer's classes (Figure 7) showed a

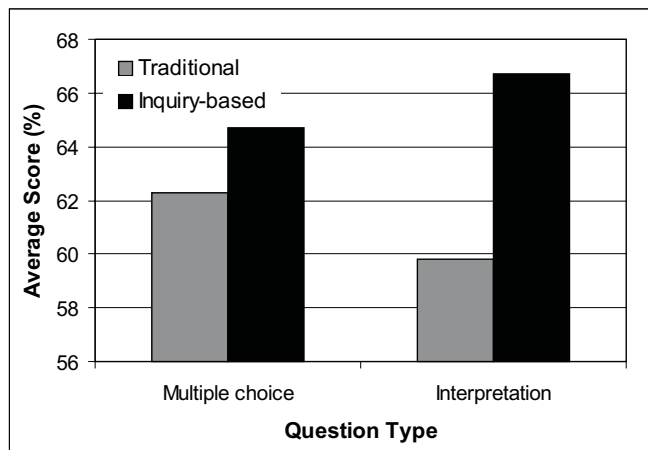


Figure 8. Average score on multiple-choice and interpretation questions on exams for traditional lecture (gray) and inquiry-based classes (black) taught by the same instructor during fall 2000. Average score was 2% higher for IBL section on multiple-choice questions even though this section received less time on content during each class period. Scores for interpretation questions were 7% higher in the inquiry-based section. Note: we attribute the relatively low scores on the multiple-choice questions to our first attempt at generating large numbers of the more challenging comprehension questions. Average scores in subsequent classes have increased by approximately 10%.

14% increase over previous years in the proportion of students who remained for the last exam as he added more IBL exercises to the course. McConnell had 8% greater retention in the IBL section in comparison to the traditional class (Figure 7) taught the same semester.

Exam Scores - Ensuring sufficient content coverage is a concern for many instructors when considering alternative teaching methods (Gold, 1988; Angelo, 1993; Ege et al., 1997). The traditional and IBL sections of McConnell's course took the same exams. Students in the IBL class slightly outperformed the traditional class on all four exams despite less direct content coverage during lecture in the IBL section (Figure 8). A more significant discrepancy was identified in the interpretation questions. Twelve short-answer interpretation questions that involved analysis, synthesis, or evaluation were distributed over the four exams. The average score on the on these questions was 7% greater in the IBL section (Figure 8).

GALT -The Group Assessment of Logical Thinking test (GALT; Roadrangka et al., 1982) is an assessment instrument that measures logical thinking skills. Higher-order thinking skills require mastery of logical operations such as proportional reasoning, controlling variables, probabilistic reasoning, combinational analysis, and correlational reasoning (Roadrangka et al.,

1982). The abbreviated form of the GALT survey contains twelve illustrated questions, a pair for each of the five logical operations listed above and another two that evaluate conservation. All questions, except those dealing with combinations, are presented in a multiple-choice format where students must select an appropriate answer (four choices) and the justification for the answer (four choices). The answer is considered wrong unless both choices are correct. The combination questions require that students identify potential groupings of different objects. Student GALT scores ranged from 1-12.

The GALT is a valid and reliable instrument for measuring logical thinking in student populations from sixth grade through college and consistently yields higher scores with increasing grade level (Roadrangka et al., 1982; Bitner, 1991; Mattheis et al., 1992). Furthermore, higher GALT scores correlate with other measures of academic achievement such as grades, SAT scores, and grade point average (Bunce and Hutchinson, 1993; Nicoll and Francisco, 2001). The GALT instrument was administered as a pre- and post-test to two IBL sections in Fall 2001. Both IBL sections showed a statistically significant 6.3% improvement in average GALT scores over the length of the semester. The same instrument showed no change in score for two traditional-format sections of Earth Science taught by different instructors.

SUMMARY

A variety of learning strategies were incorporated into large, introductory Earth Science courses for non-majors. A traditional lecture course was converted into an active learning environment through the incorporation of formative assessment methods matched to different levels of cognitive development. Such a conversion can be readily accomplished through a combination of short lecture segments and group assessment exercises. Improvements in student achievement on exams, retention in courses, and logical thinking skills were documented. A majority of students viewed the active learning methods positively.

ACKNOWLEDGMENTS

The authors thank Steve Boss, Rachel Teasdale, and Germaine Taggart for their helpful reviews of a draft of this paper. This material is based upon work supported by the National Science Foundation under Grant No. 0087894.

REFERENCES

- Allard, D.W., and Barman, C.R., 1994, The learning cycle as an alternative method for college science teaching, *Bioscience*, v.44, p. 99-101.
- American Geophysical Union, 1994, Report of the AGU Chapman Conference on scrutiny of undergraduate geoscience education; 55p.
- Anderson, L.W., and Krathwohl, D.R., 2001, *A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of educational objectives*. Longman.

- Anderson, L.W., and Sosniak, L.A., 1994, Bloom's Taxonomy: A forty-year retrospective, National Society for the Study of Education.
- Angelo, T.A., 1993, A "Teacher's dozen", American Association of Higher Education Bulletin, April, p. 3-13.
- Angelo, T. A., and Cross, K. P., 1993, Classroom assessment techniques, A handbook for college teachers (2nd ed.). San Francisco: Jossey Bass.
- Astin, A.W., 1993, What matters in college? Four critical years revisited, Jossey-Bass, 482p.
- Bitner, B.L., 1991, Formal operational reasoning modes: Predictors of critical thinking abilities and grades assigned by teachers in science and mathematics for students in grades nine through twelve, Journal of Research in Science Teaching, v.28, p. 265-274.
- Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R., 1956, Taxonomy of educational objectives: Handbook 1: Cognitive domain, David McKay.
- Brunkhorst, B.J., 1996, Assessing student learning in undergraduate geology courses by correlating assessment with what we want to teach, Journal of Geoscience Education, v. 44, p.373-378.
- Bunce, D.M., and Hutchinson, K.D., 1993, The use of the GALT (Group Assessment of Logical Thinking) as a predictor of academic success in college chemistry, Journal of Chemical Education, v.70, p. 183-187.
- Bykerk-Kauffman, A., 1995, Using cooperative learning in college geology classes, Journal of Geological Education, v.43, p. 309-316.
- Chickering, A.W., and Gamson, Z.F., 1987, Seven principles for good practice in undergraduate education, American Association of Higher Education Bulletin, March, p. 3-7.
- Collins, A., 1997, National science education standards and the college instructor. In Siebert, E.D., Caprio, M.W., and Lyda, C.M., (Eds.), Methods of effective teaching and course management for university and college science teachers, Dubuque, Iowa; Kendall/Hunt, p. 51-60.
- Crouch, C.H., and Mazur, E., 2001, Peer Instruction: Ten years of experience and results, American Journal of Physics, v. 69, p. 970-977.
- DeCaprariis, P.P., 1997, Impediments to providing scientific literacy to students in introductory survey courses, Journal of Geoscience Education, v.45, p. 207-210.
- Ebert-May, D., Brewer, C., and Allred, S., 1997, Innovation in large lectures - teaching for active learning, Bioscience, v.47, p. 601-607.
- Ege, S.N., B.P. Coppola, and R. Lawton, G., 1997, The University of Michigan Undergraduate Chemistry Curriculum: 1. Philosophy, Curriculum, and the Nature of Change, Journal of Chemical Education, v. 74, p. 74-83.
- Freedman, R.L.H., 1994, Open-ended questioning: A handbook for educators, Menlo Park, CA: Addison-Wesley.
- Fuhrman, M., 1996, Developing good multiple-choice tests and test questions, Journal of Geoscience Education, v. 44, p. 379-384.
- Gibbons, A., 1994, A time of trials and tribulations, Science, v.266, p. 844, 845, and 851.
- Gold, M., 1988, Chemical education: An obsession with content, Journal of Chemical Education, v. 65, p. 780-781.
- Hake, R.R., 1998, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, American Journal of Physics, v. 66, p. 64-74.
- King, A., 1995, Designing the instructional process to enhance critical thinking across the curriculum, Teaching of Psychology, v.22, p. 13-17.
- Lord, T.R., 2001, 101 reasons for using cooperative learning in biology teaching, American Biology Teacher, v.63, p.30-103.
- Mazur, E., 1997, Peer instruction: A user's manual, Prentice Hall, 253p.
- Macdonald, R.H. and Korinek, L., 1995, Cooperative-learning activities in large entry-level geology courses, Journal of Geological Education, v. 43, p. 341-345.
- Mattheis, F.E., Spooner, W.E., Coble, C.R., Takemura, S., Matsumoto, S., Matsumoto, K., and Yoshida, A., 1992, A study of the logical thinking skills and integrated process skills of junior high school students in North Carolina and Japan, Science Education, v.76, p. 211-222.
- McManus, D.A., 2002, The two paradigms of education and the peer review of teaching, Journal of Geoscience Education, v.49, p. 423-434.
- Murck, B., 1999, Promoting and evaluating participation in high-enrollment undergraduate courses, Journal of Geoscience Education, v. 47, p. 131-134.
- National Research Council, 1997, Science teaching reconsidered, National Academy Press, 88p.
- National Research Council, 2000, Inquiry and the national science education standards, Washington, DC, National Academy Press.
- National Science Foundation, 1996, Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology, 76p.
- Nicoll, G., and Francisco, J.S., 2001, An investigation of the factors influencing student performance in physical chemistry, Journal of Chemical Education, v.78, p. 99-102.
- Nuhfer, E.B., 1996, The place of formative evaluations in assessment and ways to reap their benefits, Journal of Geoscience Education, v. 44, p. 385-394.
- Pinet, P.R., 1995, Rediscovering geologic principles by collaborative learning, Journal of Geological Education, v.43, p. 371-376.
- Prothero, W.A., Jr., 2000, Keeping our focus; a perspective on distance learning and the large introductory science class, Computers in Geosciences, v. 26, p. 647-655.
- Roadrangka, V., Yeany, R., and Padilla, M., 1982, GALT, Group Test of Logical Thinking, Athens, Georgia: University of Georgia.
- Reynolds, S.J., and Peacock, S.M., 1998, Slide observations - Promoting active learning, landscape appreciation, and critical thinking in introductory geology courses, Journal of Geoscience Education, v.46, p. 421-426.

- Siebert, E.D., and McIntosh, W.J., 2001, College pathways to the science education standards, National Science Teachers Association Press, 192p.
- Silberman, M. 1996, Active Learning: 101 Strategies to Teach Any Subject.
- Sundburg, M.D., Dini, M.L., and Li, E., 1994, Decreasing course content improves student comprehension of science and attitudes towards science in freshman biology, *Journal of Research in Science Teaching*, v.31, p. 679-693.
- Tobias, S., 1990, They're not dumb, they're different: Stalking the second tier, Research Corporation, 94 p.
- Tobias, S., 1992, Revitalizing undergraduate education: Why some things work and most don't, Research Corporation, 192p.
- Trice, A.G. and Dey, E.L., 1997, Trends in faculty teaching goals: A longitudinal study of change, *Journal of College Student Development*, v. 38, p. 527-534.
- University of Akron Factbook, 2001 <http://www.uakron.edu/vpcio/iplan/factbook/>
- Wyckoff, S., 2001, Changing the culture of undergraduate science teaching, *Journal of College Science Teaching*: v.30, p. 306-312.
- Zeilik, M, Schau, C., Mattern, N., Hall, S., Teague, K. and Bisard, W., 1997, Conceptual astronomy: A novel model for teaching post-secondary science courses, *American Journal of Physics*, v. 65, p. 987-996.
-

**Assessment of Learning in Entry-Level Geoscience Courses: Results from the Geoscience
Concept Inventory**

Julie C. Libarkin, Dept. of Geological Sciences, 316 Clippinger Labs, Ohio University, Athens,
OH 45701; libarkin@ohio.edu; phone: 740-593-1109; fax: 740-593-0486

Steven W. Anderson, Science Department, Black Hills State University, Spearfish, SD 57799-
9102; steveanderson@bhsu.edu

Abstract (148 words; 250 allowed)

Assessment of learning in entry-level college science courses is of interest to a wide variety of faculty, administrators, and policy-makers. The question of student preparedness for college instruction, as well as the effect of instruction on student ideas, has prompted a wide range of qualitative and quantitative studies across disciplines. In the geosciences, faculty are just beginning to become aware of the importance of conceptual change in instruction. The development of the Geoscience Concept Inventory (GCI) and application to the study of learning in entry-level geoscience courses provides a common framework from which faculty can evaluate learning and teaching effectiveness. In a study of 43 courses and 2500 students, we find that students are entering geoscience courses with alternative conceptions, and in many cases are leaving the classroom with these alternative ideas intact. We find no relationship between self-reported teaching style and learning as measured by the GCI.

Keywords: Education-Science, Education-Undergraduate, Geoscience-Teaching and Curriculum, Assessment

INTRODUCTION

Learning is the goal of all instruction. Accurate assessment of learning is an important first step in determining the links between learning and teaching, and ultimately in developing instructional approaches that are effective and transferable to other classrooms and institutions. Some disciplines, primarily physics and math, have made significant headway into unraveling the complex relationships between learning and teaching, often through the application of learning research pioneered by people like Piaget and Driver (e.g., Redish, 1994). Ultimately

these efforts strive to determine how people learn, factors that can influence learning, and innovations to the teaching environment that can improve learning for all participants. Significant effort has been made to disseminate effective teaching methods for use in college level geosciences courses (e.g., Digital Library for Earth System Education), although quantitative assessment research documenting this effectiveness has been slower to evolve.

Quantitative assessment instruments for college classrooms have been used in a variety of scientific disciplines, particularly for the evaluation of attitudes or conceptual understanding (e.g., Hestenes et al., 1992; Zeilik et al., 1999; Libarkin, 2001; Yeo and Zadnick, 2001; Anderson et al., 2002). The development of the Geoscience Concept Inventory (GCI) is a first step in determining how entry-level college courses are affecting our students and in identifying factors that influence learning. The GCI is a set of conceptually based questions geared towards fundamental concepts in the Earth Sciences, including foundational concepts in physics and chemistry. The GCI was developed over a two-year period; to date, 73 questions have been evaluated and validated using item analysis techniques from both classical test theory and item response theory, particularly Rasch analysis (Libarkin and Anderson, in preparation). We report here on the analyses of pre- and post-test results from 29 GCI questions administered to ~2500 students enrolled 43 courses across the United States. These GCI questions covered concepts related to geologic time, plate tectonics, and the Earth's interior.

Previous Research

Assessment of learning in the geosciences has traditionally focused on K-12 students, with studies of college students or other adults only recently emerging (DeLaughter et al., 1998; Trend, 2000; Libarkin, 2001; Libarkin et al., in press; Dahl et al., in press). Qualitative studies

are concentrated outside of the U.S. (e.g., Happs, 1984; Marques and Thompson, 1997; Trend, 2000; Dodick and Orion, 2003), with those of American students focusing primarily on pre-college populations (Schoon, 1992; Gobert and Clement, 1999; Gobert, 2000). Existing quantitative studies have dealt with attitudes (Libarkin, 2001), visualization (e.g. Hall-Wallace and McAuliffe, 2002), and logical thinking skills (McConnell et al., 2003). Quantitative study of student conceptual understanding in the geosciences lags far behind other disciplines.

The development of the Force Concept Inventory (FCI; Hestenes et al., 1992) in the early 1990's dramatically changed the way physicists viewed teaching and learning in college level physics courses. A sharp increase in studies related to conceptual change in college-level physics (see Kurdziel and Libarkin, 2001 for a discussion) has led to significant changes in physics instruction, as well as a new perspective of the importance of physics education research in academic physics (e.g., Gonzales-Espada, 2003). Subsequent development of quantitative instruments in other disciplines followed, including development in biology (Anderson, 2002), physics (Yeo and Zadnick, 2001), astronomy (Zeilik et al., 1999), and now, the geosciences (Libarkin and Anderson, in preparation).

METHODS

Design and Procedure

The GCI was developed over several years, with item generation and validation based upon a variety of qualitative and quantitative data (Libarkin et al., in press; Libarkin and Anderson, in preparation). Determination of reliability and validity of test items evolved through qualitative means, such as validation by experts in geosciences and education, and through quantitative evaluation of student test data. Test data were analyzed using classical test theory approaches,

particularly item analysis, and through item response theory using simple Rasch models. The Rasch analysis resulted in development of a test scale that allows scaling of raw test scores to more meaningful scaled scores, and also provided information on item discrimination. One test question (of 29 original questions) was removed from the analysis based upon gender discrimination in both the pre- and post-test data. Although several questions were modified between the pre- and post-test administration based upon analytical results and expert feedback, the ordering of these questions on the Rasch scale did not change significantly, and we concluded that item revision did not dramatically impact our ability to compare pre- and post-test results (Libarkin and Anderson, in preparation).

The 29 GCI questions were distributed as two test versions of 20 questions each, with 11 common questions and 9 version-specific questions. Tests were randomly distributed to courses, with each version administered to roughly half of the students. One institution administered the test via computer and used only one version; one course from this institution also post-tested with the same version. Analysis of test data from all institutions indicated that the two versions were of similar difficulty, producing nearly identical Rasch scaling functions for conversion of raw to scaled scores.

GCI data were collected in Fall 2002 from 43 courses at 32 institutions located in 22 states across the U.S. (Fig. 1; map of U.S. with institutions marked). Tested courses were introductory level, and included physical geology, oceanography, environmental science, historical geology, and specialty topic courses. Faculty from 21 public and six private four-year institutions, four community colleges or two-year institutions, and one tribal college participated (Table 1). Individual classes ranged from nine to 210 students, with most courses falling between 35 and 75 students. 2500 students were pre-tested at the beginning of the Fall 2002 semester, and a subset

of 1295 students from 30 courses were post-tested at the end of the semester. In addition, matched pre- and post-test results from 930 individual students were obtained and compared.

Instructors of post-tested courses used a variety of teaching methods including lecture, demonstration, whole class discussions, small group activities, laboratory exercises, and technology. Each instructor in the study provided their estimated breakdown of the time spent on each of these instructional strategies, and we have made an initial comparison that relates teaching style to changes in pre- to post-test results on the GCI. Teaching approaches varied greatly, such that the reported percentage of class time devoted to lecture ranged from 0-100%, demonstration ranged from 0-30%, small group work ranged from 0-50%, lab exercises ranged from 0-60%, and use of technology ranged from 0-100%. Faculty self-reporting of teaching approaches is probably less accurate than direct classroom observation (e.g., Johnson and Roellke, 1999), although our large data set prohibited direct observation of all studied courses.

Data Analysis

Developers of multiple-choice instruments for higher education generally perform classical item analysis on test results (e.g., Hestenes et al., 1992; Anderson et al., 2002). Item analysis is primarily used to observe the statistical characteristics of particular questions and determine which items are appropriate for inclusion on a final instrument. Classical Test Theory generally drives most item analysis, with focus on item difficulty and item discrimination, and thus item characteristics are tied closely to the population sampled. Item Response Theory (IRT), an alternative item analysis technique, assumes that the characteristics of a specific item are independent of the ability of the test subjects. IRT at its foundations is the study of test and item scores based upon assumed relationships between the trait being studied (i.e. conceptual

understanding of geosciences) and item responses. Most researchers would agree that items on any test are generally not of equal difficulty, and in fact most published concept tests report "item difficulty", defined by the % of participants answering a specific item correctly. For example, Anderson et al. (2002) present a 20-item test on natural selection, with item difficulties ranging from 13-77%. In addition, discriminability reported for these items suggests a strong correlation between the difficulty of items and the overall score achieved by a student. This suggests, then, that some items are easier to answer than others. Because difficulty ranges so widely on this and most concept tests, the question of linearity must be addressed. Linearity implies that conceptual understanding is linearly correlated with raw test scores; a student answering 1/3 of items correctly has exactly half the understanding of a student answering 2/3 correctly.

Equivalent changes in raw score for multiple students may not translate to equivalent changes in conceptual understanding. Item response theory implies that not all test items are created equal, and some items will be more difficult than others. Rather than calculate a raw test score that simply reflects the number of "correct" responses, IRT allows for score scaling that more accurately reflects the difficulty of a given set of test items. Using a statistically calculated IRT scale to offset the assumption of scale linearity allows the determination of test scores that more accurately reflect "understanding". All raw GCI test results were scaled on a 0-100% scale based upon a simple IRT approach (Rasch analysis), following the methodology presented by Libarkin and Anderson (in preparation). The relationship between raw score and scaled Rasch score, as fit by the statistical package JMP, is approximately:

$$S = 3.9 + 9R - 0.71R^2 + 0.025R^3 \quad (1)$$

where S is the scaled score on a 0-100% scale and R is the raw score on a 19-item GCI.

Pre- and post-test results were then compared using simple t-tests; this comparison was conducted for the entire population of students as well as sub-groups categorized from demographic or course information. All t-tests were two-tailed and based upon $p < 0.05$, with some courses passing at the $p < 0.001$ level.

RESULTS and DISCUSSION

These data provide a unique opportunity to evaluate the pre-course conceptual frameworks of students enrolled in geoscience courses nationwide. In addition, evaluation of test data relative to course factors such as class size, institutional type, and faculty instructional approaches provides insight into the effectiveness of entry-level geoscience courses nationwide. Finally, preliminary evaluation of these data indicates that some ideas are stable across instruction, suggesting a until now unknown entrenchment of ideas (Anderson and Libarkin, 2003).

Overall, students found the test difficult, with nearly identical pre-test means of 41.5 ± 12 (43 courses; $n=2493$ students) and 42.2 ± 12 (for the 29 courses post-testing, where course 41 could not be included; $n=1498$ students). The post-test results suggest that the population of post-testing students experienced minimal learning over the course of the semester, with a mean of 45.8 ± 13 ($n=1295$ students; Fig. 2a). Results are most illuminating when pre and post-test scores are matched for individual students; in this case, 930 pre- and post-tests were matched. The pre-test mean for students with matched post-tests ($n=930$ students; 43 ± 11) is similar to all pre-tests; the matched post-test mean is also similar to overall results (47 ± 12). With the exception of one

course containing 9 students, students on average were familiar with only half of the conceptions covered by this test after completing these courses.

Comparison of matched pre- and post-tests (Fig. 2b) indicates that statistically significant improvement occurred on the post-tests, as shown on a paired, two-tailed t-test ($t_{\text{stat}} = 1.96 < t_{\text{crit}} = 12.1$). Analysis of sub-population effects indicates that students with low pre-tests (<40%, $n = 388$) dominate this effect, with a pre-test mean for this group of 32 ± 7 and a post-test mean of 41 ± 10 , and extreme significance on a t-test ($t_{\text{stat}} = 1.96 < t_{\text{crit}} = 15$). Students with intermediate scores (40-60%, $n = 489$ students) exhibited a minimal change in GCI score, with a pre-test mean of 48 ± 5 , a post-test of 50 ± 10 , and minimal significance on a t-test ($t_{\text{stat}} = 1.96 < t_{\text{crit}} = 3.5$). Students pre-testing >60% ($n = 52$ students) exhibited no change in pre- to post-test scores (average score on both tests was 67%). These data suggest that students with minimal knowledge at the beginning of an entry-level geology course are leaving with increased conceptual understanding, while students with intermediate and advanced understanding are leaving, as a population, with mixed effects. Those students with pre-test scores that are higher than their post-test scores may be using instruction to reinforce non-scientific conceptions. Interview data supports this hypothesis, suggesting that some students apply instruction in one area of geosciences to other areas. For example, the notion of Pangea is used by students to describe the Earth's surface at many different times in the past (Libarkin et al., in press); students reinforce this idea by explaining that plate tectonics causes the continents to move. Further evaluation of the underlying causes of decreasing or increasing GCI scores is needed.

Example courses

Three courses have been chosen as representative samples of the courses tested. These include courses from different types of institutions, as well as courses of differing size that reflect the instructional strategies reported by participating faculty. Of the thirty post-tested courses, only 8 showed significance on a t-test. Overall raw scores indicate that the average student gained one question on the post-test, moving from 8 correct questions to 9 out of 19.

Course 19 is a representative small course taught at a public school in the south. The instructor of this course reported using traditional lecture and laboratory pedagogical approaches, with some alternative methods. The pre-course GCI average was 47 ± 13 ($n=11$), with a post-test score of 43 ± 13 ($n=9$; Fig. 3a). Eight students were matched on pre- and post-tests; analysis of these matched tests indicates static GCI scores (Fig. 3b). This suggests that student conceptions of the content covered by the GCI questions used in this study did not change as a result of instruction.

Course 3 is a representative intermediate course taught at a public school in the mid-west. The instructor of this course reported a predominantly lecture and in-class discussion approach to teaching. The pre-course GCI average was 46 ± 12 ($n=42$), with a post-test score of 49 ± 13 ($n=38$; Fig. 4a). Matched pre- and post-tests for 28 students indicates that nearly all students experienced conceptual gain after instruction (Fig. 4b). Gains were between one and two questions per student.

Course 12 is a representative large course taught at a public institution in the north-central U.S. The instructor of this course reported lecturing 100% of the time, using a traditional approach. The pre-course GCI average was 38 ± 11 ($n=190$), with a post-test score of 42 ± 12 ($n=183$; Fig. 5a); the pre-course average for this class was much lower than the small and intermediate courses shown here. 135 students had matching pre- and post-tests; analysis of these

matched tests indicates that course effects varied, although the majority of students experienced static or positive gains (Fig. 5b). As with the overall study population (Fig. 2), those students who entered the course with pre-tests less than 40% experienced statistically significant positive gains.

Entrenchment of ideas

The student population retained several alternative ideas over the course of the semester, with remarkably consistent results across institutions (Anderson and Libarkin, 2003; Table 3). The extreme persistence of some ideas suggests that current approaches to instruction, either traditional or alternative, may not be adequate for engendering conceptual change. In particular, students have a poor idea of the scale of geologic time, the occurrence of events in geologic history, and the specifics of absolute age dating. Not surprisingly, students also ascribe a Pangea-like supercontinent to many different times in the past, including the time of Earth's formation, and as noted here, at the appearance of humans. Although entry-level geoscience textbooks universally discuss the Theory of Plate Tectonics and most faculty spend significant time on this topic, most students are exiting courses with a poor understanding of the location of tectonic plates. Previous research utilizing qualitative approaches is in agreement with these data (Libarkin et al., in press).

CONCLUSION AND IMPLICATIONS

The diverse data set collected in this study allows for a unique glimpse into entry-level geoscience courses being taught nationwide. Most notably, students are entering these courses with prior experiences in Earth Science and alternative conceptions about geologic phenomena.

Post-instructional gains in understanding at the college level are generally small, with most students exiting courses with conceptions similar to those held prior to instruction. As an exception, those students entering college geoscience courses with little familiarity or significant misconceptions (and, thus, low GCI test scores) experience significant gain across all courses and institutions, regardless of instructional approaches. This is similar to findings in physics tested with the FCI (Pollock, 2004) and with student attitudes (Libarkin, 2001); most likely, these students are simply “catching up” with their peers.

Although the geoscience community has spent significant time and energy developing and disseminating alternative instructional strategies for use in college-level classrooms, the limited conceptual gain experienced by students suggests that a different curriculum-development approach is warranted. In particular, the effects of curriculum and pedagogy on student conceptual understanding, as well as the mechanisms for conceptual change in college-level geosciences, need to be studied in detail. Qualitative and quantitative research approaches have the potential to unravel the complex relationships between teaching and learning, and implementation of research approaches into the curriculum development-testing-dissemination cycle may result in significant modification in the way faculty view entry-level instruction. Certainly, further research in all realms of conceptual change in the geosciences is needed, with potential benefits to students and faculty alike.

Acknowledgments. We thank all students and faculty who graciously participated in this research. Michaela Brewster, James Vinton, and Saunia Withers are thanked for their assistance with data entry. This study was funded by the National Science Foundation through an Assessment of Student Achievement in Undergraduate Education grant to Libarkin and

Anderson (DUE-0127765; DUE-0350395) and a Postdoctoral Fellowship in Science, Mathematics, Engineering, and Technology Education for Libarkin (DGE-9906478).

REFERENCES

- Anderson, D.L., Fisher, K.M., and Norman, G.J., 2002, Development and validation of the conceptual inventory of natural selection: *Journal of Research in Science Teaching*, v. 39, p. 952-978.
- Anderson, S., and Libarkin, J., 2003, The retention of geologic misconceptions: Alternative ideas that persist after instruction: *Eos Trans.-Fall Meeting Supplement* v. 84.
- Dahl, J., Anderson, S.W., and Libarkin, J.C., Digging into Earth Science: Alternative conceptions held by K-12 teachers: *Journal of Science Education*, *in press*.
- DeLaughter, J.E., Stein, S., and Stein, C.A., 1998, Preconceptions abound among students in an Introductory Earth Science Course: *EOS*, v. 79, p. 429-432.
- Dodick, J., and Orion, n., 2003, Cognitive factors affecting student understanding of geologic time: *Journal of Research in Science Teaching*, v 40, p. 415-442.
- 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.
- . Gobert, J.D., and Clement, J.J., 1999, Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics: *Journal of Research in Science Teaching*, v. 36, p. 39-53.
- Gonzales-Espada, W.J., 2003, Physics education research in the United States: A summary of its rationale and main findings: *Revista de Educacion en Ciencias*, v. 4, p. 5-7.
- Hall-Wallace, M.K., and McAuliffe, C.M, 2002, Design, implementation, and evaluation of GIS-based learning materials in an introductory geoscience course: *Journal of Geoscience Education*, v. 50, p. 5-14.

- Happs, J.C., 1984, Soil generation and development: view held by New Zealand students: *Journal of Geography*, v. 83, p. 177-180.
- Hestenes, D., Wells, M., and Swackhamer, G., 1992, Force Concept Inventory: The Physics Teacher, v. 30, p. 141-158.
- Johnson, S.D., and Roellke, C.F., 1999, Secondary teachers' and undergraduate Education faculty members' perceptions of teaching effectiveness criteria: A national survey: *Communication Education*, v. 48, p. 127-38.
- Kurdziel, J., and Libarkin, J.C., 2001, Research Methodologies in Science Education: Assessing Students' Alternative Conceptions, *Journal of Geoscience Education*, v. 49, p. 378-383.
- Libarkin, J.C., 2001, Development of an assessment of student conception of the nature of science: *Journal of Geoscience Education*, v. 49, p. 435-442.
- Libarkin, J.C., and Anderson, S.W., Science concept inventory development in higher education: A mixed-methods approach in the geosciences: *Journal of Research in Science Teaching*, *in preparation*.
- Libarkin, J.C., Anderson, S.W., Dahl, J., S., Beilfuss, M., Boone, W., and Kurdziel, J., Qualitative analysis of college students' ideas about the Earth: Interviews and open-ended questionnaires: *Journal of Geoscience Education*, *in press*.
- 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.
- 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 and Technological Education*, v. 15, p. 195-222.

- Pollock, S.J., 2004, No single cause: Learning gains, student attitudes, and the impacts of multiple effective reforms: Paper presented at Physics Education Research Conference, summer 2004.
- Redish, E. F. 1994, "Implications of Cognitive Studies for Teaching Physics," American Journal of Physics, v. 62, p. 796.
- Schoon, K.J., 1992, Students' alternative conceptions of Earth and space: Journal of Geological Education, v. 40, p. 209-214
- 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.
- Yeo, S., and Zadnick, M., 2001, Introductory thermal concept evaluation: Assessing Students' Understanding: The Physics Teacher, v. 39, p. 496-503.
- Zeilik, M., Schau, C., and Mattern, N., 1999, Conceptual astronomy. II. Replicating conceptual gain, probing attitude changes across three semesters: American Journal of Physics, v. 67, p. 923-927.

Table 1. Sample size, recruitment, and institutional setting.

Institutional Type	Number of schools	Number of courses	Course size (n students)
Four-year public	21	31	11 to 190
Four-year private	6	6	13 to 91
Two-year community college	4	6	15 to 82
Two-year tribal college	1	1	9

Table 2. Courses participating in GCI testing in Fall 2002.

Institution and Type	Course Code	Pre-test (n)	Post-test (n)	Course Type	Instructional Methods (%)					
					L	D	IC	A	T	LAB
A, Public	1	69	25	Physical Geology	70	30				X
A, Public	2	38	26	Historical Geology	100					
B, Public	3	42	38	Geology	50	<10	30	<10	<10	
C, Public	4	24	23	Physical Geology	60	30	<10			X
D, Public	5	81	67	Marine Science	50		50			
E, Public	6	29	25	Physical Geology	50	<10	<10	50	<10	
F, Public	7-1	57*	36	Geology	80	<10	<10			X
G, Public	7-2	---	16	Geology-online					100	X
H, Public, 2-yr	8	28	25	Physical Geology	60	<10	<10		<10	X
H, Public, 2-yr	9	21	13	Physical Geology	60	<10	<10		<10	X
I, Public	10	86	39	Unknown						
J, Public	11	108	85	Physical Geology	80		20			
K, Public	12	190	183	Earth Science	100					
L, Public	13	129	107	Physical Geology	60	<10	<10			X
M, Private	14	13	---	---	---	---	---	---	---	---
N, Public	15	40	---	---	---	---	---	---	---	---
O, Private	16	58	---	---	---	---	---	---	---	---
P, Public	17	15	16	Physical Geology	60		<10			X
P, Public	18	120	75	Unknown						
P, Public	19	11	9	Historical Geology	60	<10	<10		<10	X
Q, Navajo, 2-yr	20	9	6	Historical Geology	50	30	<10		<10	X
R, Public	21	67	57	Earth Systems	80		<10	<10		X
S, Public	22	50	54	Geology for Engineers	60	<10	<20		15	X
T, Public	23	18	17	Geology of National Parks	90	<10	<10		<10	X
U, Public, 2-yr	24	82	56	Physical Geology	65	15	10		10	X
V, Private	25	91	---	---	---	---	---	---	---	---
W, Private	26	54	---	---	---	---	---	---	---	---
X, Public	27	59	---	---	---	---	---	---	---	---
Y, Private	28-1	37*	19	Earth History	70	10	5		15	
Y, Private	28-2	---	14	Oceanography	70	5	5		20	X
Z, Private	29	24	22	Geology and Environment	45	10	40		5	X
AA, Public	30	69	---	---	---	---	---	---	---	---
BB, Public	31	31	22	Oceanography	45	5	15	25	10	X
CC, Public, 2-yr	32	24	---	---	---	---	---	---	---	---
DD, Public, 2-yr	33	39	---	---	---	---	---	---	---	---
EE, Public	34	18	---	---	---	---	---	---	---	---
FF, Public	35	75	55	Unknown	---	---	---	---	---	---
FF, Public	36	41	32	Unknown	---	---	---	---	---	X
GG, Public, 2-yr	37	15	13	Hydrogeology	90	10				X
HH, Public	38	21	---	---	---	---	---	---	---	---
II, Public	39	97	---	---	---	---	---	---	---	---
JJ, Public	40	128	---	---	---	---	---	---	---	---
KK, Public	41**	269	120	Physical Geology	---	---	---	---	---	---

*Pre-test results for courses 7-1 and 7-2 and courses 28-1 and 28-2 were combined.

**Course 41 was actually three courses that tested students via computer; pre-test results were not distinguishable by course, although only one course post-tested. L = lecture; D = demonstrations; IC = in-class discussions; A = small group activities; T = technology

Table 3. Prevalent ideas and persistence after instruction.

Topic	Conception*	Prior to Instruction	After instruction
Techniques for Calculating Earth's Age	Analyses of fossils, rock layers, or carbon are the most accurate means for calculating the Earth's age	78% (n =1377)	72% (n =669)
Location of Tectonic Plates	The Earth's surface is not the top of the tectonic plates; tectonic plates are located beneath the Earth's surface.	56% (n =2483)	46% (n=1287)
Earth's surface when humans appeared	A single continent existed when humans first appeared on Earth.	52% (n=2470)	47% (n=1284)
Life at Earth's formation	Simple, one-celled organisms existed when the Earth first formed	47% (n=2481)	43% (n=1286)
Appearance of dinosaurs	Dinosaurs came into existence about halfway through geologic time.	37% (n=1089)	40% (n=604)

*Students in the study population who did not exhibit these conceptions often held other alternative conceptions.

Figure 1. Map of the continental United States. Numbers indicate number of institutions in each state participating in this study.

Figure 2. **a)** Distribution of scaled scores for all pre- ($n = 2493$ students) and post-tests ($n = 1295$ students). The lowest individual score was 0; the highest was 100. **b)** Matched pre and post-tests for individuals. The gray line represents the zone of no change; points falling along this line represent identical pre- and post-test scores. Points falling above the line indicate an increase in score from pre- to post-test and points falling below the line indicate a decrease from pre- to post-test.

Figure 3. Course 19. **a)** Pre ($n = 11$) and post ($n = 9$) course distributions. Notice that the post-course distribution has shifted to the left, suggesting either 1) a decrease in conceptual understanding for some students; or 2) the two students who did not post-test were high scorers on the pre-test. **b)** 8 students had pre- and post-tests that could be matched. Notice that test scores do not change significantly for most individuals.

Figure 4. Course 3. **a)** Pre ($n = 42$) and post ($n = 38$) course distributions. Notice that the post-course distribution has shifted to the right, suggesting an increase in conceptual understanding for some students. As with most courses, students with the poorest performance on the pre-test experienced learning as measured by this test. **b)** 28 students had pre- and post-tests that could be matched. Notice that the majority of students experienced an increase in test score at the end of the semester.

Figure 5. Course 12. **a)** Sample population is large enough to ascertain that distribution of pre ($n = 190$) and post ($n = 183$) course scores are both normal. Notice that the post-course distribution has shifted to the right, suggesting an increase in conceptual understanding for some students. As with most courses, students with the poorest performance on the pre-test experienced learning as measured by this test. **b)** 135 students had pre- and post-tests that could be matched. Notice that the effect of this course on individual students is mixed, although almost all low-performing students experienced significant gains.



Figure 1

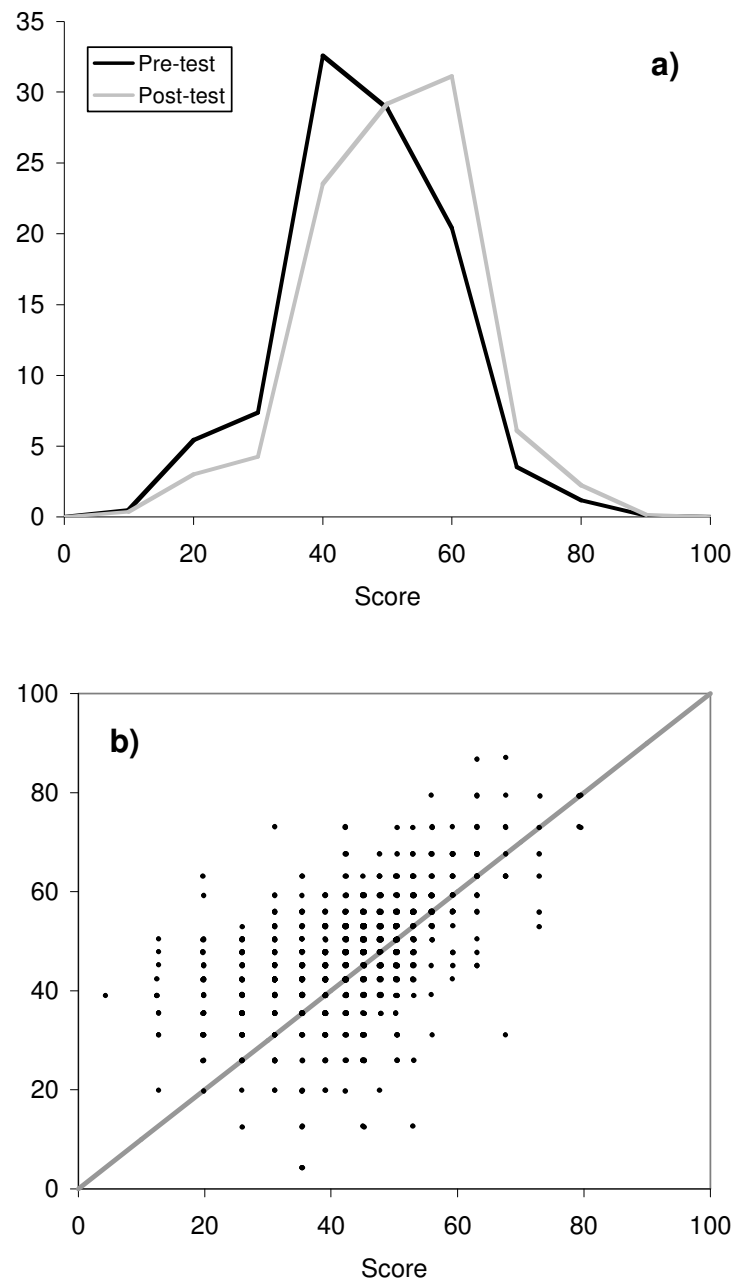


Figure 2

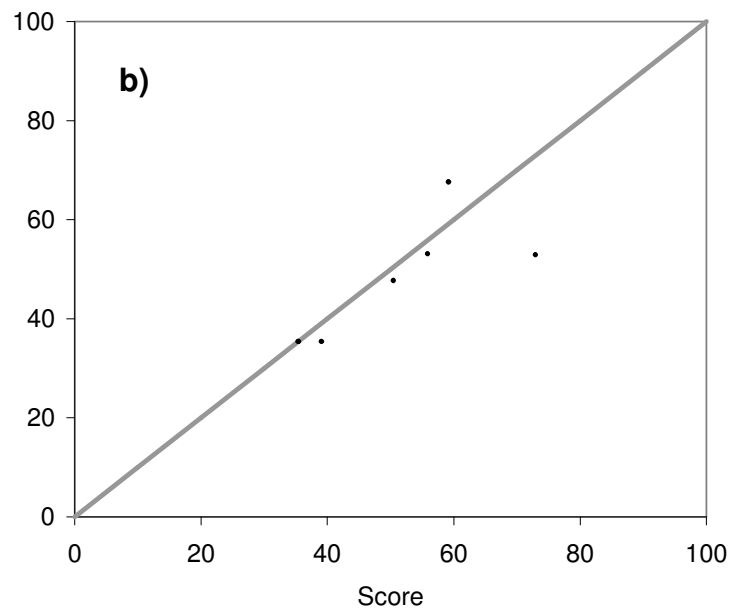
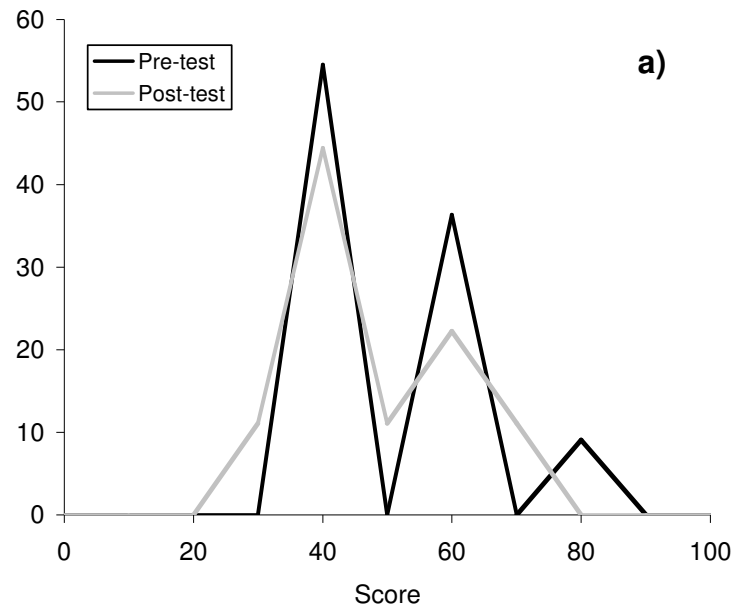


Figure 3

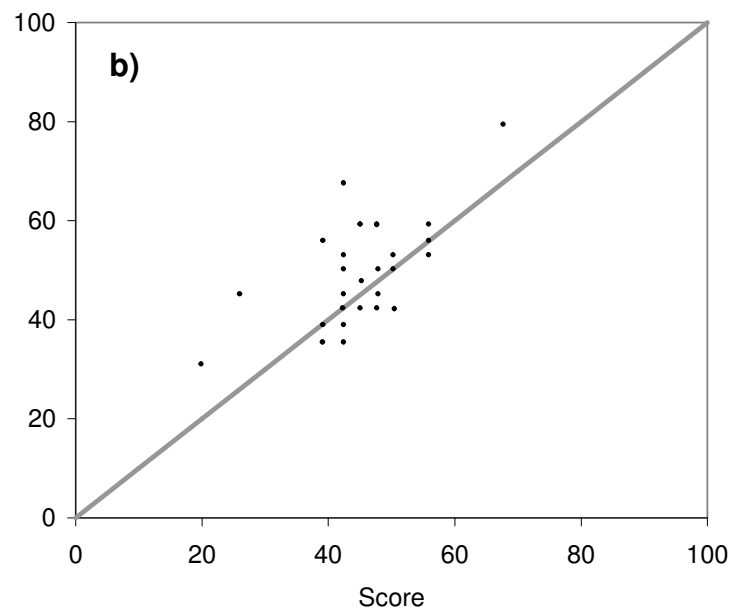
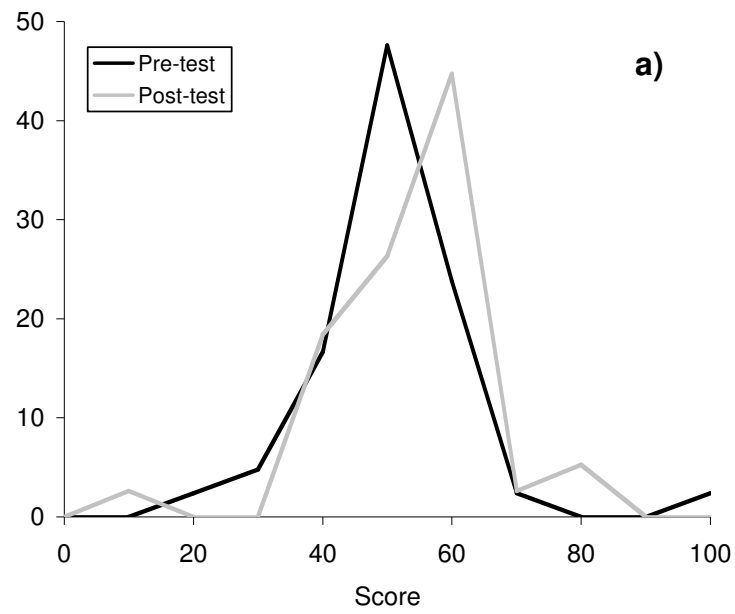


Figure 4

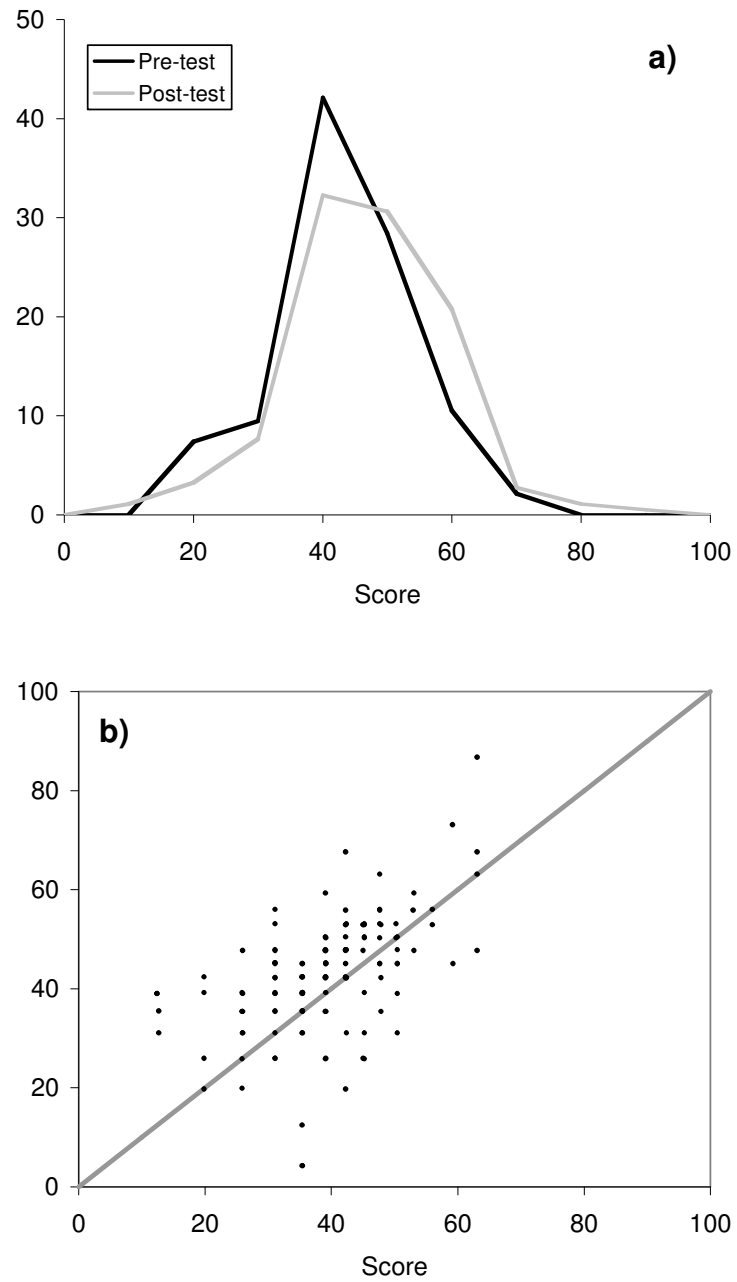


Figure 5

STUDENT ENGAGEMENT QUESTIONNAIRE

To what extent do the following behaviors, thoughts, and feelings describe *you, in this course*. Please rate each of them on the following scale:

5 = very characteristic of me 4 = characteristic of me 3 = moderately characteristic of me 2 = not really characteristic of me 1 = not at all characteristic of me
--

1. _____ Raising my hand in class
2. _____ Participating actively in small group discussions
3. _____ Asking questions when I don't understand the instructor
4. _____ Doing all the homework problems
5. _____ Coming to class every day
6. _____ Going to the professor's office hours to review assignments or tests, or to ask questions
7. _____ Thinking about the course between class meetings
8. _____ Finding ways to make the course interesting to me
9. _____ Taking good notes in class
10. _____ Looking over class notes between classes to make sure I understand the material
11. _____ Really desiring to learn the material
12. _____ Being confident that I can learn and do well in the class
13. _____ Putting forth effort
14. _____ Being organized

15. _____ Getting a good grade
16. _____ Doing well on the tests
17. _____ Staying up on the readings
18. _____ Having fun in class
19. _____ Helping fellow students
20. _____ Making sure to study on a regular basis
21. _____ Finding ways to make the course material relevant to my life
22. _____ Applying course material to my life
23. _____ Listening carefully in class

[Source: Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *Journal of Educational Research*, 98, 184-191.]