Gobert, J.D, Pallant, A.R., & Daniels, J.T.M. (2010). Unpacking inquiry skills from content knowledge in Geoscience: A research perspective with implications for assessment design. *International Journal of Learning Technologies, 5(3),* 310-334.

Unpacking inquiry skills from content knowledge in Geoscience: A research and development study with implications for assessment design

Janice D. Gobert, Social Sciences and Policy Studies, WPI
Amy R. Pallant, The Concord Consortium
Jillian T. M. Daniels, Earth and Space Sciences Dept., UCLA

Keywords
Inquiry assessment, performance assessment, separating inquiry skills from content knowledge, Geoscience assessment, learning technologies

Abstract
Recently there has been expressed need for authentic inquiry assessments that can differentiate students' content knowledge from their inquiry skills. Additionally there has been expressed need for such assessments particularly in the Geosciences, given the nature of the visualizations, the data sets, and the form of communication (i.e., topographical maps, cross-sectional drawings, etc.) used. In response to these calls, we first designed a supplemental instructional and assessment module for honing middle school students' content knowledge and inquiry skills in the domain of Geosciences, namely, Plate Boundaries. Secondly, we evaluated, using our assessment tasks, six classes of middle school students' content knowledge and inquiry skills in this domain. Thirdly, we used factor analysis to empirically demonstrate that content knowledge and inquiry skills can be assessed as separate forms of knowledge. Five factors were empirically demonstrated: some of these represent content knowledge exclusively, some of these represent inquiry skills exclusively, and some of these include both content and inquiry within the same strand. Results are discussed with regard to instructional implications and assessment design.

INTRODUCTION

For careers in the STEM disciplines in the 21st century students need critical science skills, such as how to conduct inquiry with datasets and learn with scientific visualizations such as rich, interactive simulations (Gobert, 2005a). The National frameworks for science (National Science Education Standards, 1996) emphasize **inquiry skills** also referred to as **scientific practices** by the National Assessment of Educational Progress (Champagne, Bergin, Bybee, Duschl, & Gallager, 2004) and claim that these skills are critical to science reform efforts. However, in typical classroom practice, science learning is often focused on rote learning of vocabulary, facts, and formulas for three reasons. *First*, scientific inquiry is difficult to implement in classrooms (de Jong et al, 2005); *second*, science process skills are difficult to assess (Fadel, Honey, & Pasnick, 2007), and *third*, rote knowledge is what is prioritized on high-stakes tests. However, in all sciences, and perhaps *particularly* in the Geosciences due to high reliance on visualizations (e.g., topographical maps and cross sections) and real-time data-sets (e.g., USGS data sets), it is critical that assessment activities *not* be reduced to rote memorization since these cannot accurately evaluate students' geoscience inquiry skills (Manduca, Gobert, Laws, Mogk, & S.J. Reynolds, 2005). In addition, it is important that science process skills are assessed within the contexts in which they are learned and embedded (Mislevy et al., 2002, 2003). The project described herein seeks to assess students' inquiry skills in situ in the domain of Geoscience.

We describe a project in which we developed instructional and assessment modules to supplement existing geosciences curricula that focus on understanding dynamics along Plate Boundaries; the modules were designed to support the development of students' inquiry skills in the context of Geoscience, and assess students' inquiry skills in this context. These data are used to unpack the relationship between inquiry skills and content knowledge by: 1) applying theoretical frameworks to the design and coding of rich assessment tasks, and 2) using statistical techniques, namely factor analysis to empirically demonstrate that content knowledge and inquiry skills can be coded separately, and as such represent two facets of science knowledge.

Prior Work in Geoscience

Although there has been some work conducted in the Geosciences, this compares unfavorably in amount to that in the Physical Sciences (Stofflett, 1994). The bulk of the research that has been done in this domain has been conducted on student conceptions, including: the earth as a cosmic body (Vosniadou & Brewer, 1992; Nussbaum, 1979, Nussbaum & Novak, 1976), rock-cycle processes (Stofflett, 1994), earth and space as they relate to seasons (Schneps, et al., 1989), and phases of the moon, (Schoon, 1992; Bisard et al, 1994), sea floor dynamics (Bencloski and Heyl, 1985), the earth's gravitational field (Arnold, Sarge, and Worrall, 1995), the causes of earthquakes (Ross & Shuell, 1993; Bezzi, 1989; Turner, Nigg, & Daz, 1986), and structure of the earth and interior processes (Gobert & Clement, 1999; Gobert, 2000), including Plate Tectonics (Gobert & Pallant, 2004). More recently however, Earth and Space Science are included as an integral component of Science in the NAEP Science Frameworks for 2009 (Champagne et al, 2004), thus making research in this area a priority.

Additionally, very little work has been done in the Geosciences with assessment as its goal, in particular the assessment of inquiry skills, despite the recent calls for pedagogy and assessment related specifically to the Geosciences (Manduca, Mogk, & Stillings, 2002; Manduca, Gobert, Laws, Mogk, & S.J. Reynolds, 2005). We believe that this domain provides an excellent opportunity to inform assessment design for inquiry with visualizations because visualizations are frequently used in the Geosciences for inquiry, thus this is an authentic form of inquiry in this domain. Moreover, there are likely generalizations that can be made from our project to other domains in which rich visualizations and complex data sets are used as well (Zalles, Quellmalz, Gobert & Pallant, 2007).

Overview to the DIGS Project

In the Inquiring with Geoscience Data Sets project (DIGS; NSF-GEO# 0507828) we made use of cognitively principled assessment design that relates the learning to be assessed, as specified in a student model, to a task model that specifies features of tasks and questions that would elicit the evidence of learning, to an evidence model that specifies the quality of student responses that would indicate levels of proficiency (Messick, 1994; Mislevy et al., 2003; Pellegrino et al., 2001). In addition to making use of findings from previous studies of learning in the Geosciences (Gobert & Clement, 1999; Gobert & Pallant, 2004; Gobert, 2005b), we also made use of recent work which recommended the integration of best practices in learning science with the distinctive challenges posed by using geoscience data sets and visualizations in inquiry activities, e.g., working with geologic time-referenced concepts, observing complex natural systems, using integrative and synthetic approaches (Manduca, Mogk, & Stillings, 2002; Manduca, Gobert, Laws, Mogk, & S.J. Reynolds, 2005). The instructional module described herein provides extended inquiry-based investigations employing real geoscience data sets from, for example, the USGS (United States Geological Survey) and visualizations and was designed to both reify students' content knowledge in the area of Plate Boundaries as well as hone and assess their skills for inquiry using real-time data sets and visualizations. The assessment module includes performance assessments that provide evidence of geoscience knowledge and inquiry strategies not typically captured in traditional test formats; these assessments are designed to yield evidence of students' inquiry skills specifically within the context of geoscience phenomena. The paper by Quellmalz and Zalles (this volume) further describes the development processes employed in the present project, in addition to their own, as these two papers were part of the same materials development project (Quellmalz, Zalles & Gobert, 2007; NSF-GEO# 0507828).

Goals of the Project

The first goal of this project was to design an instructional module, an assessment module, and a coding scheme, aligned with national inquiry standards (NSES, 1996), for honing and assessing students' content knowledge and inquiry skills in the domain of Geoscience, namely, the occurrence and predictability of earthquake processes along Plate Boundaries. The second goal was to evaluate, using our assessment tasks, students' content knowledge and inquiry skills in this domain. In this paper we also address a third goal, that is, to empirically demonstrate that content knowledge and inquiry skills can be assessed as separate forms of knowledge.

MATERIALS DEVELOPMENT

The Plate Boundaries Module

The plate boundaries module, *On Shaky Ground: Understanding Earthquake Activity Along Plate Boundaries*, has two components.

First, students completed a 4-5 day supplementary instructional unit on Plate boundaries. In the process, they examined authentic, publicly-available data sets with the help of appropriate software tools that permit students to select, simulate, and represent the data in different ways. Specifically, the unit engages students in a series of inquiry-based activities to explore and reify their understanding of the relationship between earthquakes and the characteristics of plate boundaries in the Earth's crust. We designed the activity such that the inquiry skills are learned and honed using, as an anchor, a time-based simulation tool called *Seismic Eruption* TM (Jones & Jones, 2004), which simulates multiple decades of three-dimensional data about earthquakes around the world. With this tool, students can also create cross-sections along plate boundaries showing the location of the earthquake epicenters below Earth's surface. In sum, the goal here was to reify students' content knowledge and hone their inquiry skills in this domain.

The *assessment component* of the unit takes two days in which we assess students' content knowledge and inquiry skills using near transfer tasks. Specifically, the tasks require that students transfer the inquiry skills practiced in the units to new, yet conceptually-related problems. This provides data on the students' interactions with and manipulation of the visualizations and data sets, which can be, in turn, used to document the development of inquiry skills (Zalles et al, 2007).

The unit's components are designed to align closely to and elicit the scientific inquiry skills identified in national science standards (NSES, 1996). Specifically, students are engaged in the following tasks; they:

- *hypothesize* about the likelihoods of earthquakes at locations around the world,
- *observe and summarize* earthquake patterns along divergent, convergent, and transform boundaries,
- *collect data* and compare earthquake depth, magnitude, frequency, and location along the different plate boundaries (convergent, divergent, transform),
- *analyze* earthquake data sets from United States Geologic Survey database in data tables and in map representations,
- *"develop"* visualizations of plate boundaries (i.e., create cross-sections using the Seismic eruption tool, draw cross-sections, etc.), and
- *relate and communicate* interactions of the plates to the emergent pattern of earthquakes.

Description of instructional module

In part A, “*Understanding Earthquake Activity Along Plate Boundaries*”, students form hypotheses about the likelihood of earthquake eruption at three different cities around the globe, each chosen for the type of plate boundary underlying them. For each, students rate the risk of a major earthquake and provide a rationale for their risk rating.

In part B, “*Current Earthquakes around the World*”, students familiarize themselves with the Seismic Eruption software (Jones & Jones, 2004). Figures are shown later,

In Part C, “*Observing the Data*”, students brainstorm about patterns and characteristics they observed at different plate boundaries (divergent, convergent, and transform).

In Part D, “*Collecting Data*”, using the Seismic Eruption software, students create cross-sections of the earth’s interior at each of the three different plate boundaries and use these to answer questions part E.

Insert figures 1 & 2 here

In Part E, “*Analysis*”, students elaborate on the magnitude, depth, frequency, and location of the earthquakes at the three different boundaries, and explain how the movements of the plates at each boundary account for the patterns in each set of earthquake data, respectively.

In Part F, “*Applying your understanding*”, students are given two tables of earthquake data and identify the type of boundary represented by each table and provide three pieces of evidence each to back up their claims.

In Part G, “*Conclude and Persuade*” students revisit the questions from part A. Here students rethink prior answers and hence demonstrate what they have learned regarding the relationship between plate boundaries and earthquake characteristics.

Description of the assessment module

In assessment module questions were designed to assess how well students could apply what they had learned in the instructional unit in terms of both content knowledge and inquiry skills.

Specifically, in part A, “*Comparing Convergent Boundaries*”, students describe what they learned about the different types of boundaries (divergent, convergent, and transform) in terms of their similarities (Item 1a) and differences (Item 2a). Students state the *evidence* they are using for their hypotheses (for each Item A1b and A2b, respectively).

In part B, “*Analyze Data*”, students are shown a map of the world with earthquakes marked at three different locations (see Figure 4); they are also shown three cross-sections depicted on the map of the world (A, B, & C). Students describe what they observe (see Figures 5 a, b, c) in terms of depth, magnitude, and location; they match the cross-sections with the locations and identify the boundary types.

Insert figures 4 & 5 a,b,c here

In part C, "*Making Conclusions*", students draw conclusions about the magnitude, depth, and location of earthquakes along three different convergent boundaries (continental-continental, oceanic-oceanic, and oceanic-continental). They sketch the three types of convergent boundaries (C2). In C3, the students describe how the processes along each boundary results in the patterns of earthquakes in the data. In C4 they view a specific location on the map (used in assessment item B) and predict the likelihood of a big earthquake (magnitude greater than 6.5) in the next 50 years, and then explain the evidence underlying their prediction.

Table 1 below provides a mapping between the instructional unit and the assessment tasks by inquiry strand. Note that we have not included the inquiry strands in the table that are not honed or assessed in our modules.

Insert Table 1 about here

Development of Assessment Scoring Rubrics for Content and Inquiry Skills

Item-specific rubrics were developed to score student responses; these were either on a two point scale (0-1) or a three-point scale (0-2). Items were scored for both the content knowledge they required as well as the inquiry skill(s) they required; exceptions were items in which the inquiry skill was not sufficiently rich to justify a separate scoring from the content scoring.

Assessment Item A1a) What similarities in earthquake patterns might you expect to find between oceanic-continental, oceanic-oceanic, and continental-continental convergent boundaries? **A1b)** What are you basing your hypothesis on?

Item A1a content scoring. Part a was scored for content. These data were scored for content as follows:

A continental-continental boundary will be different from the other two	2
Earthquakes will all be deep, or have similar patterns in terms of magnitude, or cause earthquakes and mountains because they are all convergent boundaries	1
All create earthquakes along boundary	.5
Incorrect statements	0

Item A1b inquiry scoring. This item was coded for two inquiry strands in order to provide a finer-grained lens on the inquiry components of the task.

For inquiry strand 1.1, *formulate testable hypotheses:* students were scored as to whether they provided a testable hypothesis or not, regardless of correctness of content.

Testable hypothesis	1
No hypothesis/ untestable hypothesis	0

For inquiry strand 4.1, *formulate explanations ("hypotheses") using logic and evidence:* responses were scored as to whether they included evidence for their hypothesis since formulating hypotheses is difficult for students (de Jong, 2006).

Explanation and evidence	2
Explanation, no evidence	1
No explanation	0

Assessment Item A2a) What differences in earthquake patterns might you expect to find between oceanic-continental, oceanic-oceanic, and continental-continental convergent boundaries? **A2b)** What are you basing your hypothesis on?

Item A2a content scoring. This item was scored for content as follows:

A continental-continental boundary will be different from the other two, earthquakes not as deep.	2
The characteristics of earthquakes will be different at different types of boundaries but does not specifying how	1
Incorrect statements	0

<u>Item A2b inquiry scoring.</u> See <u>Item A1b inquiry scoring description above.</u>

Assessment Item B1. Next to each picture on the next page summarize the data and describe the patterns of earthquakes along each boundary.

<u>Item B1 content scoring (for each diagram):</u>

Summarizes observed earthquake data, depth, presence or absence of subduction, magnitude and direction of subduction in relationship to map view.	3
Notes] presence or absence of subduction or other causal mechanisms; notes characteristics of earthquakes	2
Notes presence or absence of subduction or causal mechanisms or characteristics of earthquake. Identifies data in some way.	1
Incorrect statements or no comments	0

<u>Items B1a,b,c inquiry scoring (for each diagram).</u> We coded students with a "1" if they answered the question at all since they would have had to do so on the sole basis of the diagrams provided. We scored them with a "0" if they did not provide an answer.

For inquiry strand 6.2, *develop/use diagrams and charts:*

If answered, by default it means they used diagram	1
No answer	0

Assessment Item B2. Describe and label each picture with the type of convergent boundary (continental-continental, continental-oceanic, oceanic-oceanic) and the letter it corresponds to with the map above.

<u>For content scoring:</u>

Summarizes earthquake data, depth, presence or absence of subduction, magnitude and direction of subduction	3
Notes presence or absence of subduction other causal mechanism; notes the characteristics of the earthquakes	2
Notes presence or absence of subduction or causal mechanisms or characteristics of earthquakes. Identifies data in some way.	1
Incorrect statements or no comments	0

For inquiry strand, 6.2 *develop/use diagrams and charts.* Scored the same as Items B1 a,b,c.

For assessment Item C1. Compare the magnitude, depth and location of earthquake epicenters along the convergent boundaries by completing the table below:

	Magnitude (small, medium, large)	Depth (shallow, medium depth, deep)	Location (on the boundary--scattered etc.)
Continental-Continental convergent boundary	*small/medium, less than o-o.*	*shallow*	*scattered*
Continental-Oceanic convergent boundary	*medium./large, less than or equal to c-c*	*medium/deep*	*along the boundary*
Oceanic-Oceanic convergent boundary	*medium/large*	*deep*	*along the boundary*

<u>For content scoring:</u> 1 point for was given for each correct description (correct responses are in italics in the table). We did not score this item for inquiry strand 6.2 because using the chart is not a sufficiently rich inquiry task.

Assessment Item C2. Draw a sketch of the different convergent boundaries (three in total). Draw and label the location of the earthquakes along the boundaries.

<u>An example of the content rubric for oceanic-continental convergent boundaries is shown below (similar coding schemes were used for each</u> oceanic-oceanic and for continental-continental convergence):

Shows earthquakes all along the subducted plate	3
Shows the earthquakes at the contact point or deep within the plate	2
Shows earthquakes **or** shows the oceanic plate subducting	1
Obvious lack of understanding such as not showing subduction, or random doodle, shows continental plate subducting	0

For inquiry strand, 4.2, *develop models ("cross-sections") using logic and evidence:* Here we scored students' cross-sections as to whether they included information about earthquakes and plates, regardless of correctness since this type of information is critical to depicting geoscience information in cross-sectional form.

Includes earthquakes and plate references	1
Does not include reference to earthquakes or plates	0

<u>Assessment Item C3</u>. Explain how the process along each type of boundary helps describe the patterns you see with the data.

<u>Content scoring rubric:</u>

Subduction causes the characteristics of the earthquakes for oceanic-oceanic and continental-oceanic boundaries, and how lack of subduction causes the characteristics of the earthquakes for continental-continental boundaries	2
Explains fully two boundary types or explains the process but not the pattern or characteristics of earthquakes	1
Does not meet the criteria for 1 or 2 points	0

<u>Inquiry scoring for inquiry strand 4.1</u>—*formulate explanations using logic and evidence.*

Explanation uses evidence	2
Explanation given but no evidence	1
No explanation	0

<u>Assessment Item C4.</u> Look at the data for location C on the map. Predict the likelihood of big earthquakes (magnitude greater than 6.5) occurring there within the next 50 years as low risk, medium risk, or high risk. **C4b.** Explain your reasoning.

For this item we scored each student's data with two content scores: one for risk rating, and one for the correctness of their evidence for their rating.

<u>Scoring for content, part a:</u>

Low	2
Medium	1
High	0

<u>Scoring for content, part b:</u>

On a continental-continental boundary; there haven't been any large quakes since the 1960s	2
Includes one source of data but not both	1
Does not state either source of data	0

For inquiry strand 4.2, *formulate models ("make predictions") using logic and evidence:* We scored this item for inquiry because we wanted to evaluate whether students could provide a rationale for their prediction using evidence, regardless of scientific accuracy.

Prediction base with two pieces of evidence	3
Prediction with one piece of evidence	2
Prediction, no evidence	1

No prediction	0

Inter-rater reliability. Inter-rater reliability statistics were compiled from a randomly selected subset of our sample. The total number of items which were scored by two coders was 357; these were distributed all assessment items. Inter-rater reliability agreement 80%. Differences between coders were due to instances in which students' scores were coded slightly higher or lower by one coder, e.g., 1 versus 2, 2 versus 3, etc. On items for which there was disagreement in the original scoring, most were resolved through discussion.

Plate Boundaries Module Implementation

Participant Students. 15 8th grade classes in a suburban Boston middle school participated in this implementation. The school had the following profile in terms of ethnicity: 81% White, non-Hispanic; 10% Asian/Pacific Islander, 5% Hispanic, and 4% Black non-Hispanic. Each class included approximately 20 students.[1]

Participant Teachers. There were three teachers who participated in this implementation, each with five classes. One of the participating teachers had taught for nine years and the other two had been teaching for five. From these, we randomly selected six classes; these data are used in this paper.

Setting and student grouping. The students worked in pairs, assigned by their teacher, for the implementation of the instructional unit, one pair per computer. The students worked individually on the assessment.

RESULTS & DISCUSSION OF FINDINGS

One of the main goals of this study was to evaluate students' content knowledge and inquiry skills in this domain, as measured by our assessment tasks. Another goal of the analyses presented herein was to determine whether the inquiry skills could be empirically disentangled from students' content knowledge. The remainder of the paper is dedicated to describing the results of these two goals

Evaluating students' inquiry skills and content knowledge via individual items. As previously stated, the second goal of the project was to evaluate students' knowledge and inquiry skills in this domain. The table below provides the mean, standard deviation, and mean proportion correct (p-value) for each item for content and inquiry. The items are listed on the left hand side of the table. The values for each item are provided on the right hand side. The proportion correct (p-value) is provided as a means of comparing item difficulty on a standardized scale. The mean would usually suffice, but in this case interpretation is complicated by items that were scored on different scales e.g., some are

[1] National Center for Education Statistics, http://www.nces.ed.gov/globallocator/ on 9/12/2007.

out of 2 point, some are out of 3, etc., and thus the means are not directly comparable. As a measure of item difficulty, a p-value of over 0.90 indicates an item on which many people did well, while a p-value of less than 0.10 indicates an item on which few people did well. Whether these are cases of ceiling effects or floor effects, these items contribute very little variance to our assessment of student skills and so were dropped from later quantitative analyses of these skills.

Insert Table 2 here

We describe our findings for each item in turn.

In terms of the content regarding the similarities (Item 1a) and differences (Item 1b) between the three types of boundaries (o-c, c-c, o-o), students tended to obtain low scores (p values of 0.36 and 0.38). In terms of inquiry related to these items (1a & 1b), students were able to generate testable hypotheses (p values of 0.91 and 0.88); students also were able, in general, to provide evidence for their hypotheses, (p values of 0.61 and 0.60). Taken together, these results suggest that students were able to generate hypotheses but they had not yet mastered the content knowledge concerning the similarities and differences of the three types of boundaries, nor were they successful, in general, at providing evidence for their hypotheses.

Item B1 asked the students to "summarize … the patterns of earthquakes along each boundary" (Item B1a,b,c) using a map (similar to the one in Figure 4) and then to identify which type of convergent boundary each marked location referred to (Item B2a,b,c). In terms of content knowledge, students scored moderately high for both items. For inquiry, scored 0 vs. 1, students scored very high for both items. Although we acknowledge that simply "using the diagrams" to answer the question is not a high level inquiry task for this grade level, but it does reflect one of the NSES inquiry strands (6.2) and is a critical process skill in science in general and for Geoscience in particular because this is a common type of visualization used, and thus was scored for inquiry here.

For Item C1, we scored students' responses on content only since using a table was not a sufficiently rich inquiry task. Students scored moderately high on content related to oceanic-continental and continental-continental boundaries (p values of .70 and .71), but scored lower on content regarding oceanic-oceanic boundaries (p value of .53). These data indicate that students had a better understanding of oceanic-continental and continental-continental boundaries than they did oceanic-oceanic boundaries.

For Item C2, students were asked to draw cross-sections and label the location of plates and earthquakes. When scored for correctness of content, students obtained p-values of 0.57, 0.50, and 0.53, respectively, for each boundary type. When coded for inquiry strand 4.2 (develop formulate models using logic and evidence) to reflect the inclusion of geoscientific information about earthquakes and plates, students obtained p-values of 0.83, 0.77, and 0.77, respectively. Thus, students tended to include plates and earthquake labels as part of their cross-sections but regarding their content knowledge of the geoscientific processes at the three boundaries, their scores were lower.

For Item C3, students were asked to "Explain how the processes along each type of boundary helps describe the patterns you see in the data". In terms of the correctness of the content in their explanations, students scored low on this item (p-value of 0.37), indicating poor understanding of the different patterns of earthquakes. When scored in terms of inquiry strand 4.1 (formulate explanations using logic and evidence), students scored higher (p-value of 0.62) indicating that they understand (at least moderately well) the epistemic form of explanation.

For item C4a, students were asked to predict the likelihood of big earthquakes (magnitude of 6.5 or greater) occurring at a specific location on a map within the next 50 years. When scored for correctness of their risk rating, students obtained a p-value of 0.58; when scored for the correctness of their evidence, students obtained a p-value of 0.33. These data suggest that students lacked content knowledge in order to provide evidence for their prediction. We also scored this item for inquiry strand 4.2, the "formulation of models ("predictions") using logic and evidence, for which students obtained a p-value of 0.72, meaning they understood fairly well that making a prediction requires the inclusion of evidence.

Evaluating students' inquiry skills and content knowledge via aggregated strands.

From our items we selected those items that had both a content and an inquiry component and aggregated subsets of items together that reflected the same inquiry skill or similar type of content knowledge. In doing so, we came up with three inquiry skills strands, namely, hypothesizing/predicting, generating evidence, and generating models, and three content knowledge strands, namely, content for hypothesizing/predicting, content for providing evidence, and content for generating models. In Table 3 below, we provide the means, standard deviations, and proportion correct (p-values) for each of the three aggregated inquiry strand measures and the three content knowledge measures, as well as for the total aggregated score for inquiry and for content.

The results are listed on the right hand of the table. Briefly, for items in which students generated a hypothesis or a prediction, when coded for inquiry students obtained a p-value of 0.92 (Inquiry Strand 1.1). When these hypotheses and predictions were scored for accuracy, their mean score dropped to a p value 0.45. When these were scored for the inclusion of logic and evidence (Inquiry Strand 4.1), a p-value of 0.64 was obtained; however, when their evidence was scored for accuracy, the p-value dropped to 0.39. Lastly, when we scored students' models (cross-sections/predictions) for inquiry (Strand 4.2), we obtained a p-value of 0.77. When the content of their cross-sections and predictions were scored for accuracy, students' obtained a p-value of 0.49. Thus, here students were more successful at the inquiry components of these tasks than they were at the content component.

When viewing students' total inquiry skills as measured by the aggregate scores, students obtained a p-value of 0.73. For the aggregated content score, students obtained a p-value of 0.48. This is consistent with our findings for individual items above, that is, students' content knowledge, in general, lags behind their inquiry skills.

Can inquiry skills be empirically disentangled from students' content knowledge?

In order to empirically test whether the inquiry facet of the items could be statistically differentiated from the content facet of the items, we conducted exploratory factor analyses using Maximum Likelihood Estimation extraction and an Oblimin rotation with Kaiser Normalization. We conducted this analysis with 17 items and the 6 strands that we previously identified as relating to inquiry and content (see Table 4). Because of the small sample size and as the original items were interval scales containing between two to four levels, two analyses were run. The first using the original items and the second using item aggregates representing content and inquiry strands. This second analysis was conducted to minimize the splintering effects that a small sample size has on factors (using a minimum of 200 samples as a heuristic) by providing aggregates with more variance (factor analyses are calculated from covariance matrices or correlation matrices). Although the sample size was relatively low for factor analysis (n=100), the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO indicates whether partial correlations are small) was .78 for the items and .76 for the strands, both indicating that the data are suitable for factor analysis. Bartlett's Test of Sphericity was significant, χ^2 (15 df, N=107) = 181.56, $p < .001$, for items, and χ^2 (136 df, N=107) = 557.16, $p < .001$, for strands, indicating non-identity correlation matrices, a requirement for mathematically valid use of the factor model.

Insert Table 4 here

The Kaiser-Guttman retention criterion of EigenValues greater than 1.0 produced a five factor solution for items (see Table 4) and a two factor solution for the strands (see Table 5). The previously mentioned splintering effects of a small sample size are evident among the factors derived from the original items as they are more consistently grouped by problem than they are by inquiry vs. content (see Table 4); with a higher sample size it is plausible that the factors made up of inquiry problems would have been further grouped into a single factor (see factors 1, 2 and 4) with a similar result for content factors (see factors 3 and 5). Communalities were questionably high, with some values over one during iterations and so interpretation must be made carefully. Cronbach's Alpha ranged between .69 and .91 for the individual items and between .64 and .77 for the two strands, namely, total inquiry and total content, indicating acceptable internal consistency for each of the factors (when values were below the recommended 0.70, they were close which is more acceptable because of the lower sample size). The results of each factor analysis are described in turn.

In Table 4 there are several noteworthy findings with respect to our third goal in this paper, namely, to empirically test whether inquiry skills could be disentangled from content knowledge. As can be seen by looking at the bolded values listed under factors 1 through 5, in some cases there is clustering of all content or all inquiry items, and in some cases, there is clustering of both inquiry and content items within the same factor.

Factor 1 includes three items (C2a, C2b, & C2c), which all reflect the inquiry strand 4.2, *develop models (“cross-sections”)*. These are items in which students were asked to draw cross-sections of different convergent boundaries and then label the location of plates and earthquakes. In sum, factor 1 contains all items for which students’ inquiry skills were represented via their drawings of cross-sections.

On Factor 2, there are three loadings of note: one content item (C3) and two inquiry items (C3-4.1; C4-4.2). Item C3 asked students to “Explain how the process along each type of boundary helps describe the patterns you see with the data”. This item, as well as C4, was coded for the inclusion of evidence as indices of inquiry. The content score for C3 also loaded on Factor 2 (however, it obtained the lowest eigenvalue of the three items). The commonality of these items is that they all require evidence for their reasoning.

Factor 3 has two items which load on it, both reflect content knowledge, namely the risk rating assigned to their prediction (C4a) and the correctness of the evidence provided for their risk rating (C4b). These two aspects of content understanding fit together since the students are likely basing their risk rating on evidence.

Factor 4 has six items: two content items and four inquiry items. Items A1 (1.1, 4.1) and A2 (1.1, 4.1) are the inquiry components, namely generate a hypothesis (1.1) and provide evidence for your hypothesis (4.1) for content items A1 and A2 which ask students to describe the similarities (A1) and differences (A2) in earthquake patterns that they might expect to fins between o-c, o-o, and c-c boundaries. These items cluster together because hypotheses and the use of evidence to rationalize these are based on students’ understanding of the similarities and differences observed in patterns.

Lastly factor 5 includes only content items (C2a, C2b, and C2c), these items assess students’ content knowledge of the different convergent boundaries, as reflected in students’ cross-sections. As such, these three items represent the content components of the items, which are clustered on Factor 1, above.

In sum, the results of this factor analysis generated five factors: two reflecting content knowledge, one reflecting inquiry skills, and two reflecting a combination of inquiry skills and content knowledge. It is interesting to note that on the factors that reflect both content and inquiry items, the items require that students use their content knowledge to provide an explanation or rationale. These data suggest that the content knowledge and inquiry skills involved in doing explanation-type tasks are tightly linked to one another. This is intuitively plausible as well since generating an explanation requires that one know the content involved.

Table 5 below provides the data, which was generated using factor analysis on our strands, namely, aggregated items for each inquiry and content. While these aggregated items were not derived from the previous factor analysis of individual items, we found that a parallel analysis was desirable since the small range of the previous scales made it hard to differentiate whether factors were grouping items by inquiry or content or simply by problem (i.e. A’s with A’s, B’s with B’s, etc.). For instance, individual item scales ranged from 1 to 3 points while the aggregated scales ranged from 3 to 19 points. In either case it would be desirable to run confirmatory factor analysis following each exploratory analysis, but this was impractical given the small sample size and properties of the individual item covariance matrix. Finally, in the case of the aggregated items, since the exploratory analysis grouped them in the way we would force them together during a confirmatory analysis, we can infer that our model is indeed the mathematically preferable one (in the case of the aggregated items only).

Insert Table 5 here

In this factor analysis, we obtained a two-factor solution in which all the items that loaded on factor 1 are content items and all the items that loaded on factor 2 are inquiry items. It should be noted, however, that the content eigenvalue was approximately three times greater than the inquiry eigenvalue, and, further, that Cronbach's alpha for the inquiry scale was less than 0.7, although not by much.

General Discussion and Implications for Further Research and Assessment Design

The first goal of this project was to design a module and coding scheme, aligned with national inquiry standards, for honing and assessing students' content knowledge and inquiry skills in the topic of Geoscience (Plate Boundaries). Our work here served two purposes. First, to develop assessments specifically for Geoscience inquiry (Manduca et al, 2005), and secondly, to score inquiry skills separate from content knowledge since previous literature has stressed difficulty and importance of doing so (Mislevy et al, 2002, 2003).

Secondly, our approach provides significant insights into the facets and components of tasks with which students are having difficulty. In particular, our data also showed that, for many items, students scored higher on the inquiry skills used to generate their response than they did on the correctness of content scored for that item. For example, our data suggest that students understand the epistemic form of a hypothesis or a prediction, i.e., that is that they must contain two variables and a relationship between them, but that they lack a full understanding of how evidence is used to justify hypotheses and predictions. Similarly, students were able to generate explanations of data (similarities and differences), but when asked to provide evidence for these or describe their reasoning, their inquiry scores were much lower. Taken together, these data suggest that students are acquiring schematic knowledge (Shavelson, 2005) about the epistemic form of these pedagogical tasks, even when they are lacking content knowledge for the task. Here, we assume that epistemic knowledge is integral to and important in the development of inquiry skills (Gobert & Discenna, 1997). Our findings are consistent with previous studies of inquiry in which it has been shown that students lack inquiry skills (Gobert & Schunn, 2007). For example, students have difficulties drawing correct conclusions from experiments and linking hypotheses and data (de Jong, 2006); both of these are consistent with our findings.

Thirdly, our paper substantiated empirically that inquiry skills can be successfully coded as distinct from content knowledge. Specifically, factor analyses generated a set of factors that included either all content items or all inquiry items, thereby, empirically substantiating that content knowledge and inquiry skills are two distinct forms of knowledge. Two additional factors included both content and inquiry items – these included items for which we asked students to provide a detailed rationale/explanation for their response to the item(s). Given that the content facet and the inquiry facet of these items loaded on the same factor, these data suggest that one's skill at using logic and evidence relates to their understanding of the content involved.

Lessons learned from this project include the importance of designing rich tasks that can be both scored both for content and inquiry using a scale for each facet (content and inquiry) and that can reflect a full range of student skill level or content level. This is important for the psychometric aspects of the items and the statistical analyses and also important for the assessment of skills and knowledge at various levels, including nascent inquiry skills and partial content knowledge. Furthermore, it is critical that assessment items be able to distinguish students' competence from performance so that assessment schemes such as this can be generalized to lower grades in which data literacy issues (obtaining information from a graph, map, etc.) could impact students' content learning and inquiry. In the present study, tasks that required students to use diagrams to conduct inquiry and/or generate diagrams to communicate their understanding provided a challenge with respect to their inquiry assessment. It was beyond the scope of this project to collect data about students' knowledge acquisition processes from visualizations (Gobert, 1994; 1999), however, work is currently underway in which we are: 1) using think aloud protocols to collect such data in Geoscience, and 2) developing scoring rubrics for learners' cross-sectional diagrams (DePaor, Whitmeyer, & Gobert, 2007; Gobert, 2000).

Concluding Comments. Our findings are important for research on student learning because, as our factor analyses show, inquiry skills can be assessed independent from content knowledge. This also has important implications for assessment, in particular for standardized tests, which tend to focus on the assessment of rote knowledge rather than inquiry skills. Specifically, since prior research has shown that inquiry skills may help compensate for lack of content knowledge (Hulshof & de Jong, 2006), and that they make their own contribution to learning outcomes, over and above intelligence (Shute & Glaser, 1990), general reasoning ability (Schunn & Anderson, 1999), and metacognitive skills (Veenman & Elshout, 1995), these findings along with ours prescribe the need to develop standardized assessments which can evaluate students' inquiry skills. This will require either the development of items that can be more easily scored than open-response format and/or techniques for auto-scoring students' open-responses (hypotheses, predictions, explanations, etc.). Presently techniques such as C-rater (Leacock & Chodrow, 2003) or latent-semantic analysis (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990) may provide promising approaches along these lines.

Our findings also lead to an important pedagogical question, namely, whether inquiry skills can be bootstrapped in the service of content learning. As previous stated above, inquiry skills may provide compensatory strategies for students, in particular those lacking content knowledge (Hulshof & de Jong, 2006). From this literature as well as our findings, it seems clear that we should teach these skills, and in order to maximally capitalize on our efforts, however, again, we need to be able to assess inquiry skills. Data on students' inquiry skills can also be used, in turn, by teachers to both make curricular decisions as well as identify which students need help and on which specific inquiry skills. Two projects currently underway address these issues specifically (Gobert, Heffernan, Ruiz, & Kim, 2007; Gobert, Heffernan, Beck, & Koedinger, 2009).

REFERENCES

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, D.C.

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.

Champagne, A., Bergin, K., Bybee, R., Duschl, R., and Gallagher, J. (2004). NAEP 2009 science framework development: Issues and recommendations. Paper commissioned by the National Assessment Governing Board, Washington, DC.

Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W. And Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the Society for Information Science, 41*(6), 391-407.

de Jong, T., Beishuizenm J., Hulshof, C., Prins, F., van Rijn, H., van Someren, M., Veenman, M., and Wilhelm, P. (2005). Determinants of discovery learning in a complex simulation learning environment. In P. Gardenfors and P. Johansson, (Eds.), *Cognition, Education, and Communication Technology,* Mawah, NJ: Erlbaum.

de Jong, T. (2006). Computer simulations - Technological advances in inquiry learning. *Science*, *312*, 532-533.

De Paor, D., Whitmeyer, S. & Gobert, J. (2007). Emergent Models for Teaching Geology and Geophysics Using Google Earth . *Eos Trans. AGU, 89*(53), Fall Meeting, San Francisco, CA.

Fadel, C., Honey, M., and Pasnick, S. (2007). Assessment in the Age of Innovation, *Education Week, Volume 26 (38),* 34-40.

Gobert, J. (1994). *Expertise in the comprehension of architectural plans: Contributions of representation and domain knowledge.* Unpublished doctoral dissertation. University of Toronto, Toronto, Canada.

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. (2005a). Leveraging technology and cognitive theory on visualization to promote students' science learning and literacy. In *Visualization in Science Education*, J. Gilbert (Ed.), pp. 73-90. Springer-Verlag Publishers, Dordrecht, The Netherlands.

Gobert, J. (2005b). The effects of different learning tasks on conceptual understanding in science: teasing out representational modality of diagramming versus explaining. *Journal of Geoscience Education, 53(4),* 444-455.

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. 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., Heffernan, N., Ruiz, C., & Kim, R. (2007). *AMI: ASSISTments Meets Inquiry*. Proposal funded by the National Science Foundation (NSF-DRL# 0733286).

Gobert, J., Heffernan, N., Beck, J. & Koedinger, K. (2009). *Assistments Meets Science Learning*. Proposal funded by the U.S. Dept. of Education (R305A090170).

Gobert, J.D., and Pallant, A., (2004). Fostering students' epistemologies of models via authentic model-based tasks. *Journal of Science Education and Technology*. *Vol 13*(1), 7-22.

Gobert, J., & Schunn, C. (Co-chairs). (2007). *Supporting Inquiry Learning: A Comparative Look at What Matters*. Presented at the Annual Meeting of the American Educational Research Association, April 9-13, Chicago, IL.

Hulshof, C. D., and de Jong, T. (2006). Using just-in-time information to support scientific discovery learning about geometrical optics in a computer-based simulation. *Interactive Learning Environments*, *14*, 79-94.

Jones, Richard H., & Jones, A. (2004). Testing Skill in Earthquake Predictions, *Seismological Research Letters*, *74*, 753.

Leacock, Claudia & Chodorow, Martin (2003). C-rater: automated scoring of short answer questions, *Computers and the humanities, 37(4),* 389-405.

Linn, M. C., & Hsi, S. (2000). *Computers, Teachers, Peers: Science Learning Partners*. Hillsdale, NJ: Erlbaum.

Manduca, C.A., J. Gobert, P. Laws, D.W. Mogk, and S.J. Reynolds (2005). "Observing and Assessing Student Learning: A Workshop Report." *Geological Society of America Abstracts with Programs*, *37(7):* 283.

Manduca, C., Mogk, D., and Stillings, N. (2002). Bringing research on learning to the geosciences. Report from a workshop sponsored by the National Science Foundation and the Johnson Foundation. Retrieved on 9/27/2007 from http://serc.carleton.edu/files/research_on_learning/ROL0304_2004.pdf

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Mislevy, R. J., Chudowsky, N., Draney, K., Fried, R., Gaffney, T., and Haertel, G. (2002). *Design patterns for assessing science inquiry*. Unpublished manuscript, Washington, D.C.

Mislevy, R. J., Chudowsky, N., Draney, K., Fried, R., Gaffney, T., Haertel, G., Hafter, A., Hamel, L., Kennedy, C., Long, K., Morrison, A. L., Murphy, R., Pena, P., Quellmalz, E., Rosenquist, A., Songer, N., Schank, P., Wenk, A., and Wilson, M. (2003). *Design patterns for assessing science inquiry (PADI Technical Report 1)*. Menlo Park, CA: SRI International, Center for Technology in Learning.

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. and Novak, J. D. (1976). An assessment of children's concepts of the earth utilizing structured interviews. *Science Education, 60*, 535-550.

Pellegrino, J., Chudowsky, N., and Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment.* Washington, DC: National Academy Press.

Ross, K. and Shuell, T. (1993). Children's beliefs about earthquakes. *Science Education*, 77, 191-205.

Schneps. M. Sadler, P., Woll, S., and Crouse. L. (1989). *A Private universe: misconceptions that block learning*. Santa Monica, CA : Pyramid Film and Video.

Schoon, K. J. (1992). Students' Alternative Conceptions of Earth and Space. *Journal of Geological Education, 40*,. 209-214.

Schunn, C. D., and Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. Cognitive Science, 23(3), 337-370.

Shavelson, R., (2005). *Embedding Assessments in the FAST Curriculum: The Romance between Curriculum and Assessment. Final Report.* Stanford Educational Assessment Laboratory (SEAL) and Curriculum Research and Development Group (CRDG).

Shute, V. and Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments, 1*, 55-71.

Stofflett, R. (1994). Conceptual change in elementary school teacher candidate knowledge of rock-cycle processes. *Journal of Geological Education, 42*, 494-500.

Turner, R. H., Nigg, J. M., and Daz, D. H. (1986). *Waiting for disaster: Earthquake watch in California*. Berkeley, CA: University of California.

Veenman, M.V.J. and Elshout, J.J. (1995). Differential effects of instructional support on learning in simulation environments. *Instructional Science, 22*, 363-383.

Vosniadou, S. and Brewer, W. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535-585.

Zalles, D., Quellmalz, E., and Gobert, J. (2007). *Final report to the National Science Foundation on the DIGS project* (Inquiring with Geoscience Data Sets, NSF-GEO# 0507828).

Biographies

Dr. Janice Gobert is an Associate Professor of Learning Sciences and Psychology in the Department of Social Sciences and Policy Studies and the Department of Computer Science (http://www.wpi.edu/Academics/Depts/SSPS/People/jg1.html). Her Ph.D. from the University of Toronto (1994) is in Cognitive Science. Her specialty is in technology-based learning and assessment with visualizations and simulations in scientific domains. Before coming to WPI, Janice was Senior Research Scientist at the Concord Consortium, an educational research and development organization in Concord, MA (http://www.concord.org). Until December 2007, Janice served as North American Editor for the International Journal of Science Education. She has been principal investigator on many projects that address technology-based science learning and assessment.

Department of Social Sciences and Policy Studies
Atwater-Kent Laboratories
Worcester Polytechnic Institute
Worcester, MA 01609
jgobert@wpi.edu
fax: (508)831-5896

Amy Pallant is a Project Manager, educational researcher, and curriculum developer at the Concord Consortium. She holds a B.A. in Geology and M.Ed. in Science Education. She has contributed to many projects on Molecular Visualizations and Geoscience while at the Concord Consortium.

The Concord Consortium
25 Love Lane
Concord, MA 01742
apallant@concord.org
fax: (978) 405-2077

Jillian T. M. Daniels completed her B.S. in Physics in 2008 at Worcester Polytechnic Institute where she worked with Professor Janice Gobert on the DIGS project. She is currently a graduate student in Planetary Physics in the Department of Earth and Space Sciences at the University of California, Los Angeles. Her current research involves searching for lightning signals in the atmosphere of Venus.

Dept. of Earth and Space Sciences
3806 Geology Building Box 951567
595 Charles Young Drive East
University of California, Los Angeles
Loa Angeles CA 90095-1567
jtmdaniels@ucla.edu
Fax: (310) 825-2779

Acknowledgements. This project was funded by the National Science Foundation by a Materials Development grant (NSF-GEO# 0507828), awarded to Edys Quellmalz, Daniel Zalles, and Janice Gobert. Addition analyses were conducted under a second grant awarded to Janice Gobert, Neil Heffernan, Carolina Ruiz, and Ryung Kim (NSF-DRL# 0733286). The opinions expressed do not necessarily reflect those of the funding agency.

The authors wish to thank Dr. Cathleen Kennedy, editor of the special issue, for her extremely helpful comments on drafts of this document. Lastly, the authors wish to thank Mr. Jonathan Marks, Ms. Joanne Doblecki-Marks, and Ms. Vanessa Bullard, middle school teachers who provided excellent feedback on the instructional module and assessment items.

Table 1

Alignments of unit and performance assessment items to NSES Inquiry Standards and Components

NSES Inquiry Skills	Plate Boundaries Curriculum Unit	Plate Boundaries Performance Assessment
1. Identify questions and concepts that guide scientific investigations		
1.1 formulate testable hypothesis (or prediction)	Parts A, G	A1b, A2b, C4a
2. Design and conduct scientific investigations	N/A	N/A
3. Use technology and mathematics to improve investigations and communications		
3.1 use technologies to collect, organize, and display data	Parts B, D	
4. Formulate and revise scientific explanations and models using logic and evidence		
4.1 formulate explanations using logic and evidence	Parts F, G	A1b, A2b, C3
4.2 formulate models (predictions/cross-sections) using logic and evidence		C2, C4b
5. Recognize and analyze alternative explanations and models	N/A	N/A
6. Communicate and defend a scientific argument		
6.1 review, summarize, and explain information and data	Parts C, E	C1
6.2 develop/use diagrams and charts		B1, B2, C1, C2, C4

Table 2

Means, Standard Deviations, and Proportion Correct coded for content and/or Inquiry

Item Descriptions	M	SD	p-value
A1. What similarities in EQ patterns might you expect to find between o-c, o-o, and c-c boundaries? What are you basing your hypothesis on?			
A1. Content Scoring: see method section (out of 2)	0.73	0.60	0.36
A1. Inquiry Strand 1.1: formulate testable hypotheses (out of 1)	0.91	0.29	0.91
A1. Inquiry Strand 4.1: formulate explanations with evidence (out of 2)	1.21	0.77	0.61
A2. What differences in EQ patterns might you expect to find between o-c, o-o, and c-c boundaries? What are you basing your hypothesis on?			
A2. Content scoring: see method section (out of 2)	0.75	0.87	0.38
A2. Inquiry Strand 1.1: formulate testable hypotheses (out of 1)	0.88	0.33	0.88
A2. Inquiry Strand 4.1: formulate explanations with evidence (out of 2)	1.21	0.83	0.60
B1. Next to each picture on the next page summarize the data and describe the patterns of earthquakes along each boundary.			
B1a. Content Scoring: Oceanic-Continental Image (out of 2)	1.24	0.63	0.62
B1b. Content Scoring: Continental-Continental Image (out of 2)	1.17	0.58	0.58
B1c. Content Scoring: Oceanic-Oceanic Image (out of 2)	1.14	0.62	0.57
B1a. Inquiry Strand 6.2: develop/use diagrams/charts (out of 1)	0.93	0.25	0.93
B1b. Inquiry Strand 6.2: develop/use diagrams/charts (out of 1)	0.98	0.14	0.98
B1c. Inquiry Strand 6.2: develop/use diagrams/charts (out of 1)	0.97	0.17	0.97
B2. Describe and label each picture with the type of convergent boundary and the letter it corresponds to with the map above.			
B2a. Content Scoring: Oceanic-Continental Image (out of 2)	1.26	0.81	0.63
B2b. Content Scoring: Continental-Continental Image (out of 2)	1.44	0.78	0.72
B2c. Content Scoring: Oceanic-Oceanic Image (out of 2)	1.14	0.87	0.57
B2a. Inquiry Strand 6.2: develop/use diagrams/charts (out of 1)	0.97	0.17	0.97
B2b. Inquiry Strand 6.2: develop/use diagrams/charts (out of 1)	0.97	0.17	0.97
B2c. Inquiry Strand 6.2: develop/use diagrams/charts (out of 1)	0.96	0.19	0.96
C1. Compare the magnitude, depth, and location of the earthquake epicenters along the convergent boundaries by completing the table below.			
C1a. Content Scoring: Oceanic-Continental Table Row (out of 3)	2.10	0.98	0.70
C1b. Content Scoring: Continental-Continental Table Row (out of 3)	2.13	1.03	0.71
C1c. Content Scoring: Oceanic-Oceanic Table Row (out of 3)	1.59	1.15	0.53
C2. Draw a sketch of the different convergent boundaries. Draw and label the location of the earthquakes along the boundaries.			
C2a. Content Scoring: oceanic-continental sketch (out of 3)	1.72	1.20	0.57

C2b. Content Scoring: continental-continental sketch (out of 3)	1.51	1.27	0.50
C2c. Content Scoring: oceanic-oceanic sketch (out of 3)	1.58	1.13	0.53
C2a. Inquiry Strand 4.2: develop models (“cross sections”) (out of 1)	0.83	0.37	0.83
C2b. Inquiry Strand 4.2: develop models (“cross sections”) (out of 1)	0.77	0.42	0.77
C2c. Inquiry Strand 4.2: develop models (“cross sections”) (out of 1)	0.77	0.42	0.77
C3. Explain how the process along each type of boundary helps describe the patterns you see with the data.			
C3. Content Scoring: see method section (out of 2)	0.75	0.80	0.37
C3. Inquiry Strand 4.1: formulate explanations with evidence (out of 2)	1.23	0.81	0.62
C4. Look at the data for location C on the map. Predict the likelihood of big EQs (magnitude greater than 6.5) occurring there within the next 50 years as low risk, medium risk, or high risk. Explain your reasoning.			
C4. Content Scoring: Risk Assessment (out of 2)	1.15	0.91	0.58
C4. Content Scoring: Correct Evidence (out of 2)	0.66	0.73	0.33
C4. Inquiry Strand 4.2: develop models (“predictions”) (out of 3)	2.17	0.82	0.72

Table 3

Means, Standard Deviations, and Percentages for Aggregated Strands and Total Inquiry and Content Scores

Description	Items	Max. Points & # Items	M	SD	p-value
Inquiry Strands					
1.1 hypothesizing/ predicting	A1-1.1, A2-1.1, C4-4.2*	3 points; 3 items	2.77	0.55	0.92
4.1 generating evidence	A1-4.1, A2-4.1, C3-4.1	6 points; 3 items	3.81	1.85	0.64
4.2 generating models	C2a-4.2, C2b-4.2, C2c-4.2, C4-4.2	6 points; 4 items	4.62	1.44	0.77
Content Strands					
Content: hypothesizing/ Predicting	A1, A2, C4a (paired to 1.1)	6 points; 3 items	2.68	1.67	0.45
Content: evidence	A1, A2, C3 (paired to 4.1)	6 points; 3 items	2.31	1.75	0.39
Content: models	C2a, C2b, C2c, C4b (paired to 4.2)	11 points; 4 items	5.37	3.33	0.49
Total Scores					
Total Inquiry	A1-1.1, A1-4.1, A2-1.1, A2-4.1, C2a-4.2, C2b-4.2, C2c-4.2, C3-4.1, C4-4.2	14 points; 9 items	10.24	3.12	0.73
Total Content	A1, A2, C2a, C2b, C2c, C3, C4a, C4b	19 points; 8 items	9.05	4.89	0.48

*An inquiry 1.1 score reflecting whether students could generate a prediction was derived from the C4-4.2 score by coding those with “1” *or greater* as “1” and those with a “0” as “0”; thus, making a prediction was differentiated from doing so using evidence, which was scored as inquiry strand 4.2.

Table 4

Exploratory Factor Loadings and Communalities for Individual Items

	Factors					
Item Description	1	2	3	4	5	Communalities
Content A1	-.018	.184	-.019	**.316**	-.285	.316
Content A2	-.122	.112	.195	**.438**	-.199	.390
Content C2a	.077	.068	-.021	-.077	**-.735**	.597
Content C2b	.089	-.121	.008	-.017	**-.824**	.677
Content C2c	.196	.041	.193	.111	**-.428**	.479
Content C3	-.049	**.501**	.232	-.043	-.279	.590
Content C4a	.032	-.043	**.745**	-.032	.056	.511
Content C4b	.010	.041	**.848**	-.011	-.020	.765
Inquiry A1-1.1	-.033	-.104	-.023	**.508**	.012	.227
Inquiry A1-4.1	.083	.202	-.059	**.606**	-.070	.546
Inquiry A2-1.1	.165	.008	.011	**.571**	.125	.370
Inquiry A2-4.1	.127	.096	-.002	**.578**	-.021	.443
Inquiry C2a-4.2	**.748**	-.033	.011	.170	-.085	.697
Inquiry C2b-4.2	**.939**	.007	.074	.100	-.027	.999
Inquiry C2c-4.2	**.739**	.100	.029	-.097	-.116	.654
Inquiry C3-4.1	.200	**.936**	.024	-.098	.054	.905
Inquiry C4-4.2	-.076	**.326**	.032	.200	.007	.188
EigenValues	5.47	2.00	1.70	1.17	1.07	
Cronbach's Alpha	0.91	0.69	0.75	0.72	0.78	

*Extracted using Maximum Likelihood Estimation; Rotated using Oblimin with Kaiser Normalization

Table 5

Exploratory Factor Loadings and Communalities for Strands

Strand Description	Factors 1	Factors 2	Communalities
Inquiry strand: hypothesizing (A1 1.1, A2- 1.1, C4 4.2)	-.026	**.407**	.153
Inquiry strand: generating evidence (A1 4.1, A2 4.1, C3 4.1)	.046	**.845**	.766
Inquiry strand: generating models (C2a,b,c 4.2, C4 4.2)	.066	**.575**	.384
Content strand: hypothesizing/predicting (A1, A2, C4a)	**1.007**	-.140	.855
Content strand: generating evidence (A1, A2, C3)	**.747**	.189	.773
Content strand: generate models (C2a, C2b, C2c, C4b)	**.491**	.174	.379
EigenValues	3.16	1.04	
Cronbach's Alpha	0.77	0.64	

*Extracted using Maximum Likelihood Estimation; Rotated using Oblimin with Kaiser Normalization

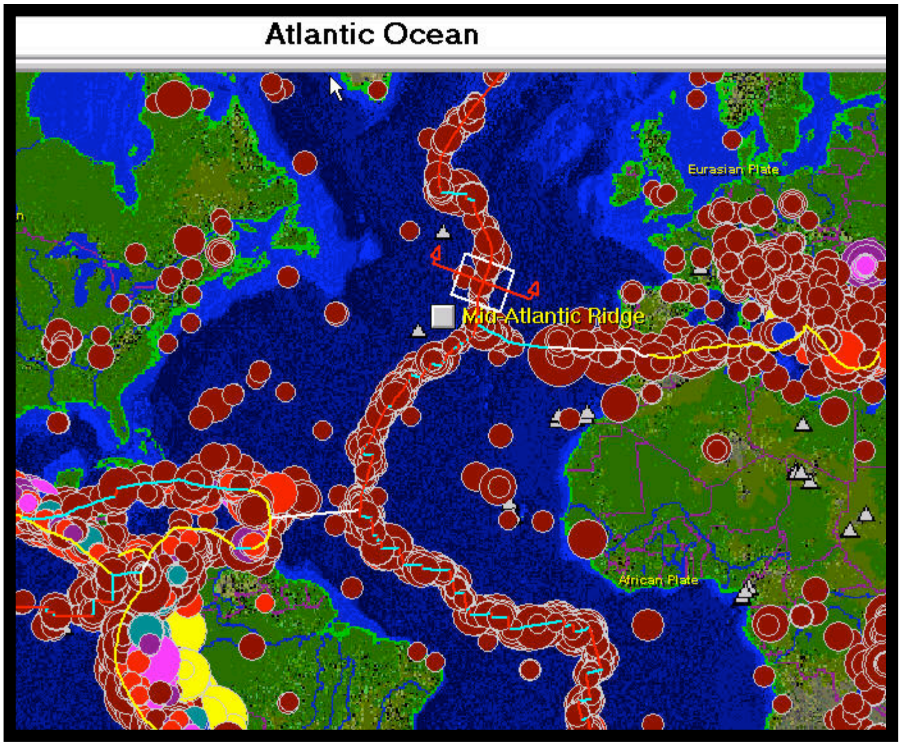

Figure 1: Plate Boundaries and Simulated Earthquake Activity

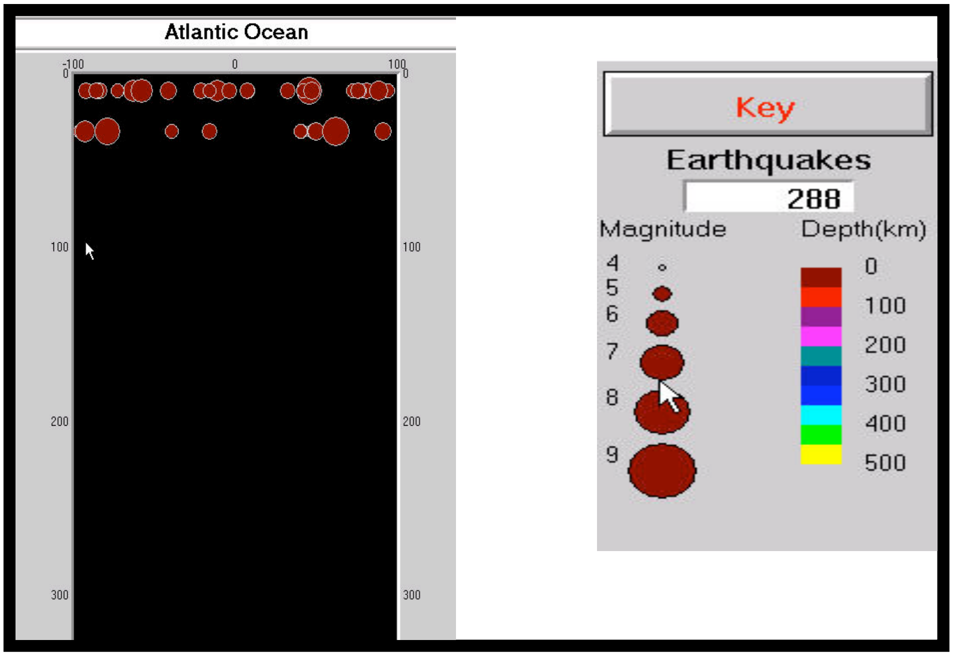

Figure 2: Cross-Sectional View of Earthquake Simulation with key

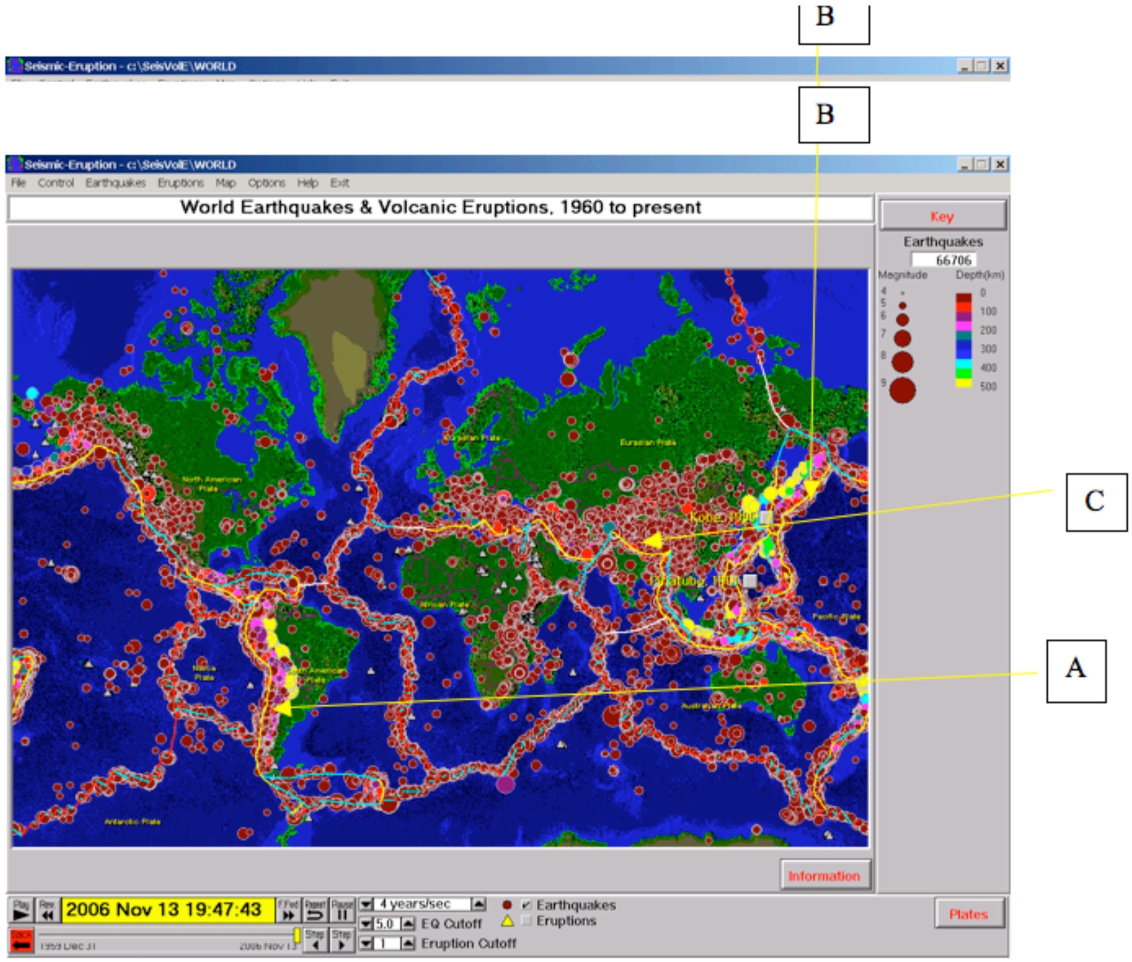

Figure 3: Map for Assessment Part B "Analyze Data"

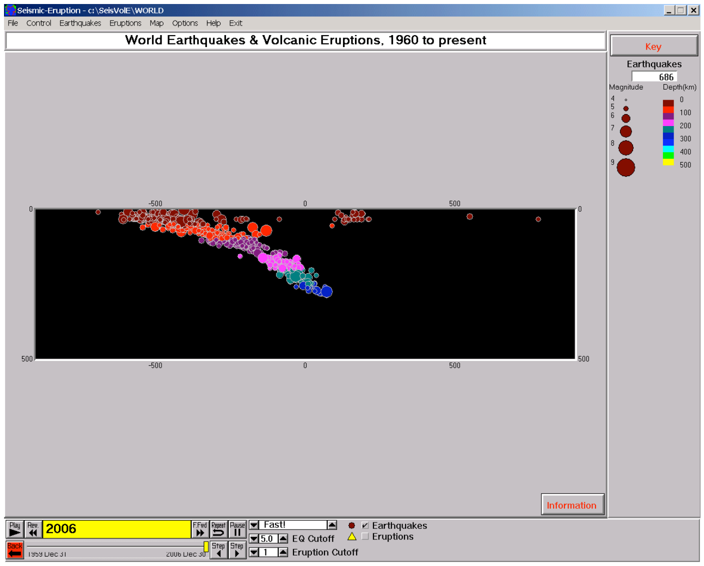

Figure 4a: Assessment Part B: Analyze Data, Cross-section 1- Location A

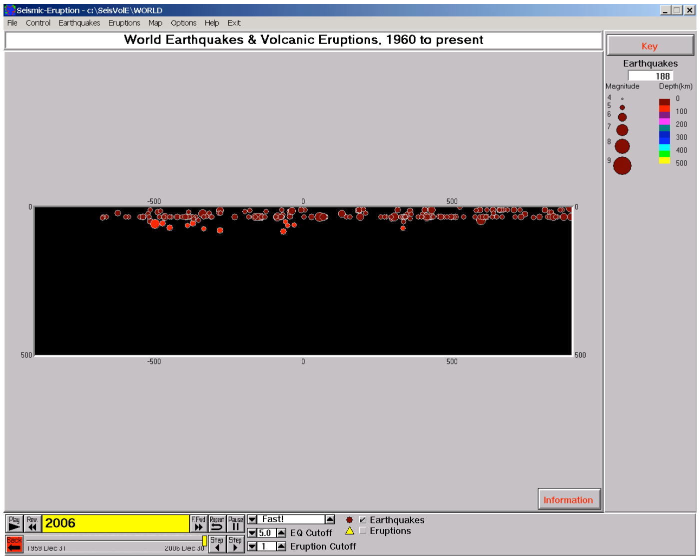

Figure 4b: Assessment Part B: Analyze Data, Cross-section 2- Location C

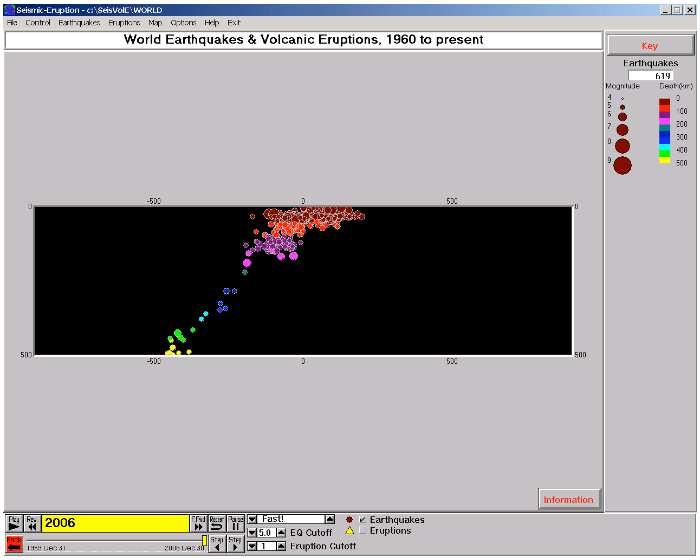

Figure 4c: Assessment Part B: Analyze Data, Cross-section 3- Location B