DESIGNS FOR ASSESSING FOUNDATIONAL DATA LITERACY

Daniel R. Zalles, Ph.D.
Center for Technology in Learning
SRI International

Introduction

Geoscience education could benefit from assessment instruments that validly and reliably assess students' foundational data literacy skills (e.g., sample size, sample selection, database structure, data distribution, central tendency, natural variability, measurement error), using appropriate item formats that provide valid and reliable evidence of different levels of skill and understanding. When geoscience educators engage students in investigating real data sets in pursuit of geoscience content objectives, lack of these fundamental skills and understandings can hinder the students' abilities to complete the geoscience tasks successfully. Conversely, data literacy problems can hide and hinder the demonstration of geoscience content understanding, leading to erroneous diagnoses of the causes of student problems when asked to carry out data-immersive geoscience tasks.

Data literacy is recognized in national standards as critical components of science, math, and social studies curricula:

> "Students…need to learn how to analyze evidence and data. The evidence they analyze may be from their investigations, other students' investigations, or databases. Data manipulation and analysis strategies need to be modeled by teachers of science and practiced by students. Determining the range of the data, the mean and mode values of the data, plotting the data, developing mathematical functions from the data, and looking for anomalous data are all examples of analyses students can perform."[1]

> "To understand the fundamentals of statistical ideas, students must work directly with data... The data analysis and statistics strand allows teachers and students to make … important connections among ideas and procedures from number, algebra, measurement, and geometry. Work in data analysis and probability offers a natural way for students to connect mathematics with other school subjects and with experiences in their daily lives."[2]

> "During the middle school years, students relate their personal experiences to happenings in other environmental contexts. Appropriate experiences will encourage increasingly abstract thought as students use data and apply skills in analyzing human behavior in relation to its physical and cultural environment."[3]

[1] National Research Council, Science Content Standards, 9-12, Science as Inquiry, Content Standard A (1996)
[2] National Council of Teachers of Mathematics (2000)
[3] National Council for the Social Studies, People Places, and Environments Thematic Strand (1994)

This essay describes assessments developed and piloted at SRI International that measure data literacy. The work has been funded through four National Science Foundation-grants and one grant commissioned by the U.S. Department of Education. Though they vary somewhat in purpose, audience, and item formats, each assessment engages students in investigating real data sets, then scores them for deep understanding. Each can be a tool for teachers and instructors who want to formatively assess their students' readiness to handle the components of their units that rely on knowledge and skills about the collection, organization, and analysis of data, as well as, in some of the assessments, other aspects of scientific inquiry.

The assessments vary in the data literacy outcomes they measure, grade levels they target, and item formats they use, which include performance tasks[4], constructed response items[5], multiple choice items, and justified multiple choice items[6]. Described below, EPA Phoenix, Solar Power, and the GLOBE Integrated Investigation Assessments are performance tasks about geoscience topics. The Thinking with Data and Foundational Tools in Data Literacy assessments present constructed response and justified multiple choice questions that require students to demonstrate data literacy skills and understandings learned and practiced in various interdisciplinary units that combine math with science and social studies.

Descriptions of the Assessments

EPA Phoenix and Solar Power

EPA Phoenix is an 8th grade assessment developed and piloted with funding from the U.S. Department of Education under a project known as Building a Foundation for a Decade of Rigorous, Systematic Educational Technology Research. Solar Power is a high school-level assessment designed for an NSF funded project called Innovative Designs for International Information Communication Technology Assessment in Science and Mathematics Education. Both assessments were designed and piloted as instantiations of a modular approach to assessing the outcomes of school ICT (information, communication, and technology) programs that have students carry out various technology data-related tasks: Internet research projects in the case of EPA Phoenix; geographic information systems in the case of Solar Power. EPA Phoenix focuses on air quality and Solar Power focuses on the feasibility of solar energy for electric power. To lessen the risk that low content knowledge in the topics of the assessments could confound the assessment of data literacy and other skills made possible in the assessment

---

[4] Sets of assessment items that revolve around common introductory materials and require longer, deeper attention than what would be required of the learner on more traditional tests are generally known as "performance tasks." The major part of most performance tasks has to do with constructing responses to items that elicit divergent thinking, though there is nothing to preclude the inclusion of items using different formats if they are relevant to the overall task.

[5] Constructed response items require constructed (as opposed to selected) answers and can range from simple fill-in-the-blank exercises or problem completion exercises to essay writing.

[6] Justified multiple choice items require that respondents construct an explanation for why they made their selections. Hence they require a multiple choice selection, followed by an explanation or by computation that led to their selection.

tasks, all the science content-specific information that the students would need to fulfill the tasks are provided in the assessments.

In EPA Phoenix, the problem posed is to help a regional soccer league determine whether air quality and temperature are optimal enough to hold championship games in Phoenix, and the best time of year in which to hold the games. Students are sent to graphs showing air quality ratings in Phoenix that have been generated from the Environmental Protection Agency's Air Quality System (AQS) database.[7] They examine trends in air quality from these data, then compare the overall air quality ratings in Phoenix to ozone ratings in nearby states, represented on a color-coded map. Data literacy outcomes addressed in EPA Phoenix include comparing trend lines on graphs, transferring relevant data about air quality from one type of representation (line graphs) to another (data table) in order to facilitate analysis, critiquing the relevance of specific data for answering a research question, and synthesizing data from different representations to formulate an overall conclusion.

In the Solar Power task, students use GIS representations to compare and contrast air temperature data, then compare and contrast model-generated data about incoming solar radiation. They also observe data about percentage of cloudy days over the course of specific periods of time and perform some calculations.

EPA Phoenix and Solar Power are both performance tasks. The items stem from a common problem and require that the students investigate data, then synthesize it in order to formulate and communicate evidence-based conclusions in the form of a report or presentation. As a culminating activity in the Solar Power task, they recommend a state that should rely more on solar energy for its electrical power. As a culminating activity in the EPA Phoenix performance task, students are asked to write an evidence-based recommendation for whether the soccer games should be held in Phoenix. By this time in the task, they have examined the data sources. The students are scored on a 4-point rubric on the basis of how well they can formulate an evidence-supported conclusion. Either a recommendation of yes or no is acceptable, as long as it is supported by evidence. Figure 1 below illustrates the range of responses that were obtained in the pilot.

[7] The EPA no longer provides public access to the visualizations.. Hence, current students who do EPA Phoenix access images of the relevant AIRS Graphics representations about Phoenix that are archived at the Integrative Performance Assessments in Technology web site (http://ipat.sri.com).

Figure 1. Examples of student responses to culminating item in EPA Phoenix

Example of score of 4:

"In our research we discovered that Phoenix would be a moderate location for the championship game competition. Phoenix has a serious problem with pollution, but in comparison with other counties, it was not the worst. With 9 years to fix their problem, it is not as bad as 20 years such as Sacramento Metro, South Coast, Ventura, and the Southeast Desert.

With recorded information from our table of highly polluted months, we figured that the months April-June are the best months to have the championship games. These months are not the least polluted but they have the best temperature range. The mean temperature for May, the middle month is 78.8 degrees F.

Some of the health affects of the ozone problem for the soccer players are as follows: Breathing problems, reduced lung function, asthma, irritated eyes, stuffy nose, reduced resistance to colds etc."

Explanation of score of 4:
The required conclusions and supporting evidence are made. There are no flaws in accuracy or relevance. The evidence is sufficiently specific to provide adequate support for the conclusions made.

……………………………

Example of score of 3:

"We think that Phoenix would be a good place to have the Soccer Tournament because the air temperature is not to hot or not to cold. The average temperature in May was 81.3° which is average for Arizona in that month. The temperature we think would be good because the temperature is just right.

If the soccer tournament were to be held in Phoenix the best months would be April-June. This is because the number of unhealthful days is the lowest in three years. That is good because if there were many unhealthful days the children would not be safe in that weather and it could cause distraction on the field. The only time that the ozone would affect the soccer players is if the pollution for air and water got really bad. That would be awful because the more pollution there is the better chance of people getting sick is more likely."

Explanation of score of 3:
The required conclusions and supporting evidence are made -- that Phoenix would be a good place for the games because it has a good temperature, and that April-June would be the best time for the games because there were fewest unhealthful days then over the course of a three-year period. There are flaws however, including an unsupported claim (that 81.3 degrees is average for Arizona in May rather than just for Phoenix), and an inaccurate association of ozone to water pollution in the second paragraph.

………………………………

Example of score of 2:

"Phoenix is hot and polluted with ozone, which can make you sick if you are not careful with your body. We recommend that you go somewhere else to play soccer. If you have to play there, play in April-June because it's less hot then than at other times during the year."

Explanation of score of 2:
Only one piece of marginal evidence is cited -- that April - June is less hot than other times of the year, but no specifics are offered about how less hot is. Hence, the requirements for evidence are not being met.

………………………………

Example of score of 1:

"Phoenix would be a good place to have the games. The scenery there is great and the pollution keeps getting better and its not too hot. Enjoy yourself!"

Explanation of score of 1:

There is a major ambiguity (e.g., that "the pollution keeps getting better"). No recommendation or supporting evidence is included about the season. The scenery is irrelevant.

EPA Phoenix and Solar Power can both be found at SRI's Integrative Performance Assessments in Technology web site (http://ipat.sri.com).

The GLOBE Integrated Investigation Assessments

The GLOBE Integrated Investigation Assessments are web-based sample student assessment tools and frameworks that provide teachers and students with evidence about progress on GLOBE program goals related to data literacy and other aspects of scientific inquiry. Students participating in GLOBE take atmosphere, hydrology, soils, and land cover/phenology measurements, post their data on the web, and create maps and graphs with the data. Content areas covered in the assessments include atmosphere, hydrology, landcover, soils, and Earth systems. In contrast to EPA Phoenix and Solar Power, the GLOBE Integrated Investigation Assessments require that students already possess significant content knowledge in order to conduct investigations with the data. Data-immersive activities posed to students include:

- examining GLOBE data/graphs and coming up with possible questions regarding the data
- finding observable trends in the data
- looking through the data for possible measurement or data entry errors and suggesting ways to avoid these types of errors in the future
- identifying relationships between two variables
- representing data in a graph or table
- using data to generate new data representations to analyze trends
- summarizing graphed data in terms of range, median, mode, and mean
- comparing and contrasting same-variable data sets from different locations
- drawing evidence-based conclusions about the data.

Like EPA Phoenix and Solar Power, the Globe Integrated Investigation Assessments are performance tasks. The assessments and supporting information can be found at http://globeassessment.sri.com.

Like EPA Phoenix and Solar Power, the Globe Integrated Investigation Assessments are performance tasks. For the sake of illustration of the sorts of responses that students provide for GLOBE assessment tasks, Figure 2 presents an example of an adequate response to an item that focuses on seeing if high school students can pose relevant research questions after examining data readings about water temperature and pH in a particular river over a two-month period.

Figure 2. Example of item and exemplar on a GLOBE Integrated Investigation Assessment

Prompt: Holcomb Elementary and Jefferson Elementary are two schools located within 5 miles of each other in Fayetteville, Arkansas. Both schools sit next to the same river, with Holcomb located upstream from Jefferson. Even though the schools are relatively close to each other, the plant and fish life appears to be different between the two sections of the riv You and several other students have been asked to report to your science class what some o the differences are and why you think they exist. To the left is data from the two schools between late November and late January to help you in your investigation. Look at the GLOBE data in the tables (not shown here). Think of two questions you might ask regarding the data. A sample question might be "What is unusual regarding water temperature between the two schools considering they take measurements from the same river?

*Example of adequate response: "*One question I might ask: is there any relationship between water temperature and pH? In other words, if temperature goes up, what happens to pH? Another question I might ask: is there a trend in how temperature changes over time (or how pH changes over time?) By this I mean since the measurements go from Nov. 22nd until Jan. 24th, is there an increase or decrease in either of the variables?"

The GLOBE Integrated Investigation Assessments and supporting information can be found at http://globeassessment.sri.com.

Thinking with Data and Foundational Tools for Data Literacy

The Thinking with Data and Foundational Tools for Data Literacy assessments assess upper elementary and middle school students' abilities to transfer data literacy skills and understandings learned in integrated math/science and math/social studies units to other data literacy tasks that are conceptually related. Students doing the units for which these assessments were designed examine real data sets having to do with water scarcity, pulse rate, and plant growth.

The Thinking with Data assessment assesses outcomes of an integrated 6th grade math/social studies unit about water scarcity in Middle Eastern nations. In the unit, students explore how to make fair comparisons among nations with different populations, water uses, and water availabilities. Assessed data literacy tasks include applying proportional reasoning and knowledge of the concept of per capita to evaluate the fairness of comparisons of data about different countries.

The Foundational Tools for Data Literacy assessments assess outcomes of integrated math/life science units about pulse rate and plant growth. In the unit on pulse, 4th grade students collect, organize, and analyze data about pulse rate from samples of people drawn from different populations. In the unit on plant growth, 6th grade students conduct experiments in which they grow sets of "fast plants" under different conditions, test hypotheses, and analyze results. In both units, the students use Tabletop, a computer-based data tool, to view the data and do analyses of results. Assessed data literacy skills and understandings include understanding which types of research questions can be addressed by collecting data; determining the appropriateness of different data representations for different analyses; and analyzing data distributions to a) evaluate the strength of relationships between variables, b) detect measurement error, c) detect central tendency, and d) critique the viability of conflicting claims about the data.

The assessments in both projects present students with tasks that require them to demonstrate transfer (Bransford, Brown, & Cocking, 2000) of the data literacy skills and understandings acquired in their units to other content. For example, in the Thinking with Data assessment, students identify whether they have enough information to fairly compare the severity of the car theft problem in France and Japan. In the Pulse Unit assessment, the students compare and contrast data about the heights of students who vary by age and gender. In the Fast Plants Unit assessment, the students examine the results of an experiment about how fast different breeds of kittens grow when fed different diets.

Item formats employed on these assessments are largely constructed response and justified multiple choice. In the Foundational Tools for Data Literacy assessments, several items focus on a common problem or other stimulus, such as an experiment or data analysis procedure. In almost all cases when selected response questions are posed, the students are asked to explain why they made their selections. On many items, they are scored on the extent to which their explanations show understanding of the underlying data literacy concepts. Many of the rubrics provide cues that the rater can use to differentiate between a response that:

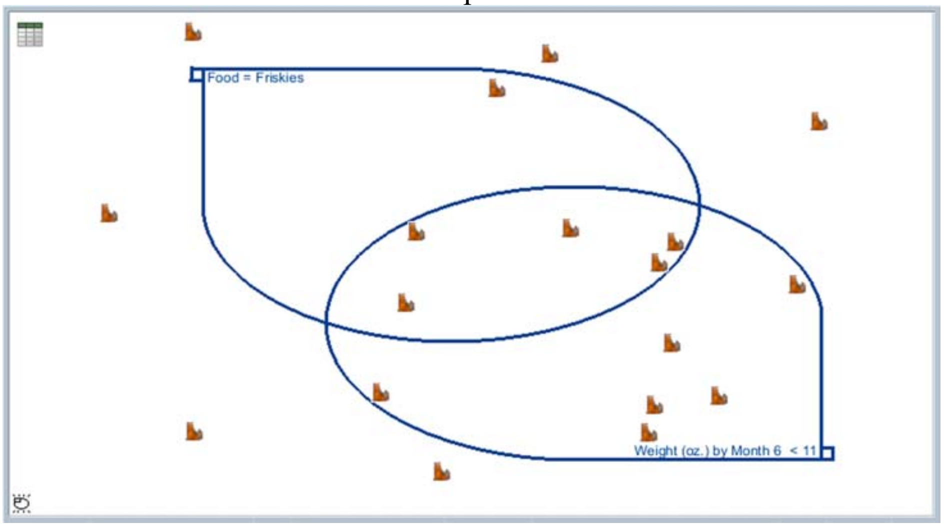
- demonstrates full understanding (Score of 3)
- is too vague to demonstrate full understanding, yet is on-track enough to infer emergent understanding (Score of 2)
- demonstrates no understanding due to it being confused, insubstantial, off-base, or inaccurate (Score of 1)

On the justified multiple choice items, student responses are coded to differentiate in the results data base between a correct selection and a rubric score for the explanation about why the particular selection was chosen. This permits examination of how many students who select correct answers also can communicate adequate explanations.

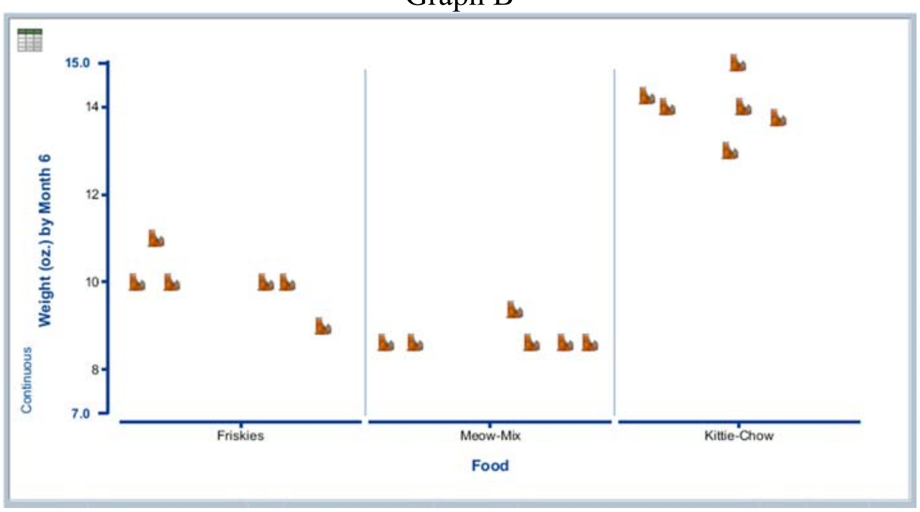
These items have proven especially worthwhile in uncovering student problems that multiple choice items alone would hide. For example, in the assessment for the Fast Plants unit, students are shown three data graphs representing data from the experiment on the different breeds of kittens and their diets. They are asked to select which of the graphs would be best for seeing if, by the end of the experiment, there was a relationship between the how much the kittens weighed and what food they ate. The graph choices are presented in Figure 3.

Figure 3. Plots of data used in Fast Plants Unit Assessment

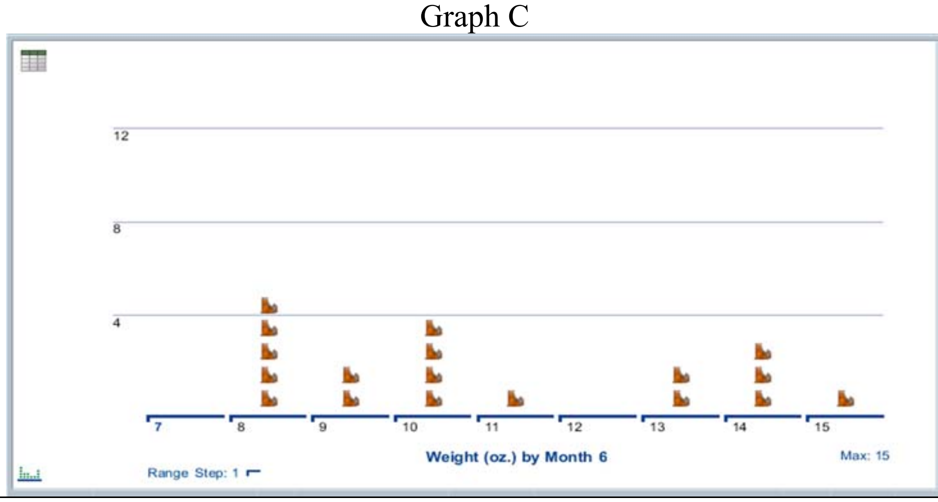
Graph A



Graph B



Graph C



The following are explanations of three students who made the correct selection of the 2nd graph, utilizing the 3-point rubric described above.

- Score of 3: "Graph B is best because it compares all weights and foods, when A & C do not."
  Explanation of score: the noting of the fact that all the weights and foods are represented on graph B demonstrates full understand of the superiority of that graph for the intended analysis
- Score of 2: "Because it is the most exact and it's easy to see the differences."
  Explanation of score: the noting that graph B is "the most exact" is on the right track, but too imprecise to qualify as a demonstration full understanding
- Score of 1: "In my opinion, graphs like that are always the easiest to read. But Graph C would have been my second choice. Graph A I couldn't understand."
  Explanation of score: Despite the fact that the student selected the correct graph in the multiple choice question, there is no evidence in this response to indicate any understanding. The explanation is insubstantial and the student admits confusion.

The Thinking with Data and Foundational Tools in Data Literacy projects are still in progress. Hence, reports on this work are not yet available.

Conclusion

Working with data is central to scientific inquiry, in the geosciences as in other sciences. Technological advances in data access and data visualization have created unprecedented opportunities for students and teachers to use real data sets as vehicles to understanding and applying the epistemologies and research practices of the different geoscientific disciplines. At the same time, immersing students in data as a vehicle for improved geoscience education has its risks. Students may or may not come to a data-immersive geoscience class prepared to handle the components of the curriculum that require foundational data literacy proficiencies. Without this background, they may be ill-equipped to carry out data-rich tasks that require them to learn and apply the proficiencies to the distinctive requirements of research in the respective geoscience disciplines.

Part of this problem stems from the fact that, in most K-12 schools, the teaching of the fundamentals of data literacy straddle content areas. Instruction about data, especially about central tendency, graph interpretation, proportional reasoning, and statistics occurs in math class, yet usually in isolation of scientific inquiry. Other data skills and understandings having to do with sampling, distribution, measurement error, natural variability and other data-related facets of inquiry receive less, if any attention, in math class. This leaves it to the science or social studies teacher to teach these other data literacy fundamentals, and many do not.

In the future, more assessments measuring foundational data literacy could be designed in a systematic, comprehensive manner. They could conform and align to a developmental set of benchmarks about what students should be able to know and do with data at different points in their education. These could support and render more systematic and explicit the school's role in building foundational data literacy skills and understandings that can then be adapted and applied to the different scientific disciplines, in the geosciences and in other areas. To develop such a resource bank of assessments, a challenge would be to differentiate between foundational aspects of data literacy and aspects that cannot be taught or assessed independently of their application in specific scientific or social scientific disciplines.

References

Bransford, J. D., Brown, A. L., and Cocking, R. R. editors (1999) *How People Learn: Brain, Mind, Experience, and School.* National Academy Press: Washington, D.C.

National Council for the Social Studies. *Expectations of Excellence, Curriculum Standards for Social Studies.* Washington: National Council for the Social Studies, 1994.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

Quellmalz, E., Kreikemeier, P., Rosenquist, A., & Hinojosa, T. (2001). *Aligning GLOBE to National and International Standards*. Paper presented at the Sixth Annual GLOBE Conference, Blaine, WA.

Quellmalz, E.S., Zalles, D. (2002). *Integrative Performance Assessments in Technology*. SRI International.

Taylor, Catherine S., Nolen, Susan Bobbit. (2004). *Classroom Assessment: Supporting Teaching and Learning in Real Classrooms*. New Jersey: Prentice-Hall, Inc.