AC 2007-1097: HAVE THEY GOT IT YET? ASSESSING STUDENT UNDERSTANDING OF DIFFICULT CONCEPTS

Paul Santi, Colorado School of Mines

Have They Got it Yet? Assessing Student Understanding of Difficult Concepts

Introduction

Some of our most challenging teaching is aimed at identifying and correcting misconceptions. Misconceptions may be defined as facts, processes, or models that are flawed or miscategorized and of which the holder is unaware[1]. They are usually considered to be robust and resistant to change, because the misconceived understanding can often provide consistent and predictable explanations when tested[2, 3]. Furthermore, ideas may be miscategorized across ontological boundaries (believing, for instance, that electricity is a substance that can be stored or leaked out, rather than a process)[2, 4].

Various teaching tools have been proposed to help students identify and correct their misconceptions, such concept tests[3, 5, 6, 7], active learning exercises[8, 9], group psychotherapy techniques[1], small group discussions[3], and tutoring sessions[10]. The effectiveness of these tools has not been clearly evaluated, in terms of addressing misconceptions. Such an assessment would be useful to identify appropriate tools for different problems, and to modify and improve tools. The purpose of this study is to report on the author's experiences using two assessment approaches to demonstrate correction of misconceptions.

Requirements for Valid Assessment

Appropriate assessment must identify improved understanding on the part of the students, and not simply improved fact retention. It is widely accepted that true misconceptions are quite difficult to identify[5, 7] and tricky to correct[2, 4]; it follows that it is probably also difficult to test for and verify that they have indeed been corrected.

Similar to tests created to identify misconceptions (see ref. 6, for example), assessments of their correction should challenge the student's model in ways that isolate weaknesses in the model to explain various observations. This is frequently done by using answers derived from misconceived notions as distracters in multiple choice questions[6]. In discussion-type assessments, probing of the student's response with deeper follow-up questions should also reveal weaknesses in their understanding.

The timing of assessment with respect to the instruction also plays a role. The assessment must take place soon enough after the instruction that students do not forget the concepts taught, yet any secondary or post-testing should isolate improvement resulting from intentional instruction aimed at repairing the misconception from improvement resulting from additional study time.

Based on these rough boundaries, the author developed and administered two different assessments to gauge the effectiveness of instruction aimed directly at specific misconceptions in two Geological Engineering classes. A written multiple choice test was given to sophomores in a Geomorphology class, and a short answer quiz followed by individual interviews was given to seniors in a Site Investigation class. Details of the tests and outcomes are described below.

Assessment with Concept Tests

The first assessment approach involved multiple choice concept tests. In four successive classes, students received initial instruction, a pre-test, revised instruction, and a post-test. Revised instruction consisted of carefully selected and explained analogies and short group exercises using Santi and Santi[1] as a guideline.

Two misconceptions were addressed. The first misconception deals with the topic of base level, which is the equilibrium level, expressed by a curvilinear surface, to which streams will erode or deposit material when adjusting to changes in sea level or land uplift or subsidence. The process is time dependent, and deposition and erosion can be occurring simultaneously along different reaches of a river, as the response to changes in land or ocean levels starts at the downstream end and migrates upstream over time. Initial instruction consisted of a discussion of the process, complete with graphical examples drawn on the chalkboard. Students' misconceptions tend to hinge on the notion that the response to change is immediate throughout the whole system. Revised instruction consisted of development of a detailed analogy, where the stream is represented by several conveyor belts in series that can be raised and lowered at either end (modeling deposition and erosion, respectively) and the terminus of the stream is represented by a large table, which may also be raised or lowered (modeling a drop or rise in sea level, respectively). The table and each conveyor belt must match the adjoining ones in elevation. Students were led through a few examples of changing conditions and asked both how the analogy model would respond and how the real world would behave similarly or differently.

The second misconception is the process of how pediment surfaces are developed. Common in the western U.S., pediments are large, smooth planar surfaces that extend from the center region of river valleys and curve up to the mountain front. Often several surfaces at different elevations and of different ages are present in the same area. They are formed partly by vertical downcutting of streams and partly by lateral sweeping erosion of streams. Initial instruction consisted of observing pediments on a field trip, with both field and classroom discussion on the processes that create them. Students' misconceptions are that pediments are created only by downcutting erosion, with no lateral cutting component from the streams; that they are exclusively an erosional feature, even though there is a substantial cap of material deposited by the sweeping stream; and that once a second set of pediments begins forming, the first set stops changing in size. Revised instruction consisted of a short in-class group exercise, where students were asked to visualize a machine to make pediments and to describe what the machine had to do.

Two concept tests were developed for testing before and after the revised instruction. Each test contained a diagram and three to five multiple choice questions for each misconception. The class was divided into two groups: the first group was given test version A (reproduced as Figure 1) as a pre-test and B (Figure 2) as a post-test, and the second group was given version B as a pre-test and A as a post-test. This was done in case one of the tests proved harder than the other, thereby confounding the interpretation of a trend in the scores.

Concept Test A

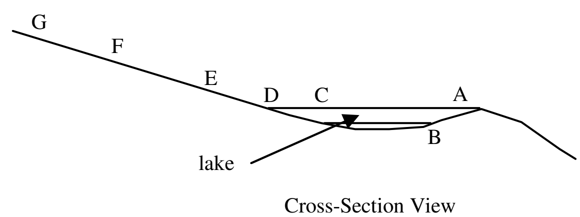
During a drought, the lake level drops from A to B.

1. Which is the first to happen?
A. erosion at D
B. erosion at G
C. deposition at C
D. deposition at E

Several wet years follow the drought and raise the lake level back to A.

2. Deposition quickly starts at C
A. True
B. False

3. Erosion continues for a short time at G
A. True
B. False



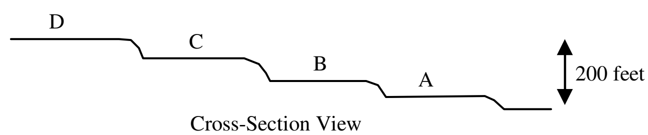
4. A nick point forms
A. True
B. False

5. A delta forms
A. True
B. False

---

6. Which pediment is the oldest?
A. A
B. B
C. C
D. D

7. Bedrock is approximately 15 feet below the surface of pediment B. How deep would you expect bedrock below the surface of pediment D?
A. 0 feet
B. 15 feet
C. 45 feet
D. 150 feet



8. In 5000 years, you would expect
A. pediment A to be larger
B. all of these pediments to be substantially lower
C. all of these pediments to be smaller in size
D. very little change

Figure 1. Concept Test A. Questions 1-5 deal with misconceptions related to stream base level, and questions 6-8 deal with misconceptions related to pediment formation.

Concept Test B

As sea level rises:

1. Active deposition on the delta shifts from B to A
A. True
B. False

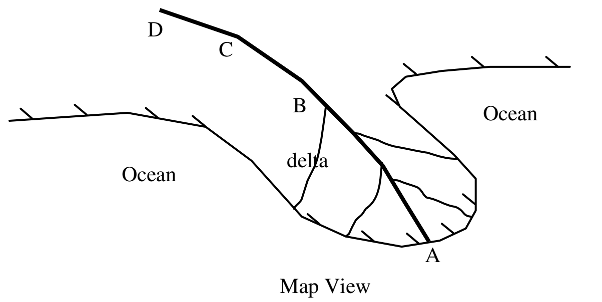
2. Erosion at D slows down
A. True
B. False

As sea level drops:

3. A nick point forms
A. True
B. False

4.The delta builds beyond A
A. True
B. False

5. Erosion at C
A. starts immediately
B. starts after some time
C. does not occur



If the surfaces A, B, C, and D are pediments:

6. Which is the youngest?
A. A
B. B
C. C
D. D

7. Which one is most likely to be getting larger?
A. A
B. B
C. C
D. D

8. What did this look like 5000 years ago?
A. all of the pediments were higher
B. all of the pediments were lower
C. A was larger
D. D was larger

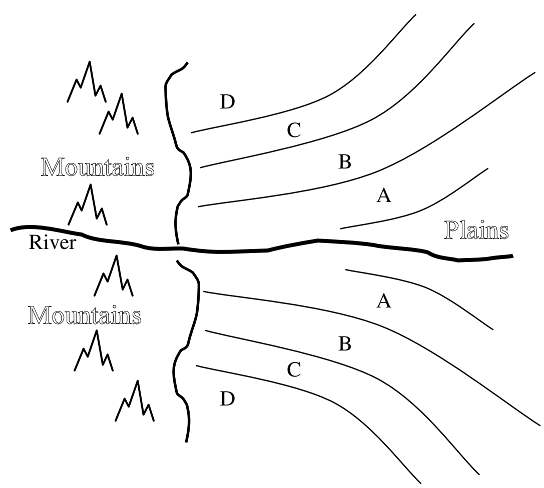


Figure 2. Concept Test B. Questions 1-5 deal with misconceptions related to stream base level, and questions 6-8 deal with misconceptions related to pediment formation.

In fact, test results showed that Test A was more difficult than Test B (Table 1), as average scores for Test A were 4.37 when given as a pre-test and 4.18 when given as a post-test, while average scores for Test B were 6.35 for either case. There is no statistical difference (at the 5% level of significance) between Test A scores when given as a pre-test or as a post-test, yet there is a difference between scores on tests A and B, regardless of which was the pre- and post-test. Further indication of the difference in difficulty of the tests is shown by the change in individual scores, tabulated in the second part of Table 1: Group 1 showed significant positive improvement when taking Test B as a post-test, yet Group 2 showed nearly identical negative impacts when taking Test A as a post-test.

Table 1. Results of multiple choice assessment testing (Figures 1 and 2). Scores are out of eight possible points. Group 1 consisted of 18 students and Group 2 consisted of 15.

	Group 1		Group 2	
	A pre-test	B post-test	B pre-test	A post-test
Average score	4.37	6.35	6.35	4.18
	Increase score	No change	Decrease score	
Group 1	14	2	2	
Group 2	0	1	14	

From this data, the author concludes that the two assessment tests were substantially different in difficulty, and could not demonstrate improvement in understanding following the revised instruction. Other possible problems that could not be judged because of the problematic test scores are that too much time may have elapsed between the original explanation of the concept and testing, or that the revised instruction method to correct the misconception was not effective or simply added complication to their understanding, rather than clarification.

Assessment with Interview or Tutoring Sessions

The second assessment approach aimed to correct the problems identified in the first assessment, and to incorporate research that indicates that what matters most in understanding is the opportunity for people to explain things to themselves (or to others)[10]. In this case, the revised instruction included more student-student interaction, and the post-test consisted of a short interview where revisions in the student's understanding were probed.

Two misconceptions were addressed in this class as well. The first involved the information needed and the nature of the questions that should be asked in order to select the proper drilling and sampling methods for subsurface drilling investigations. Over several years of teaching a class in Site Investigation, the author has noticed that students are mystified as to which components of the project tend to drive the selection of drilling and sampling methods: Is it the soil type? The purpose of the investigation? The level of detail required? As a consequence, many of them focus on the wrong parameters or improperly weight the various contributing parameters. Initial instruction consisted of several lectures on equipment and capabilities. Revised instruction included several elements: a homework exercise in which they develop a flowchart decision tree to select drilling and sampling methods, several short in-class situational problems where they select drilling and sampling methods, and a one-hour role playing exercise during which they must choose investigation methods for a specific site.

The second misconception relates to the differences between dual and borderline soil classification types, as defined by the Unified Soil Classification System (USCS). Students often do not recognize the difference between the two and do not understand how they are distinct. Initial instruction consisted of a lecture during which the USCS was described and the students practice classifying several soils. Revised instruction consisted of a think-pair-share exercise[11] on the subject, followed by a reframing homework assignment in which the students graphically plot regions representing different soils types for the USCS and one other classification system.

After initial instruction, students took a short essay quiz for each assignment. The purpose of the quizzes was not to produce a score, but to provide a baseline of their understanding to serve as discussion points later. The questions for drilling and sampling method selection were:

1. You are coordinating soil drilling for a site you know nothing about. What sort of drilling and sampling do you propose?
2. What should you ask your boss about the project goals to clarify #1?
3. What geologic characteristics would you try to find out to clarify #1?

Grades were assigned as follows:

A – answer demonstrates confidence in their own analysis and opinion. Answer is succinct and they ask incisive questions to arrive at their answer.
B – student appears to be processing the questions and developing their own answers. They ask the right questions to clarify drilling and sampling.
C – student understands the concept well enough to repeat back ideas from class. They ask thoughtful questions, but not the ideal ones to clarify drilling and sampling.
D – student gives a reasonable answer, but seems to be guessing. Some questions are not realistic or have nothing to do with drilling and sampling.

The questions for the USCS misconception are:

1. Define the dual classification for the USCS.
2. How does it differ from borderline classifications?

Grades were assigned as follows:

A – includes an accurate description of both concepts, identifies the idea of "in between" for borderline materials.
B – their description is not incorrect, but they fail to include the idea of a "range" of material types or "in between" for borderline materials.
C – student's description has some incorrect notions.
D – no correct ideas are indicated.

Post-assessment was completed after the revised instruction and consisted of individual interviews during a lab period. Students were asked why they answered as they did on the essay quizzes, and then asked to explain how their views have changed in the time since the quizzes. The instructor attempted to clarify any residual misconceptions during the interview. Although subjective, two scores were assigned to the student: their level of understanding represented by the quiz and their level of understanding at the start of the interview.

The results of the assessments for the Site Investigation class are included in Table 2. Note that both pre- and post-instruction results are shown, with pre-testing values indicated within rows and post-testing indicated within columns. For example, for drilling misconceptions (left-side of the table), 11 students received a "C" during the pre-test, but during the post-test, 10 of those students showed "B" level understanding and one student remained at the "C" level.

Both pairs of tests show a statistically significant improvement in scores (at the 5% level of significance), and it is clearly shown on Table 2 that the level of understanding improved for many students. While improvement was demonstrated, even with a small sample set, it was not possible to isolate the effect of any individual element of revised instruction, as several elements were employed.

Table 2. Results of interview assessment testing. Values on left are pre- and post-assessment grades for drilling misconceptions and values on the right are for USCS misconceptions.

Pre-Testing Scores	Post-Testing Scores			
	D	C	B	A
D	--	--	--	--
C	--	1	10	--
B	--	--	5	3
A	--	--	--	1

Pre-Testing Scores	Post-Testing Scores			
	D	C	B	A
D	1	--	1	--
C	--	1	1	2
B	--	--	4	1
A	--	--	--	7

It should also be noted that the interview assessment was free-form and did not use scripted questions to probe the students understanding, but rather used the student's own answers as a springboard for discussion. Because of this, the interviewer had to carefully word questions to avoid revealing or telegraphing the correct answer. For instance, if the student indicated preference for a drilling method that was not ideal, the interviewer should ask follow-up questions such as "why did you select this method?" or "why did you prefer this method over another method?" rather than asking "might there be some problems with this method?"

While the interview method was much more subjective and time consuming than multiple-choice testing, it offered excellent opportunities to address misconceptions. In the author's experience for these exercises, misconceptions were identified and responded to in every case where the student had not been awarded an "A" score.

Conclusions

Assessment of the quality of instruction to correct misconceptions is very sensitive to the assessment tool. Multiple choice tests are desirable, as they are quick to administer and evaluate and give a single objective score. However, we have seen that it is difficult to develop pre- and post-tests that are comparable in difficulty, and future work should re-evaluate this method of assessment, but with tests of comparable difficulty. On the other hand, individual interviews allowed a much clearer assessment of the student's understanding, even though the rating system had to be subjective, and it allowed for the instructor to address key components of the student's misconceptions. Pre-test questions for the interview assessment must be carefully worded such that correct answers can be succinct or at least easy to compare to a pre-established metric that measures whether they "get it" or not.

Improvement of understanding was possible to gauge with the interview assessment, but timing issues made in impossible to isolate the effect of a single component of the revised instruction. Future testing should use a shorter time frame, with post-instruction interviews occurring promptly after a single revised instruction period. Relying on homework assignments as the revised instruction method adds too many days during which other modes of instruction may occur, so that the positive effects of the homework are difficult to judge.

Bibiography

1) Santi, P.M. and Santi, M.M., 2006, Using Psychotherapy Techniques to Reveal Misconceptions and Improve Learning, Journal of Excellence in College Teaching, Vol. 17, No. 3, pp. 115-142.

2) Chi, M.T.H. and Roscoe, R.D., 2002, The Processes and Challenges of Conceptual Change. In M. Limon and L. Mason (Eds.), Reconsidering Conceptual Change. Issues in Theory and Practice. Netherlands: Kluwer Academic Publishers, pp. 3-27.

3) National Research Council, 1997, Science Teaching Reconsidered, A handbook. Committee on Undergraduate Science Education. Washington, DC: National Academy Press, 88 p.

4) Chi, M.T.H., 2005, Commonsense Conceptions of Emergent Processes: Why Some Misconceptions are Robust. Journal of the Learning Sciences, Vol. 14 No. 2, pp. 161-199.

5) Evans, D.L. (moderator), Midkiff, C., Miller, R.L., Morgan, J., Krause, S., Martin, et al., 2002, Tools for Assessing Conceptual Understanding in the Engineering Sciences, Panel conducted at the 32nd Annual Frontiers in Education Conference, Boston.

6) Olds, B.M., Streveler, R.A., Miller, R.L., and Nelson, M.A., 2004, June, Preliminary Results from the Development of a Concept Inventory in Thermal and Transport Science, Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition, Salt Lake City.

7) Streveler, R.A., Nelson, M.A., Olds, B.M., and Miller, R.L., 2003, Why are Some Science and Engineering Concepts so Difficult to Learn? Identifying, Assessing, and "Repairing" Student Misunderstandings of Important Concepts, 33rd Frontiers in Education Conference, Boulder, CO.

8) Felder, R.M. and Brent, R., 2004, The Intellectual Development of Science and Engineering Students. Part 1: Models and Challenges, Journal of Engineering Education, Vol. 93 No. 4, pp. 269-277.

9) Terenzini, P.T., Cabrera, A.F., Colbeck, C.L., Parente, J.M., and Bjorklund, S.A., 2001, Collaborative Learning vs. Lecture/Discussion: Students' Reported Learning Gains, Journal of Engineering Education, Vol. 90 No. 1, pp. 123-130.

10) Chi, M.T.H., DeLeeuw, N., Chiu, M.H., and Lavancher, C., 1994, Eliciting Self-Explanations Improves Understanding, Cognitive Science: A Multidisciplinary Journal, Vol. 18, No. 3, pp. 439-477.

11) Lyman, F. 1981, The Responsive Classroom Discussion, In Anderson, A. S. (Ed.), Mainstreaming Digest, College Park, MD: University of Maryland College of Education.