UNDERSTANDING UNDERGRADUATE STUDENTS' CONCEPTIONS IN SCIENCE: USING LEXICAL ANALYSIS SOFTWARE TO ANALYZE STUDENTS' CONSTRUCTED RESPONSES IN BIOLOGY

We explored the feasibility of using lexical analysis software to help identify biological concepts in students' open-ended responses. We created two items in an on-line course management system where we asked students to trace carbon during cellular respiration. Data were collected during the fall semesters of 2004 and 2006. We created a custom library of biological terms using SPSS Text Analysis for Surveys which we used to analyze students' open-ended responses. The software correctly classified 90% of student responses. Comparing pre- to post-tests, a larger proportion of students provided more accurate responses on the post-test, but also the proportion of students that provided vague responses slightly increased. Additionally, students seem to understand where the carbon is, but they are either confused about the processes by which it got there or they understand the processes, but do not know the compounds. Lexical analysis software has potential to help instructors assess students' conceptual barriers to improve science instruction.

Rosa A. Moscarella, Department of Zoology, Michigan State University

Mark Urban-Lurain, Division of Science and Mathematics Education, Michigan State University

John E. Merrill, Biological Science Program, Michigan State University

Gail Richmond, Colleges of Education and Natural Science, Michigan State University

Ronald Patterson, Department of Microbiology and Molecular Genetics, Michigan State University

Joyce M. Parker, Division of Science and Mathematics Education and Department of Teacher Education, Michigan State University

Brett W. Merritt, Department of Teacher Education, Michigan State University

Christopher D. Wilson, Biological Sciences Curriculum Study

Tammy M. Long, Biological Sciences Program, Michigan State University

Introduction

When it comes to important processes in science, many undergraduate students possess alternative conceptions, misunderstandings, or incomplete conceptions that can compromise their learning. These *conceptual barriers*, identifiable through students' use of language, are difficult to detect through standard assessments such as multiple choice questions (Ausubel, 1968, 2000; Bransford, Brown, & Cocking, 1999). In contrast, *constructed response* questions – in which students must use their own language in order to explain a phenomenon – create more meaningful opportunities for instructors to identify their students' learning obstacles.

Unfortunately, the large enrollment of many undergraduate introductory science courses prohibits the detailed evaluation of large numbers of constructed response answers. Our project team has been involved in a range of reform efforts in large, undergraduate science classes that align with Seymour's recommendations to refocus "classroom practice upon gains in student understanding, reasoning, application, and learning retention; [clarify] student learning goals and their alignment with course assessments; and [redesign] assessments to engage students in their own learning and to give feedback to teachers on the efficacy of their work" (Seymour, 2002, p. 85).

In this project, we build on these efforts, focusing primarily on the redesign of assessments. To do so, we propose a "strong" use of technology: "rethinking of our desired educational outcomes and the means of achieving these learning goals" (Garrison & Anderson, 2000, p. 28). We investigate the feasibility of using linguistic analysis software to help analyze student constructed response assessments to evaluate student understanding of various scientific concepts. For information on how this approach is being used to understand students reasoning, see (Richmond, Urban-Lurain, Parker, Merrill, & Merritt, 2008) in this issue and (Wilson et al., 2006).

The Role of Assessment

The word "assess" comes from the Latin "*assessus*" meaning "to sit beside" (The Merriam Webster Online Dictionary, 2005). This implies that assessment involves listening and gathering data in addition to evaluation. Because "the learner must construct his or her own meaning" (Anderson et al., 1994, p. 24) a teacher needs to understand students' reasoning in order to guide students in this process (Duit, 1995; Larochelle & Bednarz, 1998; Von Glasersfeld, 1994). This view of teaching defines the central role of assessment in instruction, i.e. to allow the teacher to follow the progress of students' understanding so that s/he can design and adjust activities to improve students' learning (Hake, 1998; Novak, Mintzes, & Wandersee, 2000; Pride, Vokos, & McDermott, 1998; Sadler, 2000).

In order to use assessment in the ways described above, an instructor must utilize questions that reveal students' understanding. These questions must be diagnostic of the difficulties students have making sense of scientific ideas. Studies show that instruments relying exclusively on typical objective (multiple-choice) questions are not adequate for this purpose. Students may correctly answer objective questions while still harboring conceptual barriers that are harming their learning process (Ausubel, 1968, 2000; Bransford et al., 1999).

Conceptual barriers may be more readily revealed using constructed response items. A constructed response question requires students to generate an answer, rather than selecting one from a set of alternatives. The definition encompasses a range of assessments from completion (fill-in-the-blank), through various performance-based assessments (Bennett, 1993), to essay questions, i.e., questions that allow completely free-form answers. More complex forms can reveal "higher order processing proficiencies" (Hattie, Jaeger, & Bond, 1999 , p. 433) and are better suited to revealing conceptual barriers than objective questions (Birenbaum & Tatsouka, 1987).

Assessment Logistic Constraints

Most introductory science courses have large enrollment numbers, and while it is certainly impossible "to sit beside" each student in the true sense of assessment, it is almost equally prohibitive to grade large numbers of constructed response answers, since only humans are truly capable of that. Thus, faculty resort to multiple-choice assessments, which can easily be computer-graded. Since large class sizes are likely to remain the norm, we need to explore ways that technology can help us understand and diagnose student conceptual barriers.

While computers are unable to “understand” free-form writing, they can assist humans in condensing text into easier to evaluate tokens or conceptual sections through computational text analysis. One such method is *lexical analysis*, which we investigate in this project.

To address the practical and logistical impediments of using constructed response items in large enrollment courses, we assess the feasibility of using lexical analysis software to analyze the conceptual content of constructed response items to address three research questions: 1) Can lexical analysis software assist instructors and/or researchers in the identification of conceptual barriers through the analysis of students’ open-ended responses? 2) Can the results of these analyses be used to create statistical models of student conceptual barriers? 3) Can the resulting models be used to predict conceptual barriers in future populations of students based on their constructed responses and inform pedagogical practices in order to address those barriers? The results presented in this article addressed the first of these questions.

Methods

The logistics of typical large-enrollment undergraduate classes restrict the options faculty have for moving towards more learner-focused instruction. Since large class sizes are likely to remain the norm, we need to explore ways that technology can help us understand and diagnose student conceptual barriers.

Our project team is reforming several large, introductory science courses in order to improve students’ ability to apply model-based reasoning in many disciplines. Our current research emphasizes models that focus on tracing matter and energy through biological systems. One way of assessing students’ ability to use such models is on questions involving tracking a phenomenon backwards from a known ending point. We think that by scaffolding student responses and asking them to trace a process in a reverse order, instructors can get a better sense of whether students are indeed reasoning these phenomena in a model-based fashion and not simply memorizing a series of steps.

In biology, we have found that students often confuse the flow of matter and energy in biological systems. This was revealed by asking questions such as the following: “Cells in an active muscle release CO_2. How did the carbon get into the CO_2?” We then sequentially repeat this question: “In what substance was it before that, and how did it get there?”, and so on.

We created two items in an on-line course management system so that we could capture the students’ responses for analysis. In this particular question, students were expected to trace the carbon backward indicating 4 processes (“how”) and 3 substances (“substance”), as shown in Figure 1. This number of steps, although somehow arbitrary, represents the level of detail instructors wished to be attained by students. Since these processes could be traced following different paths and step width, there was not a “right answer”. The table in Figure 1 shows the “expert answer,” which was used as reference in subsequent analysis.

During the fall semesters of 2004 and 2006, we collected data on these constructed response items in an introductory biology course that enrolls approximately 450 students per semester. Students submitted responses before instruction on this content ("pre-test") and again at the end of the course ("post-test"). Students received homework credit for completing the items, but were not graded on the correctness of their answers. We collected data from 823 students on the pre-exam test (433 from fall 2004 and 390 from fall 2006) and 631 students on the post-exam test (318 from fall 2004 and 313 from fall 2006).

Cells in an active muscle release CO_2. How did the carbon get into the CO_2?	**What substance was the carbon in?**	**How did it get there?**
Start here	*Carbon Dioxide*	Oxidation of carbon compounds
Before that…	Acetyl CoA	Pyruvate oxidation
Before that…	Pyruvate	Glycolysis
Before that…	Glucose	Photosynthesis

Figure 1. Example of an on-line item that asks students to trace the movement of carbon in cellular metabolism. The first substance (*carbon dioxide*) is provided as a starting point and the rest of the cells were blank. The table contains the "expert answers" that were used as reference for the computational text analysis.

Computational Text Analysis

We analyzed the responses with *SPSS Text Analysis for Surveys* software (SPSS, 2006) which is designed to classify survey responses into *categories*. The software provides several options that allow users to control the classification techniques and save custom libraries for particular domains or question categories. For our data set, the regular libraries provided with the software did not recognize most of the biological lexicon; therefore it was necessary to build a custom lexical library incorporating these terms. After the creation of the custom library, key words were extracted from students' responses. Those key words were recognized by a series of categories that were created using the "expert answer" as reference, getting each response falling in one or more categories. Once the students' responses have been categorized, *SPSS Text Analysis* can export the individual records and derived categorized data for subsequent analysis using other data analysis tools to model how students' concepts cluster.

Here we compared the proportion of students' responses in each category from the pre-exam test and post-exam test for the data collected in summer 2006. Also, we analyzed the pattern across the answers in an attempt to reveal how students are reasoning these questions.

Results

We created a custom project library with biological terms and a set of categories based on an "expert answer" (Figure 1) to categorize the students' responses. The "how did it get there?" component of the question generated 25 categories associated including relevant processes in cell respiration, such as respiration, Krebs cycle, glycolysis, and so on; the "substances" part

Table 1

Categories Created To Classify Student Responses For Each Component Of The Question: "How Did The Carbon Get Into The CO2?"

"How did it get there?"		"What substance was the carbon in?"	
Target Answer: Should be a *Processes*	Break down of sugars	Target Answer: Should be a *Compound*	Acetyl CoA
	Chemical reaction		ATP
	Electron transport chain		Carbon
	Glycolysis		Carbon dioxide
	Krebs cycle		Glucose
	Oxidative phosphorylation		Organic compound/molecules
	Oxidation		Other compounds
	Photosynthesis		Oxygen
	Physiological process		Pyruvate
	Respiration		Sugar
	Transition glycolysis-Krebs		
Target Answer: Should Not be a *Compound*	ATP	Target Answer: Should Not be a *Process*	Chemical reaction
	Other compounds		Metabolism
	Carbon dioxide		Oxidation
	Sugars		Physiological process
	Oxygen		Energy
	Energy related		Cell and cell organelles
	Human/animal body		Human/animal body
	Living organisms		Living organisms
	Nature		Nature
	Cell and cell organelles		Do not know
	Not pertinent/made up		Not pertinent/mad up
	Do not know		Vague/incomplete
	Vague/incomplete		Waste
	Waste		

of the question resulted in 23 categories that included many key compounds produced during cellular metabolism. These categories are fine-grained and can be collapsed depending on the sophistication of the student answers (Table 1). The custom project library and associated categories yield about 90% classification accuracy of student responses. Most of the unclassified responses are unrelated to the question, and could manually be placed into the “don’t know” category.

The results from the pre-exam test and post-exam test in both semesters are shown in Table 2.Only the most frequent ten categories resulting from "How did the carbon get there?" question are shown. In the pre-test, 14% of the students answered “Krebs cycle” (the “expert” answer) while 17% answered “respiration” and 14% answered other physiological processes or reactions. As we were not looking for a “correct answer” but rather the different pathways followed by students when thinking of biological systems, these last two responses indicate that they have the correct general idea about the processes involved in the production of the CO_2 molecule during cellular metabolism.

In the post-test, 36% of the students answered “Krebs cycle;” “respiration” declined to 11% and other physiological processes or reactions declined to 4%. Although the proportion of students giving more accurate responses increased remarkably, the proportion of students providing vague or unknown answers also increased slightly from pre to post-testing. These results also suggest that some students confuse “processes” and “substances”.

Table 2

Comparing pre/post results for the first "How did the carbon get there?" question

	Percent Student Responses	
Category	Pre-test	Post-test
Transition glycolysis-Krebs	2	2
Krebs cycle	14	36
Respiration	17	11
Photosynthesis	4	2
Do not know, made up, vague	19	22
Living organisms	9	6
Chemical reaction/physiological processes	14	4
Electron transport chain, oxidative phosphorylation	3	7
Carbon dioxide, sugar, other compounds	14	7
Glycolysis, breakdown of sugars	4	3

While it is useful to compare the responses on each of these sub-sections of these questions, there is potentially more revealing information to be gleaned from the patterns across the answers to these questions as a way of seeing how students are reasoning with scientific models. This type of analysis has the potential to provide insight into students’ reasoning processes by not only identifying individual correct or incorrect responses, but common types of wrong responses and the likelihood of them occurring.

Figure 2 shows the relationships among student responses across some of the questions for the post-test from fall, 2006. In this figure, each column of circles represents the percentage of students' answers that were classified into the six most frequently answered categories. The *How 1* column shows the same data shown in the right column in Table 2. The *Substance 2* column shows the answers to the question "what substance was the carbon in before that?" (the process in *How 1*.) The responses to *How 2*, *Substance 3* and *How 3* are shown in the remaining columns. (*Substance 4* and *How 4* are not included in the figure to improve readability.) The lines show the probabilities of a particular answer based on students' prior answer; probabilities < .10 are not shown in this figure to improve readability. The categories are color-coded, i.e. Vague/Do not know category is in canary, Krebs cycle is in green, and so on. The size of each circle is propositional to the percentage of responses in that category. Note that the categories in the first line represent the "expert answer".

As Figure 2 shows, students who answered "Krebs cycle" to *How 1* have a .53 probability of answering "Acetyl CoA" (the desired answer) to *Substance 2*, but only a .04 probability of answering "do not know." On the other hand, students who started off answering "do not know" to the *How 1* question have only a .07 probability of answering "Acetyl CoA" and a .67 probability of answering "do not know" again to *Substance 2*.

Discussion

The analyses show three main patterns: 1) the proportion of students responding with more accurate answers improved after instruction; 2) students tend to confuse "processes" and "substances;" and 3) students seem to understand the general model but do not know/remember the specific substance or process in each step.

Examination of Figure 2 provides insight into students' thinking processes. For example, students use different "step sizes" to trace the carbon in CO_2 as shown by the green lines which represent reasonable pathways of the system. Fifty-three percent go from Krebs cycle in *How 1* to Acetyl CoA in *Substance 2*. However, 15% of the students who answered Krebs cycle in *How 1* take a "bigger" step size directly to Pyruvate in *Substance 2*. Ten percent of the students took an even bigger step from Krebs cycle in *How 1* all the way to sugars in *Substance 2*. These answers may show that students are tracing matter correctly, but at different scales.

It is also possible that students may be making a recall association with a word that is familiar, but failing to trace the substances and processes backwards. Many students go from sugars to Glycolysis or from Glycolysis to Pyruvate, as shown by the red arrows.

The "expert answer" is not the only pathway possible to explain the release of CO2. For example, we could start with Respiration in *How*, as 11% of student did for *How 1*; then, the substance where the carbon was before was a sugar, as 31% said in *Substance 2*. However, no one answered correctly that the process that makes sugars is photosynthesis; instead, most went to glycolysis (which is in the incorrect order) or to vague/do not know. The problem here was that by starting with a big step students run out of options.

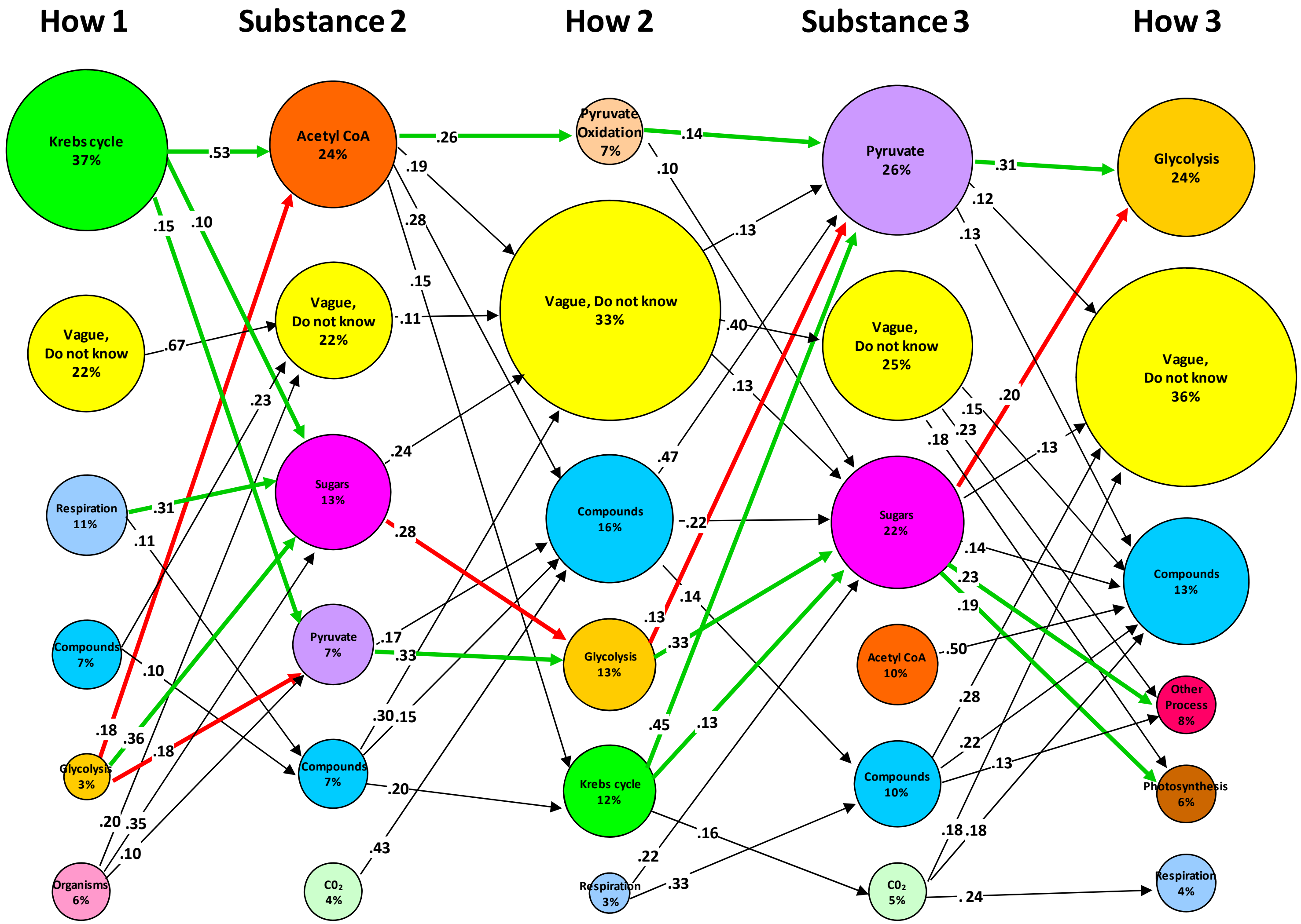

Figure 2. Relationships among answers for fall 2006 post-test.

By focusing on the path through the nodes in the figure, rather than on the individual responses, we believe we can better understand student conceptual barriers. In fact, our analysis suggests that some students understand where the carbon is, but they are confused about the processes by which it got there. Other students understand the processes, but do not know the compounds. By identifying conceptual missteps, it may be possible to change instruction to address them.

Conclusion

We found that lexical analysis software may be a powerful tool to help us assess students' conceptual barriers. The software helps instructors rapidly identify problems that students may have understanding the material. In the example of tracing matter, one of the first issues we are addressing is the distinction between processes and substances. Next, the instructor can review the exercise in class, demonstrating key steps when tracing the carbon from when is released to the point when it got into the cell.

This procedure is not only useful to analyze of data collected in the past, but also it has the potential to provide immediate formative assessment of students' learning after instruction. By having a set of appropriate categories and custom libraries, it is possible to rapidly visualize the trends in the students' responses and identify conceptual barriers that could provide insights in instructional decisions. But, more important for the objectives of our research, it makes the use of constructed response items more feasible in large enrollment courses. Since this procedure requires that responses be in electronic form, it requires a web-based course management system or other electronic means of collecting student responses. However, given the ubiquity of such systems and their use in large-enrollment courses, we do not feel that this is a barrier to adoption.

Our experience should encourage colleagues from other disciplines or institutions to experiment with lexical analysis software as a formative assessment tool. This approach may have a broad applicability, including not only undergraduate science instruction but also K-12 science classrooms.

References

Anderson, R. D., Anderson, B. L., Varanka-Martin, M. S., Romagnano, L., Bielenberg, J., Flory, M., et al. (1994). Constructivist learning and teaching. In *Issues of curriculum reform in science, mathematics and higher order thinking across the disciplines* (pp. 23-33). Washington, DC: U.S. Dept. of Education, Office of Educational Research and Improvement, Office of Research.

Ausubel, D. P. (1968). *Educational psychology; a cognitive view*. New York: Holt, Rinehart and Winston.

Ausubel, D. P. (2000). *The acquisition and retention of knowledge: A cognitive view*. Boston, MA: Kluwer Academic Publishers.

Bennett, R. E. (1993). On the meanings of constructed response. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 1-27). Hillsdale, N.J.: L. Erlbaum Associates.

Birenbaum, M., & Tatsouka, K. K. (1987). Open-ended versus multiple-choice response formats - It does make a difference for diagnostic purposes. *Applied Psychological Measurement, 11*, 329-341.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience and school*. Washington, DC: National Academy Press.

Duit, R. (1995). A constructivist view: a fashionable and fruitful paradigm for science education research and practice. In L. P. Steffe & J. Gale (Eds.), *Constructivism in Education* (pp. 3-15). Hillsdale, NJ: Lawrence Erlbaum Associates.

Garrison, R., & Anderson, T. (2000). Transforming and enhancing university teaching: Stronger and weaker technological influences. In T. D. Evans & D. Nation (Eds.), *Changing university teaching: Reflections on creating educational technologies* (pp. 24-33). London, Sterling, VA: Kogan Page Stylus Publishers.

Hake, R. R. (1998). Interactive-engagement versus traditional methods; a six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*, 64-74.

Hattie, J., Jaeger, R. M., & Bond, L. (1999). Persistent methodological questions in educational testing. In *Review of Research in Education, 24 1999* (Vol. 24, pp. 393-446).

Larochelle, M., & Bednarz, N. (1998). Constructivism and education: beyond epistemological correctness. In M. Larochelle, N. Bednarz & J. Garrison (Eds.), *Constructivism and education*. Cambridge, U.K.: Cambridge University Press.

Novak, J. D., Mintzes, J. J., & Wandersee, J. H. (2000). Learning, teaching and assessment: a human constructivist perspective. In J. J. Mintzes, J. H. Wandersee & J. D. Novak (Eds.), *Assessing science understanding: a human constructivist view* (pp. 1-13). San Diego, CA: Academic.

Pride, T. O., Vokos, S., & McDermott, L. C. (1998). The challenge of matching learning assessments to teaching goals: an example from the work-energy and impulse-momentum theorems. *American Journal of Physics, 66*, 147-157.

Richmond, G., Urban-Lurain, M., Parker, J., Merrill, J., & Merritt, B. (2008, March 30 - April 2). *Assessment-informed instructional design to support model-based reasoning* *in college-level biology*. Paper presented at the NARST 2008 Annual International Conference, Baltimore, MD.

Sadler, P. M. (2000). The relevance of multiple-choice tests is assessing science understanding. In J. J. Mintzes, J. H. Wandersee & J. D. Novak (Eds.), *Assessing science understanding: a human constructivist view* (pp. 265-296). San Diego, CA: Academic.

Seymour, E. (2002). Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology. *Science Education, 86*(1), 79-105.

SPSS. (2006). *SPSS Text analysis for surveys 2.0 user's guide*. Chicago, IL: SPSS, Inc.

Von Glasersfeld, E. (1994). A constructivist approach to teaching. In L. P. Steffe & J. Gale (Eds.), *Constructivism in Education* (pp. 3-15). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wilson, C., Anderson, C. W., Heidemann, M., Long, T., Merrill, J., Merritt, B., et al. (2006). Assessing students' ability to trace matter in dynamic systems in cell biology. *Cell Biology Education, 5*, 323-331.

Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF) under awards DUE-0243126 and EHR-0314866, Carnegie Corporation (Grant #B7458), and the Vice Provost for Libraries, Computing and Technology, Michigan State University. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Carnegie Corporation, or the Vice Provost for Libraries, Computing and Technology, Michigan State University.