Beyond Multiple Choice Exams: Using Computerized Lexical Analysis to Understand Students' Conceptual Reasoning in STEM Disciplines

Mark Urban-Lurain, Rosa A. Moscarella, Kevin C. Haudek, Emma Giese, Duncan F. Sibley, and John E. Merrill
Michigan State University, urban@msu.edu, moscarel@msu.edu, haudekke@msu.edu, gieseemm@msu.edu , sibley@msu.edu, merrill3@cns.msu.edu

Abstract - **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 than multiple choice questions. However, the realities of typical large-enrollment undergraduate classes restrict the options faculty have for moving towards more learner-focused instruction. We are exploring the use of computerized lexical analysis of students' writing in large enrollment undergraduate biology and geology courses. We have created libraries that categorize student responses with > 90% accuracy. These categories can be used to predict expert ratings of student responses with accuracy approaching inter-rater reliability among expert raters. These techniques also provide insight into students' use of analogical thinking, a fundamental part of scientific modeling. These techniques have potential for improving assessment practices across STEM disciplines.**

Index Terms – Assessment, Conceptual barriers, Constructed responses, Lexical analysis software

INTRODUCTION

There are many calls for improving STEM education at the college level, addressing teacher preparation, student motivation, college faculty preparation and rewards, class sizes, and assessment [1-8]. Assessment is crucial to all of these efforts as it is the only way we have to understand the impact of other reforms on student outcomes. Seymour points out that the central focus of STEM education reform should concentrate on moving from teaching to learning by "… redesigning assessments to engage students in their own learning and to give feedback to teachers on the efficacy of their work." [2, p. 85].

Assessments should provide information to instructors about their students' thinking so that instructors are able to help students better understand a given subject [9-11]. Constructed response assignments in which students have to use their own language to demonstrate their knowledge can provide good insight into student thinking. However, the large enrollment of many introductory courses makes it challenging to apply these kinds of assessments because evaluating them can be extremely time consuming. Because of these constraints, multiple-choice questions are often the norm, especially in large enrollment, introductory classes that are common in most STEM curricula. However, there is evidence that students may correctly answer multiple-choice questions but still harbor misconceptions, or *conceptual barriers*, that seriously compromise their learning [12-14].

The logistics of typical large-enrollment undergraduate introductory 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 uncover and understand student conceptual barriers. We face two issues that seem incompatible: using constructed response assessments *versus* time and resource constraints. Recent developments in computer software may be the answer to this dilemma. 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. We have been researching the use of computerized lexical analysis software to study students' conceptual understanding in biology, chemistry and geology by analyzing their written responses to a variety of open-ended assessments.

This paper outlines the techniques we have used, summarizes our results and discusses lessons we have learned regarding not only students' learning process but also about improving the assessment. These techniques may be applied to a wide range of STEM disciplines and should be of particular interest to faculty who teach large enrollment courses.

METHODS

We are using *SPSS Text Analysis for Surveys* v2.1[15] to study students' conceptual understanding in biology, chemistry and geology by analyzing students' written responses to a variety of open-ended items. This software is designed to classify written text 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. It has two ways to create categories: by linguistic analysis using semantic networks, term co-occurrence and term inclusion and exclusion; and by term or type frequency. We have created the *categories* for these projects combining both strategies, based on an "expert answer" in each case.

The standard libraries provided with the software do not recognize most of the technical lexicon of those disciplines; therefore it was necessary to build custom lexical libraries with those terms, as well as synonyms, abbreviations, and spelling variations or misspellings. After we created a set of custom libraries, the software extracted key words from students' responses. The key words were classified into *categories* that we created using the "expert answer" as reference, so that each response was classified into one or more categories. Once the students' responses were categorized, the individual responses and the associated categorized data were exported for subsequent data analysis. This procedure was followed in three different projects, described in the next section.

RESULTS AND DISCUSSION

Case 1. Biology: Cell Respiration

The complete description of this study can be found in Moscarella et al. [16]. Our main objective was to investigate students' ability to apply model-based reasoning in biology. We have been focusing on models to help students trace matter and energy during cell respiration [17]. We asked students to trace the CO_2 released at the end of cell respiration backwards. By asking them to trace this process in a reverse order, instructors can get a better sense of whether students are indeed applying model based reasoning and not simply memorizing a series of steps. Students were expected to trace the carbon backward indicating 4 processes ("how") and 3 substances ("substance") alternately.

We collected data during the fall semesters of 2004 and 2006 in an introductory cellular biology course that enrolls approximately 450 students per semester. Students submitted responses before instruction ("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 and 631 students on the post-exam test.

We created a custom project library with biological terms and a set of categories based on an "expert answer" to categorize the students' responses. The "how" 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 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. The custom project library and associated categories produce 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.

We compared the proportion of students' responses in each category from the pre-test to post-test and examined patterns across the answers to understand how students are reasoning about these questions. Results from the pre-test and post-test in both semesters are shown in Table I. Only the most frequent ten categories resulting from "how" component question are shown. For this question the "expert" answer was "Krebs cycle". As we were not looking for a "correct answer" but rather the different pathways followed by students when thinking of biological systems, other responses like "respiration" or other physiological processes or reactions indicate that they have the correct general idea about the processes involved in the production of the CO_2 molecule during cellular metabolism.

TABLE I
COMPARISON OF PRE/POST RESULTS FOR THE FIRST "HOW" COMPONENT OF THE QUESTION

Category	Percent Student Responses	
	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

By performing a computerized text analysis of these open-ended responses we were able to see that although the proportion of students giving more accurate responses increased substantially, 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".

We also analyzed the patterns across the answers to these questions to understand 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. By focusing on the path through the cycles, rather than on the individual responses, we believe we can better understand student conceptual barriers. 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 or forget to trace backwards. By identifying conceptual barriers, we can change instruction to address them.

Case 2. Chemistry: Acid/Base chemistry

The complete description of this project can be found in Haudek et al. [18]. The main objective of this study was to evaluate students' understanding of the basic chemistry that may be related to conceptual problems students have in cellular and molecular biology. We used text analysis to evaluate students' understanding of important biological chemistry topics and use those results to predict how experts would score the student responses.

Data were collected in fall 2008 from an introductory cellular biology course. Students were asked to answer online question sets for homework credit. We designed a question set to assess topics common to the introductory chemistry and biology courses after these topics were addressed in both courses. The first question was about acid/base chemistry: students were asked to give examples of weak and strong acids and explain the characteristics of each. The second question was about pH and functional groups (amino/hydroxyl chemistry). We asked students to answer the following multiple choice question and explain their answer choice:

Consider two small organic molecules in the cytoplasm of a cell, one with a hydroxyl group (-OH) and the other with an amino group (-NH2). Which of these small molecules (either, both or neither) is most likely to have an impact on the cytoplasmic pH?
A. Compound with amino group (correct response)
B. Compound with hydroxyl group
C. Both
D. Neither

We collected 382 student explanations. To determine if computerized lexical analysis can be used to develop reliable scoring functions for students' responses, two members of our research group with expertise in chemistry and biology independently evaluated a sample of the answers and rated them according to the following rubrics:

For the strong/weak acid questions:

Level 1: Correct definitions of strong and weak acids (e.g., Strong acids ionize completely in solution; weak acids only partially ionize in solution.)

Level 2: Correct definitions with minor errors in additional facts or reasoning (e.g., Strong acids ionize completely in solution and have very low pH; weak acids don't dissociate completely in water.)

Level 3: Correct definition for one acid, incorrect for the other (e.g., Strong acids ionize completely; weak acids do not ionize.)

Level 4: Totally incorrect/irrelevant response for both acids (e.g. Strong acids have a lower pH; weak acids have a higher pH.)

For the amino/hydroxyl questions:

Level 1: Correct description of basic nature of amino (e.g., Amino groups act as weak bases in a cell.)

Level 2: Partially correct explanation/irrelevant to question (e.g., Amino groups are molecules with a higher pH level than the hydroxyl.)

Level 3: Totally incorrect/irrelevant explanation (e.g., Amino group has two H atoms it may give up, but hydroxyl has only one OH molecule it may give up.)

For the rating of the responses for strong/weak acid question, both explanations were rated, regardless of whether the examples were correct or not. For the amino/hydroxyl question, only students who selected the correct multiple choice response (***compound with amino group***) had their explanations evaluated further using the above rubric ($N = 129$).

For the strong/weak acid question, the 150 student responses that were independently scored by the experts with very good inter-rater agreement (Cronbach's Alpha = .973, $p < .000$). The two raters were in absolute agreement on 136 of the 150 responses. Only student responses scored identically by both raters were used for further analysis.

The lexical analysis generated 27 categories for the strong and weak acid explanations. These categories included relevant terms that can potentially reveal students' understanding of acid-base principles applied to biological systems such as: pH, ionization, conjugation, etc. We used the lexical classifications of each students' answers as the independent variables in a discriminant analysis [19], with the expert classification of the student answers as the dependent variable. Stepwise analysis resulted in 12 of the categories being selected for prediction. These categories aligned with the expert rubric, providing good evidence of content and face validity.

To test the utility of the classification function, we used a cross-validation classification in which each case is classified by the functions derived from all cases other than that case. We used prior probabilities of group membership based on group size, though the results with equal prior probabilities are very similar. The results are shown in Table II. Correctly classified cases are shown in bold on the diagonal. The function classified 83.8% of the cases correctly; by chance ($p < .001$), we would expect to classify 25% of the cases correctly.

Most cases that were incorrectly classified were still within one category of the expert raters. The worst cases were the level 2 responses, where only 33.3% were correctly classified. This is likely due to that fact that only 9 of the original student responses were rated as level 2 by the experts. The low number of responses in a given rubric level makes it more difficult for the software to uncover patterns in the response categories. Therefore, the fewer response examples at a certain level, the more difficult it is for the software to correctly predict new responses at that level. This suggests that it may be reasonable to collapse levels 2 and 3 to form a three level rating scheme.

TABLE II
CLASSIFICATION PERCENTAGES OF CROSS-VALIDATED STUDENT RESPONSES FOR STRONG/WEAK ACID QUESTION CLASSIFIED AT EACH LEVEL

	Computer Predicted Rating			
Expert Rating	1	2	3	4
1	**93.4**	3.3	3.3	0.0
2	33.3	**33.3**	22.2	11.1
3	5.6	5.6	**66.7**	33.3
4	11.1	0.0	11.1	**77.8**

We used a similar procedure to analyze the amino/hydroxyl questions. Analysis of the raters' scoring of the student responses showed very good inter-rater agreement (Cronbach's Alpha = .961, $p < .000$). The two raters were in absolute agreement on 113 of the 129 responses. Only student responses rated the same by both raters were used for further analysis.

The lexical analysis generated 29 categories for the amino explanations. These categories included relevant terms that can potentially reveal students' understanding of acid-base principles applied to biological systems, such as: pH, ionization, conjugation, etc. Stepwise analysis resulted in six of the categories being selected for prediction. The cross-validation classification classified 77.0% ($p < .001$) of the cases correctly; by chance, we would expect to classify 33% of the cases correctly (Table III). Correctly classified cases are shown in bold on the diagonal.

TABLE III
PERCENT OF CROSS-VALIDATED STUDENT RESPONSES FOR AMINO QUESTION CLASSIFIED AT EACH LEVEL

	Computer Predicted Rating		
Expert Rating	1	2	3
1	**82.9**	12.2	4.9
2	21.4	**42.9**	35.7
3	6.9	12.1	**81.0**

Again, the worst cases were the level 2 responses, where only 42.9% were correctly classified. This is likely due to that fact that only 14 of the original student responses were rated as level 2 by the experts.

Case 3. Geology: Analogical reasoning

The main objective of this study is to test the hypothesis that we can make students smarter through explicit instruction and practice using analogical reasoning in a general education-science course on Global Change. We define smarter as being able to do what experts do: learn new content and apply previous knowledge to new situations. We focus on analogical transfer processes because 1) analogical reasoning is fundamental to human cognition [20-23]; 2) analogical reasoning makes us smart [24]; 3) scientific models are analogs or generate analogs that scientists use to understand natural systems [25-27]; 4) scientists use analogies to generate hypotheses and to solve research problems [28]; and 5) analogies are a major source of intellectual creativity [29]. We used computerized lexical analysis to examine the analogies generated by students.

Data for this study were obtained from analogical reasoning exercises from non-science major students enrolled in a general education Global Change class in fall 2008 (N = 180). We instructed students to develop analogies about processes in the water, rock and carbon cycles. Students practiced drawing analogies during in-class exercises and as online homework assignments. Figure 1 shows a typical online question. The students received full credit for completing the online analogy work.

For this problem we'll assume you know are trying to explain condensation to someone who is not familiar with this process. Fill in each of the boxes below. This question has 6 parts (A-F). You must submit all six in order to receive credit.

A. Condensation is like______

B. Describe one way what you wrote in Part A is different from condensation?

C. Describe a second way what you wrote in Part A is different from condensation.

D. Describe one way what you wrote in Part A is similar to condensation?

E. Describe second way what you wrote in Part A is similar to condensation?

F. Based on these similarities and differences, what can you infer about what you wrote in part A (the target)?

FIGURE 1.
A TYPICAL ONLINE ANALOGY EXERCISE. STUDENTS COMPLETE EACH QUESTION THEN SUBMIT THE ANSWER.

Students' descriptions of similarities between targets and the analogs they proposed were classified as scientific or non-scientific by two investigators. The two investigators agreed on ~80% of the classifications. A similarity was considered scientific if it involved causes or explanations of how or why something happens. For example, a student who used evaporation as an analog for photosynthesis stated that both use solar energy, a scientific similarity. Another student used an analogy of a factory and stated that in both cases there is a product. We classify this as non-scientific. We used SPSS Text Analysis for Surveys to analyze vocabulary commonly associated with the similarities that were categorized as scientific.

The percentage of students who wrote about scientific similarities between analogs and targets for five online questions increased throughout the semester from targets early in the semester, such as such as condensation and subduction, to targets late in the semester, such as degassing and carbon cycle (Figure 2). The data also suggest a potentially strong relationship between the particular target and whether or not students recognize scientific similarities as evidenced by the fact that students described a significantly higher percentage of scientific similarities when the target was photosynthesis.

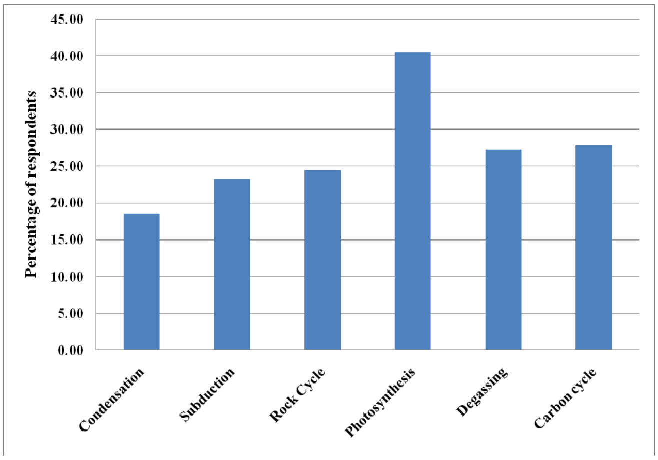

FIGURE 2.
THE PERCENTAGE OF STUDENTS DESCRIBING SCIENTIFIC SIMILARITIES BETWEEN SIX TARGET CONCEPTS AND THE ANALOGS THEY CHOSE. THE STUDENTS WERE PRESENTED THE TARGETS THROUGHOUT THE SEMESTER WITH THOSE ON THE LEFT (I.E. CONDENSATION) AT THE BEGINNING OF THE SEMESTER AND THOSE ON THE RIGHT TOWARD THE END OF THE SEMESTER.

Vocabulary commonly associated with the similarities that were categorized as scientific is shown in Table IV. Examining Table IV shows for example, 4 students mentioned energy when describing similarities between condensation and their analog, 2 mentioned energy when describing subduction, and 8 mentioned it when describing the rock cycle, etc. These results suggest that 1) non-science majors can draw scientifically useful analogies, but the majority of students draw analogies that do not afford cause or effect inferences; 2) students' analogies become more scientific throughout the semester but the effect is small; and 3) students' scientific analogies can be recognized by relatively few terms.

TABLE IV.
TERMS USED BY STUDENTS WHEN WRITING SCIENTIFIC SIMILARITIES BETWEEN TARGETS, UPPER ROW AND ANALOGS THAT STUDENTS CHOSE.

	Con-den-sa-tion	Sub-duc-tion	Rock Cycle	Photo-syn-thesis	De-gass-ing	Car-bon cycle
Energy	4	2	8	54	1	4
Heat	4	4	3	4	8	2
Bonds	4					2
Temperature	18		5		21	1
Weathering			17			3
Density		22				
State Change						19
Dissolve					3	
Gravity		1				
Pressure		2	2			2

LESSONS LEARNED

As a result of our research, we have not only gained insights into students' understanding of various scientific subjects, but we also are 1) learning how to ask questions so that responses are better structured for analyses; 2) learning effective ways to build custom libraries; 3) gaining insights into optimal numbers and specificity of categories; and 4) learning about optimal granularity in classification rubrics used to rate student responses. Specifically, some of the lesson we have learned are:

- Questions need to be stated in a way that correct and incorrect responses can be discriminated.
- Question statements should not include key words that may be important to classify responses because the students usually repeat the terms in their answers.
- Responses need to be long enough (but not too long) to provide enough key words that would allow a more accurate classification.
- It is more efficient to include inflections of verbs when they are included in custom libraries.
- Categorization should be fine grained. Specific categories can be collapsed if required by further analysis but broad categories are more difficult to disaggregate.
- A 3-level classification rubric (roughly good, regular, and bad) works well for a discriminant analysis to predict the score of students' responses

CONCLUSIONS

Constructed response assessments provide greater insight into student thinking than closed-form assessments such as multiple choice questions. However, using open-ended response items in large enrollment courses is a logistic challenge. Lexical analysis software may help us overcome these logistic barriers. This procedure is not only useful for research into student understanding, 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 to support Just-in-Time-Teaching [30]. The ability to see the term usage and category connections for an entire class at once glance is valuable for the instructor to assess class understanding as well as modify instruction when necessary. Further analysis can help identify vital concepts in student explanations and/or misconceptions. Coupling this lexical analysis with statistical classification functions of student responses opens the possibility for the critical evaluation of a large number of constructed response items.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under awards DUE-0736952, DUE-0243126 and EHR-0314866; the 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 funding agencies.

REFERENCES

[1] J. Gess-Newsome, A. Johnston, and S. Woodbury, "Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching," *American Educational Research Journal,* vol. 40, pp. 731-767, Fall 2003.

[2] E. Seymour, "Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology," *Science Education,* vol. 86, pp. 79-105, January 2002.

[3] M. A. Ruiz-Primo, R. J. Shavelson, L. Hamilton, and S. Klein, "On the evaluation of systemic science education reform: Searching for instructional sensitivity," *Journal of Research in Science Teaching,* vol. 39, pp. 369-393, April 2002.

[4] C. M. Kardash and M. L. Wallace, "The perceptions of science classes survey: What undergraduate science reform efforts really need to address," *Journal of Educational Psychology,* vol. 93, pp. 199-210, March 2001.

[5] Committee on Undergraduate Science Education, *Transforming undergraduate education in science, mathematics, engineering, and technology*. Washington, DC: National Academy Press, 1999.

[6] E. Seymour and N. M. Hewitt, *Talking about leaving: why undergraduates leave the sciences*. Boulder, Colo.: Westview Press, 1997.

[7] National Science Foundation, "Shaping the future: New expectations for undergraduate education in science, mathematics, engineering and technology," National Science Foundation, Directorate for Education and Human Resources, Washington, DC 1996.

[8] S. Tobias, *They're not dumb, they're different: Stalking the second tier*, 94 ed. Tucson, AZ: Research Corporation, 1990.

[9] E. Von Glasersfeld, "A constructivist approach to teaching," in *Constructivism in Education*, L. P. Steffe and J. Gale, Eds. Hillsdale, NJ: Lawrence Erlbaum Associates, 1994, pp. 3-15.

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

[11] M. Larochelle and N. Bednarz, "Constructivism and education: beyond epistemological correctness," in *Constructivism and education*, M. Larochelle, N. Bednarz, and J. Garrison, Eds. Cambridge, U.K.: Cambridge University Press, 1998.

[12] J. D. Bransford, A. L. Brown, and R. R. Cocking, "How people learn: Brain, mind, experience and school," Washington, DC: National Academy Press, 1999, p. 319.

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

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

[15] SPSS, *SPSS Text analysis for surveys 2.0 user's guide*. Chicago, IL: SPSS, Inc., 2006.

[16] R. A. Moscarella, M. Urban-Lurain, B. Merritt, T. Long, G. Richmond, J. Merrill, J. Parker, R. Patterson, and C. Wilson, "Understanding undergraduate students' conceptions in science: Using lexical analysis software to analyze students' constructed responses in biology," in *NARST 2008 Annual International Conference*, Baltimore, MD, 2008.

[17] C. Wilson, C. W. Anderson, M. Heidemann, T. Long, J. Merrill, B. Merritt, G. Richmond, D. Sibley, and J. Parker, "Assessing students' ability to trace matter in dynamic systems in cell biology," *Cell Biology Education,* vol. 5, pp. 323-331, 2006.

[18] K. Haudek, R. A. Moscarella, M. Urban-Lurain, J. Merrill, R. Sweeder, and G. Richmond, "Using lexical analysis software to understand student knowledge transfer between chemistry and biology," in *National Association of Research in Science Teaching Annual Conference*, Garden Grove, CA, 2009.

[19] J. Spicer, *Making sense of multivariate data analysis*. Thousand Oaks, Calif.: Sage Publications, 2005.

[20] D. Hofstadter, "Epilogue: analogy as the core of cognition," in *The Analogical Mind*, D. Gentner, K. J. Holyoak, and B. N. Kokinov, Eds. Cambridge, MA: MIT Press, 2001, pp. 499-538.

[21] D. Hofstadter, *I am a strange loop*. New York: Basic Books, 2007.

[22] K. J. Holyoak, "Analogy," in *The analogical mind*, D. Gentner, K. J. Holyoak, and B. N. Kokinov, Eds. Cambridge, MA: MIT Press, 2005, pp. 117-142.

[23] S. Pinker, *The stuff of thought*. New York: Penguin, 2007.

[24] D. Gentner, "Why we're so smart," in *Language in mind: Advances in the study of language and thought*, D. Gentner and S. Goldin-Meadow, Eds. Cambridge, MA: MIT Press, 2003, pp. 195-235.

[25] R. Frodeman, "Geological reasoning: Geology as an interpretive and historical science," *Geological Society of America Bulletin,* vol. 107, pp. 960-968, 1995.

[26] M. Hesse, *Models and analogies in science*. Illinois: University of Dotre Dame Press, 1966.

[27] M. Hesse, "Models and analogies," in *A companion to the philosophy of science*, Newton-Smith, Ed. Malden, MA: Blackwell Publication, 2000, pp. 299-307.

[28] K. Dunbar, "How scientists think in the real world: implications for science education," *Journal of Applied Developmental Psychology,* vol. 21, pp. 49-58, 2000.

[29] M. Root-Bernstein and R. Root-Bernstein, *Sparks of genius*. Boston, MA: Houghton Mifflin, 1999.

[30] G. M. Novak, A. Gavrini, W. Christian, and E. Patterson, *Just-in-time teaching: Blending active learning with web technology*. Upper Saddle River NJ: Prentice Hall, 1999.

AUTHOR INFORMATION

Mark Urban-Lurain, Director of Instructional Technology Research & Development, Division of Science and Mathematics Education, Michigan State University, urban@msu.edu.

Rosa A. Moscarella, Graduate Research Assistant, Division of Science and Mathematics Education, Michigan State University, moscarel@msu.edu.

Kevin C. Haudek, Research Associate, Center for Research on College Science Teaching and Learning, Michigan State University, haudekke@msu.edu.

Emma Giese, Undergraduate Research Assistant, Center for Research on College Science Teaching and Learning, Michigan State University, gieseemm@msu.edu.

Duncan F. Sibley, Director of Center for Research on College Science Teaching and Learning, Michigan State University, sibley@msu.edu.

John E. Merrill, Director of Bioscience Program, Michigan State University, merrill3@msu.edu.