Examining student constructed explanations of thermodynamics using lexical analysis

Luanna Prevost, Kevin Haudek, Mark Urban-Lurain
Center for Engineering Education Research
Michigan State University
East Lansing, MI
prevostl@msu.edu, haudekke@msu.edu, urban@msu.edu

John Merrill
Biological Sciences Program
Michigan State University
East Lansing, MI
merrill3@msu.edu

***Abstract*—Thermodynamics can be challenging to students, thus improving thermodynamics instruction and assessment is an important area of science and engineering education research. Constructed response assessments can reveal the complexity of students' ideas about thermodynamics. We investigate the use of lexical analysis software for examining students' constructed responses using a group of three questions related to reaction thermodynamics. These questions were administered to students in a large enrollment undergraduate introductory science course and examined learning at two levels of Bloom's Taxonomy: comprehension and application.**

Our results show that students are able to identify correct statements about thermodynamics in a multiple choice context but fail to construct correct explanations using thermodynamic concepts. Lexical analysis revealed that students who provided correct explanations incorporated more correct concepts and/or made more connections among these concepts than did students with incorrect explanations. Lexical analysis provided insight into student understanding by revealing heterogeneous ideas that were masked in multiple choice versions of the assessment.

Keywords-component; lexical analysis, text analysis, thermodynamics, assessment, Bloom's taxonomy

I. INTRODUCTION

A basic understanding of thermodynamics is essential to learning in STEM disciplines in order to understand the relations between different forms of energy in physical, chemical and biological systems. Science and engineering education research has found thermodynamics concepts to be particularly challenging for students. Students have difficulty distinguishing among endothermic and exothermic reactions [1], and struggle with concepts such as entropy and reversible and irreversible reactions [2] and heat transfer [3].

The need for improved learning continues to drive instruction and assessment of thermodynamics. Current instructional research focuses on conceptual change through active learning including the use of models and simulations [3–5]. Assessment research has focused on the development of concept inventories such as the Thermodynamics Concept Inventory [6], the Heat Transfer Concept Inventory [7], and the Thermal and Transport Concept Inventory (TTCI) ([8], [9]. For example, the recently developed TTCI is a 2-tiered multiple choice diagnostic test that seeks to identify student misconceptions and the reasons for these misconceptions.

Another approach to assessment in thermodynamics is the use of writing to allow students to represent their understanding in their own words and reveal the heterogeneity of their ideas [10], [11]. These written formative assessment give instructors insight into students' thinking by allowing students to generate ideas [12] and to present the heterogeneity of their ideas [13]. For example, students may hold more than one misconception along with correct ideas of a concept, all of which can be present in a written or constructed response. However, one obstacle to the use of written assessments has been the time and expense involved in grading these assessments.

The Automated Analysis of Constructed Responses (AACR) research group at Michigan State University explores student understanding expressed in constructed responses using computerized lexical and statistical analyses. Linguistic or text analysis can be conducted by vector space (such as latent semantic analysis), linguistic structure and feature-based methods[14]. We employ a feature-based approach that involves extracting and categorizing ideas from student responses, giving instructors insight into student thinking. This method can facilitate formative assessment and feedback in large-enrollment courses [15]. AACR has been exploring lexical analysis in evolution, genetics, cell metabolism and acid-base chemistry assessments [15–17].

In this study, we demonstrate the use of automated text analytics software to investigate students' understanding of thermodynamics in an introductory biology course. We examined student writing using questions targeted at two level's of Bloom's Taxonomy of the Cognitive Domain [18]: comprehension and application. This approach can be applied to other STEM disciplines and should be of interest to faculty who would like to use written assessments in large enrollment courses.

II. METHODS

A. Study Questions

Our study was conducted in an introductory cell and molecular biology course. At least one semester of general chemistry is required as a prerequisite for the course and students were expected to have a basic understanding of thermodynamics. We have collected thermodynamics data using different questions over several semesters. Data for this study was collected from questions administered during Fall semester 2008 and Spring semester 2012. Prior to receiving this assignment, the students also revisited this topic in their current course. Students were given online homework assignments including a set of three questions relating to reaction thermodynamics – two closed-form items followed by explanation, and one open response question. Students received credit for completion of the assignment regardless of the accuracy of their answer.

B. *Text Analysis*

We use IBM SPSS Text for Surveys and IBM SPSS Modeler 14.2 Text Mining node to analyze constructed responses. These software identify *terms* from custom-built libraries, similar to dictionaries. Terms are classified into *categories* by predefined computer algorithms which are subsequently modified by the researcher. Each response can contain multiple terms, with each term belonging to one or more categories. The software also displays *web diagrams*, similar to those in Figure 2, illustrating the connections among categories within groups of responses. (For more details on the operation of the software see references 20 and 21).

C. *Statistical Analysis*

We conducted discriminant analysis to determine categories that predict correct and incorrect post-instruction responses for Question 1. The discriminant analysis used a stepwise-forward, Wilk's method with an F-in of 3.84 and F-out of 2.71 [19]. We used group sizes for prior classification probabilities and a leave-one-out cross validation. Discriminant analysis is similar to linear regression and results in a linear function that expresses the relationship between dependent and independent variables. However, in the case of discriminant analysis, the dependent variables are categorical instead of interval. For this analysis, we have a series of binary independent variables (presence or absence of a student's response in each lexical category) which are combined to predict categorical dependent variables (expert rating). In this analysis the dependent variable has two categories (correct and incorrect) which results in a single linear function. Discriminant analysis analyzes the covariance between independent variables, or whether the variables change together or not. Because of this, it is not the values of individual independent variables but the relationship among them that is critical in determining the discriminant functions.

For Question 2 and 3, we used cluster analysis [20], another classification approach, to determine how responses aggregate into distinct groups based on the combination of categories to which they are assigned. Cluster analysis differs from discriminant in that the identity of groups (clusters) is unknown prior to classification. Clusters are later examined for similar properties shared by responses in the cluster. We use *k*-means clustering which finds *k* clusters of cases (responses). Clusters are mutually exclusive. The responses are classified into the cluster that is closest to that cluster mean or centroid than to the centroid of any other cluster. *K*-means cluster analysis allows the recombination of cases and *k* user-defined cluster over repeated iterations. Recombinations are iterated until no further change occurs. The algorithm computes the F-statistic using analysis of variance and specifies which lexical categories contribute most to differentiating among the clusters.

We use web diagrams to illustrate the frequency of and relationships among categories within a discriminant class or within a cluster.

III. RESULTS AND DISCCUSION

A. *Question 1*

The main objective of this part of our study was to investigate how students *describe* (Bloom's comprehension level) free energy change during an exergonic reaction using the graphic representation provided in Question 1 (Figure 1). We compared students' multiple choice and written explanations to Question 1. We analyzed 168 explanations from students who selected the correct multiple choice response. All responses were independently scored by two raters with expertise in biology and chemistry. Complete responses described the components of the system and the change in free energy (see below). The raters scored responses that had only correct ideas as 1. Responses with incorrect, incomplete or mixed (both right and wrong) ideas were scored 0 (this scoring group is hereafter referred to as "incorrect").. Raters demonstrated high scoring agreement (Cronbach's alpha = 0.88) and scoring disagreements were resolved by consensus.

Question 1.Study the graph that illustrates the chemical reaction A +B → C + D. Does this graph represent an exergonic or endergonic reaction?

A. Endergonic
B. Exergonic
C. Either

How can you tell?

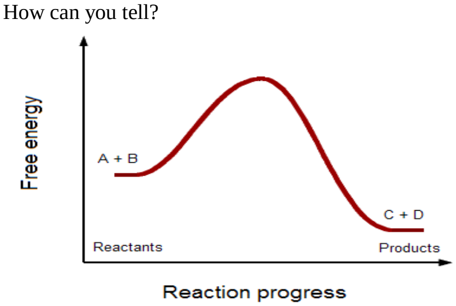

Figure 1. Question 1

The majority of students (84.9%) selected the correct multiple choice answer. However, 49% of these students had "incorrect" explanations for the type of reaction in the post test. Correct and complete responses compared the free energy of the reactants and products or referred to the net negative change in free energy for the reaction. The following are examples of correct responses:

Correct response 1: "energy is released as the reaction proceeds, with the products having less energy than the reactants"

Correct response 2 "The change in G is negative which implies the reaction is giving off energy"

Incorrect responses were either incomplete, completely wrong, or contained both wrong and correct ideas. The following are examples of responses scored as incorrect:

Incorrect Response 1 - Incomplete response: "releases energy" (This was the entire explanation submitted.)

Incorrect Response 2 - Only wrong ideas: "The energy is transferred from the reactants to the products"

Incorrect Response 3: Both correct and incorrect ideas, "You can tell because the reactants have more potential energy than the products. Also because there is a loss of free energy this reaction gives off heat."

Text analysis successfully categorized all 168 responses into one or more categories. Categories (*in italics*) were first automatically generated by software and refined by researchers. For example, the category *products* was automatically created by the software to extract the word product and phrases containing the word product. This category was further modified by researchers to include the phrase "C+D". This was important to capture, for in lieu of the word "products" several students mentioned C+D, the symbols representing products in the question diagram. The categories identified included

- components of the reaction (*reactants* and *products*),
- energy (*free energy* and *energy*),
- reaction types (*exergonic*, *endergonic*) and
- energy changes (*lower*, *release*, *higher*).

We used discriminant analysis to validate the categories from lexical analysis as the independent variables and expert scores (correct or incorrect) as the dependent variables. The use of a stepwise model allows only categories that are significant to the model to be included. We used a leave-one-out classification for cross-validation. The resulting discriminant function was significant (Wilk's Lambda = .629, Chi-square 47.111, df = 5, p <0.001). Discriminant analysis of post-instruction responses identified five categories that significantly predicted correct and incorrect responses. These categories and their standardized discriminant coefficients, which are similar to beta weights in regression, indicating relative weights among the variables, are reported in Table I. These categories were used to construct web diagrams illustrating the frequency of responses and connections among the categories (Figure 2).

TABLE I. STANDARDIZED DISCRTIMINANT COEFFECIENTS FOR CATEGORIES PREDICTION OF HUMAN RATINGS

Category	**Standardized Discriminant Coefficient**
Delta G	0.525
Energy of the products	0.502
Products	0.492
Lower	0.469
Reactants	0.360
Computer- Human Scoring Inter-rater reliability	Cronbach's Alpha = 0.75

Web diagrams illustrated distinct differences between correct and incorrect responses (Figure 2 a and b, respectively). Correct responses had considerably larger node sizes indicating the more correct responses contained these predictive ideas (i.e. *reactants*, *products*, *energy of products*, *lower* (energy). Correct responses also contained terms in the category *delta G* while incorrect responses did not.

We also observed more co-occurrences among categories in the correct responses than among categories in the incorrect responses web diagram. These co-occurrences are also more frequent as represented by the solid line between the nodes *reactants* and *lower* in the correct response web diagram.

These connections reflect the comparisons made by students giving correct responses, such as the response "…with the *products* having *less* energy than the *reactants*" which contained 3 ideas. In contrast, the fewer or weaker connections among incorrect responses were indicative of several incomplete responses such as "the *products* have *less* free energy", in which students expressed fewer ideas.

Although most students could select the correct multiple-choice option post-instruction, only half of these provided complete and correct explanations for their choice. Problems understanding endothermic/exothermic process can be prevalent among students learning thermodynamics[10], [13]. Our data show that this difficulty can extend to exergonic/endergonic processes as well. Additionally, the large number of students unable to describe the change in free energy in the system may have difficulty between the system and surroundings[1]. Our results demonstrate that this gap in student understanding could go undetected with only

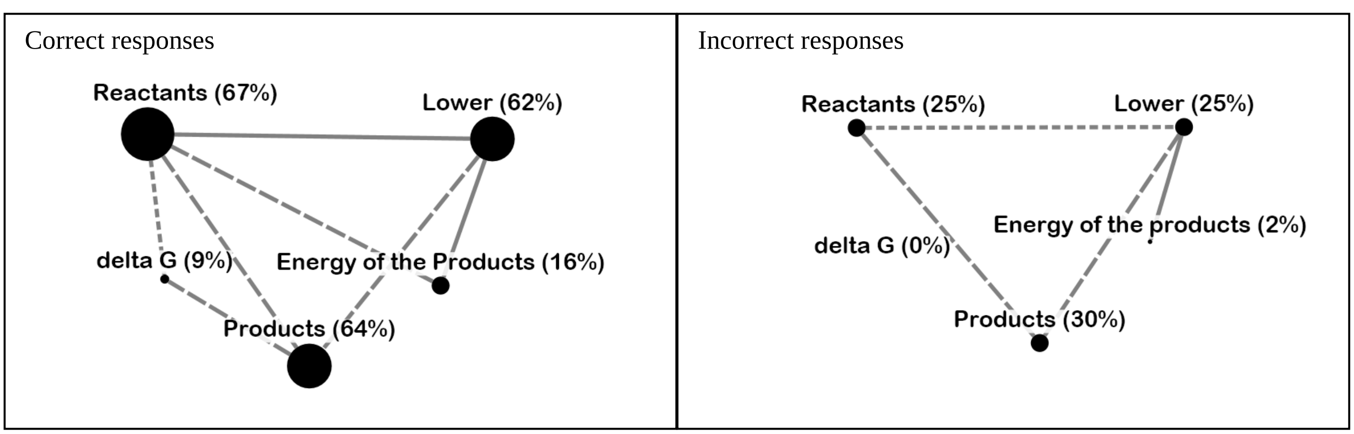

Figure 2. Web diagram of categories and links for a) Correct and b) Incorrect responses. Each category is represented by a node. The size of the node corresponds to the frequency of responses containing a category. Lines indicate the percentage of shared responses. Solid lines indicate that 75% shared responses; dashed lines 50-74%, dotted lines 25-49%. Nodes with fewer than 25%share responses were not linked.

multiple choice testing, but was revealed through written assessment, as has been observed in other fields [21], [22].

B. *Question 2*

The main objective of Question 2 was to examine how students can use their knowledge of the definition of Gibbs free energy to *identify* (Bloom's comprehension level) conditions under which a reaction may become spontaneous.

> Question 2. Can a non-spontaneous chemical reaction become spontaneous under different conditions? Explain your reasoning for your choice above.

Seventy five percent (n= 329) of students responded "Yes" to Question 2. We analyzed these 329 explanations using text and cluster analyses. Text analysis categorized the responses into 32 categories. Some of the most common categories were terms found in the question stem: *spontaneous* (70% of responses), *reaction* (45%) and *non-spontaneous* (36%). Other categories with high frequencies included *temperature* (48%), *change* (35%) and *conditions* (16%).

To identify groups among the responses, we used *k*-means cluster analysis. Values of *k* between 2 and 5 were examined. Categories from the question stem (listed in the previous paragraph) were removed from the cluster analysis as these ideas were not generated independently by students but may impact the results. With *k*= 2, the ANOVA indicated that the category *temperature* contributed the most to differentiating among the clusters. However, examination of the responses revealed a good deal of heterogeneity within the two clusters.

We decided to use three clusters, as trials with *k*=4 and *k*=5 showed little improvement in the distances among centroids and did not add much homogeneity to newly formed clusters. The *k*=3 cluster ANOVA identified the categories *change*, *energy*, *high*, *reactants* and *temperature* as contributing the most to differentiating among the models.

Cluster 1 identified 46 responses (14%) which demonstrated high quality explanations. These majority of the responses indicated that a non-spontaneous reaction could occur spontaneously at sufficiently high temperatures. This is observed in the web diagram (Figure 3a) showing that 58% of the responses in used the categories *temperature* and *high* (which in includes terms such as increased, greater and higher). These categories and the connection between them dominate the web diagram allowing a quick visualization of the main idea that is representative of Cluster 1.

A few students in Cluster 1 referenced the equation or definition for free energy, for example, *"If temperature is increased in the delta g = delta h - t delta s then the g can go from positive to negative and become spontaneous."*.

However, some students also gave less detail, e.g. *"Increasing the overall temperature of the reactants will cause a non-spontaneous reaction to become spontaneous."*

Cluster 2 contained 112 responses (34%) in which students affirmed that under changed conditions, particularly temperature, a non-spontaneous reaction may become spontaneous. However, these students did not specify that a higher temperature was required and were less likely to refer to the definition for Gibb's free energy. The strong association of *change* and *temperature* in Cluster 2 responses is shown in Figure 3b. We observe a change in the dominant variable and connection from *high* and *temperature* in Cluster 1 to *change* and *temperature* in Cluster 2. The following are representative of Cluster 2 responses

"Many times a non-spontaneous reaction can become spontaneous if the temperature is changed."
"You can change some reactions to become spontaneous by changing the temperature."

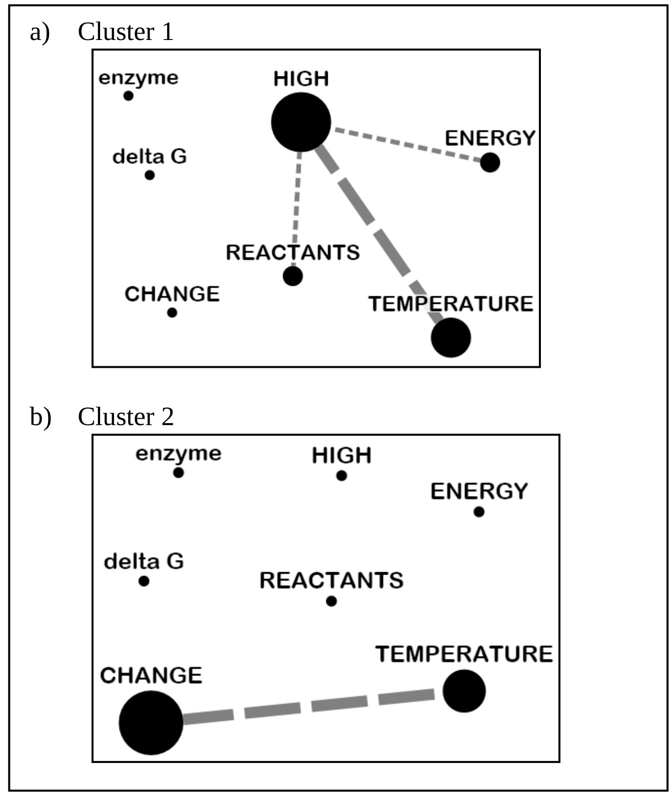

Figure 3. Web diagram of categories and links for a) Cluster 1 and b) Cluster 2 for Question 2: Spontaneous Reactions See Figure 2 caption for legend details.

Cluster 3 contained 171 responses (52%) that suggested reversing the reaction and selecting addition of an enzyme as a new condition. The use of k=4 analyses did not further divide these Cluster 3 into two distinct groups. Students who suggested reversing the reaction demonstrated knowledge of the free energy of the reaction but did not apply it correctly to the question. Responses suggesting the addition of an enzyme revealed a misconception held by students that enzymes alter the spontaneity of the reaction.

For example students stated that

"*Added enzymes can make a non-spontaneous reaction spontaneous but not the other way around."*
"catalysts lower the energy reaction pathway. An example of a catalysts [sic] is an enzyme. An enzyme reduces the transition energy pathway."

Keeping in mind that 75% of the students selected "Yes" - a change in reaction spontaneity is possible, we gain detailed information about student thinking from the use of the constructed responses. Coupling text analysis and cluster analysis, we were able to identify three groups of responses of varying content and quality. A subset of students in Cluster 1 presented a detailed explanation which included the definition or equation for free energy change. However, the majority of students did not reference this concept in their responses. Additionally, students hold misconceptions about the role of enzymes in reaction thermodynamics and kinetics.

C. Question 3

To observe how students *apply* (Bloom's application level) thermodynamics concepts in biology contexts, we analyzed responses to question 3. Students were expected to use thermodynamic concepts to explain carbohydrate decomposition rate.

> Question 3. A carbohydrate is composed of a string of covalently linked monosaccharides. Breaking those bonds between the monosaccharides is a chemically spontaneous reaction (ΔG for this reaction is -3.7 kcal/mol). However, this reaction occurs very slowly at room temperature. Why do you think this is so?

We analyzed 386 student responses. Text analysis generated 44 categories of terms. The most frequent categories containing terms not found in the stem were *temperature* (excluding room temperature) (34%), *energy* (25%), *high* (19%), *heat* (18%), *activation energy* (16%) and *catalyst* (13%).

For Question 3, we used *k*-means cluster analysis to identify groups, testing values of *k* between 2 and 5. We present the results for k= 3 which maximized the distance among clusters. The categories *enzyme* and *high* contributed most to differentiating among clusters.

Cluster 1 was the least well defined cluster. The 278 responses in this category were very heterogeneous. Some were vague, for example,

"This is because, the room temperature and the temperature of the reaction are very similar"
"Just because a reaction is spontaneous, does not mean it occurs fast."

Other responses contained concepts such as phospholipids and descriptions of state changes of carbohydrates at room temperature that were unrelated to the topic.

Surprisingly, some correct explanations mentioning the high activation energy as a rate-limiting process were also included in this category. It is possible that the low number of responses (~10 %) using activation energy as the only or main idea made these cases difficult to classify. These did not form a separate cluster with k = 4 or 5.

The web diagram for cluster 1 is not shown as all connections observed shared fewer than 25% of responses for students connected few ideas.

Cluster 2 contained 71 responses (18%) expressing ideas about bond stability and the need for higher temperatures to increase the rate of reaction (Figure 4). For example, students stated that:

"it is a more complex molecule with strong bonds. Though some bonds may break at room temperature, there simply isnt enough energy to make it occur rapidly."
"Since it is a spontaneous reaction it will happen without any added energy. A higher temperature might speed up the reaction because it would drive the reaction more than the lower temperature."

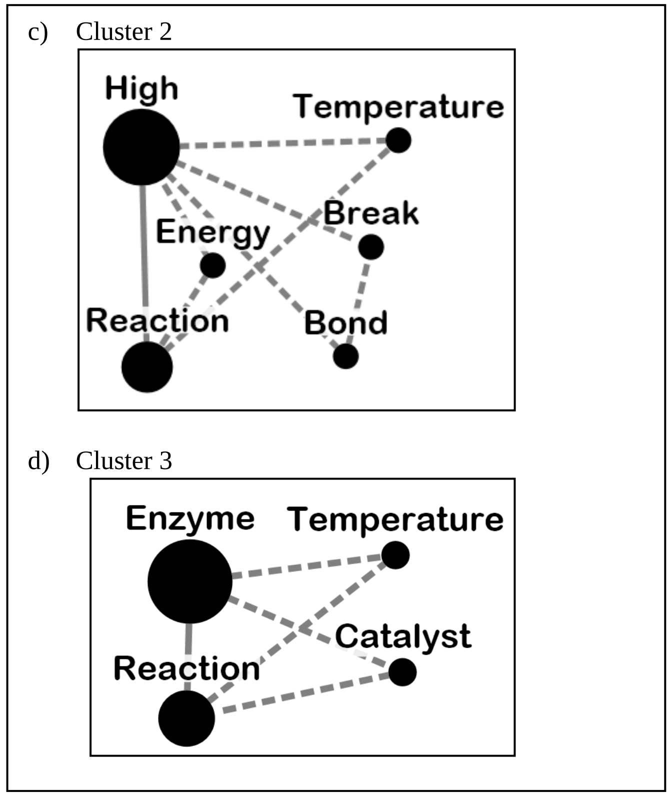

Figure 4. Web diagram of categories and links for a) Cluster 2 and b) Cluster 3 for Question 3: Reaction Rate. See Figure 2 caption for legend details.

Responses in Cluster 3 centered on the theme of enzymes and catalysts. About 10 percent of students described that a catalyst would help increase the rate. However, not all students detailed how the catalyst would change the thermodynamics or kinetics of the reaction. Some representative responses include

"because the room temperature is not high enough or it needs enzyme to speed up."
"because in the body, where most starch is broken down, the temperature is greater than room temp, it has a different pH, and has enzymes that catalyze the reaction"

We observed in Question 3 that students give varied explanations of the scenario presented. Though instructors and/or researchers apply thermodynamics concepts to indicate why spontaneity does not determine reaction rate, students often gave vague or irrelevant surface-level responses. We also observed that some students drew only from their biology knowledge and did not use thermodynamic concepts in answering this question. Students may require more explicit prompts to draw on knowledge from outside the domain of biology when answering problems within a biological context.

IV. CONCLUSIONS

Our analysis of students' written responses revealed that students held varied ideas that could not be observed from their multiple-choice responses. Though some students were able to articulate well thought-out responses incorporating thermodynamic concepts, others failed to identify, describe, and/or apply their knowledge of thermodynamics when asked to provide an explanation. Furthermore, when asked to apply thermodynamics concepts in a biological context, few students approached this question from an application level and most relied on surface-level descriptions, consistent with other research on conceptual change in novices [23].

Written assessment is an important component of thermodynamics and other STEM instruction [10], [21]. We have demonstrated that computerized lexical analysis, coupled with classification analyses, can be used to investigate large volumes of student writing. Lexical analysis allows patterns in student responses in the form of categories and their connections to be easily visualized. Statistical classification, with or without human scoring, can organize student ideas into related groups. One goal of the AACR group is to develop these assessment resources (questions, categories and scoring models) for STEM disciplines [15], [16]. As this body of work grows, it will be made available to instructors to facilitate rapid feedback and tailor their lessons to address student misunderstandings. Resources developed in this project are freely available at http://aacr.crcstl.msu.edu/.

ACKNOWLEDGMENT

This work was funded through a grant from the National Science Foundation (DUE 0736952 and DUE 1022653). Any opinions, findings and conclusions or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

[1] Johnstone, A. H., Macdonald, J. J. Webb, G. 1977. "Misconceptions in School Thermodynamics." *Physics Education*. Vol. 12, pp. 248–251.

[2] Carson, E., Watson, J. 2002. "Undergraduate Students' Understandings of Entropy and Gibbs Free Energy," *University Chemistry Education*. Vol. 6, pp. 4–12.

[3] Chi, M. T. H. "Three types of conceptual change." In Vosniadou (ed.). 2008 *International Handbook of Research on Conceptual Change*., New York: Taylor & Francis, pp. 61–81.

[4] Clark, D., Jorde, D. "Helping Students Revise Disruptive Experientially Supported Ideas about Thermodynamics: Computer Visualizations and Tactile Models." *Journal of Research in Science Teaching*. Vol. 41, pp. 1–23.

[5] Anderson, E. E., Taraban, R., Sharma M. P. 2005. "Implementing and Assessing Computer-Based Active Learning Materials in Introductory Thermodynamics." *International Journal of Engineering Education*. Vol. 21, pp. 1168-1176.

[6] Midkiff, K. C., Litzinger, T. A., Evans, D. L. "Development of Engineering Thermodynamics Concept Inventory Instruments." In Frontiers in Education Conference. 2001. *31st Annual ASEE/IEEE Frontiers in Education Conference.* 2001. Vol. 2, Reno, Nevada: pp. F2A–F23.

[7] Jacobi, A., Martin, J., Mitchell, J., Newell, T. "A Concept Inventory for Heat Transfer." In Frontiers in Education. 2003. *33rd ASEEAEEE Frontiers in Education Conference*, Vol. 1. Boulder, CO, pp. T3D–12.

[8] Miller, R. L., Streveler, R. A., Nelson, M. A., Geist, M. R., Olds, B. M. "Concept Inventories Meet Cognitive Psychology: Using Beta Testing as a Mechanism for Identifying Engineering Student Misconceptions." In *Proceedings of the American Society for Engineering Education Annual Conference*. 2005. Portland, OR.

[9] Streveler, R. A., Miller, R. L., Santiago-Román, A.I., Nelson, M. A., Geist, M. R., Olds, B. M. R. A. 2011. "Rigorous Methodology for Concept Inventory Development: Using the 'Assessment Triangle' to Develop and Test the Thermal and Transport Science Concept Inventory (TTCI)." *International Journal of Engineering Education.* Vol. 27, pp. 968-984.

[10] Beall, H. 1994. "Probing Student Misconceptions in Thermodynamics with In-Class Writing," *Journal of Chemical Education*. Vol. 71, pp. 1056-1057.

[11] Turns, S. R., Van Meter, P. N. "Applying Knowledge from Educational Psychology and Cognitive Science to a First Course in Thermodynamics." In *Proceedings of the American Society for Engineering Education Annual Conference*. 2011. Vancouver, British Colombia.

[12] Ackerman, T. A., Smith, P. L. 1988 "A Comparison of the Information Provided by Essay, Multiple-Choice, and Free-Response Writing Tests." *Applied Psychological Measurement*. Vol. 12, pp. 117–128.

[13] Birenbaum, M., Tatsuoka, K. K. 1987. "Open-Ended Versus Multiple-Choice Response Formats—It Does Make a Difference for Diagnostic Purposes." *Applied Psychological Measurement*. Vol. 11, pp. 385 –395.

[14] Deane, P. "Strategies for Evidence Identification through Linguistic Assessment of Textual Responses." In Williamson, D. M., Bejar, I. I., Mislevy, R. J. (ed.). 2006. Automated Scoring of Complex Tasks in Computer Based Testing. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 313–371.

[15] Haudek, K. C., Kaplan, J. J., Knight, J., Long, T., Merrill, J., Munn, A., Nehm, R., Smith, M., Urban-Lurain, M. 2011. "Harnessing Technology to Improve Formative Assessment of Student Conceptions in STEM: Forging a National Network," *CBE-Life Sciences Education*. Vol. 10, pp. 149–155.

[16] Ha, M., Nehm, R. H., Urban-Lurain, M., Merrill, J. E. 2011. "Applying Computerized-Scoring Models of Written Biological Explanations across Courses and Colleges: Prospects and Limitations," *CBE-Life Sciences Education*. Vol. 10, pp. 379 –393.

[17] Haudek, K. C., Prevost, L. B, Moscarella, Merrill, J.E, and Urban-Lurain, M. In press. "What Are They Thinking? Automated Analysis of Student Writing about Acid/Base Chemistry in Introductory Biology," *CBE-Life Sciences Education*.

[18] Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., Wittrock, M. C. 2000. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Complete Edition. Boston, MA: Allyn & Bacon.

[19] Spicer, J. 2004. *Making Sense of Multivariate Data Analysis: An Intuitive Approach*. First edition. Thousand Oaks, CA: Sage Publications, Inc.

[20] Tan, P. N., Steinbach, M., Kumar, V. 2006. *Introduction to Data Mining*. Boston, MA: Pearson Addison Wesley.

[21] Nehm, R. H., Schonfeld, I. S. 2008. "Measuring Knowledge of Natural Selection: A Comparison of the CINS, an Open-Response Instrument, and an Oral Interview." *Journal of Research in Science Teaching*. Vol. 45, pp. 1131–1160.

[22] Kuechler W. L., Simkin, M. G. 2010. "Why Is Performance on Multiple- Choice Tests and Constructed- Response Tests Not More Closely Related? Theory and an Empirical Test." *Decision Sciences Journal of Innovative Education*. Vol. 8, pp. 55–73.

[23] Chi, M. T. H., Feltovich, P. J. Glaser, R. 1981. "Categorization and Representation of Physics Problems by Experts and Novices." *Cognitive Science*. Vol. 5, pp. 121–152.