2011

A Valid and Reliable Instrument for Cognitive Complexity Rating Assignment of Chemistry Exam Items

Karen Knaus
University of Colorado Denver

Kristen Murphy
University of Wisconsin - Milwaukee

Anja Blecking
University of Wisconsin - Milwaukee

Thomas Holme
Iowa State University, taholme@iastate.edu

A Valid and Reliable Instrument for Cognitive Complexity Rating Assignment of Chemistry Exam Items

Abstract

The design and use of a valid and reliable instrument for the assignment of cognitive complexity ratings to chemistry exam items is described in this paper. Use of such an instrument provides a simple method to quantify the cognitive demands of chemistry exam items. Instrument validity was established in two different ways: statistically significant correlations between expert-based cognitive complexity ratings and student performance (as measured through statistical difficulty of items), and statistically significant correlations between expert-based cognitive complexity ratings and student mental effort ratings. Key benefits associated with instrument use include an enhanced understanding of the cognitive complexity of chemistry assessment tasks and as a means for characterizing exam content for the measurement of cognitive development.

Keywords

first-year undergraduate/general, chemical education research, testing/assessment

Disciplines

Curriculum and Instruction | Educational Assessment, Evaluation, and Research | Higher Education | Other Chemistry | Science and Mathematics Education

Comments

Reprinted (adapted) with permission from J. Chem. Educ., 2011, 88 (5), pp 554–560.

JOURNAL OF CHEMICAL EDUCATION

ARTICLE

pubs.acs.org/jchemeduc

A Valid and Reliable Instrument for Cognitive Complexity Rating Assignment of Chemistry Exam Items

Karen Knaus,[†] Kristen Murphy,[‡] Anja Blecking,[‡] and Thomas Holme[*,§]

[†]Department of Chemistry, University of Colorado—Denver, Denver 80217-3364, United States

[‡]Department of Chemistry and Biochemistry, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53211, United States

[§]Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States

ABSTRACT: The design and use of a valid and reliable instrument for the assignment of cognitive complexity ratings to chemistry exam items is described in this paper. Use of such an instrument provides a simple method to quantify the cognitive demands of chemistry exam items. Instrument validity was established in two different ways: statistically significant correlations between expert-based cognitive complexity ratings and student performance (as measured through statistical difficulty of items), and statistically significant correlations between expert-based cognitive complexity ratings and student mental effort ratings. Key benefits associated with instrument use include an enhanced understanding of the cognitive complexity of chemistry assessment tasks and as a means for characterizing exam content for the measurement of cognitive development.

Cognitive Complexity Rating Rubric

Rating	Count and relative difficulty of component concepts or skills needed to master item		
	Easy	Medium	Difficult
1	1		
2	2	1	
3	3–4	2	
4	5–6	3–4	1
5		5–6	2
6			3–4
7			5–6

KEYWORDS: First-Year Undergraduate/General, Chemical Education Research, Testing/Assessment

FEATURE: Chemical Education Research

Chemistry is considered a complex domain of learning. As such, assessment of learning in chemistry presents a variety of challenges. Most instructors have been surprised by poor student performance on a test item that they expected to be straightforward. When this happens, educators may wonder whether it was the learning or the testing that hindered student performance. While many factors play a role in the measurement that is made when a student takes a test, one critical component lies in the complexity of the item itself. Items derive their complexity from the content and the manner in which it is presented, the item construct. When both aspects of item complexity are reliably estimated, the construction of exams can be improved. Additionally, for established exams such as those constructed by the ACS Examinations Institute, a reliable complexity measure can provide an additional way to characterize student performances.

Complexity theory provides a conceptual template for understanding any system with multiple interrelated parts. For example, complexity theory can be applied in theoretical computer science, where one looks at a task in terms of the computational resources required to solve it.[1,2] Complexity research also has multidisciplinary applications, including areas such as decision-making, goal setting and task design,[3−7] subjective workload,[8] and more recently, science education and nonlinear dynamics.[9] Complexity is conceptually central to the work of Johnstone and co-workers in the development of the Information Processing Model (IPM).[10−12]

Task characterization represents a key component of objective complexity.[13,14] Such characterization may be accomplished by using experts who are familiar with the cognitive demands of the task. Furthermore, studies have shown that objective complexity can be used to predict performance on tasks, including assessment tasks.[6,15−18]

Complexity theory may be applied to chemistry education by considering the cognitive demands that students experience while answering chemistry test items. Key factors in determining complexity can be elaborated, importantly including the degree of element interactivity.[19−22] In a conceptual sense, the intrinsic complexity of a task can be enumerated in terms of the amount of memory resources used to accomplish it. This enumeration is closely tied to concepts of cognitive load theory. The basic premise of cognitive load theory is that there are limits on the working memory used for analytical thinking, and the "load" on that cognitive capacity affects learning and testing. The connection between complexity and cognitive load has been previously noted by Sweller and co-workers, "intrinsic cognitive load refers to the internal complexity of the task being attempted".[23] Niaz[24] has applied the concept of M-capacity to the analysis of test items and this theory is also closely related to the complexity concepts noted here. The cognitive complexity rating assignment instrument described in this research was designed through a combination of emergent ideas found in both complexity theory (i.e., the construct of objective complexity) and cognitive load theory (i.e., the idea of intrinsic cognitive load or "element interactivity").

There are challenges inherent in establishing a reliable rating for a trait as potentially elusive as task complexity. In an ideal measurement, the cognitive complexity would correlate strongly

Published: March 18, 2011

Cognitive Complexity Rating Rubric

Rating	Count and relative difficulty of component concepts or skills needed to master item		
	Easy	Medium	Difficult
1	1		
2	2	1	
3	3–4	2	
4	5–6	3–4	1
5		5–6	2
6			3–4
7			5–6

Rating	Concept/skill interactivity
0	Nonsignificant
1	Basic
2	Complex

Figure 1. Cognitive complexity rating rubric. Domain experts estimate the number of knowledge elements (concepts and/or skills) needed to accomplish an exam task. They also judge the relative difficulty or complexity of each component concept or skill from the perspective of a student. Having enumerated all components needed to complete the test item, they can then look up a rating value (e.g., 2 "medium difficulty" tasks in a test item would add 3 points to the rating), using the rubric to obtain a numerical cognitive complexity rating. Concept/skill interactivity is an "add on" that may increase the overall cognitive complexity rating for an item in which students must use the interdependence of components in order to complete the task.

with student performance. Most test items, however, have a number of features between the content knowledge and the construct that ultimately affect student performance on the measure. Thus, item complexity alone will not predict all variation in student performance, but it should correlate with student performance, particularly over populations of students that take nationally standardized exams. It is also important to realize that any complexity rating system will rely on expert judgments, and thereby insert a significant subjective component into the rating process. The crafting of a rubric, therefore, must be carried out so that the results are reliable and valid, and thus allow the expert rating method for complexity to serve as a proxy for objective complexity.

COGNITIVE COMPLEXITY ANALYSIS OF CHEMISTRY EXAM ITEMS

Figure 1 presents the rubric by which cognitive complexity of an exam item or other chemistry task may be assigned. Item cognitive complexity analysis is achieved through a four-step process:

1. Counting the number of pieces of knowledge (i.e., chemistry concepts and skills) needed to complete a task, such as an exam item
2. Estimating from the perspective of a student the relative difficulty rating to each of the component concepts or skills needed
3. Using the rubric to add up the component complexities to determine a numerical cognitive complexity rating
4. Increasing the overall complexity rating of an item by estimating how interrelated or interactive the chemistry knowledge must be for accomplishing the task.

Note that the rubric itself is constructed with substantial overlap between ratings with easy, medium, or difficult tasks. This feature is critical to the ability of this instrument to serve as a proxy for objective complexity. When different experts identify tasks in different ways, the rubric is designed to have results end up with a similar or the same rating. One expert may parse a required task into two "easy" steps, where another would identify the same task as one "medium" step. Both judgments would lead to the same complexity rating with this rubric. This feature of the rubric serves to improve inter-rater reliability as noted later, and it also is consistent with learning theories that suggest that as students learn more, they "chunk" information in their long-term memory.[25,26]

An example application of this rubric for the cognitive complexity rating of a general chemistry exam item is provided in Figure 2. A rater has identified the pieces of knowledge required for completion of the cognitive task (a chemistry exam item), which are listed in step 2 of Figure 2. Furthermore, the relative cognitive difficulties for acquiring each of these pieces of knowledge is considered from the perspective of a general chemistry student, and assigned by the rater in this example. Then, in step 3, the component difficulty and complexity assignments are added up to determine a numerical cognitive complexity rating using the rating scale from the rubric. Finally, an additional factor, in this case, +2 rating points, is added to factor in element interactivity. This particular test item is considered by the rater to have complex interactivity because of the need for a student to compare item responses both to each other and to information derived from the graph of the phase diagram.

Interactivity in this rubric can assume values of 0, 1, or 2. In Figure 3, two additional examples are shown with these values. The item shown in Figure 3A can be assigned as having two elements (the definition of ionization energy and the concept of the periodic trend associated with ionization energy). Presentation of the connectedness of these two ideas is fundamental in general chemistry and this concept is typically understood in tandem, so the interrelatedness is assigned as 0. For Figure 3B, the item can be assigned as having three elements (definition of isotopes and relative abundance, and a conceptual understanding of the weighted average). Because a student must be able to use the periodic table to determine the actual molar mass, and then judge how the weighted average produces it, the interrelatedness is assigned as 1.

INSTRUMENT RELIABILITY

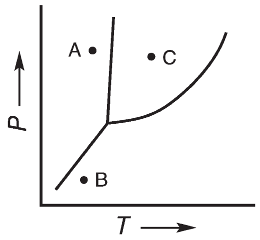
This instrument is designed as an expert-based rating system for which any test item must be rated by several individuals and the inter-rater agreement determined. Active instructors of college general chemistry courses who were not part of the original design team were recruited as content experts. Several such experts independently assigned cognitive complexity ratings for a large number of chemistry test items, including items from two practice examinations. The experts received a brief, 20-min training session on using the instrument before they did their own rating. Inter-rater reliability for the cognitive complexity rating instrument was established using ratings from 45-items of

dx.doi.org/10.1021/ed900070y |J. Chem. Educ. 2011, 88, 554–560

Example Cognitive Complexity Rating Task

Step 1.
Analyze the item.

Consider the phase diagram of a pure compound. Which statement applies?



(A) The path A ⟶ C represents sublimation.

(B) Following the path A ⟶ B ⟶ C, the compound would first liquefy and then vaporize.

(C) If the compound is in state A, continued reduction of the pressure (at constant temperature) will cause it to melt.

(D) None of these statements is correct.

Step 2A.
Count the concepts and skills necessary for a student to answer the question and assign a corresponding relative difficulty value to each one.

Components	Description	Level of Difficulty (Count)
Concept	Basic knowledge of the concept of a phase	Easy (1)
Skill	Interpret phase diagram	Medium (1)
Concept	Knowledge of phase-change-related terminology: *sublimation, liquefy/melt, vaporize*	Easy (3)
Concept	Basic knowledge of effects of pressure changes on solids	Easy (1)

Step 2B.
Next, consider the concepts and skills identified in Step 2A and determine their level of interactivity (nonsignificant, basic, or complex: see Figure 1).

Component	Description	Interactivity (Rating)
Interactivity	Conceptual/skill interactivity: intercomparisons of response items and with diagram	Complex (2)

Step 3.
Use the rubric to determine the overall objective complexity rating for this item.

Components	Count	Component Set Rating (Figure 1)
Easy Concepts	5	4
Medium Skill	1	2
Complex Concept/Skill Interactivity	NA	2
Overall cognitive complexity rating (sum of component ratings) for this item		**8**

Figure 2. Example of a rater assignment of cognitive complexity of a chemistry exam item. Component concepts and skills (knowledge elements) needed to accomplish the item are identified and the relative level of difficulty (from the perspective of a student) was determined. The interactivity of the cognitive demands is estimated to be "complex". Finally, the component complexity assignments add up to an overall cognitive complexity rating for the exam item, a rating of 8.

a practice exam for general chemistry under development by the ACS Exams Institute (hereafter this exam is referred to as the ACS general chemistry practice exam) and 72-items of a preparatory chemistry practice exam. (See Table 1.) The SPSS statistical software package (version 11.0) was used to calculate inter-rater reliability statistics. The two-way, mixed intraclass correlation (ICC) model with the consistency definition was used for the reliability analysis.[27] The methodology results in estimates that mirror Cronbach's α model for internal consistency. Thus, in exploratory research, ICC values should be at least 0.70 or higher to retain an "adequate" scale; many researchers require a cutoff of 0.80 for a "good scale".[28] Scales and instruments with a greater number of items are more reliable, which is revealed in the higher ICC values for the 72-items in comparison to the 45-items. The average calculated intraclass correlation coefficients for ratings collected from both chemistry practice exams indicate an inter-rater agreement of approximately 82%. When item ratings are standardized, the average inter-rater agreement is approximately 83%.

■ VALIDATION STUDIES

In addition to reliability, the validity of the cognitive complexity rating instrument was established. Because the expert ratings determined via this rubric are considered a proxy for objective

Item A	Item B
Which element has the lowest first ionization energy?	Only two isotopes exist for copper, copper-63 and copper-65. Which statement is true?
(A) Ca	(A) Copper-63 is more abundant than copper-65.
(B) K	(B) Copper-65 is more abundant than copper-63.
(C) Mg	(C) Both isotopes are equally abundant.
(D) Na	(D) Relative abundance cannot be approximated.

Figure 3. Additional examples of interrelatedness. (A) An item with nonsignificant interrelatedness (numerical value = 0). (B) An item with basic interrelatedness (numerical value = 1).

Table 1. Inter-Rater Reliability Statistics for ICC Cognitive Complexity Ratings

Measure[a]	ICC Values[b]	Lower Bound	Upper Bound	F Values[c]	Significance
Single rater (ACS)	0.3367	0.2308	0.4725	5.5682	0.0000
Average of raters (ACS)	0.8204	0.7297	0.8897	5.5682	0.0000
9 raters (45 items rated)					
Standardized item, α = 0.8310					
Single rater (Prep-chem)	0.5455	0.4312	0.6570	5.8014	0.0000
Average of raters (Prep-chem)	0.8276	0.7520	0.8845	5.8014	0.0000
4 raters (72 items rated)					
Standardized item, α = 0.8386					

[a] Two different exams were used to collect ratings, an ACS general chemistry practice exam and a preparatory chemistry practice exam. [b] ICC (intraclass correlation coefficients) were evaluated at the 95% confidence interval. [c] Two-way mixed-effects model was used (consistency definition).

How much mental effort did you expend on question #1?

(A) Very little

(B) Little

(C) Moderate amounts

(D) Large amounts

(E) Very large amounts

Figure 4. Example of mental effort item inserted into the practice exam format.

complexity as defined in the literature, the values derived from this instrument should be predictive for both the complexity a student perceives (subjective complexity) and student performance (measured as item difficulty). Thus, the validity of the cognitive complexity instrument described here was established by determining the correlation between it and these measures. These studies were carried out by the administration of practice exams with students in general chemistry and preparatory chemistry.

Performance and mental effort data were collected from the administration of two different chemistry practice exams at a large, Midwestern university. A general chemistry practice exam was used in one single-semester preengineering general chemistry course (n = 75) and two second-semester general chemistry courses (n = 83). A preparatory chemistry practice exam was used in two preparatory chemistry classes (n = 175). Data included in the study are derived from students who signed consent forms as part of the relevant IRB approval process. The practice exam format and validation is described elsewhere;[29] the key feature is that students complete the tasks by answering each test item and also estimating their mental effort for that item. The mental effort is estimated using a 5-point quasi-interval scale as illustrated in Figure 4.

VALIDITY MEASURES

Instrument validity was established in two different ways in this study. First, the correlation was determined between cognitive complexity ratings and student performance for items, as shown in Figures 5 and 6. Second, the expert rating cognitive complexity was correlated with student mental effort ratings, depicted in Figures 7 and 8. Previous research has demonstrated a high correlation between perceived (subjective) task complexity and objective task complexity.[16,17,30] Therefore, if the expert complexity ratings are a proxy for the objective task complexity, these two measures should also correlate.

Results from both validity tests demonstrate a strong, statistically significant correlation between variables. For example, P-sig < 0.001 for all tests, as noted in Figures 5−8. Another measure of the statistical veracity of the measures reported is to obtain the Cohen's effect size, f^2, which by convention, are termed *small*, *medium*, and *large*, when values are above 0.02, 0.15, and 0.35, respectively.[31] Again, as noted in Figures 5−8, effect sizes lie from medium to large for the data provided. In the second validity test (student mental effort vs cognitive complexity), a somewhat higher correlation between the variables is observed for the preparatory chemistry practice exam in comparison to the ACS general chemistry practice exam (r = 0.492 vs r = 0.650). There are several possible explanations for this observation. First and foremost, the preparatory general chemistry practice exam was administered to a group of students who had used the mental effort ratings throughout the semester, whereas those taking the

dx.doi.org/10.1021/ed900070y | J. Chem. Educ. 2011, 88, 554–560

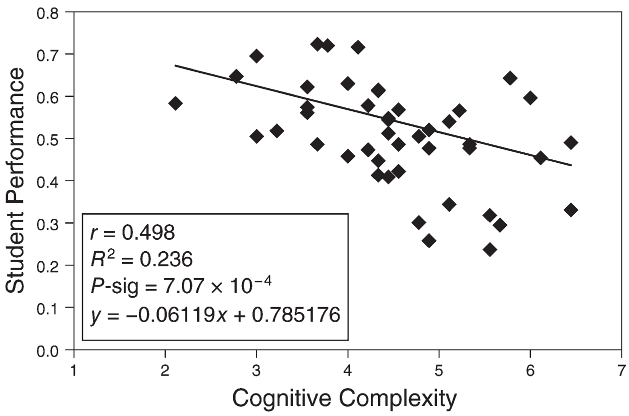


Figure 5. Correlation of student performance (using statistical item difficulty) with cognitive complexity (using average item ratings) for the ACS general chemistry practice exam. The graph shows a statistically significant correlation between the variables (P-sig =7.07 × 10^{-4}; F = 13.3; f^2 = 0.309; n = 158 students).

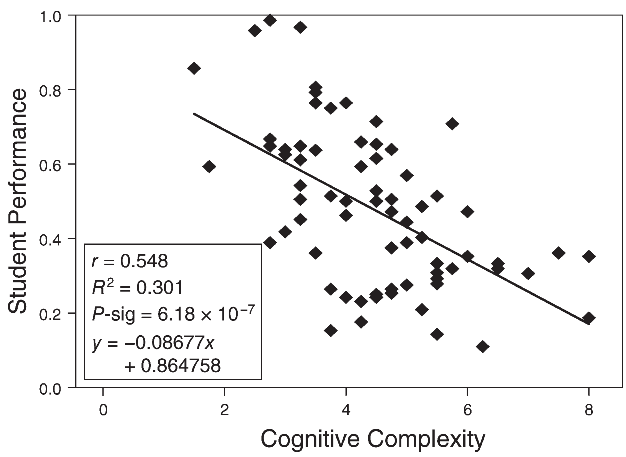


Figure 6. Correlation of student performance (using statistical item difficulty) with cognitive complexity (using average item ratings) for the preparatory chemistry practice exam. The graph shows a statistically significant correlation between the variables (P-sig = 6.18 × 10^{-7}; F = 30.1; f^2 = 0.431; n = 175 students).

ACS general chemistry practice exam had not. Note that in Figure 8 students used a wider range of the quasi-interval scale, especially on the lower end of the scale (average mental effort ratings <2.0), suggesting that the practice they had in the classroom served as a form of training in using this scale. Both correlations are significant, so training in the use of mental effort quasi-interval scales is not required, but it appears to improve the mental-effort results for the students, and subsequently the correlation with the expert ratings of complexity. Second, the content coverage on the preparatory chemistry practice exam spans a wider range of cognitive complexity ratings, a feature of the practice exam design more than the course offering itself. Finally, the ACS general chemistry practice exam was tested in a course in which the contents of a full-year general chemistry course are covered in a single semester (for engineering students). These students encountered items on the practice exam regarding content that had received somewhat cursory coverage, as might be expected for such a survey. This represents one effect that can help explain how some items with low

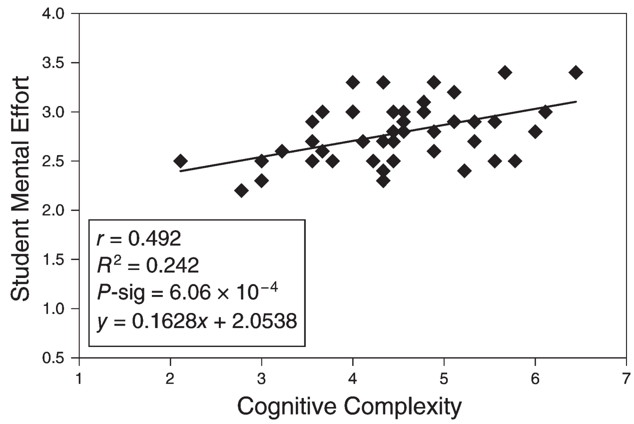


Figure 7. Correlation of student mental effort (using average student ratings) with cognitive complexity (using average item ratings) for the ACS general chemistry practice exam. The graphs show a statistically significant correlation between the variables (P-sig = 6.06 × 10^{-4}; F = 13.7; f^2 = 0.319; n = 158).

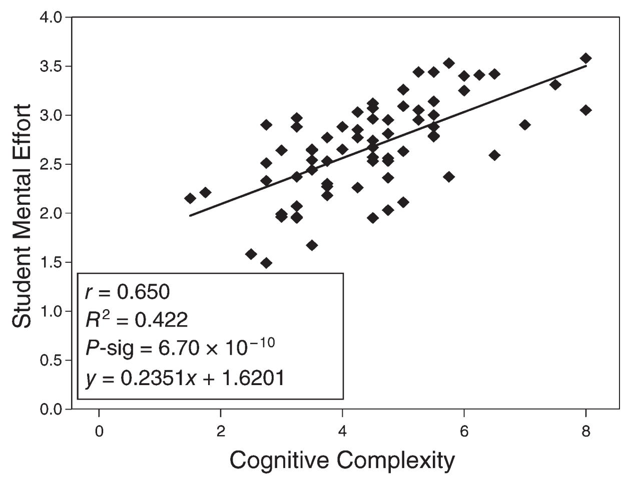


Figure 8. Correlation of student mental effort (using average student ratings) with cognitive complexity (using average item ratings) for the preparatory chemistry practice exam. The graph shows a statistically significant correlation between the variables (P-sig = 6.70 × 10^{-10}; F = 51.1; f^2 = 0.730; n = 175 students).

cognitive complexity can nonetheless result in high mental effort because students may use a relatively large amount of mental resources on recall of information for which they received only modest instruction.

CONCLUSIONS

This study has reported a reliable and valid instrument for the cognitive complexity rating assignment of chemistry tasks, such as exam items. The instrument itself is capable of providing good inter-rater reliability, even when a majority of the raters have only a modest (20-min) training period. The validity of the instrument is established by virtue of the correlations it has with student subjective ratings derived from taking the same items, and, importantly, with student performance on the items.

Understanding the cognitive complexity of chemistry learning and assessment tasks shows promise in informing improvements in chemistry education. First and foremost, it is important to recognize that learning is subject to fundamental constraints associated with cognition. In particular, when the cognitive load of a task, either learning or assessment gets too high, the student is simply not capable of successful learning. It may be possible to use intuition to speculate on when material has become very complex, but the use of this instrument helps establish a well-defined procedure for quantifying such intuition. As such, the rubric will help allow chemical educators to have greater awareness of the relative complexity of the material they are assessing. This fundamental understanding can inform teachers about circumstances that require care to address complexity explicitly, both in teaching and assessment of learning. For example, by being able to reliably establish the cognitive complexity of a task within the context of a specific course, chemistry instructors will have a better gauge for matching instruction with the learning needs of their students. These needs include matching the depth and breadth of chemistry content covered in the classroom with working memory load (i.e., cognitive load) at a level that promotes optimal learning even in large lecture classrooms. Research has provided considerable evidence that high levels of working memory load, that is, high cognitive load, are detrimental to learning.[19,24,32] Thus, with regard to chemistry instruction, cognitive complexity analysis in combination with student performance measures provides a new window into exploring the relationship between the complexity of content taught and student cognition and learning.

Cognitive complexity analysis of chemistry exam items also serves to enhance the design of chemistry assessment materials. Once again, the key new capacity is associated with the provision of a sound method for quantification of instructor intuition. Most instructors are aware of the fact that some test items are more complex than others. Being able to replace this innate intuition with a more reliable method for quantification of cognitive demands associated with exam content allows instructors to be more aware of how their instruction promotes complex learning in chemistry. Because chemistry is inherently complex, instruction necessarily seeks to enhance students' progress to more complex understandings characterized by greater knowledge integration, which requires higher-order cognition. By designing exams that span the complexity variable of cognition, instructors can gain a handle on both the content recall within a course and the growth in cognitive maturity of students. This concept is similar to the one being used by Stacy and co-workers in the ChemQuery project.[33]

Finally, in terms of research projects designed to assess student learning, this rubric will provide a new method for quantifying chemical knowledge addressed in chemistry exams. This capacity will allow for two benefits. First, this type of analysis of exam items may enhance the comparison of different chemistry exams with respect to their content. In one sense, this will help clarify psychometric properties of the exam. More importantly, it suggests how exams throughout the curriculum can be designed to touch on transferable traits of cognitive development. Insofar as students develop their ability to digest more complex content as they proceed in an undergraduate curriculum, making more explicit note of task complexity of tests can provide a method to assess this cognitive development. Second, the use of complexity in the design of new assessment materials promises to provide a more robust measure of learning within any individual course. For example, materials designed with complexity in mind will possess a more nuanced measurement of the range of abilities students achieve during the learning of chemistry. While norm-referenced exams that differentiate students along a distribution of performances are commonplace, being able to reference that performance to cognitive criteria related to complexity would be a novel development worth pursuing. Indeed, such a scenario has recently been suggested for statewide exams.[34]

AUTHOR INFORMATION

Corresponding Author
*E-mail: taholme@iastate.edu.

ACKNOWLEDGMENT

We gratefully acknowledge the efforts of expert raters who participated in workshops on complexity rating with the author team. This work is based on work supported by the National Science Foundation (NSF) under grant numbers CHE-0407378 and DUE-0618600, 0717769. Any opinions, findings, and conclusions communicated in this material are those of the authors and do not necessarily reflect the views of NSF.

REFERENCES

(1) Fortnow, L.; Homer, S. A Short History of Computational Complexity. In *The History of Mathematical Logic*; van Dalen, D., Dawson, J., Kanamori, A., Eds.; North-Holland Publishing Co.: Amsterdam, 2002.
(2) Goldreich, O. Computational Complexity: A Conceptual Perspective Page. http://www.wisdom.weizmann.ac.il/~oded/cc-book.html (accessed Feb 2011).
(3) Larichev, O. I.; Moshkovich, H. M. *Organ. Behav. Hum. Decis.* **1988**, *42*, 217–233.
(4) Locke, E. A.; Shaw, K. N.; Saari, L. M.; Latham, G. P. *Psychol. Bull.* **1981**, *90*, 125–152.
(5) Campbell, D. J.; Gingrich, K. *Organ. Behav. Hum. Decis.* **1986**, *38*, 162–180.
(6) Earley, P. C. *J. Appl. Psychol.* **1985**, *70*, 481–491.
(7) Locke, E. A.; Frederick, E.; Lee, C.; Bobko, P. *J. Appl. Psychol.* **1984**, *69*, 241–251.
(8) Braarud, P. O. *Int. J. Cogn. Ergon.* **2001**, *5*, 261–273.
(9) Stamovlasis, D.; Tsaparlis, G. *Nonlinear Dyn. Psychol. Life Sci.* **2003**, *7* (3), 221–244.
(10) El-Banna, H.; Johnstone, A. H. *Educ. Chem.* **1986**, *23*, 80–84.
(11) Al-Naeme, F.; Johnstone, A. H. *Int. J. Sci. Educ.* **1991**, *13*, 187–192.
(12) Johnstone, A. H. *Chem. Educ. Res. Pract.* **2006**, *7*, 49–63.
(13) Campbell, D. J. *Acad. Manage. Rev.* **1988**, *13*, 40–52.
(14) Nadolski, R. J.; Kirschner, P. A.; van Merriënboer, J. J. G.; Wöretschofer, J. *Educ. Res. Eval.* **2005**, *11* (1), 1–27.
(15) Boggs, D. H.; Simon, R. J. *J. Appl. Psychol.* **1968**, *52*, 148–152.
(16) Kernan, M. C.; Bruning, M. S.; Miller-Guhde, L. *Hum. Perform.* **1994**, *7*, 273–289.
(17) Scott, W. E.; Fahr, J.; Podsakoff, P. M. *Organ. Behav. Hum. Decis.* **1988**, *41*, 405–425.
(18) Maynard, D. C.; Hackel, M. D. *Hum. Perform.* **1997**, *10* (4), 303–330.
(19) Sweller, J. *Learn. Instr.* **1994**, *4*, 295–312.
(20) Sweller, J.; Chandler, P. *Cognition Instruct.* **1994**, *12*, 185–233.
(21) Paas, F.; Renkl, A.; Sweller, J. *Educ. Psychol.* **2003**, *38*, 1–4.
(22) Ayres, P. *Learn Instr.* **2006**, *16*, 389–400.
(23) Pawley, D.; Ayres, P.; Cooper, M.; Sweller, J. *Educ. Psychol.* **2005**, *25* (1), 75–97.
(24) Niaz, M. *J. Chem. Educ.* **1987**, *64*, 502–505.
(25) Simon, H. A. *Science* **1974**, *183*, 482–488.
(26) Munyofu, M.; Swain, W. J.; Ausman, B. D.; Lin, H.; Kidwai, K.; Dwyer, F. *Learn. Media Technol.* **2007**, *32* (4), 407–419.

(27) Shrout, P. E.; Fleiss, J. L. *Psychol. Bull.* **1979**, *86*, 420–428.
(28) Nunnelly, J. C. *Psychometric Theory*, 2nd ed.; McGraw Hill: New York, 1978.
(29) Knaus, K. J.; Murphy, K. L.; Holme, T. A. *J. Chem. Educ.* **2009**, *86*, 827–832.
(30) Huber, V. L. *J. Appl. Psychol.* **1985**, *70*, 492–504.
(31) Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Mahwah, NJ, 1988.
(32) Sweller, J.; van Merriënboer, J. J. G.; Paas, F. G. W. C. *Educ. Psychol. Rev.* **1998**, *10*, 251–296.
(33) Claesgens, J.; Scalise, K.; Wilson, M.; Stacy, A. M. *Assess. Update* **2008**, *20*, 6–8.
(34) Florida Department of Education. Cognitive Complexity Classification of FCAT Test Items. http://fcat.fldoe.org/pdf/cog_complexity-fv31.pdf (accessed Feb 2011).