Chemistry Publications

Chemistry

2013

Adaptation of an Instrument for Measuring the Cognitive Complexity of Organic Chemistry Exam Items

Jeffrey R. Raker
Iowa State University

Jaclyn M. Trate
University of Wisconsin - Milwaukee

Thomas Holme
Iowa State University, taholme@iastate.edu

Kristen Murphy
University of Wisconsin - Milwaukee

Follow this and additional works at: http://lib.dr.iastate.edu/chem_pubs

Part of the Educational Assessment, Evaluation, and Research Commons, Higher Education Commons, Organic Chemistry Commons, Other Chemistry Commons, and the Science and Mathematics Education Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/chem_pubs/433. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

Adaptation of an Instrument for Measuring the Cognitive Complexity of Organic Chemistry Exam Items

Abstract

Experts use their domain expertise and knowledge of examinees' ability levels as they write test items. The expert test writer can then estimate the difficulty of the test items subjectively. However, an objective method for assigning difficulty to a test item would capture the cognitive demands imposed on the examinee as well as be assignable by any domain expert familiar with the examinee group. One such instrument for assigning objective complexity of general chemistry exam items has already been reported. A revised instrument for assigning objective complexity of organic chemistry exam items is presented including the reliability and validity studies of the rubric.

Disciplines

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

Adaptation of an Instrument for Measuring the Cognitive Complexity of Organic Chemistry Exam Items

Jeffrey R. Raker,[†] Jaclyn M. Trate,[‡] Thomas A. Holme,[†] and Kristen Murphy*[,‡]

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

[‡]Department of Chemistry and Biochemistry, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53201, United States

ABSTRACT: Experts use their domain expertise and knowledge of examinees' ability levels as they write test items. The expert test writer can then estimate the difficulty of the test items subjectively. However, an objective method for assigning difficulty to a test item would capture the cognitive demands imposed on the examinee as well as be assignable by any domain expert familiar with the examinee group. One such instrument for assigning objective complexity of general chemistry exam items has already been reported. A revised instrument for assigning objective complexity of organic chemistry exam items is presented including the reliability and validity studies of the rubric.

KEYWORDS: *Testing/Assessment, Organic Chemistry, Chemical Education Research*

FEATURE: Chemical Education Research
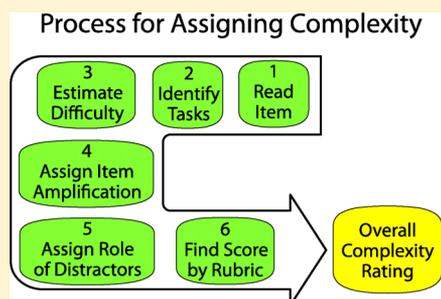
■ INTRODUCTION

A fundamental premise in any form of content testing is that the test items constitute a set of cognitive tasks required of students. In 2011, Knaus et al. reported the development of an instrument for the expert-based, objective determination of cognitive complexity for general chemistry exam items.[1] The basis of the instrument was the identification of concepts or skills needed to answer an exam item, the relative difficulty (i.e., easy, medium, or hard) for each concept or skill, and the interactivity of the concepts and skills (i.e., nonsignificant, basic, or complex). Each concept or skill and the interactivity were assigned a numerical value that when summed represents a numerical measure of objective exam item complexity. Knaus et al. further report inter-reliability statistics for the complexity measures and associated their objective complexity with general chemistry exam item performance and student mental effort ratings (i.e., subjective complexity), thereby establishing validity and reliability for the complexity rubric for general chemistry exam items.

Cognitive complexity is a concept that was first proposed by Bieri[2] where "cognitive complexity—simplicity" designated "the degree of differentiation of the construct system" (ref 2, p 263), or a more cognitively complex system differentiates well where a cognitively simple system poorly differentiates. Cognitive complexity was further quantified by examining the number of independent constructs as proposed by Crockett.[3] Cognitive complexity has been examined with mathematics items,[4] on science[5] and chemistry[6] assessments, between multiple-choice and constructed-response formats,[7] and in relation to spatial tasks,[8] to name a few. Within cognitive complexity one can specifically focus on the complexity related to a task where task complexity is predicated on considering the cognitive demand a

test item (the cognitive task) imposes on students. This idea expands on the foundational concepts of cognitive load theory wherein the cognitive demand on students manifests primarily in the working memory.[9−13] The fundamental concepts of cognitive load have factored into a number of theories of cognition and learning;[14] therefore, the combination of information processing and cognitive load can serve as a useful organizational theme for understanding task complexity of chemistry test items.[6,15,16]

The mental workload or cognitive demand on the working memory is separated into three components: the intrinsic, extraneous, and germane cognitive load. The extraneous cognitive load is the component that can be altered, depending on the learner and the environment presented to the learner. For example, presenting multiple representations to explain a concept that require a learner to integrate information can pose a higher extraneous cognitive load as opposed to presenting a single, integrated figure with explanation. Intrinsic cognitive load is the inherent difficulty of the material from the learner perspective and cannot be altered. The germane cognitive load is the portion of the load remaining for processing in the working memory. The "intrinsic cognitive load refers to the internal complexity of the task being attempted" and this should be able to be estimated by experts both in the domain and with the level of learner and their understanding.[17] The tasks must then be considered in terms of the elements of knowledge required to complete the task from the perspective of the learner. These elements have varying levels of difficulty and may be related to one another in the successful completion of the task. Therefore, both the estimation of the difficulty of

Table 1. Chronology of the Rubric Development Process

Time	Location	Intent of Focus Group
Spring 2010	ACS National Meeting, San Francisco, CA	Discuss visual−spatial components of common organic chemistry items and experts ability to relate to what students interpret from these representations
Fall 2010	ACS National Meeting, Boston, MA	(1) Discuss the general chemistry rubric and how to adapt for organic chemistry
		(2) Trial the new ideas for assigning complexity with organic chemistry items
Winter 2010	Combined Southwestern and Southeastern Regional Meetings of the ACS, New Orleans, LA	Testing of the complexity rubric with two groups of participants (faculty and graduate students). Rated components of approximately 50 organic chemistry exam items
Spring 2011	ACS National Meeting, Anaheim, CA	Testing of the complexity rubric with faculty. Rated components of approximately 50 organic chemistry exam items
Fall 2011	ACS National Meeting, Denver, CO	Testing of the complexity rubric with faculty; Rated components of approximately 50 organic chemistry exam items.
Winter 2011	Southwestern Regional Meetings of the ACS, Austin, TX	Testing of the complexity rubric with faculty. Rated components of approximately 50 organic chemistry exam items
Spring 2012	ACS National Meeting, San Diego, CA	Testing of the numeric complexity rubric with faculty; Rated components of approximately 50 organic chemistry exam items
Summer 2012	Biennial Conference on Chemical Education, Pennsylvania State University	Testing of the numeric complexity rubric with faculty. Rated components of approximately 50 organic chemistry exam items

the elements of knowledge and their interactivity are key to a successful estimation of the cognitive complexity of an exam item. Fundamentally, this estimation is based on the expertise of the raters' knowledge of the knowledge of the learner.

Initial attempts to use an unmodified version of the general chemistry complexity rubric on multiple-choice organic chemistry exam items found that expert raters were unable to directly apply this rubric to exam items in organic chemistry. Through a series of focus groups with organic chemistry instructors, a portion of those who use and write organic chemistry examinations for the ACS Examinations Institute, a modified complexity rubric has been designed that offers a valid and reliable instrument for multiple-choice organic chemistry exam items. This paper will describe: (i) the process used to develop the modified instrument, (ii) a method for using the instrument to evaluate complexity of exam items, and (iii) studies to establish instrument validity and reliability.

■ INSTRUMENT DEVELOPMENT

Over the course of three years, eight focus groups were held at national and regional meetings of chemists and chemical educators. The location, time, and details about the various focus groups are listed in Table 1.

Two major themes related to ways to enhance the complexity rubric emerged during these activities. First, the "concept/skill interactivity" factor of the general chemistry complexity rubric was not supported by the organic chemistry workshop participants. The interactivity between the concepts and skills necessary to answer an exam item was associated with an idea participants defined as "amplification". The term amplification attempts to capture the sense that students confront the learning of the many chemical reactions they encounter in organic chemistry using strategies that cluster similar reactions that may be quickly recognized, or require more effort to recognize. As such the rankings were set as easy, medium, and hard levels rather than nonsignificant, basic, and complex, which were used in the general chemistry complexity rubric. Discussions during the development of the instrument rarely included methods of translating qualitative information into a numeric rating system. As such, when the discussion of "amplification" began at the first focus group, the association of this concept as what students must do with elements of knowledge to successfully solve exam items was the predominate feature of this component, rather than any mathematical operation or translation into a numeric rating.

Participants, drawing on their instructional experiences, may have inferred a multiplicative factor to describe the interactivity and this differs from the additive factor as used in the general chemistry complexity rubric.

The second rubric enhancement was related to ways to characterize answer options for the multiple-choice items. This theme emerged spontaneously at the second focus group and was routinely reaffirmed by subsequent raters as an important inclusion in a rubric for assigning complexity to organic chemistry test items. For general chemistry, multiple-choice exam items consist of a task prompt and possibly a table, graph, or representation; an answer can often be derived independent of the exam item's multiple answer options. In other words, a student most often can develop an answer prior to looking for that answer in the item's answer options. This is not as often the case with multiple-choice organic chemistry exam items. Some exam items could be solved with the "selection" process used in general chemistry; however, some items required an "elimination" of answer options to determine the correct answer and some items required an "evaluation" of every answer option to determine the correct, best answer. Elimination and evaluation of options may represent multiple-choice test taking strategies for any content topic when a "selection" answer is not known, but elimination and evaluation are not optional test-taking strategies for the evaluated multiple-choice organic chemistry exam items; they were required.

As an example of an elimination item on an organic chemistry test, consider a question that requires a student to determine which compound (from the multiple-choice responses) produces a given infrared spectrum. To answer this question, an examinee generally needs to determine the functional groups present (or not present) in the infrared spectrum and eliminate answer options based on this information. Answer options are eliminated based on the information inferred from the spectrum until only one answer option, the correct option, remains.

For an example of an evaluation item, consider a question that expects a student to determine which of a series of reactions leads to the formation of a desired product. To answer this question, an examinee must normally evaluate each of the series of reactions to determine which would lead to the formation of the desired product. In comparing evaluation to elimination, elimination is the use of information in the task prompt to eliminate answers, whereas evaluation is the use of

information in the answer options to determine the correct answer.

To produce a more functional complexity rubric for applications to organic chemistry test items, each of these concerns must be considered. While the workshop participants tended to view an "amplification factor" as multiplying the complexity (e.g., "this makes the questions twice as hard") in terms of the concepts of working memory and cognitive load, this impression is not consistent with learning theories, so the amplification factor is considered as an additive component of complexity. This choice may be at odds with the linguistic prompts associated with amplification, but the label was chosen by a group of experts within the content domain. The idea of making an exam item more difficult by including more elements maps to their cognitive structure of the domain; therefore, maintaining a word that prompts a useful categorization is arguably more important than the specific mathematical construct this factor holds within the model for assigning the complexity. Next, the adjustment of the complexity rubric model to account for the differing role of distractors is also considered an additive component of complexity. Items that can be solved using only selection will have an additive value of "0", thus maintaining coherence with the general chemistry rubric. Items that use elimination have an additive value of "1", and items that required evaluation of responses have an additive value of "2". This ranking of the role of distractors assumes that the cognitive demand of eliminating answer options is less onerous than evaluating answer options. Each of these factors will be discussed in more detail in the next section on using the instrument to evaluate exam items.

Figure 1 presents the rubric by which the cognitive complexity of an organic chemistry exam item or other

Using the Instrument

Difficulty of Subtasks

Rating	Easy	Medium	Hard
1	1		
2	2	1	
3	3	2	
4	4	3	1
5	5	4	2
6	6	5	3
7	7	6	4
8	8	7	5

Amplification

Rating	Amplification
1	Easy
2	Medium
3	Hard

Role of Distractors

Rating	Role of Distractors
0	Selection
1	Elimination
2	Evaluation

Figure 1. Cognitive complexity rating rubric for multiple-choice organic chemistry test items.

chemistry task can be determined. Analysis of an exam item is achieved through a seven-step process in which the rater must

1. Analyze the item.
2. Determine the subtasks (i.e., student processes) used in solving the problem as "what the student needs to do" and "what the student needs to recognize".
3. Estimate from the perspective of a student the relative difficulty (i.e., easy, medium, or hard) of each of the subtasks.
4. Estimate from the perspective of a student how difficult (i.e., easy, medium, or hard) the extent of the interactivity of the subtasks required to answer the problem and assigns an amplification score from the rubric.
5. Determine the role of the distractors (i.e., selection, elimination, or evaluation) and assign a score from the rubric.
6. Use the rubric to assign a score to each subtask process.
7. Determine the overall complexity rating by summing all of the values from steps 4–6.

The design of the "difficulty of subtasks" component of this rubric follows that of the previously published rubric in general chemistry[1] where overlap exists between the easy, medium, and hard subtask ratings. This feature allows for two or more raters to come to similar complexity values through different parsing of the subtasks. For example, one rater may state that a given problem is composed of "three" easy tasks, whereas another rater may state that a given problem is composed of "two" medium tasks. In both these instances, the overall subtask complexity score would be 3. Both sets of subtasks and corresponding difficulty ratings lead to the same complexity score. This feature of the rubric allows for greater inter-rater reliability, as will be discussed in the next section.

As noted earlier, when compared to the general chemistry complexity rubric, this new organic chemistry complexity rubric renames the "interactivity" factor to the "amplification" factor. Because both factors are additive, the difference in labels can be viewed as semantic, as is justified by the difference in language usage of groups of chemists with different subdisciplinary specialization. In determining instrument reliability, several numeric models of the rubric were tested to determine whether the amplification factor should be considered a multiplicative factor; these models provided similar reliability statistics. Therefore, in an effort to keep the organic chemistry complexity rubric similar to the general chemistry complexity rubric, the amplification factor adds a value of 1, 2, or 3 to the overall problem complexity corresponding to easy, medium, or hard. Note that the organic chemistry complexity rubric does not include a 0 value corresponding to "nonsignificant" amplification as was the case with the general chemistry rubric. This difference suggests a difference in perspective among instructors in the two types of courses, where organic chemists perceive that test items in their subject inevitably lead to having students piecing together concepts.

Within this template, Figure 2 provides an example of the assignment of complexity for a multiple-choice organic chemistry item. For this item, which seeks to determine the ability of students to identify an aromatic molecule, the example shows the first analysis step as one way that the required cognitive tasks can be parsed. It is important to realize that, while this example is drawn from workshop participants, it is

Step 1. Analyze the item.

Which compound or ion is aromatic?

A   B   C   D

Steps 2 and 3. Determine the subtasks necessary for a student to answer the question (Step 2) and assign a corresponding relative difficulty value to each one (Step 3).

Subtasks	Level of Difficulty
Recall the definition of an aromatic compound	Easy
Determine if each structure has a cyclic array of p-orbitals	Medium
Determine if each structure has requisite "$4n + 2$" electrons in the p-orbitals	Medium

Step 4. Next, consider the subtasks identified in Step 2 and determine their level of amplification (easy, medium, or hard).

Description	Amplification (Rating)
These subtasks together are the basic algorithm for determining aromaticity of a compound	Easy

Step 5. Next, consider the distractors in the problem and determine the role of the distractors in answering the problem (selection, elimination, or evaluation)

Description	Role of Distractors (Rating)
Distractors are eliminated based on the definition of an aromatic compound	Elimination

Steps 6 and 7. Use the rubric to determine the rating of the subtasks (Step 6) and the overall objective complexity rating for this item (Step 7).

Components	Count	Component Set Rating
Easy Subtasks	1	1
Medium Subtasks	2	3
Hard Subtasks	0	0
Amplification	NA	1
Role of the Distractors	NA	1
Overall cognitive complexity rating (sum of ratings) for this item	**NA**	**6**

Figure 2. Example of a rater assignment of cognitive complexity of an organic chemistry item.

not unique and the rubric does not require uniqueness. Once the student steps are identified and classified in terms of their challenge level, the amplification and distractor role aspects are considered. When all steps are taken, a numerical value can be determined.

■ INSTRUMENT RELIABILITY

This instrument is designed to be an expert-based rating system whereby multiple organic chemistry instructors rate a given test item. For any such instrument, measures of inter-rater agreement must be determined. Faculty who write and use ACS Organic Chemistry Examinations, who were not part of the rubric design team, were recruited to serve as expert raters. These faculty members were asked to rate the complexity of all exam items from a 50-question online Organic Chemistry Practice Exam offered by the ACS Examinations Institute; of these 50 items, 42 items were rated by all raters (and thus included in the analysis). This practice exam was independently administered to a trial group of students in Spring 2012. The faculty received a 25-min training session on using the instrument; the training included the rating of two exam items as a workshop group. Independent ratings of exam item complexity were then made. Inter-rater reliability for the cognitive complexity rating rubric was established using ratings from 42-items of the practice exam. Two-way, mixed intraclass correlations were calculated using Stata 12 statistical analysis software.[18] Additionally, Cronbach α values were calculated. 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. Table 2 provides the data of the calculated intraclass correlation coefficients for ratings

Table 2. Inter-Rater Reliability Statistics for Revised Cognitive Complexity Ratings

Measure	ICC Values[a]	Lower Bound	Upper Bound	F Values	Significance
Single Rater	0.3801	0.2630	0.5246	5.91	<0.001
Average of Raters	0.8307	0.7406	0.8982	5.91	<0.001

[a]Two-way, mixed intraclass correlation coefficient values for 8 raters (42 items rated); Cronbach α = 0.8349.

collected from the practice exam. These data establish an inter-rater agreement of approximately 83%. This value is sufficient, per accepted cutoffs, for establishing internal reliability.

■ INSTRUMENT VALIDITY

To determine the instrument validity, correlation studies were performed between the complexity ratings derived from the rubric and both student performance on exam items and average student mental effort on the exam items. Item complexity, as determined by this rubric, is a proxy for objective complexity because ratings are derived from experts who parse the cognitive steps a student must likely take independent of the action of any particular student. Because a more complex item has a greater cognitive demand, a negative correlation between objective complexity, thus measured, and student performance is hypothesized. In a similar way, unless most students use an unexpected pathway that subverts the estimated objective complexity, a positive correlation between objective complexity and average student mental effort (i.e., subjective complexity) can be predicted. Similar correlations were observed in a study on general chemistry practice exams.[19] These hypotheses were tested by recording the performance (correctness of item responses) and load on working memory (measured by student reports of mental effort) from 80 students who participated in an online trial offering of an ACS Examinations Institute Organic Chemistry Practice Exam during the Spring 2012 semester. During Spring 2012, students who purchased "ACS Organic Chemistry Exams—The Official Guide" received an enclosed card that offered the use of the online practice exam. Additionally, one instructor of organic chemistry was provided with codes for the students in the organic chemistry course to participate in the online practice exam in preparation for their final exam. Therefore, the student data collected were presumably from students preparing to take a final examination or similar in organic chemistry.

To process this information into validity evidence, student data were included in this study only if they responded to all 50 exam questions and provided mental effort ratings for each of those items. Performance was calculated for each exam item as the fraction of students answering the question correctly, which defines the item difficulty in classical test theory.[20] Students were also asked to rate their mental effort on each exam item by

responding to the prompt "how much mental effort did you expend on this question?" that immediately followed each item. Students could respond with "very little", "little", "moderate amounts", "large amounts", or "very large amounts". To put these responses onto a numerical scale for comparison basis, they were scored as 1, 2, 3, 4, and 5, respectively. Average student mental effort ratings were calculated for each exam item as the sum of mental effort values divided by the total number of students.

Once the data were collected, several steps were taken to process that data. First, student performance versus cognitive (objective) complexity was plotted, as shown in Figure 3. As
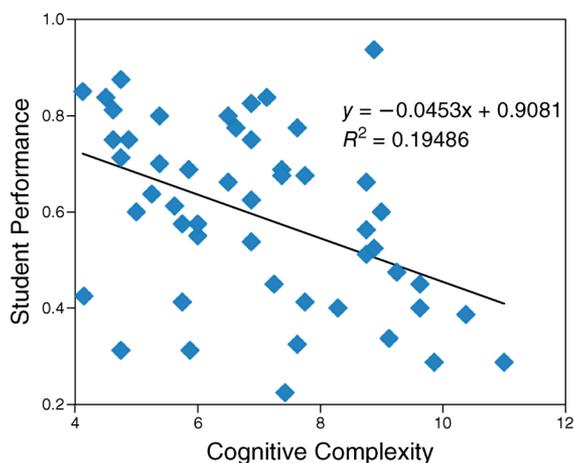
Figure 3. Correlation of student performance (using statistical item difficulty) with cognitive complexity (using average item ratings) for 80 students on the ACS organic chemistry online practice exam (Pearson correlation $r = -0.4414$; $p = 0.0013$).

hypothesized, a negative relationship exists between these two variables; lower cognitive complexity values, in general, correspond to better student performance (i.e., values closer to 1). A Pearson correlation between these two measures was determined to be -0.4414 ($p = 0.0013$); this is considered to be a moderate and significant correlation. Visual inspection of the scatter plot suggests a handful of items for which student performance is lower than might be predicted based on task complexity alone. Items that include common misconceptions as distractors, for example, can fall into this category. The complexity of the item is not increased by the availability of these incorrect answers, but they nonetheless collect a relatively high fraction of students, thereby lowering student performance on such items. Faculty members who regularly teach organic chemistry and serve on exam writing committees have developed expertise in identifying the common distractors that students may select based on incorrect processes. These distractors can include misconceptions, although whether a distractor specifically addresses a misconception is not requested of the committee when the exam items are constructed. Additionally, through the editing process, exam items are routinely altered as the committee contributes their collective expertise, thus making it difficult to pinpoint one expert's knowledge of students' misconceptions and how this is translated into particular distractors.[21] It is also important to note that it is not inherent in the nature of the distractor that represents a misconception whether or not the student "finds" that distractor with more, less, or about the same amount of mental effort, that is, experiences similar or different cognitive

complexity. Further investigation of outliers, particularly with qualitative methods, to elucidate this could provide more insight into what a student is experiencing while working through these items; however, this was beyond the scope of this project.

Second, student performance versus average mental effort was plotted, as shown in Figure 4. Once again, as predicted, a
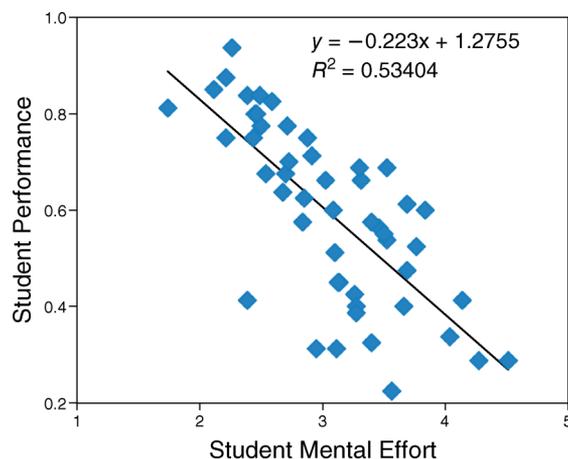

Figure 4. Correlation of student performance (using statistical item difficulty) with cognitive complexity (using average item ratings) for 80 students on the ACS organic chemistry online practice exam (Pearson correlation $r = -0.7308$; $p < 0.0001$).

negative relationship exists between these two variables; lower average mental effort ratings, in general, correspond to better student performance (i.e., values closer to 1). A Pearson correlation between these two measures was determined to be -0.7308 ($p < 0.0001$); this is considered to be a high and significant correlation.

Lastly, cognitive complexity versus average mental effort was plotted as shown in Figure 5. In this case, the prediction is that the two forms of complexity, objective and subjective, that are postulated to be measured should mirror each other. As predicted, a positive relationship exists between these two variables; high cognitive complexity ratings tend to correspond
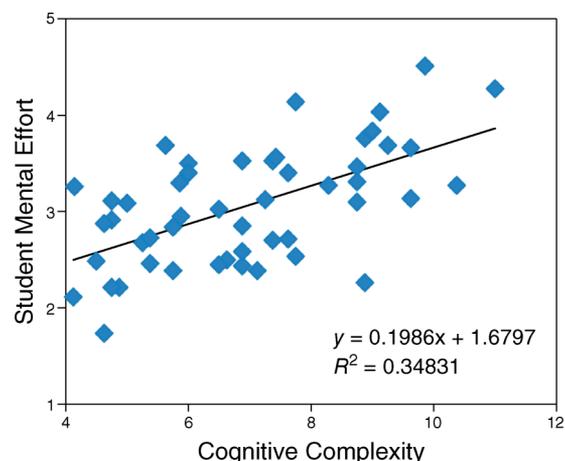

Figure 5. Correlation of student mental effort (using average student ratings) with cognitive complexity (using average item ratings) for 80 students on the ACS organic chemistry online practice exam (Pearson correlation $r = 0.5902$; $p < 0.0001$).

to high average mental effort ratings. A Pearson correlation between these two measures was determined to be 0.5902 ($p < 0.0001$); this is considered to be a moderate and significant correlation.

These moderate to high correlations between student performance, cognitive complexity, and student mental effort suggest that the developed rubric is valid. The correlation values obtained with this study are higher than those found in the previously described general chemistry complexity rubric.[19] This observation suggests that the revised rubric is an improved tool for measuring the cognitive complexity and is more inclusive of a broader set of examination items.

CONCLUSION AND IMPLICATIONS FOR INSTRUCTION

This study has reported a revised instrument for the assignment of cognitive complexity of organic chemistry exam items. A rubric that helps content experts parse cognitive demands on students of multiple-choice items is presented. High inter-rater reliability measures were obtained for this revised rubric instrument. The cognitive complexity assigned to a set of organic chemistry items using this tool correlates at moderate to high levels with similar constructs (i.e., student performance on exam items and average mental effort ratings). As with the initial publication of the general chemistry complexity rubric, this revised rubric will provide an enhanced window into "exploring the relationship between the complexity of content taught and student cognition and learning".[1] In addition, studies of complexity can contribute to the development and design of chemistry assessment materials.

Beyond the research and development of an instrument to assign cognitive complexity, this instrument can also be useful for individual instructors in developing assessments. An experience that is quite likely for many instructors is the construction of a test thought to be easy and yet the students perform poorly on the items. One possible consideration for the "easy" items would be to conduct a complexity analysis of the items. Through this analysis, the instructor can gain a clearer picture of what the students must successfully recognize and do in order to perform well on the test. This analysis can be remarkably revealing in how difficult items may actually be for students. Beyond this, instructors can evaluate possible test items prior to giving an assessment for a better distribution of easy, medium, and hard test items. Although this process will always involve the human element and is not envisioned to be automated, fluidity in regular use of the rubric can underlie the process of test development. Indeed the rubric is used in similar efforts of the ACS-EI to assign difficulty of items by groups of raters as easy, medium, or hard. Ultimately, a better sensitivity of the difficulty of test items from the test-taker perspective can assist instructors in building and using better assessment tools.

AUTHOR INFORMATION

Corresponding Author
*E-mail: kmurphy@uwm.edu.

Notes
The authors declare no competing financial interest.

REFERENCES

(1) Knaus, K. J.; Murphy, K. L.; Blecking, A.; Holme, T. A. *J. Chem. Educ.* **2011**, *88*, 554.
(2) Bieri, J. *J. Abnorm. Soc. Psychol.* **1955**, *51*, 263.
(3) Crockett, W. H. *Prog. Exp. Personality Res.* **1965**, *2*, 47.
(4) Daniel, R. C.; Embretson, S. E. *Appl. Psychol. Meas.* **2010**, *34*, 348.
(5) Baxter, G. P.; Glaser, R. *Educ. Meas: Issues Pract.* **1998**, *17*, 37.
(6) Johnstone, A. H. *Chem. Educ. Res. Pract.* **2006**, *7*, 49.
(7) Hancock, G. R. *J. Exp. Educ.* **1994**, *62*, 143.
(8) Thompson, E. G.; Mann, I. T.; Harris, L. *J. Br. J. Psychol.* **1981**, *72*, 249.
(9) Kalyuga, S.; Sweller, J. *J. Educ. Psychol.* **2004**, *96*, 558.
(10) Nadolski, R. J.; Kirshner, P. A.; van Merrienboer, J. J. G.; Woretschofer, J. *Educ. Res. Eval.* **2005**, *11*, 1.
(11) Paas, F.; Tuovinen, J. E.; Tabbers, H.; VanGerven, P. W. M. *Educ. Psychol.* **2003**, *38*, 63.
(12) Pelligrino, J. W.; Chudowsky, N.; Glaser, R. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academy Press: Washington, DC, 2001.
(13) van Merrienboer, J. J. G.; Sweller, J. *J. Educ. Psychcol.* **2005**, *97*, 147.
(14) Mayer, R. E. *Multimedia Learning*, 2nd ed.; Cambridge Univ. Press: New York, 2009.
(15) Benbasat, I.; Todd, P. *Decision Support Systems* **1996**, *17*, 241.
(16) Campbell, D. J. *The Academy of Management Review* **1988**, *13*, 40.
(17) Pawley, D.; Ayres, P.; Cooper, M.; Sweller, J. *J. Educ. Psychol. U.K.* **2005**, *25*, 75.
(18) StataCorp. *Stata Statistical Software: Release 12*; StataCorp LP: College Station, TX, 2011.
(19) Knaus, K. J.; Murphy, K. L.; Holme, T. A. *J. Chem. Educ.* **2009**, *86*, 827.
(20) Crocker, L.; Algira, J. *Introduction to Classical and Modern Test Theory*; Holt, Reinhart and Wilson: New York, 1986.
(21) Holme, T. A. *J. Chem. Educ.* **2003**, *80*, 594−598.