Are we keeping the promise? Investigation of students' critical thinking growth

Journal of College Science Teaching, May-June 2013

Author abstract

College instruction aims not only to expand students ' content knowledge, but also to help students develop practical skills, such as the ability to think critically. This study was conducted in a chemistry course for nonscience majors offered as part of a liberal education core curriculum at a large public university in the Midwest. Students enrolled in the class were given the Lawson Classroom Test of Scientific Reasoning as a pre- and postcourse measure of their scientific reasoning (SR) ability, a subset of critical thinking. Although average gains in students ' SR ability were significant, the improvement was quite modest when compared with where students should be developmentally, according to Piaget. Literature suggests the need for instruction that markedly improves intellectual development, which may be attained through the use of more student-centered practices. Data presented support this need for instruction.

In the field of education, the use of the phrase critical thinking is ubiquitous. Doing a keyword search in this journal results in more than 30 articles from February 1998 through March 2012 that have some connection to this very important but ill-defined construct. The National Council for Excellence in Critical Thinking (1987) defined critical thinking as "the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action." Gill and Burke (1999) stated that critical thinking is the critical evaluation of evidence, whereas Herreid (2004) mentioned that critical thinking is "skepticism, flexibility, and the ability to see alternative approaches" (p. 12).

Clearer definitions begin to appear with Lauer (2005), who defined the construct using the top three levels of Bloom's Taxonomy (analysis, synthesis, and evaluation), and with White et al. (2011), citing a definition of "purposeful, self-regulatory judgment" (p. 102). All of these definitions look different and emphasize different facets of the construct. It is not surprising then, that assessing critical thinking is challenging (Ennis, 1993), especially with some of the popular tests being quite long and focusing on skills that are not necessarily aligned with skills valued by scientists (Facione et al., 1990). In order to assess the construct, it is necessary to operationalize the definition of critical thinking and specify what evidence would show different levels of proficiency in this skill (Ennis & Millman, 1985; Facione, 1990; Facione & Facione, 1992).

Aim of classes to improve critical thinking

According to the Association of American Colleges and Universities (AAC&U), one of the aims of liberal education is to develop intellectual and practical skills through a set of core classes that all students will take throughout their studies. Of these intellectual skills, critical thinking is explicitly stated (AAC&U, 2012). Each university develops its own "core" curriculum in order to address the intellectual and practical skills mentioned by the AAC&U. Although curricula are designed to fill the need of critical thinking development and this is promised in many syllabi, the assessment is challenging because of the breadth of the construct.

Critical thinking in chemistry

In the preface to the textbook Chemistry in Context, which was designed by the American Chemical Society for chemistry classes for nonscience majors, the development of critical thinking skills is particularly emphasized. The book's contextual approach is said to help students to develop "critical thinking ability, the chemical knowledge and

competence to better assess risks and benefits, and the skills that enable them to make informed and reasonable decisions about technology-based issues" (American Chemical Society, 2006, p. xi). To assess the claims that courses, texts, and curricula are developing students' critical thinking skills, the definition of critical thinking needs to be narrowed to focus on specific skills that all scientists--not just chemists--value, use, and aim to develop in science students.

Scientific reasoning

Scientific reasoning (SR) focuses on the reasoning processes that help students identify and evaluate evidence to support or reject hypothetical causal propositions (Lawson, 1978). Looking for patterns in scientific data and deciding whether hypotheses are supported on the basis of experimental data are keys to success in learning science. As scientists, we engage in these processes and use these reasoning skills with a high level of fidelity and autonomy. Students, however, need coaching to be able to do this effectively. If it is the goal of introductory science classes to teach students to think more like scientists and thus to become better consumers of science, then it is important to measure the effectiveness of this mission. This matter becomes more important when we consider that for nonscience majors, introductory courses are most likely the only exposure these students will have to formal science experiences at the college level.

Theoretical framework

The Piagetian developmental model is used as a framework for the theory of student learning that guides data collection and analysis in this study. With the construction of the Lawson Classroom Test of Scientific Reasoning (LCTSR) being based on Piagetian tasks, student performance on the test is a portrayal of their Piagetian developmental level. Piaget provides a continuum of stages in students' intellectual development on the basis of their age and skills, moving from the sensorimotor stage and preoperational thought to concrete and formal operational thought (Inhelder & Piaget, 1958). At the sensorimotor and preoperational stages, reasoning based on the rules of formal logic does not take place. In the concrete and formal operational stages, logical thought is present, the difference being that concrete thought requires physical manipulation of objects, whereas formal thought uses abstract thoughts and ideas.

Although the ages Piaget associated with each developmental level have been criticized (Driver, 1978; Lourenco & Machaldo, 1996), such critiques do not diminish the application of his theory to the study presented here. This study aims to examine growth in reasoning ability rather than focusing on absolute measures of developmental stages. Piaget's tasks provide the theoretical basis for a scale on which to measure students' SR abilities to address these research questions:

1. What are students' SR abilities on entering a chemistry course for non-science majors?

2. What effect does typical instruction have on students' SR ability?

3. What strategies do students use to answer LCTSR items?

Methods

Setting

This study takes place at a large public university in the Midwest. The university is a primarily undergraduate institution that serves approximately 16,000 students. As part of their degree requirements, students must

complete a series of liberal education "core" courses in addition to fulfilling the courses within their major. To fulfill the science requirement of the core curriculum, students must take at least 9 credit hours of natural science: 3 hours must be in a biological science and another 3 hours must be in a physical science. At least one of the classes used to fulfill the requirement must be a laboratory course.

Sample

The sample is comprised of students enrolled in a chemistry course for nonscience majors: Chemistry in Modern Society. This is a 3-credit, one-semester course with no prerequisites and is offered every semester. The course is lecture based and meets three times per week with three semester exams and a final exam comprising a vast majority of the course grade. The course also has a laboratory component that is offered concurrently, but simultaneous enrollment is not required for completion of the students' physical science requirement. The course serves a wide range of majors, as can be seen in Figure 1, with the majority of students being business or education majors. The distribution of students' ACT scores are shown in Figure 2, with the mean ACT score for the university (26) highlighted in purple. The students enrolled in the course have had sufficient science and math courses, with most taking first-year college calculus and 4 years of science in high school, as shown in Figures 3 and 4.

Because these students are not science majors, they are often described as "second tier"; however, their demographic data demonstrating high achievement and persistence in secondary math and science (as shown by the number of courses taken in Figures 3 and 4) reinforce the notion that the "second tier is not second rate" (Tobias, 1990, p. 15).

Instrument

In order to measure SR ability, the LCTSR was used as a pretest and as a posttest (Lawson, 1978, 2000). This test has been used in multiple studies across the science disciplines (Coletta & Phillips, 2005; Johnson & Lawson, 1998; Lawson, Alkhoury, Benford, Clark, & Falconer, 2000; Lawson, Banks, & Logvin, 2007; Lawson, Clark, et al., 2000; Norman, 1997), often as a measure to determine skills of incoming students for the purposes of grouping them, rather than to measure their growth in the course (Bao et al., 2009). The LCTSR is a 12-item, two-tiered test that requires a correct answer and a correct reason for each item. The test is designed to measure students' abilities with respect to six different SR skills: conservation of weight and volume, proportional reasoning, probabilistic reasoning, correlational reasoning, control of variables, and hypothetico-deductive reasoning. The test is multiple choice, with each of the distracters being a correct answer for an incorrect reasoning pattern. An example item from the test is shown in Figure 5. Although it appears that items could be mapped to each of the aforementioned reasoning types, a review of previous analyses shows that the LCTSR, along with many other tests of formal reasoning, has a one-factor structure (Lawson, 1985).

Data collection

The university's Institutional Review Board granted approval for this study prior to any data collection. This study uses a one-group, pretest-posttest design (Shadish, Cook, & Campbell, 2002). A total of 314 students were given the pretest on the first day of class and the posttest within the last week of instruction. Students were given at least 25 minutes to complete the test, with most finishing within 15 to 20 minutes. The students were provided with scratch paper to work out any problems, and their answers were collected using a Scantron sheet.

Students who did not complete the test or who were obviously not taking the test seriously (e.g., making a pattern on the Scantron, having an answer for the question tier of "unable to be determined" and a reasoning response that gave an actual reason, etc.) were discarded from the final analysis, as well as those who did not take both the

pretest and posttest. Those who did take both tests but did not give consent for their data to be used were also removed from the data set. The final sample size was 177 students: 118 from the fall and 59 from the spring. Semistructured interviews were conducted on a small sample (N = 18) of the students from the fall semester to evaluate test-retest reliability and to further investigate students' reasoning patterns. Students who participated in the interviews were asked to solve the problems from the test aloud. Interviews lasted between 30 and 60 minutes and were transcribed verbatim. A Livescribe pen was also used in the interviews to record students' drawings and problem-solving procedures in real time. The interview participants are referred to using pseudonyms to protect their identities.

Data analysis

Tests were scored using a dichotomous scale (correct/incorrect), requiring a correct answer and correct reasoning to score 1 point for each item, making the maximum score for the assessment a 12. Descriptive statistics were run on the total scores for both the pretest and posttest and checked for normality. Difficulty and discrimination values were determined for each item of the assessment and the internal consistency of the instrument was tested using Cronbach [alpha]. The two populations (fall and spring) were tested for equivalence through the use of two one-tailed t-tests (Lewis & Lewis, 2005). To investigate differences between pretest and posttest data, a repeated measures analysis of variance (ANOVA) was used. Interview audio data were transcribed verbatim and analyzed using NVivo 9. The data were coded for instances of high-, medium-, and low-level reasoning to look for similarities and differences in the reasoning patterns of individuals at different reasoning levels. The qualitative data were also used to validate the test with this sample of students.

Results and discussion

Instrument performance

Each item on the pretest and the posttest was analyzed for both difficulty (the percentage of students correct) and discrimination (how well the item distinguishes between the high and low performers). Items that have an item difficulty below 0.25 are considered to be difficult, between 0.25 and 0.80 are ideal, and above 0.80 are considered to be easy (Popham, 2005). The discrimination of each item can be considered ideal if the value falls above 0.30 (Popham, 2005). A visual representation of this performance is shown in Figure 6.

The majority of the items fell within the ideal range for both the pretest and the posttest, with those that did not discriminate well falling in the easy and difficult categories. It would be expected that the easiest and hardest items would discriminate poorly because of the amount of high-performing students and low-performing students answering in the same way (all correct or all incorrect, respectively). Examples of an easy item (Item 1) and a difficult item (Item 11) are shown in Figures 7 and 8, respectively. Item 1 focuses on conservation of weight, which most of the students should be able to answer correctly as it is a question that focuses on preoperational thought, the Piagetian primer for concrete and formal operational. The data support this item as a low-level skill because a high percentage of both high- and low-performing students answered the questions correctly. Item 11 focuses on a construct that is unique to the LCTSR--hypothetico-deductive reasoning, the ability to deductively trace hypotheses and arrive at valid conclusions (Lawson, 2004). The process of hypothetico-deductive reasoning requires a substantial level of formal operational reasoning requiring students to analyze an experiment that they may or may not have experience with and answer questions about a hypothetical modification to the experiment. This higher level reasoning skill is not present in most students in this population. This can be inferred because of the high percentages of both high- and low-performing students who answered this question incorrectly.

FIGURE 5 Example of LCTSR questions. 2) To the right are drawings of two cylinders filled to the same level with water. The cylinders are identical in size and shape.

```
Also shown at the right are two marbles, one glass and one steel. The marbles are
the same size but the steel one is much heavier than the glass one. When the glass
marble is put into Cylinder 1 it sinks to the bottom and the water level rises to
the 6th mark. If we put the steel marble into Cylinder 2. the water will rise a. to
the same level as it did in Cylinder 1 b. to a higher level than it did in Cylinder
1 c. to a lower level than it did in Cylinder 1 because a. the steel marble will
sink faster. b. the marbles are made of different materials. c. the steel marble is
heavier than the glass marble. d. the glass marble creates less pressure c. the
marbles are the same size.
```

Internal consistency (Cronbach [alpha]) of the pretest was 0.599 and the posttest was 0.701. With alpha values near 0.7, the instrument can be said to be internally consistent (Murphy & Davidshofer, 2001). Interviews with study participants were conducted to evaluate the test-retest reliability of the instrument. For the 18 students who were interviewed, the stability coefficient (a correlation of test answers with retest answers) for the students' answers was strong and significant ($p = .676$, $p < .01$), suggesting that the test's results are stable. In further analysis of the students' test and retest answers, it was found that 74% of the answers given on the interview were consistent in whether they were correct or incorrect on the pretest.

This evidence suggests that the skills measured by the tests do not fluctuate on a short time scale (2-3 weeks), meaning that the authors can regard the scores on the test as a stable measure. In the analysis of the interview transcripts, students interpreted the questions accurately, lending validity to the data generated by the LCTSR. This is not surprising, given the wide use and multiple revisions to the test as recently as 2000.

Student performance

A test of equivalence between the two groups of students (fall and spring) suggested that the populations are in fact similar with respect to their LCTSR score as evidenced by results of analysis with two one-tailed t-tests (Lewis & Lewis, 2005). With this evidence of equivalence, the samples were pooled for further analyses to increase the statistical power. The pretest and posttest data were very close to being normally distributed. The pretest had a nonsignificant skew and kurtosis, whereas the posttest had a nonsignificant skew but a significant kurtosis, showing a slightly platykurtic distribution. The distributions of pretest and posttest total scores can be found in Figures 9 and 10.

As shown in Table 1, student scores on the pretest ranged from 1 to 10 out of a possible 12 points, with a mean of 5.6 and a standard deviation of 2.1. Scores on the posttest had a range of 1 to 12, with a mean of 6.3 and standard deviation of 2.5. The mean scores on the pretest and posttest highlight that most students are in the concrete stage, exhibiting mostly concrete thought, or in a transitional phase, having concrete thought with the beginnings of formal operational reasoning. A repeated-measures ANOVA was conducted with the within-groups variable being the pretest or posttest condition and the dependent variable being total LCTSR score. The data analyzed met each of the underlying assumptions of the ANOVA, except for that of having a normally distributed dependent variable. The test is still appropriate; however, as it is suggested that sample sizes greater than 30 per group will still result in trustworthy p-values (Green & Salkind, 2011). The difference between the means of the pretest and the posttest was 0.7 out of 12 (6% of the total score) and was statistically significant, Wilks' $A = 0.91$, $F(1, 177) = 18.109$, $p < .001$. Although instruction seems to have a medium effect size, [[eta].sup.2] = .093, the increase is modest and calls into question how much the course is truly improving students' SR skills.

[FIGURE 6 OMITTED]

Alternative incorrect reasoning strategies

A preliminary analysis of the interview data revealed that there are consistencies among the reasoning patterns of students who answer specific questions incorrectly. This important discovery will inform the authors' decision as to which reasoning skills to focus on through classroom activity design. In examining the incorrect responses for one proportional reasoning item, it is possible to illustrate the wide range of abilities of the students in the study (see item in Figure 11).

FIGURE 7 LCTSR Item 1. 1) Suppose you are given two clay balls of equal size and shape. The two clay balls also weigh the same. One ball is flattened into a pancake-shaped piece. Which of these statements is correct? a. The pancake-shaped piece weighs more than the hall b. The two pieces still weigh the same c. The ball weighs more than the pancake-shaped piece because a. the flattened piece covers a larger area, b. the ball pushes down more on one spot C. when something is flattened it loses weight d. clay has not been added or taken away. c. when something is flattened it gains weight. FIGURE 8 LCTSR Item 1. 11) The figure below at the left shows a drinking glass and a burning birthday candle stuck in a small piece of clay standing in a pan of water When the glass is turned upside down, put over the candle, and placed in the water, the candle quickly goes cut and water rushes up into the glass (as shown at the right). This observation raises an interesting question: Why does the water rush up into the glass? Here is a possible explanation. The flame converts oxygen into carbon dioxide. Because oxygen docs not dissolve rapidly into water but carbon dioxide does, the newly formed carbon dioxide dissolves rapidly into the water, lowering the air pressure inside the glass. Suppose you have the materials mentioned above plus some matches and some dry ice (dry ice is frozen carbon dioxide). Using some or all of the materials, how could you text this possible explanation? a. Saturate the water with carbon dioxide and redo the experiment noting the amount of water rise. b. The water rises because oxygen is consumed, so redo the experiment in exactly the same way to show water rise due to oxygen loss. c. Conduct a controlled experiment varying only the number of candles to see if that makes a difference. d. Suction is responsible for the water rise, so put a balloon over the top of an open-ended cylinder and place the cylinder over the burning candle. e. Redo the experiment, but make sure it is controlled by holding all independent variables constant, then measure the amount of water rise. What result of your test (mentioned above) would show that your expla- nation is probably wrong? a. The water rises to the same level as it did before. b. The water rises less than it did before. c. The balloon expands out d. The balloon is sucked in.

Those considered to have a high level of reasoning ability were able to arrive at the correct answer because of their correct use of proportions to address the task at hand. Sara, a high-reasoning student, shows how to correctly reason through the problem: "for [the first part], B, because 4/6 is 2/3 and 6/9 is 2/3 ... C, it goes up 3 in the narrow for every 2 in the wide." The Livescribe image of her work is shown in Figure 12a.

Brian was unable to arrive at the correct answer because of his use of additive reasoning, the Piagetian precursor to proportional reasoning (Inhelder & Piaget, 1958). Brian states: "I said A for [both parts], oh sorry, A for [the first part] and B for [the second part], because if it was 4 in the first one and 6 in the second one, it went up by 2, so if it goes up by 2 on A, it should go up by 2 on B again." Brian's work is shown in Figure 12b. What is also interesting is that students who answer with additive reasoning will go on to the next question, an identical task with different numbers, and state that the reason for their answers was because the ratios had to stay the same, when a true ratio was not used to solve the problem. This task seems to show a student's progression along the Piagetian

developmental spectrum, as there are some students who solve the problem using a hybrid of both additive and proportional reasoning. For example, Kelvin says, "okay, so you get up to the 6th mark so it would be 2 times 6 = 12, which equals 18, and, that's kinda ... so then just simplify it, divide by 2 and you get 6 for the wide, 9 for the narrow." In looking at the Livescribe data (Figure 12c) with his work, it appears that he begins his work thinking that a ratio will lead him to the correct answer; however he uses a series of addition steps to arrive at his final answer. Although Kelvin's answer to the question is correct, there is evidence of naive SR patterns. In addition to validating students' responses, it is apparent that the qualitative data are helpful in identifying particular student reasoning problems for some of the reasoning types. Emergent qualitative findings demonstrate that the distracters for each item expertly capture incomplete and incorrect patterns, making it more efficient to rely on the quantitative data to apply these findings to instruction--to ultimately target specific incorrect reasoning pattern through specific activities to build the skills in the correct way.

Limitations

Discussion of student performance and development is constrained to those types of reasoning evaluated by the LCTSR. Other methods to evaluate reasoning ability exist; however, they are inefficient for use with large samples. Although the course syllabus explicitly states critical thinking development as a goal, the skills measured by the LCTSR are not taught explicitly. The LCTSR provided a means to narrow critical thinking skills to a measurable and valued construct. As such, a limitation of the findings is that they may not capture all facets of students' critical thinking that may be developed in the course.

It is also important to mention that study results and implications described are only generalizable to the population studied: nonscience majors enrolled in a specific chemistry class. Other schools in the authors' home state, both large research universities and smaller liberal arts colleges, have similar courses, suggesting that studies done elsewhere may yield similar results. The modest but significant gain in LCTSR scores over the semester may be attributed to instruction or effects such as maturation and testing, typical threats to internal validity associated with the one-group pretest-postest design (Shadish et al., 2002).

Implications for teaching

The information gathered from this portion of the study has been used to design and implement guided inquiry classroom activities that not only focus on the content of the course, but also engage students in the reasoning skills that still have room for development: proportional reasoning and correlational reasoning in particular. Student performance on the proportional reasoning items (3 and 4) and the correlational reasoning item (10) can be seen in Figure 6. The guided inquiry framework for the activities was selected because of its documented success in assisting the improvement of students' SR skills in biology (Benford & Lawson, 2001). Using such activities responds to the call for more student-centered, collaborative learning opportunities in undergraduate science courses (National Research Council, 2012). Teaching for the development of critical thinking is important in a science course for nonmajors (Gill & Burke, 1999), and the traditional "stand and deliver" method often used in these courses is not effective in developing the reasoning skills that are promised. Building on Ben ford and Lawson (2001), using more student-centered approaches that use the learning cycle provides students with more opportunities to improve reasoning skills as they interact with course content.

[FIGURE 9 OMITTED]

[FIGURE 10 OMITTED]

Conclusions and future work

The aim of this study was to determine students' SR abilities at the beginning of a chemistry class for nonscience majors, how they reason through items on a SR test, and how instruction affected their abilities. If the aim of science courses for non-majors is to teach students content relevant to their lives while developing skills that help them to weigh evidence and draw sound conclusions, then it is necessary to assess both content and reasoning skills. From the current work, it has been shown that students' SR skills, although increasing, are not making adequate progress to where they should be by developmental standards. It is the authors' hope that through use of targeted interventions, students will engage in the learning cycle and develop the skills necessary to apply and reason with scientific information.

Future work related to this study includes an investigation of the effect of content-driven, reasoning skill centered, inquiry-based classroom activities on students' SR skills, as well as an exploration of the factor structure of the LCTSR using structural equation modeling (Raykov & Marcoulides, 2006). As previously mentioned, the instrument results in a single-factor structure, making it difficult to understand the relationships among the tasks evaluated by the LCTSR. A clearer understanding of the structure underlying the reasoning skills assessed by this instrument will not only be valuable for instruction, but also inform a path for growth in SR ability.

FIGURE 11 Proportional reasoning item. To the right are drawings of a wide and a narrow cylinder. The cylinders have equally spaced marks on them Water is poured into the wide cylinder up to the 4th mark (see A). This water rises to the 6th mark when poured into the narrow cylinder (see B). 3) Both cylinders are emptied (not shown I and water is poured into the wide cylinder up to the 6th mark How high would this water rise if it were poured into the empty narrow cylinder? a. to 8 b. to 9 c. to 10 d. to 12 e. none of these answers is correct because a. the answer cannot be determined with the information given. b. it went up 2 more before, so it will go up 2 more again c. it goes up 3 m the narrow for every 2 in the wide. d. the second cylinder is narrower. e. for every 2 in the wide it goes up 1 more in the narrow. [ILLUSTRATION OMITTED]

[FIGURE 12 OMITTED]

Acknowledgments

We thank the Chemistry in Modern Society students and their instructor for participating in this study. We appreciate the feedback received from the Yezierski and Bretz research groups at Miami University as well as the insightful comments from reviewers.

References

American Chemical Society. (2006). Chemistry in context: Applying chemistry to society (6th ed.). New York, NY: McGraw-Hill.

Association of American Colleges and Universities. (2012). Liberal education resources. Retrieved from http://www.aacu.org/resources/ liberaleducation/index.cfm

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., ... Wu, N. (2009). Learning and scientific reasoning. Science, 323, 586-587.

Benford, R., & Lawson, A. E. (2001). Relationships between effective inquiry use and the development of scientific reasoning skills in college biology labs (Report to the National Science Foundation). (ERIC Document Reproduction Service No. ED456157)

Coletta, V. P., & Phillips, J. A. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. American Journal of Physics, 73, 1172-1182.

Driver, R. (1978). When is a stage not a stage? A critique of Piaget's theory of cognitive development and its application to science education. Educational Research, 21, 54-61.

Ennis, R. H. (1993). Critical thinking assessment. Theory Into Practice, 32, 179-186.

Ennis, R. H., & Millman, J. (1985). Cornell Critical Thinking Tests, LevelX and Level Z: Manual. Pacific Grove, CA: Midwest Publications.

Facione, P. (1990). The California Critical Thinking Skills Test (CCTST). Millbrae, CA: California Academic Press.

Facione, P., & Facione, N. (1992). The California Critical Thinking Dispositions Inventory (CCTDI). Millbrae, CA: California Academic Press.

Gill, R. A., & Burke, I. C. (1999). Using an environmental science course to promote scientific literacy: Expanding critical-thinking skills beyond the environmental sciences. Journal of College Science Teaching, 29(2), 105-110.

Green, S. B., & Salkind, N. J. (2011). Using SPSS for Windows and Macintosh: Analyzing and understanding data (6th ed.). Columbus, OH: Prentice Hall. Herreid, C. F. (2004). Can case studies be used to teach critical thinking? Journal of College Science Teaching, 33(6), 12-14.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking: From childhood to adolescence. An essay on the construction of formal operational structures. New York, NY: Basic Books.

Johnson, M. A., & Lawson, A. E. (1998). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? Journal of Research in Science Teaching, 35, 89-103.

Lauer, T. (2005). Teaching critical-thinking using course content material: A reversal of roles. Journal of College Science Teaching, 34(6) 34-37.

Lawson, A. E. (1978). The development and validations of a classroom test of formal reasoning. Journal of Research in Science Teaching, 15, 11-24.

Lawson, A. E. (1985). A review of research on formal reasoning and science teaching. Journal of Research in Science Teaching, 22, 569-617.

Lawson, A. E. (2000). Classroom Test of Scientific Reasoning: Revised. Tempe, AZ: Arizona State University.

Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. International Journal of Science and Mathematics Education, 2, 307-338.

Lawson, A. E., Alkhoury, S., Benford, R., Clark, B. R., & Falconer, K. A. (2000). What kinds of scientific concepts exist? Concept construction and intellectual development in college biology. Journal of Research in Science Teaching, 37, 996-1018.

Lawson, A. E., Banks, D. L, & Logvin, M. (2007). Self-efficacy, reasoning ability and achievement in college biology. Journal of Research in Science Teaching, 44(5), 706-724.

Lawson, A. E., Clark, B., Cramer-Meldrum, E., Falconer, K. A., Sequist, J. M., & Kwon, Y. (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? Journal of Research in Science Teaching, 37, 81-101.

Lewis, S. E., & Lewis, J. E. (2005). The same or not the same: Equivalence as an issue in educational research. Journal of Chemical Education, 82, 1408-1412.

Lourenco, O., & Machaldo, A. (1996). In defense of Piaget's theory: A reply to 10 common criticisms. Psychological Review, 103, 143-164. Murphy, K. R., & Davidshofer, C. O. (2001). Psychological testing: Principles and applications (5th ed.). Upper Saddle River, NJ: Prentice Hall.

National Council on Excellence in Critical Thinking. (1987). Defining critical thinking. Retrieved from http://www.criticalthinking.org/ pages/defining-critical-thinking/766

National Research Council. (2012). Discipline-based education research: Understanding and improving learning in undergraduate science and engineering. Washington, DC: Author.

Norman, O. (1997). Investigating the nature of formal reasoning in chemistry: Testing Lawson's multiple hypothesis theory. Journal of Research in Science Teaching, 34, 1067-1081.

Popham, W. J. (2005). Classroom assessment: What teachers need to know. New York, NY: Pearson.

Raykov, T., & Marcoulides, G. A. (2006). A first course in structural equation modeling (2nd ed.). Mahwah, NJ: Erlbaum. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasiexperimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.

Tobias, S. (1990). They're not dumb, they're different: Stalking the second tier. Tucson, AZ: Research Corporation.

White, B., Stains, M., Escriu-Sune, M., Medaglia, E., Rostamnjad, L., Chinn, C., & Sevian, H. (2011). A novel instrument for assessing students' critical thinking abilities. Journal of College Science Teaching, 40(5), 102-107.

Justin H. Carmel

is a doctoral student Department of Chemistry and Biochemistry at Miami University in Oxford, Ohio.

Ellen J. Yezierski

is an associate professor, Department of Chemistry and Biochemistry at Miami University in Oxford, Ohio.

(yeziere@miamiOH.edu)

TABLE 1 Descriptive statistics for LCTSR. Pretest Posttest (N = 177) (N = 177) Mean 5.6 6.3 SD 2.1 2.5 Minimum 1 1 Maximum 10 12 FIGURE 1 Distribution of majors (N= 177). Engineering 2% Undeclared 4% Arts 3% Social Science 6% No Response 4% Humanities 12% Education 18% Business 51% Note: Table made from pie chart. FIGURE 2 Distribution of ACT scores (N = 131). ACT Number Score of Students 18 2 19 0 20 0 21 2 22 7 23 7 24 6 25 12 26 13 27 21 28 28 29 9 30 11 31 7 32 2 33 2 34 2 Note: Table made from bar graph. FIGURE 3 Highest math level completed (N = 169). Number of Students Geometry 1 Algebra 2/Trig 6 Precale/Stats 47 AP Calc/AP Stats/ 47 Math for Teachers Calc 1 [or equivalent] 57 Calc 2 9 Calc 3 1 Discrete Math 1 Note: Table made from bar graph. FIGURE 4 Number of science classes in high school (N = 169). Number of students 1 3 2 6 3 52 4 72 5+ 36 Note: Table made from bar graph.

Carmel, Justin H.^Yezierski, Ellen J.