

# Understanding the Complex Relationship between Critical Thinking and Science Reasoning among Undergraduate Thesis Writers

Jason E. Dowd,<sup>1\*</sup> Robert J. Thompson, Jr.,<sup>2</sup> Leslie A. Schiff,<sup>3</sup> and Julie A. Reynolds<sup>1</sup>

<sup>1</sup>Department of Biology and <sup>2</sup>Department of Psychology and Neuroscience, Duke University, Durham, NC 27708; <sup>3</sup>Department of Microbiology and Immunology, University of Minnesota, Minneapolis, MN 55455

## ABSTRACT

Developing critical-thinking and scientific reasoning skills are core learning objectives of science education, but little empirical evidence exists regarding the interrelationships between these constructs. Writing effectively fosters students' development of these constructs, and it offers a unique window into studying how they relate. In this study of undergraduate thesis writing in biology at two universities, we examine how scientific reasoning exhibited in writing (assessed using the Biology Thesis Assessment Protocol) relates to general and specific critical-thinking skills (assessed using the California Critical Thinking Skills Test), and we consider implications for instruction. We find that scientific reasoning in writing is strongly related to *inference*, while other aspects of science reasoning that emerge in writing (epistemological considerations, writing conventions, etc.) are not significantly related to critical-thinking skills. Science reasoning in writing is not merely a proxy for critical thinking. In linking features of students' writing to their critical-thinking skills, this study 1) provides a bridge to prior work suggesting that engagement in science writing enhances critical thinking and 2) serves as a foundational step for subsequently determining whether instruction focused explicitly on developing critical-thinking skills (particularly *inference*) can actually improve students' scientific reasoning in their writing.

## INTRODUCTION

Critical-thinking and scientific reasoning skills are core learning objectives of science education for all students, regardless of whether or not they intend to pursue a career in science or engineering. Consistent with the view of learning as construction of understanding and meaning (National Research Council, 2000), the pedagogical practice of writing has been found to be effective not only in fostering the development of students' conceptual and procedural knowledge (Gerdeman *et al.*, 2007) and communication skills (Clase *et al.*, 2010), but also scientific reasoning (Reynolds *et al.*, 2012) and critical-thinking skills (Quitadamo and Kurtz, 2007).

Critical thinking and scientific reasoning are similar but different constructs that include various types of higher-order cognitive processes, metacognitive strategies, and dispositions involved in making meaning of information. Critical thinking is generally understood as the broader construct (Holyoak and Morrison, 2005), comprising an array of cognitive processes and dispositions that are drawn upon differentially in everyday life and across domains of inquiry such as the natural sciences, social sciences, and humanities. Scientific reasoning, then, may be interpreted as the subset of critical-thinking skills (cognitive and metacognitive processes and dispositions) that 1) are involved in making meaning of information in scientific domains and 2) support the epistemological commitment to scientific methodology and paradigm(s).

John Coley, *Monitoring Editor*

Submitted March 17, 2017; Revised October 19, 2017; Accepted October 20, 2017

CBE Life Sci Educ March 1, 2018 17:ar4

DOI:10.1187/cbe.17-03-0052

\*Address correspondence to: Jason E. Dowd (jason.dowd@duke.edu).

© 2018 J. E. Dowd *et al.* CBE—Life Sciences Education © 2018 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

Although there has been an enduring focus in higher education on promoting critical thinking and reasoning as general or “transferable” skills, research evidence provides increasing support for the view that reasoning and critical thinking are also situational or domain specific (Beyer *et al.*, 2013). Some researchers, such as Lawson (2010), present frameworks in which science reasoning is characterized explicitly in terms of critical-thinking skills. There are, however, limited coherent frameworks and empirical evidence regarding either the general or domain-specific interrelationships of scientific reasoning, as it is most broadly defined, and critical-thinking skills.

The Vision and Change in Undergraduate Biology Education Initiative provides a framework for thinking about these constructs and their interrelationship in the context of the core competencies and disciplinary practice they describe (American Association for the Advancement of Science, 2011). These learning objectives aim for undergraduates to “understand the process of science, the interdisciplinary nature of the new biology and how science is closely integrated within society; be competent in communication and collaboration; have quantitative competency and a basic ability to interpret data; and have some experience with modeling, simulation and computational and systems level approaches as well as with using large databases” (Woodin *et al.*, 2010, pp. 71–72). This framework makes clear that science reasoning and critical-thinking skills play key roles in major learning outcomes; for example, “understanding the process of science” requires students to engage in (and be metacognitive about) scientific reasoning, and having the “ability to interpret data” requires critical-thinking skills. To help students better achieve these core competencies, we must better understand the interrelationships of their composite parts. Thus, the next step is to determine which specific critical-thinking skills are drawn upon when students engage in science reasoning in general and with regard to the particular scientific domain being studied. Such a determination could be applied to improve science education for both majors and nonmajors through pedagogical approaches that foster critical-thinking skills that are most relevant to science reasoning.

Writing affords one of the most effective means for making thinking visible (Reynolds *et al.*, 2012) and learning how to “think like” and “write like” disciplinary experts (Meizlish *et al.*, 2013). As a result, student writing affords the opportunities to both foster and examine the interrelationship of scientific reasoning and critical-thinking skills within and across disciplinary contexts. The purpose of this study was to better understand the relationship between students’ critical-thinking skills and scientific reasoning skills as reflected in the genre of undergraduate thesis writing in biology departments at two research universities, the University of Minnesota and Duke University.

In the following subsections, we discuss in greater detail the constructs of scientific reasoning and critical thinking, as well as the assessment of scientific reasoning in students’ thesis writing. In subsequent sections, we discuss our study design, findings, and the implications for enhancing educational practices.

### Critical Thinking

The advances in cognitive science in the 21st century have increased our understanding of the mental processes involved

in thinking and reasoning, as well as memory, learning, and problem solving. Critical thinking is understood to include both a cognitive dimension and a disposition dimension (e.g., reflective thinking) and is defined as “purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based” (Facione, 1990, p. 3). Although various other definitions of critical thinking have been proposed, researchers have generally coalesced on this consensus: expert view (Blattner and Frazier, 2002; Condon and Kelly-Riley, 2004; Bissell and Lemons, 2006; Quitadamo and Kurtz, 2007) and the corresponding measures of critical-thinking skills (August, 2016; Stephenson and Sadler-McKnight, 2016).

Both the cognitive skills and dispositional components of critical thinking have been recognized as important to science education (Quitadamo and Kurtz, 2007). Empirical research demonstrates that specific pedagogical practices in science courses are effective in fostering students’ critical-thinking skills. Quitadamo and Kurtz (2007) found that students who engaged in a laboratory writing component in the context of a general education biology course significantly improved their overall critical-thinking skills (and their analytical and inference skills, in particular), whereas students engaged in a traditional quiz-based laboratory did not improve their critical-thinking skills. In related work, Quitadamo *et al.* (2008) found that a community-based inquiry experience, involving inquiry, writing, research, and analysis, was associated with improved critical thinking in a biology course for nonmajors, compared with traditionally taught sections. In both studies, students who exhibited stronger presemester critical-thinking skills exhibited stronger gains, suggesting that “students who have not been explicitly taught how to think critically may not reach the same potential as peers who have been taught these skills” (Quitadamo and Kurtz, 2007, p. 151).

Recently, Stephenson and Sadler-McKnight (2016) found that first-year general chemistry students who engaged in a science writing heuristic laboratory, which is an inquiry-based, writing-to-learn approach to instruction (Hand and Keys, 1999), had significantly greater gains in total critical-thinking scores than students who received traditional laboratory instruction. Each of the four components—*inquiry, writing, collaboration, and reflection*—have been linked to critical thinking (Stephenson and Sadler-McKnight, 2016). Like the other studies, this work highlights the value of targeting critical-thinking skills and the effectiveness of an inquiry-based, writing-to-learn approach to enhance critical thinking. Across studies, authors advocate adopting critical thinking as the course framework (Pukkila, 2004) and developing explicit examples of how critical thinking relates to the scientific method (Miri *et al.*, 2007).

In these examples, the important connection between writing and critical thinking is highlighted by the fact that each intervention involves the incorporation of writing into science, technology, engineering, and mathematics education (either alone or in combination with other pedagogical practices). However, critical-thinking skills are not always the primary learning outcome; in some contexts, scientific reasoning is the primary outcome that is assessed.

### Scientific Reasoning

Scientific reasoning is a complex process that is broadly defined as “the skills involved in inquiry, experimentation, evidence evaluation, and inference that are done in the service of conceptual change or scientific understanding” (Zimmerman, 2007, p. 172). Scientific reasoning is understood to include both conceptual knowledge and the cognitive processes involved with generation of hypotheses (i.e., inductive processes involved in the generation of hypotheses and the deductive processes used in the testing of hypotheses), experimentation strategies, and evidence evaluation strategies. These dimensions are interrelated, in that “experimentation and inference strategies are selected based on prior conceptual knowledge of the domain” (Zimmerman, 2000, p. 139). Furthermore, conceptual and procedural knowledge and cognitive process dimensions can be general and domain specific (or discipline specific).

With regard to conceptual knowledge, attention has been focused on the acquisition of core methodological concepts fundamental to scientists’ causal reasoning and metacognitive distancing (or decontextualized thinking), which is the ability to reason independently of prior knowledge or beliefs (Greenhoot *et al.*, 2004). The latter involves what Kuhn and Dean (2004) refer to as the coordination of theory and evidence, which requires that one question existing theories (i.e., prior knowledge and beliefs), seek contradictory evidence, eliminate alternative explanations, and revise one’s prior beliefs in the face of contradictory evidence. Kuhn and colleagues (2008) further elaborate that scientific thinking requires “a mature understanding of the epistemological foundations of science, recognizing scientific knowledge as constructed by humans rather than simply discovered in the world,” and “the ability to engage in skilled argumentation in the scientific domain, with an appreciation of argumentation as entailing the coordination of theory and evidence” (Kuhn *et al.*, 2008, p. 435). “This approach to scientific reasoning not only highlights the skills of generating and evaluating evidence-based inferences, but also encompasses epistemological appreciation of the functions of evidence and theory” (Ding *et al.*, 2016, p. 616). Evaluating evidence-based inferences involves epistemic cognition, which Moshman (2015) defines as the subset of metacognition that is concerned with justification, truth, and associated forms of reasoning. Epistemic cognition is both general and domain specific (or discipline specific; Moshman, 2015).

There is empirical support for the contributions of both prior knowledge and an understanding of the epistemological foundations of science to scientific reasoning. In a study of undergraduate science students, advanced scientific reasoning was most often accompanied by accurate prior knowledge as well as sophisticated epistemological commitments; additionally, for students who had comparable levels of prior knowledge, skillful reasoning was associated with a strong epistemological commitment to the consistency of theory with evidence (Zeineddin and Abd-El-Khalick, 2010). These findings highlight the importance of the need for instructional activities that intentionally help learners develop sophisticated epistemological commitments focused on the nature of knowledge and the role of evidence in supporting knowledge claims (Zeineddin and Abd-El-Khalick, 2010).

### Scientific Reasoning in Students’ Thesis Writing

Pedagogical approaches that incorporate writing have also focused on enhancing scientific reasoning. Many rubrics have been developed to assess aspects of scientific reasoning in written artifacts. For example, Timmerman and colleagues (2011), in the course of describing their own rubric for assessing scientific reasoning, highlight several examples of scientific reasoning assessment criteria (Haaga, 1993; Tariq *et al.*, 1998; Topping *et al.*, 2000; Kelly and Takao, 2002; Halonen *et al.*, 2003; Willison and O’Regan, 2007).

At both the University of Minnesota and Duke University, we have focused on the genre of the undergraduate honors thesis as the rhetorical context in which to study and improve students’ scientific reasoning and writing. We view the process of writing an undergraduate honors thesis as a form of professional development in the sciences (i.e., a way of engaging students in the practices of a community of discourse). We have found that structured courses designed to scaffold the thesis-writing process and promote metacognition can improve writing and reasoning skills in biology, chemistry, and economics (Reynolds and Thompson, 2011; Dowd *et al.*, 2015a,b). In the context of this prior work, we have defined scientific reasoning *in writing* as the emergent, underlying construct measured across distinct aspects of students’ written discussion of independent research in their undergraduate theses.

The Biology Thesis Assessment Protocol (BioTAP) was developed at Duke University as a tool for systematically guiding students and faculty through a “draft–feedback–revision” writing process, modeled after professional scientific peer-review processes (Reynolds *et al.*, 2009). BioTAP includes activities and worksheets that allow students to engage in critical peer review and provides detailed descriptions, presented as rubrics, of the questions (i.e., dimensions, shown in Table 1) upon which such review should focus. Nine rubric dimensions focus on communication to the broader scientific community, and four rubric dimensions focus on the accuracy and appropriateness of the research. These rubric dimensions provide criteria by which the thesis is assessed, and therefore allow BioTAP to be used as an assessment tool as well as a teaching resource (Reynolds *et al.*, 2009). Full details are available at [www.science-writing.org/biotap.html](http://www.science-writing.org/biotap.html).

In previous work, we have used BioTAP to quantitatively assess students’ undergraduate honors theses and explore the relationship between thesis-writing courses (or specific interventions within the courses) and the strength of students’ science reasoning in writing across different science disciplines: biology (Reynolds and Thompson, 2011); chemistry (Dowd *et al.*, 2015b); and economics (Dowd *et al.*, 2015a). We have focused exclusively on the nine dimensions related to reasoning and writing (questions 1–9), as the other four dimensions (questions 10–13) require topic-specific expertise and are intended to be used by the student’s thesis supervisor.

Beyond considering individual dimensions, we have investigated whether meaningful constructs underlie students’ thesis scores. We conducted exploratory factor analysis of students’ theses in biology, economics, and chemistry and found one dominant underlying factor in each discipline; we termed the factor “scientific reasoning in writing” (Dowd *et al.*, 2015a,b, 2016). That is, each of the nine dimensions could be

**TABLE 1. Theses assessment protocol dimensions**

Scientific community focused	Science reasoning	1 Is the writing appropriate for the target audience?
		2 Does the thesis make a compelling argument for the significance of the student's research within the context of the current literature?
		3 Does the thesis clearly articulate the student's research goals?
		4 Does the thesis skillfully interpret the results?
		5 Is there a compelling discussion of the implications of findings?
Scientific community focused	Mechanical	6 Is the thesis clearly organized?
		7 Is the thesis free of writing errors?
		8 Are the citations presented consistently and professionally throughout the text and in the list of works cited?
		9 Are the tables and figures clear, effective, and informative?
Topic specific		10 Does the thesis represent the student's significant scientific research?
		11 Is the literature review accurate and complete?
		12 Are the methods appropriate, given the student's research question?
		13 Is the data analysis appropriate, accurate, and unbiased?

understood as reflecting, in different ways and to different degrees, the construct of scientific reasoning in writing. The findings indicated evidence of both general and discipline-specific components to scientific reasoning in writing that relate to epistemic beliefs and paradigms, in keeping with broader ideas about science reasoning discussed earlier. Specifically, scientific reasoning in writing is more strongly associated with formulating a compelling argument for the significance of the research in the context of current literature in biology, making meaning regarding the implications of the findings in chemistry, and providing an organizational framework for interpreting the thesis in economics. We suggested that instruction, whether occurring in writing studios or in writing courses to facilitate thesis preparation, should attend to both components.

### Research Question and Study Design

The genre of thesis writing combines the pedagogies of writing and inquiry found to foster scientific reasoning (Reynolds *et al.*, 2012) and critical thinking (Quitadamo and Kurtz, 2007; Quitadamo *et al.*, 2008; Stephenson and Sadler-McKnight, 2016). However, there is no empirical evidence regarding the general or domain-specific interrelationships of scientific reasoning and critical-thinking skills, particularly in the rhetorical context of the undergraduate thesis. The BioTAP studies discussed earlier indicate that the rubric-based assessment produces evidence of scientific reasoning in the undergraduate thesis, but it was not designed to foster or measure critical thinking. The current study was undertaken to address the research question: How are students' critical-thinking skills related to scientific reasoning as reflected in the genre of undergraduate thesis writing in biology? Determining these interrelationships could guide efforts to enhance students' scientific reasoning and writing skills through focusing instruction on specific critical-thinking skills as well as disciplinary conventions.

## METHODS

To address this research question, we focused on undergraduate thesis writers in biology courses at two institutions, Duke University and the University of Minnesota, and examined the extent to which students' scientific reasoning in writing, assessed in the undergraduate thesis using BioTAP, corresponds to students' critical-thinking skills, assessed using the California Critical Thinking Skills Test (CCTST; August, 2016).

### Study Sample

The study sample was composed of students enrolled in courses designed to scaffold the thesis-writing process in the Department of Biology at Duke University and the College of Biological Sciences at the University of Minnesota. Both courses complement students' individual work with research advisors. The course is required for thesis writers at the University of Minnesota and optional for writers at Duke University. Not all students are required to complete a thesis, though it is required for students to graduate with honors; at the University of Minnesota, such students are enrolled in an honors program within the college. In total, 28 students were enrolled in the course at Duke University and 44 students were enrolled in the course at the University of Minnesota. Of those students, two students did not consent to participate in the study; additionally, five students did not validly complete the CCTST (i.e., attempted fewer than 60% of items or completed the test in less than 15 minutes). Thus, our overall rate of valid participation is 90%, with 27 students from Duke University and 38 students from the University of Minnesota. We found no statistically significant differences in thesis assessment between students with valid CCTST scores and invalid CCTST scores. Therefore, we focus on the 65 students who consented to participate and for whom we have complete and valid data in most of this study. Additionally, in asking students for their consent to participate, we allowed them to choose whether to provide or decline access to academic and demographic background data. Of the 65 students who consented to participate, 52 students granted access to such data. Therefore, for additional analyses involving academic and background data, we focus on the 52 students who consented. We note that the 13 students who participated but declined to share additional data performed slightly lower on the CCTST than the 52 others (perhaps suggesting that they differ by other measures, but we cannot determine this with certainty). Among the 52 students, 60% identified as female and 10% identified as being from under-represented ethnicities.

### Critical Thinking

In both courses, students completed the CCTST online, either in class or on their own, late in the Spring 2016 semester. This is the same assessment that was used in prior studies of critical thinking (Quitadamo and Kurtz, 2007; Quitadamo *et al.*, 2008; Stephenson and Sadler-McKnight, 2016). It is "an objective measure of the core reasoning skills needed for reflective decision making concerning what to believe or what to do" (Insight Assessment, 2016a). In the test, students are asked to read and consider information as they answer multiple-choice questions. The questions are intended to be appropriate for all users, so there is no expectation of prior disciplinary knowledge in biology (or any other subject). Although actual test items are

protected, sample items are available on the Insight Assessment website (Insight Assessment, 2016b). We have included one sample item in the Supplemental Material.

The CCTST is based on a consensus definition of critical thinking, measures cognitive and metacognitive skills associated with critical thinking, and has been evaluated for validity and reliability at the college level (August, 2016; Stephenson and Sadler-McKnight, 2016). In addition to providing overall critical-thinking score, the CCTST assesses seven dimensions of critical thinking: analysis, interpretation, inference, evaluation, explanation, induction, and deduction. Scores on each dimension are calculated based on students' performance on items related to that dimension. Analysis focuses on identifying assumptions, reasons, and claims and examining how they interact to form arguments. Interpretation, related to analysis, focuses on determining the precise meaning and significance of information. Inference focuses on drawing conclusions from reasons and evidence. Evaluation focuses on assessing the credibility of sources of information and claims they make. Explanation, related to evaluation, focuses on describing the evidence, assumptions, or rationale for beliefs and conclusions. Induction focuses on drawing inferences about what is probably true based on evidence. Deduction focuses on drawing conclusions about what must be true when the context completely determines the outcome. These are not independent dimensions; the fact that they are related supports their collective interpretation as critical thinking. Together, the CCTST dimensions provide a basis for evaluating students' overall strength in using reasoning to form reflective judgments about what to believe or what to do (August, 2016). Each of the seven dimensions and the overall CCTST score are measured on a scale of 0–100, where higher scores indicate superior performance. Scores correspond to superior (86–100), strong (79–85), moderate (70–78), weak (63–69), or not manifested (62 and below) skills.

### Scientific Reasoning in Writing

At the end of the semester, students' final, submitted undergraduate theses were assessed using BioTAP, which consists of nine rubric dimensions that focus on communication to the broader scientific community and four additional dimensions that focus on the exhibition of topic-specific expertise (Reynolds *et al.*, 2009). These dimensions, framed as questions, are displayed in Table 1.

Student theses were assessed on questions 1–9 of BioTAP using the same procedures described in previous studies (Reynolds and Thompson, 2011; Dowd *et al.*, 2015a,b). In this study, six raters were trained in the valid, reliable use of BioTAP rubrics. Each dimension was rated on a five-point scale: 1 indicates the dimension is missing, incomplete, or below acceptable standards; 3 indicates that the dimension is adequate but not exhibiting mastery; and 5 indicates that the dimension is excellent and exhibits mastery (intermediate ratings of 2 and 4 are appropriate when different parts of the thesis make a single category challenging). After training, two raters independently assessed each thesis and then discussed their independent ratings with one another to form a consensus rating. The consensus score is not an average score, but rather an agreed-upon, discussion-based score. On a five-point scale, raters independently assessed dimensions to be within 1 point of each

other 82.4% of the time before discussion and formed consensus ratings 100% of the time after discussion.

In this study, we consider both categorical (mastery/non-mastery, where a score of 5 corresponds to mastery) and numerical treatments of individual BioTAP scores to better relate the manifestation of critical thinking in BioTAP assessment to all of the prior studies. For comprehensive/cumulative measures of BioTAP, we focus on the partial sum of questions 1–5, as these questions relate to higher-order scientific reasoning (whereas questions 6–9 relate to mid- and lower-order writing mechanics [Reynolds *et al.*, 2009]), and the factor scores (i.e., numerical representations of the extent to which each student exhibits the underlying factor), which are calculated from the factor loadings published by Dowd *et al.* (2016). We do not focus on questions 6–9 individually in statistical analyses, because we do not expect critical-thinking skills to relate to mid- and lower-order writing skills.

The final, submitted thesis reflects the student's writing, the student's scientific reasoning, the quality of feedback provided to the student by peers and mentors, and the student's ability to incorporate that feedback into his or her work. Therefore, our assessment is not the same as an assessment of unpolished, unrevised samples of students' written work. While one might imagine that such an unpolished sample may be more strongly correlated with critical-thinking skills measured by the CCTST, we argue that the complete, submitted thesis, assessed using BioTAP, is ultimately a more appropriate reflection of how students exhibit science reasoning in the scientific community.

### Statistical Analyses

We took several steps to analyze the collected data. First, to provide context for subsequent interpretations, we generated descriptive statistics for the CCTST scores of the participants based on the norms for undergraduate CCTST test takers. To determine the strength of relationships among CCTST dimensions (including overall score) and the BioTAP dimensions, partial-sum score (questions 1–5), and factor score, we calculated Pearson's correlations for each pair of measures. To examine whether falling on one side of the nonmastery/mastery threshold (as opposed to a linear scale of performance) was related to critical thinking, we grouped BioTAP dimensions into categories (mastery/nonmastery) and conducted Student's *t* tests to compare the means scores of the two groups on each of the seven dimensions and overall score of the CCTST. Finally, for the strongest relationship that emerged, we included additional academic and background variables as covariates in multiple linear-regression analysis to explore questions about how much observed relationships between critical-thinking skills and science reasoning in writing might be explained by variation in these other factors.

Although BioTAP scores represent discreet, ordinal bins, the five-point scale is intended to capture an underlying continuous construct (from inadequate to exhibiting mastery). It has been argued that five categories is an appropriate cutoff for treating ordinal variables as pseudo-continuous (Rhemtulla *et al.*, 2012)—and therefore using continuous-variable statistical methods (e.g., Pearson's correlations)—as long as the underlying assumption that ordinal scores are linearly distributed is valid. Although we have no way to statistically test this assumption, we interpret adequate scores to be approximately halfway

between inadequate and mastery scores, resulting in a linear scale. In part because this assumption is subject to disagreement, we also consider and interpret a categorical (mastery/nonmastery) treatment of BioTAP variables.

We corrected for multiple comparisons using the Holm-Bonferroni method (Holm, 1979). At the most general level, where we consider the single, comprehensive measures for BioTAP (partial-sum and factor score) and the CCTST (overall score), there is no need to correct for multiple comparisons, because the multiple, individual dimensions are collapsed into single dimensions. When we considered individual CCTST dimensions in relation to comprehensive measures for BioTAP, we accounted for seven comparisons; similarly, when we considered individual dimensions of BioTAP in relation to overall CCTST score, we accounted for five comparisons. When all seven CCTST and five BioTAP dimensions were examined individually and without prior knowledge, we accounted for 35 comparisons; such a rigorous threshold is likely to reject weak and moderate relationships, but it is appropriate if there are no specific pre-existing hypotheses. All *p* values are presented in tables for complete transparency, and we carefully consider the implications of our interpretation of these data in the *Discussion* section.

### RESULTS

CCTST scores for students in this sample ranged from the 39th to 99th percentile of the general population of undergraduate CCTST test takers (mean percentile = 84.3, median = 85th percentile; Table 2); these percentiles reflect overall scores that range from moderate to superior. Scores on individual dimensions and overall scores were sufficiently normal and far enough from the ceiling of the scale to justify subsequent statistical analyses.

The Pearson's correlations between students' cumulative scores on BioTAP (the factor score based on loadings published by Dowd *et al.*, 2016, and the partial sum of scores on questions 1–5) and students' overall scores on the CCTST are presented in Table 3. We found that the partial-sum measure of BioTAP was significantly related to the overall measure of critical thinking ( $r = 0.27, p = 0.03$ ), while the BioTAP factor score was marginally related to overall CCTST ( $r = 0.24, p = 0.05$ ). When we looked at relationships between comprehensive BioTAP measures and scores for individual dimensions of the CCTST (Table 3), we found significant positive correlations between the both BioTAP partial-sum and factor scores and CCTST inference ( $r = 0.45, p < 0.001$ , and  $r = 0.41, p < 0.001$ , respectively). Although some other relationships have *p* values

TABLE 2. Descriptive statistics of CCTST dimensions<sup>a</sup>

	Minimum	Mean	Median	Maximum
Analysis	70	88.6	90	100
Interpretation	74	89.7	87	100
Inference	78	87.9	89	100
Evaluation	63	83.6	84	100
Explanation	61	84.4	87	100
Induction	74	87.4	87	97
Deduction	71	86.4	87	97
Overall	73	86	85	97

<sup>a</sup>Scores correspond to superior (86–100), strong (79–85), moderate (70–78), weak (63–69), or not manifested (62 and lower) skills.

below 0.05 (e.g., the correlations between BioTAP partial-sum scores and CCTST induction and interpretation scores), they are not significant when we correct for multiple comparisons.

When we expanded comparisons to include all 35 potential correlations among individual BioTAP and CCTST dimensions—and, accordingly, corrected for 35 comparisons—we did not find any additional statistically significant relationships. The Pearson's correlations between students' scores on each dimension of BioTAP and students' scores on each dimension of the CCTST range from –0.11 to 0.35 (Table 3); although the relationship between discussion of implications (BioTAP question 5) and inference appears to be relatively large ( $r = 0.35$ ), it is not significant ( $p = 0.005$ ; the Holm-Bonferroni cutoff is 0.00143). We found no statistically significant relationships between BioTAP questions 6–9 and CCTST dimensions (unpublished data), regardless of whether we correct for multiple comparisons.

The results of Student's *t* tests comparing scores on each dimension of the CCTST of students who exhibit mastery with those of students who do not exhibit mastery on each dimension of BioTAP are presented in Table 4. Focusing first on the overall CCTST scores, we found that the difference between those who exhibit mastery and those who do not in discussing implications of results (BioTAP question 5) is statistically significant ( $t = 2.73, p = 0.008, d = 0.71$ ). When we expanded *t* tests to include all 35 comparisons—and, like above, corrected for

TABLE 3. Correlations between dimensions of CCTST and dimensions of BioTAP<sup>a</sup>

CCTST	BioTAP rubric dimensions (five-point scale)						
	Audience	Argument	Goals	Interpret results	Discuss implications	BioTAP <sup>b</sup> <sub>subtotal</sub>	BioTAP <sup>c</sup> <sub>factor</sub>
Analysis	0.05	–0.03	0.03	0.21	0.10	0.14	0.13
	<i>0.72</i>	<i>0.81</i>	<i>0.81</i>	<i>0.09</i>	<i>0.43</i>	<i>0.27</i>	<i>0.31</i>
Interpretation	0.26	0.01	0.25	0.08	0.20	0.27	0.23
	<i>0.03</i>	<i>0.92</i>	<i>0.04</i>	<i>0.53</i>	<i>0.12</i>	<i>0.03</i>	<i>0.07</i>
Inference	0.27	0.19	0.20	0.30	0.35	<b>0.45</b>	<b>0.41</b>
	<i>0.03</i>	<i>0.14</i>	<i>0.11</i>	<i>0.01</i>	<i>0.005</i>	<b>0.0002</b>	<b>0.0006</b>
Evaluation	0.05	–0.03	–0.06	0.15	0.10	0.09	0.07
	<i>0.67</i>	<i>0.79</i>	<i>0.63</i>	<i>0.24</i>	<i>0.43</i>	<i>0.50</i>	<i>0.59</i>
Explanation	0.15	–0.06	–0.03	0.16	0.16	0.15	0.12
	<i>0.23</i>	<i>0.65</i>	<i>0.84</i>	<i>0.21</i>	<i>0.20</i>	<i>0.25</i>	<i>0.35</i>
Induction	0.20	–0.01	0.07	0.22	0.24	0.26	0.23
	<i>0.11</i>	<i>0.93</i>	<i>0.57</i>	<i>0.08</i>	<i>0.05</i>	<i>0.04</i>	<i>0.06</i>
Deduction	0.16	0.07	0.10	0.20	0.15	0.23	0.21
	<i>0.21</i>	<i>0.56</i>	<i>0.43</i>	<i>0.10</i>	<i>0.25</i>	<i>0.06</i>	<i>0.10</i>
Overall	0.21	0.03	0.09	0.24	0.20	<b>0.27</b>	0.24
	<i>0.10</i>	<i>0.80</i>	<i>0.49</i>	<i>0.05</i>	<i>0.10</i>	<b>0.03</b>	<i>0.05</i>

<sup>a</sup>In each cell, the top number is the correlation, and the bottom, italicized number is the associated *p* value. Correlations that are statistically significant after correcting for multiple comparisons are shown in bold.

<sup>b</sup>This is the partial sum of BioTAP scores on questions 1–5.

<sup>c</sup>This is the factor score calculated from factor loadings published by Dowd *et al.* (2016).

TABLE 4. The *t* statistics and effect sizes of differences in dimensions of CCTST across dimensions of BioTAP<sup>a</sup>

CCTST	BioTAP (mastery vs. nonmastery)				
	Audience	Argument	Goals	Interpret results	Discuss implications
<i>n</i> <sub>master</sub>	29	40	54	35	23
<i>n</i> <sub>non</sub>	36	25	11	30	42
Analysis	0.92	-0.36	-0.49	1.04	1.41
	<i>0.36</i>	<i>0.72</i>	<i>0.62</i>	<i>0.30</i>	<i>0.16</i>
	0.23	-0.09	-0.16	0.26	0.37
Interpretation	1.18	-0.49	1.41	1.34	1.79
	<i>0.24</i>	<i>0.63</i>	<i>0.16</i>	<i>0.19</i>	<i>0.08</i>
	0.29	-0.12	0.47	0.33	0.47
Inference	2.32	0.87	0.39	2.46	<b>3.41</b>
	<i>0.02</i>	<i>0.39</i>	<i>0.70</i>	<i>0.02</i>	<b>0.001</b>
	0.58	0.22	0.13	0.61	<b>0.88</b>
Evaluation	0.90	-0.47	-1.38	1.14	2.00
	<i>0.37</i>	<i>0.64</i>	<i>0.17</i>	<i>0.26</i>	<i>0.05</i>
	0.23	-0.12	-0.46	0.28	0.52
Explanation	1.47	-0.99	-1.31	1.72	2.60
	<i>0.15</i>	<i>0.32</i>	<i>0.20</i>	<i>0.09</i>	<i>0.01</i>
	0.37	-0.25	-0.43	0.43	0.67
Induction	1.67	-0.63	-0.36	1.87	3.26
	<i>0.10</i>	<i>0.53</i>	<i>0.72</i>	<i>0.07</i>	<i>0.002</i>
	0.42	-0.16	-0.12	0.47	0.84
Deduction	1.86	0.19	-0.37	1.31	1.72
	<i>0.07</i>	<i>0.85</i>	<i>0.71</i>	<i>0.19</i>	<i>0.09</i>
	0.46	0.05	-0.12	0.33	0.45
Overall	2.06	-0.27	-0.47	1.86	<b>2.73</b>
	<i>0.04</i>	<i>0.79</i>	<i>0.64</i>	<i>0.07</i>	<b>0.008</b>
	0.52	-0.07	-0.16	0.46	<b>0.71</b>

<sup>a</sup>In each cell, the top number is the *t* statistic for each comparison, and the middle, italicized number is the associated *p* value. The bottom number is the effect size. Correlations that are statistically significant after correcting for multiple comparisons are shown in bold.

35 comparisons—we found a significant difference in inference scores between students who exhibit mastery on question 5 and students who do not ( $t = 3.41$ ,  $p = 0.0012$ ,  $d = 0.88$ ), as well as a marginally significant difference in these students' induction scores ( $t = 3.26$ ,  $p = 0.0018$ ,  $d = 0.84$ ; the Holm-Bonferroni cutoff is  $p = 0.00147$ ). Cohen's *d* effect sizes, which reveal the strength of the differences for statistically significant relationships, range from 0.71 to 0.88.

Finally, we more closely examined the strongest relationship that we observed, which was between the CCTST dimension of inference and the BioTAP partial-sum composite score (shown in Table 3), using multiple regression analysis (Table 5). Focusing on the 52 students for whom we have background information, we looked at the simple relationship between BioTAP and inference (model 1), a robust background model including multiple covariates that one might expect to explain some part of the variation in BioTAP (model 2), and a combined model

TABLE 5. Partial sum (questions 1–5) of BioTAP scores ( $n = 52$ )

Variable	Model 1	Model 2	Model 3
CCTST inference	0.536***		0.491**
Grade point average		0.176	0.092
Independent study courses		-0.087	0.001
Writing-intensive courses		0.131	0.021
Institution		0.329	0.115
Male		0.085	0.041
Underrepresented group		-0.114	-0.060
Adjusted $R^2$	0.273	-0.022	0.195

\*\* $p < 0.01$ .

\*\*\* $p < 0.001$ .

including all variables (model 3). As model 3 shows, the covariates explain very little variation in BioTAP scores, and the relationship between inference and BioTAP persists even in the presence of all of the covariates.

## DISCUSSION

The aim of this study was to examine the extent to which the various components of scientific reasoning—manifested in writing in the genre of undergraduate thesis and assessed using BioTAP—draw on general and specific critical-thinking skills (assessed using CCTST) and to consider the implications for educational practices. Although science reasoning involves critical-thinking skills, it also relates to conceptual knowledge and the epistemological foundations of science disciplines (Kuhn *et al.*, 2008). Moreover, science reasoning *in writing*, captured in students' undergraduate theses, reflects habits, conventions, and the incorporation of feedback that may alter evidence of individuals' critical-thinking skills. Our findings, however, provide empirical evidence that cumulative measures of science reasoning in writing are nonetheless related to students' overall critical-thinking skills (Table 3). The particularly significant roles of inference skills (Table 3) and the discussion of implications of results (BioTAP question 5; Table 4) provide a basis for more specific ideas about how these constructs relate to one another and what educational interventions may have the most success in fostering these skills.

Our results build on previous findings. The genre of thesis writing combines pedagogies of writing and inquiry found to foster scientific reasoning (Reynolds *et al.*, 2012) and critical thinking (Quitadamo and Kurtz, 2007; Quitadamo *et al.*, 2008; Stephenson and Sadler-McKnight, 2016). Quitadamo and Kurtz (2007) reported that students who engaged in a laboratory writing component in a general education biology course significantly improved their inference and analysis skills, and Quitadamo and colleagues (2008) found that participation in a community-based inquiry biology course (that included a writing component) was associated with significant gains in students' inference and evaluation skills. The shared focus on inference is noteworthy, because these prior studies actually differ from the current study; the former considered critical-thinking skills as the primary learning outcome of writing-focused interventions, whereas the latter focused on emergent links between two learning outcomes (science reasoning in writing and critical thinking). In other words, inference skills are *impacted by writing* as well as *manifested in writing*.

Inference focuses on drawing conclusions from argument and evidence. According to the consensus definition of critical thinking, the specific skill of inference includes several processes: querying evidence, conjecturing alternatives, and drawing conclusions. All of these activities are central to the independent research at the core of writing an undergraduate thesis. Indeed, a critical part of what we call “science reasoning in writing” might be characterized as a measure of students’ ability to infer and make meaning of information and findings. Because the cumulative BioTAP measures distill underlying similarities and, to an extent, suppress unique aspects of individual dimensions, we argue that it is appropriate to relate inference to *scientific reasoning in writing*. Even when we control for other potentially relevant background characteristics, the relationship is strong (Table 5).

In taking the complementary view and focusing on BioTAP, when we compared students who exhibit mastery with those who do not, we found that the specific dimension of “discussing the implications of results” (question 5) differentiates students’ performance on several critical-thinking skills. To achieve mastery on this dimension, students must make connections between their results and other published studies and discuss the future directions of the research; in short, they must demonstrate an understanding of the bigger picture. The specific relationship between question 5 and inference is the strongest observed among all individual comparisons. Altogether, perhaps more than any other BioTAP dimension, this aspect of students’ writing provides a clear view of the role of students’ critical-thinking skills (particularly inference and, marginally, induction) in science reasoning.

While inference and discussion of implications emerge as particularly strongly related dimensions in this work, we note that the strongest contribution to “science reasoning in writing in biology,” as determined through exploratory factor analysis, is “argument for the significance of research” (BioTAP question 2, not question 5; Dowd *et al.*, 2016). Question 2 is not clearly related to critical-thinking skills. These findings are not contradictory, but rather suggest that the epistemological and disciplinary-specific aspects of science reasoning that emerge in writing through BioTAP are not completely aligned with aspects related to critical thinking. In other words, science reasoning in writing is not simply a proxy for those critical-thinking skills that play a role in science reasoning.

In a similar vein, the content-related, epistemological aspects of science reasoning, as well as the conventions associated with writing the undergraduate thesis (including feedback from peers and revision), may explain the *lack* of significant relationships between some science reasoning dimensions and some critical-thinking skills that might otherwise seem counterintuitive (e.g., BioTAP question 2, which relates to making an argument, and the critical-thinking skill of argument). It is possible that an individual’s critical-thinking skills may explain some variation in a particular BioTAP dimension, but other aspects of science reasoning and practice exert much stronger influence. Although these relationships do not emerge in our analyses, the lack of significant correlation does not mean that there is definitively no correlation. Correcting for multiple comparisons suppresses type 1 error at the expense of exacerbating type 2 error, which, combined with the limited sample size, constrains statistical power and makes weak relationships more difficult to

detect. Ultimately, though, the relationships that *do* emerge highlight places where individuals’ distinct critical-thinking skills emerge most coherently in thesis assessment, which is why we are particularly interested in unpacking those relationships.

We recognize that, because only honors students submit theses at these institutions, this study sample is composed of a selective subset of the larger population of biology majors. Although this is an inherent limitation of focusing on thesis writing, links between our findings and results of other studies (with different populations) suggest that observed relationships may occur more broadly. The goal of improved science reasoning and critical thinking is shared among all biology majors, particularly those engaged in capstone research experiences. So while the implications of this work most directly apply to honors thesis writers, we provisionally suggest that all students could benefit from further study of them.

There are several important implications of this study for science education practices. Students’ inference skills relate to the understanding and effective application of scientific content. The fact that we find no statistically significant relationships between BioTAP questions 6–9 and CCTST dimensions suggests that such mid- to lower-order elements of BioTAP (Reynolds *et al.*, 2009), which tend to be more structural in nature, do not focus on aspects of the finished thesis that draw strongly on critical thinking. In keeping with prior analyses (Reynolds and Thompson, 2011; Dowd *et al.*, 2016), these findings further reinforce the notion that disciplinary instructors, who are most capable of teaching and assessing scientific reasoning and perhaps least interested in the more mechanical aspects of writing, may nonetheless be best suited to effectively model and assess students’ writing.

The goal of the thesis writing course at both Duke University and the University of Minnesota is not merely to improve thesis scores but to move students’ writing into the category of mastery across BioTAP dimensions. Recognizing that students with differing critical-thinking skills (particularly inference) are more or less likely to achieve mastery in the undergraduate thesis (particularly in discussing implications [question 5]) is important for developing and testing targeted pedagogical interventions to improve learning outcomes for all students.

The competencies characterized by the Vision and Change in Undergraduate Biology Education Initiative provide a general framework for recognizing that science reasoning and critical-thinking skills play key roles in major learning outcomes of science education. Our findings highlight places where science reasoning–related competencies (like “understanding the process of science”) connect to critical-thinking skills and places where critical thinking–related competencies might be manifested in scientific products (such as the ability to discuss implications in scientific writing). We encourage broader efforts to build empirical connections between competencies and pedagogical practices to further improve science education.

One specific implication of this work for science education is to focus on providing opportunities for students to develop their critical-thinking skills (particularly inference). Of course, as this correlational study is not designed to test causality, we do not claim that enhancing students’ inference skills will improve science reasoning in writing. However, as prior work shows that science writing activities influence students’ inference skills (Quitadamo and Kurtz, 2007; Quitadamo *et al.*,



2008), there is reason to test such a hypothesis. Nevertheless, the focus must extend beyond inference as an isolated skill; rather, it is important to relate inference to the foundations of the scientific method (Miri *et al.*, 2007) in terms of the epistemological appreciation of the functions and coordination of evidence (Kuhn and Dean, 2004; Zeineddin and Abd-El-Khalick, 2010; Ding *et al.*, 2016) and disciplinary paradigms of truth and justification (Moshman, 2015).

Although this study is limited to the domain of biology at two institutions with a relatively small number of students, the findings represent a foundational step in the direction of achieving success with more integrated learning outcomes. Hopefully, it will spur greater interest in empirically grounding discussions of the constructs of scientific reasoning and critical-thinking skills.

## CONCLUSION

This study contributes to the efforts to improve science education, for both majors and nonmajors, through an empirically driven analysis of the relationships between scientific reasoning reflected in the genre of thesis writing and critical-thinking skills. This work is rooted in the usefulness of BioTAP as a method 1) to facilitate communication and learning and 2) to assess disciplinary-specific and general dimensions of science reasoning. The findings support the important role of the critical-thinking skill of inference in scientific reasoning in writing, while also highlighting ways in which other aspects of science reasoning (epistemological considerations, writing conventions, etc.) are not significantly related to critical thinking. Future research into the impact of interventions focused on specific critical-thinking skills (i.e., inference) for improved science reasoning in writing will build on this work and its implications for science education.

## ACKNOWLEDGMENTS

We acknowledge the contributions of Kelaine Haas and Alexander Motten to the implementation and collection of data. We also thank Mine Çetinkaya-Rundel for her insights regarding our statistical analyses. This research was funded by National Science Foundation award DUE-1525602.

## REFERENCES

- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC. Retrieved September 26, 2017, from <https://visionandchange.org/files/2013/11/aaas-VISchange-web1113.pdf>
- August, D. (2016). *California Critical Thinking Skills Test user manual and resource guide*. San Jose: Insight Assessment/California Academic Press.
- Beyer, C. H., Taylor, E., & Gillmore, G. M. (2013). *Inside the undergraduate teaching experience: The University of Washington's growth in faculty teaching study*. Albany, NY: SUNY Press.
- Bissell, A. N., & Lemons, P. P. (2006). A new method for assessing critical thinking in the classroom. *BioScience*, 56(1), 66–72. [https://doi.org/10.1641/0006-3568\(2006\)056\[0066:ANMFAC\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)056[0066:ANMFAC]2.0.CO;2)
- Blattner, N. H., & Frazier, C. L. (2002). Developing a performance-based assessment of students' critical thinking skills. *Assessing Writing*, 8(1), 47–64.
- Clase, K. L., Gundlach, E., & Pelaez, N. J. (2010). Calibrated peer review for computer-assisted learning of biological research competencies. *Biochemistry and Molecular Biology Education*, 38(5), 290–295.
- Condon, W., & Kelly-Riley, D. (2004). Assessing and teaching what we value: The relationship between college-level writing and critical thinking abilities. *Assessing Writing*, 9(1), 56–75. <https://doi.org/10.1016/j.asw.2004.01.003>
- Ding, L., Wei, X., & Liu, X. (2016). Variations in university students' scientific reasoning skills across majors, years, and types of institutions. *Research in Science Education*, 46(5), 613–632. <https://doi.org/10.1007/s11165-015-9473-y>
- Dowd, J. E., Connolly, M. P., Thompson, R. J. Jr., & Reynolds, J. A. (2015a). Improved reasoning in undergraduate writing through structured workshops. *Journal of Economic Education*, 46(1), 14–27. <https://doi.org/10.1080/00220485.2014.978924>
- Dowd, J. E., Roy, C. P., Thompson, R. J. Jr., & Reynolds, J. A. (2015b). "On course" for supporting expanded participation and improving scientific reasoning in undergraduate thesis writing. *Journal of Chemical Education*, 92(1), 39–45. <https://doi.org/10.1021/ed500298r>
- Dowd, J. E., Thompson, R. J. Jr., & Reynolds, J. A. (2016). Quantitative genre analysis of undergraduate theses: Uncovering different ways of writing and thinking in science disciplines. *WAC Journal*, 27, 36–51.
- Facione, P. A. (1990). *Critical thinking: a statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. Newark, DE: American Philosophical Association. Retrieved September 26, 2017, from <https://philpapers.org/archive/FACCTA.pdf>
- Gerdeman, R. D., Russell, A. A., Worden, K. J., Gerdeman, R. D., Russell, A. A., & Worden, K. J. (2007). Web-based student writing and reviewing in a large biology lecture course. *Journal of College Science Teaching*, 36(5), 46–52.
- Greenhoot, A. F., Semb, G., Colombo, J., & Schreiber, T. (2004). Prior beliefs and methodological concepts in scientific reasoning. *Applied Cognitive Psychology*, 18(2), 203–221. <https://doi.org/10.1002/acp.959>
- Haaga, D. A. F. (1993). Peer review of term papers in graduate psychology courses. *Teaching of Psychology*, 20(1), 28–32. [https://doi.org/10.1207/s15328023top2001\\_5](https://doi.org/10.1207/s15328023top2001_5)
- Halonen, J. S., Bosack, T., Clay, S., McCarthy, M., Dunn, D. S., Hill, G. W., ... Whitlock, K. (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology*, 30(3), 196–208. [https://doi.org/10.1207/s15328023TOP3003\\_01](https://doi.org/10.1207/s15328023TOP3003_01)
- Hand, B., & Keys, C. W. (1999). Inquiry investigation. *Science Teacher*, 66(4), 27–29.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Holyoak, K. J., & Morrison, R. G. (2005). *The Cambridge handbook of thinking and reasoning*. New York: Cambridge University Press.
- Insight Assessment. (2016a). California Critical Thinking Skills Test (CCTST). Retrieved September 26, 2017, from [www.insightassessment.com/Products/Products-Summary/Critical-Thinking-Skills-Tests/California-Critical-Thinking-Skills-Test-CCTST](http://www.insightassessment.com/Products/Products-Summary/Critical-Thinking-Skills-Tests/California-Critical-Thinking-Skills-Test-CCTST)
- Insight Assessment. (2016b). Sample thinking skills questions. Retrieved September 26, 2017, from [www.insightassessment.com/Resources/Teaching-Training-and-Learning-Tools/node\\_1487](http://www.insightassessment.com/Resources/Teaching-Training-and-Learning-Tools/node_1487)
- Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, 86(3), 314–342. <https://doi.org/10.1002/scce.10024>
- Kuhn, D., & Dean, D. Jr. (2004). Connecting scientific reasoning and causal inference. *Journal of Cognition and Development*, 5(2), 261–288. [https://doi.org/10.1207/s15327647jcd0502\\_5](https://doi.org/10.1207/s15327647jcd0502_5)
- Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, 23(4), 435–451. <https://doi.org/10.1016/j.cogdev.2008.09.006>
- Lawson, A. E. (2010). Basic inferences of scientific reasoning, argumentation, and discovery. *Science Education*, 94(2), 336–364. <https://doi.org/10.1002/scce.20357>
- Meizlish, D., LaVaque-Manty, D., Silver, N., & Kaplan, M. (2013). Think like/write like: Metacognitive strategies to foster students' development as disciplinary thinkers and writers. In Thompson, R. J. (Ed.), *Changing the conversation about higher education* (pp. 53–73). Lanham, MD: Rowman & Littlefield.

- Miri, B., David, B.-C., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education, 37*(4), 353–369. <https://doi.org/10.1007/s11165-006-9029-2>
- Moshman, D. (2015). *Epistemic cognition and development: The psychology of justification and truth*. New York: Psychology Press.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: National Academies Press.
- Pukkila, P. J. (2004). Introducing student inquiry in large introductory genetics classes. *Genetics, 166*(1), 11–18. <https://doi.org/10.1534/genetics.166.1.11>
- Quitadamo, I. J., Faiola, C. L., Johnson, J. E., & Kurtz, M. J. (2008). Community-based inquiry improves critical thinking in general education biology. *CBE—Life Sciences Education, 7*(3), 327–337. <https://doi.org/10.1187/cbe.07-11-0097>
- Quitadamo, I. J., & Kurtz, M. J. (2007). Learning to improve: Using writing to increase critical thinking performance in general education biology. *CBE—Life Sciences Education, 6*(2), 140–154. <https://doi.org/10.1187/cbe.06-11-0203>
- Reynolds, J. A., Smith, R., Moskovitz, C., & Sayle, A. (2009). BioTAP: A systematic approach to teaching scientific writing and evaluating undergraduate theses. *BioScience, 59*(10), 896–903. <https://doi.org/10.1525/bio.2009.59.10.11>
- Reynolds, J. A., Thaiss, C., Katkin, W., & Thompson, R. J. (2012). Writing-to-learn in undergraduate science education: A community-based, conceptually driven approach. *CBE—Life Sciences Education, 11*(1), 17–25. <https://doi.org/10.1187/cbe.11-08-0064>
- Reynolds, J. A., & Thompson, R. J. (2011). Want to improve undergraduate thesis writing? Engage students and their faculty readers in scientific peer review. *CBE—Life Sciences Education, 10*(2), 209–215. <https://doi.org/10.1187/cbe.10-10-0127>
- Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. <https://doi.org/10.1037/a0029315>
- Stephenson, N. S., & Sadler-McKnight, N. P. (2016). Developing critical thinking skills using the science writing heuristic in the chemistry laboratory. *Chemistry Education Research and Practice, 17*(1), 72–79. <https://doi.org/10.1039/C5RP00102A>
- Tariq, V. N., Stefani, L. A. J., Butcher, A. C., & Heylings, D. J. A. (1998). Developing a new approach to the assessment of project work. *Assessment and Evaluation in Higher Education, 23*(3), 221–240. <https://doi.org/10.1080/0260293980230301>
- Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a “universal” rubric for assessing undergraduates’ scientific reasoning skills using scientific writing. *Assessment and Evaluation in Higher Education, 36*(5), 509–547. <https://doi.org/10.1080/02602930903540991>
- Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education, 25*(2), 149–169. <https://doi.org/10.1080/713611428>
- Willison, J., & O’Regan, K. (2007). Commonly known, commonly not known, totally unknown: A framework for students becoming researchers. *Higher Education Research and Development, 26*(4), 393–409. <https://doi.org/10.1080/07294360701658609>
- Woodin, T., Carter, V. C., & Fletcher, L. (2010). Vision and Change in Biology Undergraduate Education: A Call for Action—Initial responses. *CBE—Life Sciences Education, 9*(2), 71–73. <https://doi.org/10.1187/cbe.10-03-0044>
- Zeineddin, A., & Abd-El-Khalick, F. (2010). Scientific reasoning and epistemological commitments: Coordination of theory and evidence among college science students. *Journal of Research in Science Teaching, 47*(9), 1064–1093. <https://doi.org/10.1002/tea.20368>
- Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20*(1), 99–149. <https://doi.org/10.1006/drev.1999.0497>
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*(2), 172–223. <https://doi.org/10.1016/j.dr.2006.12.001>