### Active Learning in Flipped Life Science Courses Promotes Development of Critical Thinking Skills

Melanie L. Styers,<sup>†\*</sup> Peter A. Van Zandt,<sup>†</sup> and Katherine L. Hayden<sup>‡</sup>

<sup>1</sup>Department of Biology and <sup>1</sup>Department of Chemistry and Physics, Birmingham-Southern College, Birmingham, AL 35254

#### ABSTRACT

Although development of critical thinking skills has emerged as an important issue in undergraduate education, implementation of pedagogies targeting these skills across different science, technology, engineering, and mathematics disciplines has proved challenging. Our goal was to assess the impact of targeted interventions in 1) an introductory cell and molecular biology course, 2) an intermediate-level evolutionary ecology course, and 3) an upper-level biochemistry course. Each instructor used Web-based videos to flip some aspect of the course in order to implement active-learning exercises during class meetings. Activities included process-oriented guided-inquiry learning, model building, case studies, clicker-based think-pair-share strategies, and targeted critical thinking exercises. The proportion of time spent in active-learning activities relative to lecture varied among the courses, with increased active learning in intermediate/upper-level courses. Critical thinking was assessed via a pre/posttest design using the Critical Thinking Assessment Test. Students also assessed their own learning through a self-reported survey. Students in flipped courses exhibited gains in critical thinking, with the largest objective gains in intermediate and upper-level courses. Results from this study suggest that implementing active-learning strategies in the flipped classroom may benefit critical thinking and provide initial evidence suggesting that underrepresented and first-year students may experience a greater benefit.

### **INTRODUCTION**

Over the past decade, development of critical thinking skills has emerged as a fundamental goal for undergraduate education. Recent, highly publicized studies have suggested that most undergraduates make only minimal gains in critical thinking and analytical skills during their time in college (Arum and Roksa, 2011; Pascarella et al., 2011). Although the conclusions from these studies may be debated (Huber and Kuncel, 2016), it remains unclear what types of curricular and pedagogical changes best support development of critical thinking skills in undergraduates (Niu et al., 2013; Huber and Kuncel, 2016). Multiple stakeholders, including higher education organizations (Rhodes, 2008; Association of American Colleges and Universities [AAC&U], 2013), employers (AAC&U, 2013; Korn, 2014), and governmental organizations (Duncan, 2010; National Academy of Sciences, National Academy of Engineering, and Institute of Medicine [NAS et al., 2010]) have called for mechanisms to be put in place to ensure that undergraduates develop these skills as a part of baccalaureate education. To identify the most effective pedagogical strategies and develop comprehensive curricula targeting critical thinking, it is essential that we investigate the effects of specific interventions across a variety of courses using well-characterized instruments designed to assess critical thinking.

Identification of optimal educational strategies to enhance critical thinking is hampered by its complexity, with no single, agreed-upon definition among educators. A wide variety of definitions and attributes have emerged, most of which incorporate the use of evidence and/or logic to draw and evaluate conclusions. For example, Erin L. Dolan, Monitoring Editor

Submitted Dec 15, 2016; Revised Apr 10, 2018; Accepted Apr 18, 2018

CBE Life Sci Educ September 1, 2018 17:ar39 DOI:10.1187/cbe.16-11-0332

\*Address correspondence to: Melanie L. Styers (mstyers@bsc.edu).

© 2018 M. L. Styers *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. Rowe and colleagues defined critical thinking as "the ability to draw reasonable conclusions based on evidence, logic, and intellectual honesty" (Rowe *et al.*, 2015). Similarly, Ennis defines critical thinking as "reasonable reflective thinking focused on deciding what to believe or do" (Ennis, 2013, p. 1). He further describes a set of skills or abilities associated with critical thinking that include analyzing arguments, judging the credibility of a source, making and judging inductive inferences and arguments, and using existing knowledge, among others (Ennis, 2015).

As a prelude to developing the Critical Thinking Assessment Test (CAT), Stein and colleagues found that faculty agreed on an assortment of 12 skills that were important aspects of critical thinking. These included separating factual information from inferences, identifying evidence that might support or contradict a hypothesis, separating relevant from irrelevant information, and analyzing and integrating information from separate sources to solve a problem (Stein et al., 2007a). Both Ennis and Stein and colleagues also included basic mathematical reasoning and interpreting graphical information as important critical thinking abilities, underscoring the need for fundamental mathematical skills (Stein et al., 2007a; Ennis, 2015). Importantly, aside from quantitative reasoning, none of the skills identified by Stein and colleagues were discipline specific, suggesting that they could be applied and assessed independent of context or discipline (Stein et al., 2007a). Although the importance of context versus domain-general critical thinking remains a topic of debate (Perkins and Salomon, 1989), because of the demand for critical thinking skills across diverse array of professions, it is crucial that students in all disciplines hone these skills as part of preparation for postgraduate success (AAC&U, 2013).

Although critical thinking skills align very closely with scientific practices, science, technology, engineering, and mathematics (STEM) courses have been some of the most highly criticized regarding development of these skills (Handelsman et al., 2004). Reliance on traditional lecture instead of student-centered instructional strategies and a tendency to focus on content rather than practice have been raised as criticisms of undergraduate science education (Johnson and Pigliucci, 2004; Johnson, 2007; Alberts, 2005, 2009; Momsen et al., 2010). In response, recent recommendations have emphasized that both content and skills (or competencies) should be included in a comprehensive science education curriculum and that evidence-based pedagogical strategies should be employed and assessed to maximize student learning (American Association for the Advancement of Science [AAAS], 2009; Alberts, 2009; Aguirre et al., 2013; Arum et al., 2016; Laverty et al., 2016).

Many lines of evidence suggest that active-learning strategies enhance student performance more than passive, content-focused, traditional, lecture-based strategies (Knight and Wood, 2005; Ruiz-Primo et al., 2011; Freeman et al., 2014), though the quality of implementation appears to be a key factor (Andrews et al., 2011). Unfortunately, evidence assessing direct effects of active learning on critical thinking is limited (Tsui, 1999, 2002), though the impact of student-centered pedagogies on critical thinking is under active investigation (Quitadamo and Kurtz, 2007; Quitadamo et al., 2008; Kim et al., 2013; Greenwald and Quitadamo, 2014; Carson, 2015; Goeden et al., 2015; Snyder and Wiles, 2015). Evidence suggests that active learning is an essential component across pedagogies that enhance critical thinking (Udovic et al., 2002; Prince, 2004; Knight and Wood, 2005; Michael, 2006; Jensen et al., 2015), and many student-centered, active-learning strategies have been reported to improve these skills (Table 1). Repeated interventions are beneficial, as practicing critical thinking skills through student-centered strategies significantly enhances both short- and long-term gains (Holmes et al., 2015). However, despite the evidence supporting use of these active-learning methods, objective assessment of critical thinking gains associated with specific course structures designed to promote active learning is lacking.

Traditional instructor-centered classrooms still predominate in STEM, and lack of faculty implementation of active-learning strategies remains an issue for improving student learning (Ebert-May et al., 2011; Brownell and Tanner, 2012; Fraser et al., 2014). Frequently, instructors feel that they do not have time to implement time-intensive active-learning strategies due to the need for content delivery (Michael, 2007). This dichotomy has led to increasing use of the flipped classroom model for instructors who want to integrate active learning and provide a mechanism for students to review content (Bergmann and Sams, 2012; Rutherfoord and Rutherfoord, 2013). In a lecture-based course, in-class time is focused on content delivery through lecture by the instructor, while out-of-class time is focused on application of the content through homework assignments. In the flipped classroom, students approach content through videos or reading assignments completed before class, and class time is focused on active-learning strategies (Bishop and Verleger, 2013; Herreid and Schiller, 2013). The flipped-class pedagogy has been shown to improve student exam scores and satisfaction (Rossi, 2015; Eichler and Peeples, 2016), likely as a result of increased time spent in active learning (Jensen et al., 2015). However, multiple flipped courses in

TABLE 1.	Active-learning	strategies	shown to	improve	critical	thinkina

	-
Strategy	Reference
Case-based learning	Rowe <i>et al.</i> , 2015
Problem-based learning	Allen and Tanner, 2003; Eberlein et al., 2008
Peer-led team learning/collaborative learning	Gosser and Roth, 1998; Crouch and Mazur, 2001; Quitadamo <i>et al.</i> , 2009; Smith <i>et al.</i> , 2009; Snyder and Wiles, 2015
Process-oriented guided-inquiry learning (POGIL)	Moog et al., 2006; Minderhout and Loertscher, 2007
Student response systems/clickers	Knight and Wood, 2005; Smith et al., 2011
Writing and analysis of the scientific literature (the CREATE method)	Quitadamo and Kurtz, 2007; Gottesman and Hoskins, 2013
Integration of authentic research into courses/labs	Quitadamo et al., 2008; Gasper and Gardner, 2013

a curriculum may be required to achieve persistent learning gains (van Vliet *et al.*, 2015).

Another challenge that hampers development and use of pedagogical strategies that enhance critical thinking is the lack of awareness and infrequent use of validated and reliable measures of critical thinking skills. In many cases, studies have used exam scores as a measure of higher-order thinking skills (e.g., Jensen et al., 2015). However, exam format may influence gains in critical thinking skills, and some question structures have been associated with gender bias (Stanger-Hall, 2012), contributing to the variability of these measures. Self-reported surveys are another commonly used assessment instrument, but it is unclear how well gains from these surveys correlate with objective measures (Bowman, 2010). To address these deficiencies and provide objective means to measure critical thinking, a number of instruments have emerged and have been implemented in the context of undergraduate education. These include the Watson-Glaser Critical Thinking Appraisal (Sendag and Odabasi, 2009), the College Learning Assessment (Arum and Roksa, 2011), the Collegiate Assessment of Academic Proficiency Critical Thinking Test (Pascarella et al., 2011), the Critical Thinking Test (Kenyon et al., 2016), the California Critical Thinking Skills Test (Facione, 1990), the recently developed HEIghten outcomes assessment (Liu et al., 2016), and the CAT (Stein et al., 2016).

The CAT is unique among measures of critical thinking, because this broadly applicable, free-response exam is scored by faculty at the participating institution, providing an opportunity for faculty to better understand the critical thinking skills of their own students (Stein et al., 2007a; Stein and Haynes, 2011). Faculty collaborate on scoring their students' exams using a defined rubric, providing a vehicle for discussing how critical thinking is approached across courses and how different pedagogies affect student learning. These discussions of innovative pedagogies used by colleagues can be beneficial, as exposing faculty to these types of development opportunities has been shown to improve student critical thinking outcomes (Rutz et al., 2012). In addition, because the CAT was developed to assess a discipline-independent set of critical thinking skills (Stein et al., 2007a), it has been used to assess critical thinking in a wide variety of STEM and non-STEM disciplines (e.g., Bielefeldt et al., 2010; Shannon and Bennett, 2012; Gasper and Gardner, 2013; Alvarez et al., 2015; Ransdell, 2016). Application of the CAT across a sample of courses at Tennessee Technological University (TTU) has revealed that the majority of courses do not result in significant gains in critical thinking, suggesting that the CAT can be used to identify courses with interventions that positively impact critical thinking (Harris et al., 2014).

The primary goal of this study was to assess the impact of flipped instruction and active learning on development of critical thinking skills in undergraduate life science students using a valid and reliable measure of critical thinking. We also aimed to compare these results with students' perceptions of their own learning gains to determine the extent to which students could accurately assess their own learning gains. Each of the three courses analyzed adopted a partially or fully flipped teaching strategy, in which lecture videos were used for some portion of content delivery and a variety of active-learning strategies were employed during all or part of class meeting times. Critical thinking skills were measured using the CAT (Stein *et al.*, 2007b), and CAT results were compared with self-reported perceived learning gains through a Likert-style survey implemented through the Student Assessment of their Learning Gains (SALG) site (www.salgsite.org; Seymour *et al.*, 2000), both in a pretest/posttest model. Subgroup analysis was used to identify specific groups of students who derived the greatest benefit from active learning in the flipped classroom.

### **METHODS**

### Institutional Review Board Approval

This study was reviewed and approved by the Institutional Review Board of Birmingham-Southern College (BSC) and was conducted during the Fall semester of 2015. Assessment results and demographic data were deidentified before analysis. Students were informed that assessments (CAT and SALG; see *The CAT*) were to be used for an independent study to assess gains in critical thinking and would not negatively impact their grades. Students were also informed that all data would be deidentified and would not be analyzed until after final grades were submitted and posted. All participants provided informed consent and were presented with an opportunity to decline participation each time an assessment was administered.

### Institution and Courses

BSC is an undergraduate, largely residential, private liberal arts college located in Birmingham, Alabama. BSC belongs to the Associated Colleges of the South (https://colleges.org), a consortium of 16 liberal arts colleges that includes Centre College, Rhodes College, Davidson, Millsaps College, the University of Richmond, Morehouse College, Spelman College, and Centenary, among others. Of the approximately 1300 students at BSC, 19% receive income-based Pell Grants, and approximately 30% are first-generation college students. Additional demographic data are shown in Table 2.

This study analyzed the effects of flipped teaching in three STEM courses: BI 125 Cell and Molecular Biology, BI 225 Evolutionary Ecology, and BI/CH 308 Biochemistry. BI 125 is an introductory cell and molecular biology course that primarily targets STEM majors and pre-health students (Table 2). The Fall 2015 section enrolled 52 students, with a majority being freshmen (68%) or sophomores (25%). This section of the course was slightly weighted toward women (62%) and included a low number of students from underrepresented minorities (6%). The lecture portion of the course employed a "partially flipped" approach, with approximately 20% of class time dedicated to active-learning activities based on the instructor's estimate. Quizzes were used to hold students responsible for watching brief (~10 min) introductory videos associated with each chapter before class, in addition to reading the assigned text. Active-learning strategies included brief, clicker-based, think-pair-share discussions; group problem solving; and hands-on model building. The course also included a required inquiry-based laboratory that met for 3 hours each week. BI 125 was designed to introduce many of the core concepts and competencies recommended in the Vision and Change report (AAAS, 2009), and analysis of course examinations revealed that approximately 44% of points on course assessments were associated with evaluation of higher-order cognitive skills (Table 3).

### TABLE 2. Demographics by course and sample

		BI	BI 125	BI	BI 225 CAT	BI/CH	BI/CH 308
Demographic <sup>a</sup>	BSC	125	CAT sample <sup>b</sup>	225	sample <sup>ь</sup>	308	CAT sample <sup>b</sup>
Ν	1337	52	29	27	19	21	18
Woman	664 (50%)	32 (62%)	17 (59%)	13 (48%)	8 (42%)	8 (38%)	5 (28%)
URM <sup>c</sup>	206 (15%) <sup>d</sup>	3 (6%)	2 (7%)	3 (11%)	3 (16%)	2 (10%)	2 (11%)
Hispanic	28 (2%)	2 (4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
ACT (mean ± SD)	26.0	$28.3\pm3.4$	$28.1\pm3.7$	$27.2\pm3.6$	$26.7\pm3.7$	$29.8\pm3.4$	$30.0 \pm 3.4$
Freshman	504 (38%)	35 (67%)	21 (72%)	1 (4%)	1 (5%)	0 (0%)	0 (0%)
Sophomore	334 (25%)	13 (25%)	7 (24%)	8 (30%)	6 (32%)	0 (0%)	0 (0%)
Junior	305 (23%)	4 (8%)	1 (3%)	14 (52%)	8 (42%)	18 (82%)	14 (78%)
Senior	198 (15%)	0 (0%)	0 (0%)	4 (15%)	4 (21%)	4 (18%)	4 (22%)
Declared STEM major <sup>e</sup>	$\sim 25\%^{\rm f}$	42 (81%)	N/A <sup>g</sup>	24 (89%)	N/A <sup>g</sup>	22 (100%)	N/A <sup>g</sup>

<sup>a</sup>Demographics based on total enrollment at the beginning of the Fall 2015 semester.

<sup>b</sup>Paired pretests and posttests were selected at random from each class.

<sup>c</sup>URM, underrepresented minority.

<sup>d</sup>Institutional data do not include Pacific Islander students due to consolidation with Asian students in the original data set.

<sup>e</sup>STEM majors included biology, chemistry, mathematics, physics, and interdisciplinary urban environmental studies.

Approximately 25% of first-year student express an interest in STEM fields according to student surveys.

<sup>g</sup>N/A, data not available.

BI 225 is an intermediate-level evolutionary ecology course with no laboratory component that was designed for students majoring in biology and in BSC's interdisciplinary urban environmental studies major. This study included two sections of the course enrolling a total of 27 students, the majority of whom were sophomores (30%) or juniors (52%). Men and women (48%) were equally represented in these sections, and underrepresented minority (URM) students comprised a total of 11% of students (Table 2). Approximately 75% of class time was spent in active-learning activities based on the instructor's estimate. To facilitate active-learning strategies, instructors used preclass videos and computer simulations coupled with preclass quizzes to introduce content before most of the class meetings. This mechanism allowed class meetings to be used

#### TABLE 3. Core concepts, competencies, and cognitive skills addressed in BI 125, BI 225, and BI/CH 308

	BI 125	BI 225	BI/CH 308
Core concepts for biological literacy <sup>a</sup>			
Evolution	Х	Х	
Structure and function	Х		Х
Information flow, exchange, and storage	Х	Х	
Pathways and transformations of energy and matter	Х	Х	Х
Systems		Х	Х
Core competencies and disciplinary practice <sup>a</sup>			
Ability to apply the process of science	Х	Х	Х
Ability to use quantitative reasoning	Х	Х	Х
Ability to use modeling and simulation		Х	Х
Ability to tap into the interdisciplinary nature of science	Х	Х	Х
Ability to communicate and collaborate with other disciplines	Х	Х	Х
Ability to understand the relationship between science and society	Х	Х	Х
Course examinations			
Lower-order cognitive skills (LOCS) <sup>b</sup>			
Knowledge	34%	16%	18%
Comprehension	22%	17%	17%
Total LOCS	56%	33%	35%
Higher-order cognitive skills (HOCS) <sup>b</sup>			
Application	16%	30%	10%
Analysis	10%	12%	21%
Synthesis	10%	12%	15%
Evaluation	8%	13%	19%
Total HOCS	44%	67%	65%

<sup>a</sup>Vision and Change core concepts and competencies (AAAS, 2009) addressed in each course were evaluated by each instructor.

<sup>b</sup>Percentages represent average point values associated with each level of Bloom's taxonomy on course examinations as assessed using the Blooming Biology Tool (Crowe *et al.*, 2008).

for case studies, think–pair–share, small-group problem sets, and other active-learning strategies. Approximately 67% of points on course assessments were associated with evaluation of higher-order cognitive skills, and the course addressed most of the concepts and all of the competencies from *Vision and Change* (Table 3; AAAS, 2009).

BI/CH 308 is an intermediate/upper-level biochemistry course with no laboratory component that enrolls both biology and chemistry students, most of whom are interested in health-related professions. The Fall 2015 section of the course enrolled 21 students. Women were underrepresented (38%), and 10% of students came from URMs. All students were juniors (82%) or seniors (18%; Table 2). Content-focused lectures were delivered online and included embedded and graded quizzes. Social media was used to encourage student questions; to identify, share, and discuss pseudo-science-based stories; and to identify content gaps that were addressed in brief "muddiest point" lectures during class meetings. The process-oriented guided-inquiry learning (POGIL) workbook Foundations of Biochemistry (3rd edition) by Loertscher and Minderhout served as a guide for active learning, which comprised more than 90% of class time based on the instructor's estimate. Approximately 65% of points on course assessments were associated with evaluation of higher-order cognitive skills, and the course addressed many of the concepts and all of the competencies from Vision and Change (Table 3; AAAS, 2009).

One pedagogical approach employed in all three courses was the use of targeted critical thinking exercises modeled after CAT Applications (CAT Apps). The CAT App structure was developed by researchers at TTU to provide instructors with a model for developing discipline-specific modules that allow students to practice the critical thinking skills assessed by the CAT (for further information, see www.tntech.edu/cat/cat-applications-in -the-discipline). We designed exercises to help students approach critical thinking in a stepwise manner by presenting students with a problem or data set, then following up with a series of questions meant to guide them through the analysis. Critical thinking exercises were used both as developmental in-class activities and as questions on quizzes and examinations. Sample discipline-specific and non–discipline specific critical thinking exercises are included in the Supplemental Material.

The three courses all differed from one another with respect to content (Table 3), level, and instructor. The enrollment of all three courses was largely STEM majors, and students had similar mean ACT scores (Table 2). Examinations in BI 225 and BI/ CH 308 addressed similar levels of cognitive skills, while examinations in BI 125 focused more on lower-order cognitive skills compared with the intermediate- and upper-level courses (Table 3). Sample course materials and the syllabus for each course are available in the Supplemental Material. Consistent with national trends, a majority of students (71%) enrolled in the three courses self-reported as planning careers in healthcare after graduation, including medicine, dentistry, and allied health fields, on the SALG (www.salgsite.org) pretest described below.

*Vision and Change* core concepts and competencies (AAAS, 2009) addressed in each course were evaluated by each instructor (Table 3). For evaluation of alignment of course assessments with the stated goals of the study, examinations were evaluated based on Bloom's taxonomy using the "Blooming Biology Tool" as previously described (Crowe *et al.*, 2008). As individual

exam questions typically involve multiple levels of Bloom's taxonomy, each level used was tallied, and the highest level achieved was then recorded. Percentages were calculated based on point values for each question. The percentages across the course exams were then averaged to yield the overall percentage for each level (Table 3). Lower-order cognitive skills (LOCS) and higher-order cognitive skills (HOCS) were then summed for each course. Application, which was defined by Crowe *et al.* (2008) as the transition between LOCS and HOCS, was grouped with HOCS in this analysis.

### Assessment of Critical Thinking

Gains in critical thinking were assessed with a matched pretest/ posttest model using two different measures: 1) the CAT (Stein *et al.*, 2007a,b) and 2) a Likert-style survey implemented through the SALG site (www.salgsite.org; Seymour *et al.*, 2000).

### The CAT

The CAT was designed as an interdisciplinary instrument to provide an objective measure of critical thinking by college students (Stein *et al.*, 2007b). The test is composed of 15 free-response and short-answer questions and is designed to be completed within 1 hour by students without any foundational or background knowledge requirements. Validation of the instrument has revealed no cultural biases (Stein *et al.*, 2007b).

Pre- and posttests were completed during the first and last weeks of classes, respectively. The tests were given during the first and last laboratory meetings for BI 125, outside class (pretest) and during class (posttest) for BI 225, and during class meetings for BI/CH 308. Pretests and posttests were paired by student ID number. For BI 125, a subset of matched pairs was selected at random for scoring due to the large size of the class. All matched pairs available were scored for BI 225 and BI/CH 308.

Deidentified tests were scored in two workshops on different days by volunteer faculty and staff from BSC, other local area institutions, and institutions belonging to the Associated Colleges of the South. Each question was scored by a minimum of two reviewers, and by a third reviewer if there was a discrepancy between the first two scores. Interrater reliability was ensured by calibrating scorers before each question using highly detailed rubrics provided by TTU. Scoring sessions were led by BSC faculty trained in scoring the CAT. For assessment of institutional reliability, a random sample of tests were rescored by TTU staff and were found to fall within a 6% error in overall scoring accuracy.

Gains (or losses) represent differences in mean pretest and posttest scores for each course or subgroup within a course. Significant differences were assessed using paired, two-tailed, Student's *t* tests. Effect size was defined as the mean difference of the pretest and posttest scores divided by the pooled SD. Graphical representations of scores are presented as the mean  $\pm$  the SEM. Four students enrolled in both BI 225 and BI/CH 308 were not included in the analysis in order to minimize confounding effects between courses.

### SALG

For alignment of student perceptions of learning with objective differences measured by the CAT, SALG surveys (Seymour *et al.*, 2000) were employed for each course, also in a pretest/posttest model. The survey (included in the Supplemental Material)

required students to rate their own abilities regarding different aspects of critical thinking targeted by the CAT (Stein *et al.*, 2007b). The survey utilized a Likert scale with the following responses: "not applicable" (1), "not at all" (2), "just a little" (3), "somewhat" (4), "a lot" (5), or "a great deal" (6). Precourse and postcourse surveys were available for 5-10 days at the beginning and end of the term. Students in each course were incentivized to complete the surveys for extra credit to increase participation. Results are presented graphically as mean preand posttest scores for each course. Significant differences in student perceptions between pretests and posttests were assessed using unpaired Student's *t* tests.

### **RESULTS**

## Does Active Learning Result in Objective Gains in Critical Thinking?

Scores on the CAT exam were analyzed both individually by question (Table 4) and collectively by total score on the pretest

and posttest (Table 4 and Figure 1A). There was a consistent pattern of higher posttest mean scores across all three courses, although the gain for BI 125 (1.51 points) was not significant (p = 0.06; Table 4). For each of these courses, the mean pretest score for one course was similar to the mean posttest score for the lower-level course (Figure 1A). In line with previously reported results (Stein *et al.*, 2007b), we observed no ceiling effect; students in upper-division courses demonstrated significant critical thinking gains (with effect sizes of 0.52 and 0.46 for BI 225 and BI/CH 308, respectively), and their scores ranged from 7 to 36 (out of a maximum possible score of 40; Table 4).

Each of the questions on the CAT is associated with specific skill(s) or aspect(s) of critical thinking that can be grouped into four major categories: 1) evaluating and interpreting information (E), 2) problem solving (P), 3) creative thinking (C), and 4) effective communication (Co). The distribution of these skills across the exam is shown in Table 4. Importantly, the different courses appeared to target different aspects of

### TABLE 4. Critical Thinking Assessment Test (CAT) results by course

		kill c	atego	ory <sup>a</sup>	E	BI 125	BI 225		BI/CH 308	
Question	E	Р	С	Со	Delta <sup>b</sup>	Effect size <sup>c</sup>	Delta <sup>b</sup>	Effect size <sup>c</sup>	Delta <sup>b</sup>	Effect size <sup>c</sup>
Summarize the pattern of results in a graph without making inappropriate inferences.	Х				0.28*	0.67	0.21		0.39**	1.10
Evaluate how strongly correlational-type data support a hypothesis.	Х			Х	0.11		0.83*	0.78	0.17	
Provide alternative explanations for a pattern of results that has many possible causes.			Х	Х	0.1		0.26		0.62**	0.68
Identify additional information needed to evaluate a hypothesis.		Х	Х	Х	-0.1		-0.06		-0.15	
Evaluate whether spurious information strongly supports a hypothesis.	Х				0		0.06		-0.05	
Provide alternative explanations for spurious associa- tions.			Х	Х	0.2		0.16		0.06	
Identify additional information needed to evaluate a hypothesis.		Х	Х	Х	0.03		-0.05		$-0.28^{\dagger}$	-0.51
Determine whether an invited inference is supported by specific information.	Х				0.04		$0.15^{\dagger}$	0.32	0.05	
Provide relevant alternative interpretations for a specific set of results.			Х	Х	0.14		0.11		0.06	
Separate relevant from irrelevant information when solving a real-world problem.	Х	Х			-0.03		0.1		0.44*	0.57
Use and apply relevant information to evaluate a problem.	Х	Х		Х	$0.28^{\dagger}$	0.40	0.59*	0.79	0	
Use basic mathematical skills to help solve a real-world problem.		Х			0.01		0.15		0.22*	0.73
Identify suitable solutions for a real-world problem using relevant information.	Х	Х			0.1		0.63*	0.59	0.67*	0.61
Identify and explain the best solution for a real-world problem using relevant information.	Х	Х	Х		0.42		0.47		0.34	
Explain how changes in a real-world problem situation might affect the solution.		Х	Х	Х	-0.06		0		0.22	
CAT total score					$1.51^{\dagger}$	0.29	3.56*	0.52	2.74*	0.46

Bold and italic indicate significant differences (p < 0.05). Italic only indicates marginal differences (0.05 ).

<sup>a</sup>Skills categories assessed by each question: E, evaluate and interpret information; P, problem solving; C, creative thinking; Co, effective communication.

<sup>b</sup>Gains or losses (deltas) are reported as the difference between pretest and posttest means for each question or total score. Significant differences were assessed by paired, two-tailed Student's *t* tests.

<sup>c</sup>Mean difference divided by pooled group SD (0.1–0.3, small effect; 0.3–0.5, moderate effect; >0.5, large effect).

\*p < 0.05.

\*\**p* < 0.01.

 $^{\dagger}0.05$ 



FIGURE 1. Implementing active-learning strategies in the flipped classroom significantly improves critical thinking skills. (A) Comparison of mean pre- and posttest CAT scores for students enrolled in BI 125, BI 225, and BI/CH 308 during the Fall semester of 2015. (B) Comparison of mean pre- and posttest CAT scores for men (n = 36), women (n = 30), non-URM (n = 59), and URM students (n = 7) enrolled in BI 125, BI 225, and BI/CH 308 in Fall 2015. (C) Comparison of mean pre- and posttest CAT scores for first-year (freshman [Fr]; n = 21) and second- and third-year (sophomore and junior [So/Jr]; n = 8) students enrolled in BI 125 during the Fall semester of 2015. Error bars represent mean ± SEM. \*, p < 0.05; \*\*, p < 0.01; <sup>†</sup>, 0.05 < p < 0.10.

critical thinking, although some (e.g., summarizing the pattern of results in a graph, identifying suitable solutions for a real-world problem) were shared between two courses. Intriguingly, none of these life science courses resulted in significant gains in identifying additional information needed to evaluate a hypothesis or evaluating whether spurious information strongly supports a hypothesis, which was surprising in light of the similarity of these critical thinking skills to scientific practices (Table 4).

# Do Some Groups of Students Disproportionately Benefit from Active Learning?

Active learning has previously been shown to promote greater learning gains in some subpopulations, including women, first-generation college students, and URM students (Lorenzo et al., 2006; Eddy and Hogan, 2014). Therefore, we performed subgroup analyses to determine whether some groups of students selectively benefited from implementation of active-learning strategies. Both men and women exhibited significant differences in mean pre- and posttest scores on the CAT (Figure 1B; p < 0.01); however, comparison of gains between men and women revealed no differences between the two groups (p =0.75). Though the number of URM students in our sample was small (n = 7), our data suggest that URM students exhibit a gain in mean posttest scores compared with pretest scores (Figure 1B; p < 0.05). Furthermore, the observed trend suggests that these students experienced a larger gain in CAT scores than non-URM students (Figure 1B; 5.19 vs. 2.32; p = 0.11), similar to results from previous studies (Lorenzo et al., 2006; Eddy and Hogan, 2014).

Consistent with the decreased time spent in active learning and decreased emphasis on HOCS on assessments (Table 3), the smallest gains in critical thinking skills were observed in students enrolled in BI 125 (Table 4). Because entering students are the primary target audience for this course, a substantial amount of time in the course is dedicated to exploration of the basic concepts and vocabulary of cell and molecular biology (see BI 125 syllabus in the Supplemental Material). Because the course is designed to accommodate first-year, first-semester students, we next analyzed whether this subgroup experienced gains in critical thinking skills. Subgroup analysis by class standing revealed that the BI 125 first-year cohort, but not upper-level students, exhibited a gain of 2.17 points in critical thinking skills (Figure 1C and Table 5; effect size = 0.41). Similarly, first-year students had significant gains for three questions focused on summarizing the pattern in a graph, providing alternative explanations for spurious associations, and applying relevant information to evaluate a problem (Table 5).

Can Students Accurately Assess Their Own Critical Thinking? Owing to the difficulty and cost of obtaining and implementing objective measures of critical thinking, studies have often used student self-assessment as a measure of learning gains. Thus, one of our goals was to compare the accuracy of these measures with the more objective CAT instrument and to determine whether implementation of active learning affected student attitudes toward critical thinking and/or their course work. Analysis of responses to a SALG survey provided interesting insight into students' perceptions of their own critical thinking (Figure 2). First, in general, many students appear to overestimate their own gains in critical thinking when compared with the objective CAT results, especially in the 100- and 200-level courses. Students in BI 125 reported significant gains in five of 12 critical thinking skills, and students in BI 225 reported significant gains in 11 of 12 skills (Figure 2A). In contrast, students in the upper-division BI/CH 308 course tended to underestimate their own skills, self-reporting significant gains in only two skill categories, in contrast to demonstrating significant gains on five questions on the CAT (compare Figure 2A and Table 4). Furthermore, students in all three courses were not

### TABLE 5. Subgroup analysis of CAT results for BI 125 by class standing

	Skill category <sup>a</sup>			ory <sup>a</sup>	BI 125	First-year	BI 125 So/Jr <sup>b</sup>	
Question	E	Р	С	Со	Deltac	Effect size <sup>d</sup>	Deltac	Effect sized
Summarize the pattern of results in a graph without making inappropriate inferences.	Х				0.33*	0.80	0.13	
Evaluate how strongly correlational-type data support a hypothesis.	Х			Х	-0.09		$0.63^{\dagger}$	0.78
Provide alternative explanations for a pattern of results that has many possible causes.			Х	Х	0.34		-0.50	
Identify additional information needed to evaluate a hypothesis.		Х	Х	Х	0.11		-0.63	
Evaluate whether spurious information strongly supports a hypothesis.	Х				0.05		-0.13	
Provide alternative explanations for spurious associations.			Х	Х	0.38*	0.60	-0.25	
Identify additional information needed to evaluate a hypothesis.		Х	Х	Х	-0.09		$0.38^{\dagger}$	0.76
Determine whether an invited inference is supported by specific information.	Х				0		0.13	
Provide relevant alternative interpretations for a specific set of results.			Х	Х	0.19		0.00	
Separate relevant from irrelevant information when solving a real-world problem.	Х	Х			0		-0.13	
Use and apply relevant information to evaluate a problem.	Х	Х		Х	0.38*	0.56	0.00	
Use basic mathematical skills to help solve a real-world problem.		Х			0.06		-0.13	
Identify suitable solutions for a real-world problem using relevant informa- tion.	Х	Х			0.09		0.13	
Identify and explain the best solution for a real-world problem using relevant information.	Х	Х	Х		0.38		0.50	
Explain how changes in a real-world problem situation might affect the solution.		Х	Х	Х	0.07		-0.38	
CAT total score					2.17*	0.41	-0.25	

Bold and italic indicate significant differences (p < 0.05). Italic only indicates marginal differences (0.05 ).<sup>a</sup>Skills categories assessed by each question: E, evaluate and interpret information; p, problem solving; c, creative thinking; co, effective communication.

<sup>b</sup>So, sophomore; Jr, junior.

<sup>c</sup>Gains or losses (deltas) are reported as the difference between pretest and posttest means for each question or total score. Significant differences were assessed by paired, two-tailed Student's *t* tests.

<sup>d</sup>Mean difference divided by pooled group SD (0.1–0.3, small effect; 0.3–0.5, moderate effect; >0.5, large effect).

\**p* < 0.05.

 $^{\dagger}0.05$ 

confident about using mathematical skills to solve real-world problems (Figure 2A); however, BI/CH 308 students improved their CAT scores on this topic by a large effect size of 0.73 (Table 4). Students in BI 225 generally reported the lowest pretest scores and the highest posttest scores for most categories on the SALG survey, resulting in the largest self-perceived gains (Figure 2A). While these data are consistent with the objective data showing the largest gains for that course, the pattern of gains differs substantially (compare Table 4 and Figure 2A). For example, BI 225 students reported strong gains in identifying new information that might support or contradict a hypothesis, but showed no significant gains in identifying additional information needed to evaluate a hypothesis or evaluating whether spurious information strongly supports a hypothesis on the CAT. Similarly, BI 125 students accurately predicted gains in interpreting graphs, but self-reported some of their strongest gains in understanding the limitations of correlational data, which did not show a significant gain on the CAT (compare Table 4 and Figure 2A). Based on these data, it does not appear that student perceptions are an accurate predication of measurable, objective gains in critical thinking skills.

The SALG survey data also show insight into students' perceptions of the transferability of critical thinking skills (Figure 2B). In general, students reported increases in using critical approaches to solve problems in their daily lives and outside class. Although these data are consistent with the idea that critical thinking skills are transferable, these self-reported data may suffer from the same issues with student perceptions noted above, and more objective measures are needed. Students in BI 225 and BI/CH 308 reported significant increases in their confidence in understanding the material in these courses (Figure 2C), suggesting that students did not perceive that an emphasis on critical thinking replaced important course content.

One of the barriers that keeps instructors from implementing the flipped-classroom approach is the fear of poor evaluations from students (Bernot and Metzler, 2014; Van Sickle, 2016). Because student evaluations often factor into tenure and promotion decisions, implementing these types of pedagogies could negatively affect career progression. In this study, students in all three courses reported a fairly high level of enthusiasm for each of the courses at the beginning of the semester, with no significant changes in level of enthusiasm at the end of the semester (Figure 2C). Students also reported slight increases in confidence in their own performance and in comfort in working with complex ideas (Figure 2C), suggesting that active learning does have some benefit for student attitudes.

### DISCUSSION

Our results are consistent with the interpretation that implementing active-learning strategies through a partially or fully flipped–classroom model can improve critical thinking skills in life science students at different levels and across a variety



FIGURE 2. Self-reported gains in critical thinking skills for students enrolled in STEM courses. Comparison of mean pre- and posttest delta values from the SALG survey for students enrolled in BI 125, BI 225, and BI/CH 308. Self-assessment of critical thinking skills (A), application of critical thinking skills (B), and attitudes toward critical thinking and course work (C). Likert scale responses: 1, not applicable; 2, not at all; 3, just a little; 4, somewhat; 5, a lot; 6, a great deal. \*, p < 0.05; \*\*, p < 0.02; \*\*\*, p < 0.001 comparing subgroup pretest vs. posttest via paired Student's t test. For BI 125, N = 37 pre, 33 post; for BI 225, N = 20 pre, 25 post; for BI/CH 308, N = 22 pre, 21 post. Survey data are included in the Supplemental Material.

of biological subdisciplines. Objective measurement of critical thinking using the CAT revealed pre-to-post gains in critical thinking skills across three courses at BSC: BI 125 Cell and Molecular Biology, BI 225 Evolutionary Ecology, and BI/ CH 308 Biochemistry. Despite our small sample size and lack of matched control courses (see *Limitations and Recommendations for Future Studies*), the gains observed in this study suggest that the interventions employed may benefit students, as the majority of sampled courses do not result in significant gains in critical thinking on the CAT (Harris *et al.*, 2014).

# Implementing Active Learning in the Flipped Classroom Was Associated with Gains in Critical thinking Skills

Over the past several years, the flipped-classroom model has increased in popularity with faculty interested in creating a student-centered classroom with increased time spent in active-learning strategies. Previous studies have shown that the flipped classroom enhances student exam scores and satisfaction (Rossi, 2015; Eichler and Peeples, 2016). These gains have been attributed to increased time spent in active learning (Jensen *et al.*, 2015), increased student preparation (Gross *et al.*, 2015), and increased time spent on-task during class meetings (McLean *et al.*, 2016). However, flipped and activelearning strategies have not yet been widely implemented by faculty, largely due to concerns of workload and the potential for poor student evaluations (Bernot and Metzler, 2014; Miller and Metz, 2014; Van Sickle, 2016).

Our goal was to evaluate the impact of flipped teaching and active learning on critical thinking skills across multiple life science courses targeting different levels of students at the same institution. Comparison of mean pre- and posttest scores across all three courses revealed gains consistent with the interpretation that active learning in the flipped classroom promotes development of critical thinking skills. Of course, the observed gains in critical thinking skills cannot solely be attributed to these courses, as the pre/posttest design of this study is reflective of gains achieved across a semester, when BSC students typically take four courses. However, due to the wide variety of other courses taken by these students, specific effects of individual courses are likely to be minimized. This confounding factor is one issue that arises from the use of an interdisciplinary instrument, although it can be argued that mastery of transferable critical thinking skills, as opposed to discipline-specific problem-solving strategies, is more important for societal goals (Pascarella et al., 2011).

The Vision and Change report highlighted the importance of student-centered pedagogies in improving participation and increasing the confidence of the diverse pool of students entering higher education today (AAAS, 2009). Although our population was too small for broader generalization, findings from our demographic subgroup analysis (Figure 1B) add to the body of evidence suggesting that implementing evidence-based course structures or pedagogies can successfully close achievement gaps between majority and URM students (Lorenzo *et al.*, 2006; Eddy and Hogan, 2014).

Surprisingly, our subgroup analysis also revealed that firstyear students may experience greater gains in critical thinking skills than other students enrolled in the same introductory course. First-year, first-semester students in BI 125 exhibited pre/post gains in a number of critical thinking skills, while sophomores and juniors did not (see Table 5 and Figure 1C). These results suggest that first-year students may be more sensitive to active learning, even in introductory courses characterized by less time spent in active-learning activities and a greater focus on lower-level cognitive skills. In addition to lecture-associated active-learning activities, BI 125 students were also enrolled in an inquiry-based lab that was likely to have contributed to the observed gains (Quitadamo et al., 2008; Gasper and Gardner, 2013). The BI 125 lab was likely the first college-level lab for first-year students, while the majority of upper-level students would have already taken one or more courses with labs, potentially contributing to differential observed gains in CAT scores between these groups.

### Self-Assessment of Student Learning

Our results from SALG surveys confirm previous findings (Boud and Falchikov, 1989; Falchikov and Boud, 1989; Kruger and Dunning, 1999; Dunning *et al.*, 2003) showing that self-perception of student learning, in this case critical thinking skills, has little to no relationship with measured learning gains (compare Figure 2A and Table 4). This lack of coherence has been sug-

gested to be worse for students enrolled in introductory courses (Falchikov and Boud, 1989; Bowman, 2010), consistent with our results for BI 125 and BI 225 versus BI/CH 308. However, we were able to identify some positive attitudinal shifts in response to the interventions used in these courses, indicating that, while students may not be accurate in critical self-assessment, they do experience an increase in confidence. As an example, students, especially those in BI 225, reported greater confidence in their ability to apply critical thinking skills outside class (Figure 2B). Broad applicability and transfer of critical thinking skills are particularly relevant, given the emphasis on critical thinking skills by employers (AAC&U, 2013; Korn, 2014) and the federal government (Duncan, 2010; NAS *et al.*, 2010).

### Implications for Course Design

Owing to the medium-to-large enrollments typically associated with introductory life science courses, implementation of active learning in this setting can be challenging, frequently requiring both technology and thoughtful planning. Oftentimes, the majority of active learning is relegated to smaller lab sections. However, our results suggest that even small amounts of active learning dispersed throughout the lecture may positively impact critical thinking skills, particularly for first-year students. Consistent with this observation, small, brief interventions, such as a few 1-minute papers, have been shown to have substantial impacts on student performance in large, introductory courses (Adrian, 2010). Therefore, for faculty not willing or able to dedicate the time or effort needed to fully flip an introductory course, small, incremental increases in active learning may prove beneficial and should be encouraged. In addition, active learning in the flipped classroom may also be beneficial in upper-level electives, which are frequently smaller and easier to target. These observations provide further support for incorporation of highimpact, student-centered practices, such as undergraduate research and senior capstone experiences (Kilgo et al., 2015), as a part of upper-division course work. In this way, faculty can maximize gains by focusing their efforts in ways most likely to benefit student learning. As more data become available, educators will be in a better position to design courses that efficiently help their students improve their critical thinking skills.

### Limitations and Recommendations for Future Studies

While the results of this study align with previous literature suggesting that active learning improves students' critical thinking skills, the authors acknowledge that, due to the size and nature of our institution, this study was designed with a lack of synchronous, more traditionally taught control sections for each course. In addition, evidence consistent with a greater benefit of flipped teaching for URM students in this study (Figure 1B) is merely suggestive due to the low number of students and high level of variation in the data. Future studies comparing sections of the same course with comparable student demographics, taught by the same faculty using more traditional teaching pedagogies would allow for a more objective analysis and would help to confirm whether active learning in the context of flipped courses is associated with objectively demonstrated gains in specific critical thinking skills.

Additionally, the use of a larger number of comparable course sections and statistical methods to help normalize for student-level differences would better demonstrate whether the benefits of flipped teaching are greater for some groups of students, including URM students and first-year college students. The interpretation of our results is unfortunately limited by our small sample sizes and lack of control sections, particularly for subgroup analyses.

As a result of our study design, we also cannot exclude other differences between the three courses as potential factors affecting gains in critical thinking skills. Differences between the courses analyzed, such as different types of critical thinking emphasized based on course content and level, varying enrollments, the presence or absence of a course-associated lab, and individual differences between instructors, may have contributed to differences in the observed skill-specific gains on the CAT. For example, BI 125 is the largest of the three courses, enrolling more than 50 students in lecture, compared with approximately 20 students per section for BI 225 and BI/CH 308. The effects of class size on student learning outcomes are unclear, with some studies reporting positive effects of smaller class size (Glass and Smith, 1979), others reporting mixed effects (Chingos, 2013), and some reporting differences based on the discipline and level of difficulty of the course (De Paola et al., 2013). These effects are further confounded by the fact that positive impacts of small class size on student learning outcomes, including critical thinking, have been attributed to increased use of active learning in these courses (Cuseo, 2007). The design of our study also did not account for differences in the quality of instruction or active-learning activities, which have been suggested to vary widely in practice (Andrews et al., 2011). CAT scores also could have been impacted by self-selection bias, as not all students participated in the study, as well as differences in motivation and ability-related beliefs, which have been suggested to impact critical thinking (Miele and Wigfield, 2014).

In the future, similar studies should be conducted in a variety of disciplines and class sizes and across cohorts of students. As our study represents a cross-sectional view, it would be helpful to employ a longitudinal design to evaluate the same cohort of students as they progress through a curriculum. Furthermore, because we were unable to separate the specific contributions of different types of active learning used in our study, dissection of how different types of active learning contribute to critical thinking would be beneficial to help faculty identify classroom interventions that produce the greatest gains. As this study demonstrates, the use of objective measures of critical thinking, like the CAT (Stein *et al.*, 2007b), will be essential to provide objective assessment of critical thinking in future studies.

### ACKNOWLEDGMENTS

We thank the Center for Assessment and Improvement of Learning at TTU for their assistance with training, implementation, and analysis of results for the CAT. We also thank Jessica Grunda for her assistance with administering the CAT and Sean McCarthy and many faculty and staff members for volunteering their time to assist with scoring the CAT exam. Funding for this study was provided by a Blended Learning Grant from the Associated Colleges of the South.

### REFERENCES

- Adrian, L. M. (2010). Active learning in large classes: Can small interventions produce greater results than are statistically predictable? *Journal of General Education*, 59, 223–237.
- Aguirre, K. M., Balser, T. C., Jack, T., Marley, K. E., Miller, K. G., Osgood, M. P., ... Romano, S. L. (2013). PULSE Vision & Change rubrics. *CBE–Life Sci*ences Education, 12, 579–581.
- Alberts, B. (2005). A wakeup call for science faculty. Cell, 123, 739-741.
- Alberts, B. (2009). Redefining science education. Science, 323, 437.
- Allen, D., & Tanner, K. (2003). Approaches to cell biology teaching: Learning content in context—problem-based learning. *Cell Biology Education*, 2, 73–81.
- Alvarez, C., Taylor, K., & Rauseo, N. (2015). Creating thoughtful salespeople: Experiential learning to improve critical thinking skills in traditional and online sales education. *Marketing Educational Review*, 25, 233–243.
- American Association for the Advancement of Science (AAAS). (2009). Vision and change in undergraduate biology education: A call to action. Retrieved November 15, 2016, from https://visionandchange.org/ files/2011/03/VC-Brochure-V6-3.pdf
- Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning "not" associated with student learning in a random sample of college biology courses. CBE-Life Sciences Education, 10, 394–405.
- Arum, R., & Roksa, J. (2011). Academically adrift: Limited learning on college campuses. Chicago: University of Chicago Press.
- Arum, R., Roksa, J., & Cook, A. (2016). Improving quality in American higher education: Learning outcomes and assessments for the 21st century. San Francisco: Jossey-Bass.
- Association of American Colleges and Universities (AAC&U). (2013). It takes more than a major: Employer priorities for college learning and student success. Retrieved November 15, 2016, from https://www.aacu.org/ sites/default/files/LEAP/2013\_EmployerSurvey.pdf
- Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. Eugene, OR: International Society for Technology in Education.
- Bernot, M. J., & Metzler, J. (2014). A comparative study of instructor and student-led learning in a large nonmajors biology course: Student performance and perceptions. *Journal of College Science Teaching*, 44, 48–55.
- Bielefeldt, A. R., Paterson, K. G., & Swan, C. W. (2010). Measuring the value added from service learning in project-based engineering education. *International Journal of Engineering Education*, 26, 535–546.
- Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. Paper presented at: 120th ASEE Annual Conference and Exposition, 23–26 June (Atlanta, GA).
- Boud, D., & Falchikov, N. (1989). Quantitative studies of self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18, 529–549.
- Bowman, N. A. (2010). Can 1st-year college students accurately report their learning and development? *American Educational Research Journal*, 47, 466–496.
- Brownell, S. E., & Tanner, K. D. (2012). Barriers to faculty pedagogical change: Lack of training, time, incentives, and ... tensions with professional identity? CBE—Life Sciences Education, 11, 339–346.
- Carson, S. (2015). Targeting critical thinking skills in a first-year undergraduate research course. *Journal of Microbiology & Biology Education*, 16, 148–156.
- Chingos, M. M. (2013). Class size and student outcomes: Research and policy implications. Journal of Policy Analysis and Management, 32, 411–438.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. American Journal of Physiology, 69, 970–977.
- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE–Life Sciences Education*, 7, 368–381.
- Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *Journal of Faculty Development*, *21*, 5–21.
- De Paola, M., Ponzo, M., & Scoppa, V. (2013). Class size effects on student achievement: Heterogeneity across abilities and fields. *Education Economics*, 21, 135–153.

- Duncan, A. (2010). Back to school: Enhancing U.S. education and competitiveness. Foreign Affairs, 89, 65–74.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83–87.
- Eberlein, T., Kampmeier, J., Minderhout, V., Moog, R. S., Platt, T., Varma-Nelson, P., & White, H. B. (2008). Pedagogies of engagement in science: A comparison of PBL, POGIL, and PLTL. *Biochemistry and Molecular Biology Education*, 36, 262–273.
- Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., & Jardeleza, S. E. (2011). What we say is not what we do: Effective evaluation of faculty professional development programs. *BioScience*, 61, 550–558.
- Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE–Life Sciences Education*, *13*, 453–468.
- Eichler, J. F., & Peeples, J. (2016). Flipped classroom modules for large enrollment general chemistry courses: A low barrier approach to increase active learning and improve student grades. *Chemical Education Research and Practice*, *17*, 197–208.
- Ennis, R. H. (2013). The nature of critical thinking: Outlines of general critical thinking dispositions and abilities. Retrieved June 29, 2018, from http://faculty.education.illinois.edu/rhennis/documents/ TheNatureofCriticalThinking\_51711\_000.pdf
- Ennis, R. H. (2015). Critical thinking: A streamlined conception. In Davies, M.,
  & Barnett, R. (Eds.), A handbook of critical thinking in higher education (pp. 31–47). New York: Palgrave Macmillan.
- Facione, P. A. (1990). The California Critical Thinking Skills Test—College level (Technical report #1. Experimental validation and content validity). Millbrae, CA: California Academic Press.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59, 395–430.
- Fraser, J. M., Timan, A. L., Miller, K., Dowd, J. E., Tucker, L., & Mazur, E. (2014). Teaching and physics education research: Bridging the gap. *Reports on Progress in Physics*, 77, 032401.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA*, 111, 8410–8415.
- Gasper, B. J., & Gardner, S. M. (2013). Engaging students in authentic microbiology research in an introductory biology laboratory course is correlated with gains in student understanding of the nature of authentic research and critical thinking. *Journal of Microbiology & Biology Education*, 14, 25–34.
- Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis*, 1, 2–16.
- Goeden, T. J., Kurtz, M. J., Quitadamo, I. J., & Thomas, C. (2015). Community-based inquiry in allied health biochemistry promotes equity by improving critical thinking for women and showing promise for increasing content gains for ethnic minority students. *Journal of Chemical Education*, 92, 788–796.
- Gosser, D. K., & Roth, V. (1998). The workshop chemistry project: Peer-led team learning. *Journal of Chemical Education*, *75*, 185–187.
- Gottesman, A. J., & Hoskins, S. G. (2013). CREATE cornerstone: Introduction to scientific thinking, a new course for STEM-interested freshmen, demystifies scientific thinking through analysis of scientific literature. *CBE–Life Sciences Education*, *12*, 59–72.
- Greenwald, R. R., & Quitadamo, I. J. (2014). A mind of their own: Using inquiry-based teaching to build critical thinking skills and intellectual engagement in an undergraduate neuroanatomy course. *Journal of Undergraduate Neuroscience Education 12*, A100–106.
- Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. *CBE–Life Sciences Education*, 14, ar36.
- Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., ... Wood, W. B. (2004). Scientific teaching. *Science*, *304*, 521–522.
- Harris, K., Stein, B., Haynes, A., Lisic, E., & Leming, K. (2014). Identifying courses es that improve students' critical thinking using the CAT instrument: A case study. In Proceedings of the 10th Annual International Joint Conferences in Computer, Information, System Sciences, and Engineering.

- Holmes, N. G., Wieman, C. E., & Bonn, D. A. (2015). Teaching critical thinking. Proceedings of the National Academy of Sciences USA, 112, 11199–11204.
- Huber, C. R., & Kuncel, N. R. (2016). Does college teach critical thinking: A meta-analysis. Review of Educational Research, 86, 431–468.
- Jensen, J. L., Kummer, T. A., & Godoy, P. D. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE–Life Sciences Education*, 14, ar5.
- Johnson, A. C. (2007). Unintended consequences: How science professors discourage women of color. *Science Education*, *91*, 805–821.
- Johnson, M., & Pigliucci, M. (2004). Is knowledge of science associated with higher skepticism of pseudoscientific claims? *American Biology Teacher*, *66*, 536.
- Kenyon, K. L., Onorato, M. E., Gottesman, A. J., Hoque, J., & Hoskins, S. G. (2016). Testing CREATE at community colleges: An examination of faculty perspectives and diverse student gains. *CBE—Life Sciences Education*, 15, ar8.
- Kilgo, C. A., Ezell Sheets, J. K., & Pascarella, E. T. (2015). The link between high-impact practices and student learning: Some longitudinal evidence. *Higher Education*, 69, 509–525.
- Kim, K., Sharma, P., Land, S. M., & Furlong, K. P. (2013). Effects of active learning on enhancing student critical thinking in an undergraduate general science course. *Innovative Higher Education*, 38, 223–235.
- Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. Cell Biology Education, 4, 298–310.
- Korn, M. (2014, October 22). Bosses seek critical thinking, but what is it? Wall Street Journal, B6.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–1134.
- Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., ... Cooper, M. M. (2016). Characterizing college science assessments: The three-dimensional learning assessment protocol. *PLoS ONE*, 11, e0162333.
- Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016). Assessing critical thinking in higher education: The HEIghten approach and preliminary validity evidence. Assessment & Evaluation Higher Education, 41, 677–694.
- Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physiology*, 74, 118–122.
- McLean, S., Attardi, S. M., Faden, L., & Goldszmidt, M. (2016). Flipped classrooms and student learning: Not just surface gains. Advances in Physiology Education, 40, 47–55.
- Michael, J. (2006). Where's the evidence that active learning works? Advances in Physiology Education, 30, 159–167.
- Michael, J. (2007). Faculty perceptions about barriers to active learning. *College Teaching*, 55, 42–47.
- Miele, D. B., & Wigfield, A. (2014). Quantitative and qualitative relations between motivation and critical-analytic thinking. *Educational Psychology Review*, 26, 519–541.
- Miller, C. J., & Metz, M. J. (2014). A comparison of professional-level faculty and student perceptions of active learning: Its current use, effectiveness, and barriers. Advances in Physiology Education, 38, 246–252.
- Minderhout, V., & Loertscher, J. (2007). Lecture-free biochemistry: A process oriented guided inquiry approach. *Biochemistry and Molecular Biology Education*, 35, 172–180.
- Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. CBE–Life Sciences Education, 9, 435–440.
- Moog, R. S., Creegan, F. J., Hanson, D. M., Spencer, J. N., & Straumanis, A. R. (2006). Process-oriented guided inquiry learning: POGIL and the POGIL project. *Metropolitan Universities*, 17, 41–52.
- National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2010). *Rising above the Gathering Storm, Revisited: Rapidly Approaching Category 5*. Washington, DC: National Academies Press.
- Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Education Research Reviews*, 9, 114–128.

- Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How robust are the findings of "Academically Adrift"? *Change: The Magazine of Higher Learning*, 43, 20–24.
- Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? Education Researcher, 18, 16–25.
- Prince, M. (2004). Does active learning work? A review of the research. Journal of Engineering Education, 93, 223–231.
- Quitadamo, I. J., Brahler, C. J., & Crouch, G. J. (2009). Peer-led team learning: A prospective method for increasing critical thinking in undergraduate science courses. *Science Education*, *18*, 29–39.
- 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, 327–337.
- 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, 140–154.
- Ransdell, M. (2016). Design process rubrics: Identifying and enhancing critical thinking in creative problem solving. In 2016 Proceedings of the Interior Design Educators Council Conference, March 9–12, 2016, Portland, OR.
- Rhodes, T. L. (2008). VALUE: Valid Assessment of Learning in Undergraduate Education. New Directions for Institutional Research, 2008, 59–70.
- Rossi, R. D. (2015). ConfChem conference on flipped classroom: Improving student engagement in organic chemistry using the inverted classroom model. *Journal of Chemical Education*, *92*, 1577–1579.
- Rowe, M. P., Gillespie, B. M., Harris, K. R., Koether, S. D., Shannon, L. J., & Rose, L. A. (2015). Redesigning a general education science course to promote critical thinking. *CBE–Life Sciences Education*, 14. ar30.
- Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science*, 331, 1269–1270.
- Rutherfoord, R. H., & Rutherfoord, J. K. (2013). Flipping the classroom: Is it for you? In *SIGITE '13 Proceedings of the 14th annual ACM SIGITE Conference on Information Technology Education* (pp. 19–22). New York: Association for Computing Machinery.
- Rutz, C., Condon, W., Iverson, E. R., Manduca, C. A., & Willett, G. (2012). Faculty professional development and student learning: What is the relationship? *Change: The Magazine of Higher Learning*, 44, 40–47.
- Sendag, S., & Odabasi, H. F. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education*, 53, 132–141.
- Seymour, E., Wiese, D., Hunter, A., & Daffinrud, S. M. (2000). Creating a better mousetrap: On-line student assessment of their learning gains. In *National Meeting of the American Chemical Society, August 20–24,* 2000, Washington, DC.

- Shannon, L., & Bennett, J. (2012). A case study: Applying critical thinking skills to computer science and technology. *Information Systems Education Journal*, 10, 41–48.
- Smith, M. K., Trujillo, C., & Su, T. T. (2011). The benefits of using clickers in small-enrollment seminar-style biology courses. CBE—Life Sciences Education, 10, 14–17.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323, 122–124.
- Snyder, J. J., & Wiles, J. R. (2015). Peer led team learning in introductory biology: Effects on peer leader critical thinking skills. *PLoS ONE*, 10, e0115084.
- Stanger-Hall, K. F. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. CBE–Life Sciences Education, 11, 294–306.
- Stein, B., & Haynes, A. (2011). Engaging faculty in the assessment and improvement of students' critical thinking using the critical thinking assessment test. Change: The Magazine of Higher Learning, 43, 44–49.
- Stein, B., Haynes, A., & Redding, M. (2007a). Project CAT: Assessing critical thinking skills. In Deeds, D., & Callen, B. (Eds.), Proceedings of the National STEM Assessment Conference (pp. 290–299). Alexandria, VA: National Science Foundation.
- Stein, B., Haynes, A., & Redding, M. (2016). National dissemination of the CAT instrument: Lessons learned and implications. In Proceedings of the AAAS/NSF Envisioning the Future of Undergraduate STEM Education: Research and Practice Symposium, April 27–29, 2016, Washington, DC.
- Stein, B., Haynes, A., Redding, M., Ennis, T., & Cecil, M. (2007b). Assessing critical thinking in STEM and beyond. In Iskander, M. (ed.), *Innovations in E-learning, instruction technology, assessment, and engineering education* (pp. 79-82). Springer.
- Tsui, L. (1999). Courses and instruction affecting critical thinking. *Research in Higher Education*, *40*, 185–200.
- Tsui, L. (2002). Fostering critical thinking through effective pedagogy: Evidence from four institutional case studies. *Journal of Higher Education*, 73, 740–763.
- Udovic, D., Morris, D., Dickman, A., Postlethwait, J., & Wetherwax, P. (2002). Workshop biology: Demonstrating the effectiveness of active learning in an introductory biology course. *BioScience*, 52, 272–281.
- Van Sickle, J. (2016). Discrepancies between student perception and achievement of learning outcomes in a flipped classroom. *Journal of the Scholarship of Teaching and Learning*, *16*, 29–38.
- van Vliet, E. A., Winnips, J. C., & Brouwer, N. (2015). Flipped-class pedagogy enhances student metacognition and collaborative-learning strategies in higher education but effect does not persist. *CBE–Life Sciences Educa-tion*, 14, ar26.