Scholarship of Teaching and Learning in Psychology
2016, Vol. 2, No. 4, 245–260

2332-2101/16/$12.00 http://dx.doi.org/10.1037/stl0000070

Strengthening Scientific Reasoning Skills in Introductory Psychology: Evidence From Community College and Liberal Arts Classrooms

Courtney Stevens, Melissa R. Witkow, and Brenna Smelt
Willamette University

This research represents an extension of our prior article, published in *Teaching of Psychology* (Stevens & Witkow, 2014), which detailed the development and evaluation of a single module for training scientific reasoning skills in introductory psychology. Here we report on the development of a larger set of 8 modules, along with evaluation data from a 4-year baccalaureate liberal arts college ($N = 195$ students) and a 2-year community college ($N = 94$ students). Each module required 30–45 min of class time and was keyed to a different major content topic covered in introductory psychology. Participating instructors, who were uninvolved in the development of materials, chose how many and which modules to implement during the term. Gains on a scientific reasoning assessment were compared from the beginning to the end of the term in students in treatment sections versus comparison sections taught as usual. In the liberal arts and community college settings, students in treatment sections showed significantly greater gains in scientific reasoning relative to students in comparison sections (Cohen's $d = +0.66$ at the liberal arts college and $d = +1.06$ at the community college). At the community college setting, an additional within-instructor comparison demonstrated that students in the treatment section also made greater gains in scientific reasoning than students taught by the same instructor, but in the term before receiving the modules (Cohen's $d = +1.05$). Taken together, these results provide evidence for the use of modular activities embedded throughout the term to improve scientific reasoning in introductory psychology courses.

Keywords: scientific reasoning, critical thinking, introductory psychology

The American Psychological Association (APA) identifies "scientific inquiry and critical thinking" as one of five key learning goals for the undergraduate psychology major (APA, 2013, 2016). By the time a student is partway through the major, it is expected that she or he will be able to interpret simple graphs and statistical findings, describe fundamental principles of research design, and discuss the value of controlled experiments in drawing causal conclusions (APA, 2013, 2016). The APA further advocates that beginning at the introductory level, students should be introduced to psychology as a scientific discipline (APA, 2011). Indeed, in the report of an APA task force on the introductory course, the first recommendation for conceptual consistency calls for strong coverage of the scientific method, stating "students in Intro Psych should learn skills involving the development of scientific reasoning and problem solving, including effective research methods" (APA, 2014, p. 16). Thus, it is not surprising that research methods are considered the base of the ideal introductory psychology course on which the study of different content domains can be anchored (APA, 2014; Gurung et al., 2016).

The Medical College Admission Test (MCAT) offers additional, perhaps unexpected, support for the focus on scientific reasoning in introductory psychology. Introductory psychology is now considered part of the premedical

Courtney Stevens, Melissa R. Witkow, and Brenna Smelt, Department of Psychology, Willamette University.

This research was supported by National Science Foundation Grant 1505060 from the Division of Undergraduate Education.

Correspondence concerning this article should be addressed to Courtney Stevens, Department of Psychology, Willamette University, 900 State Street, Salem, OR 97301. E-mail: cstevens@willamette.edu

curriculum, with the latest revision of the MCAT including a new section on psychological, social, and biological foundations of behavior (Association of American Medical Colleges, 2011, 2012; Mitchell, Lewis, Satterfield, & Hong, 2016). More than half of the questions on this new section are based on introductory psychology and, along with testing psychology content, several questions require students to show they can "do" science—in other words, to apply scientific reasoning principles (Association of American Medical Colleges, 2011, 2012; Mitchell et al., 2016). This includes testing of data-based and statistical reasoning (MCAT Skill 4; e.g., using data to draw conclusions, interpreting data presented in figures, graphs, or tables) as well as reasoning about the design and execution of research (MCAT Skill 3; e.g., identifying dependent and independent variables, critiquing conclusions that can be drawn from particular types of studies). Questions tapping these scientific reasoning skills generally take the form of passage sets, in which students read a summary of a research study, including graphs and/or tables of study results, and then answer a series of multiple choice questions based on that information. Beyond their relevance to future physicians, these types of scientific reasoning questions represent meaningful learning outcomes for all introductory psychology students (Frazer & Twohig, 2012; Mitchell et al., 2016). Therefore, it is perhaps not surprising that some textbook publishers have begun including MCAT-style passage set questions in their test banks, sometimes labeled as "MCAT scenarios," which are identified as aligning with the APA learning outcome of scientific inquiry and critical thinking (e.g., see associated test bank for Schacter, Gilbert, Wegner, & Nock, 2014).

Although the new APA guidelines and MCAT questions provide direction for the importance and assessment of scientific reasoning skills in introductory psychology, what is less clear is how introductory psychology courses could be structured to effectively train these skills. The inclusion of research projects or dedicated laboratory sections is one natural pedagogical strategy, and there is evidence that students who take introductory psychology with an accompanying laboratory section (Thieman, Clary, Olson, Dauner, & Ring, 2009) or a supplementary course that focuses on scientific reasoning (Penningroth, Despain, & Gray, 2007) show larger improvements in the ability to identify flaws in research studies relative to students taking introductory psychology without these supplements. However, a recent review of introductory psychology courses nationwide indicates that fewer than 10% of introductory psychology courses include a laboratory component (Norcross et al., 2016), making this strategy difficult to implement without major accompanying changes in course structure and staffing resources. Accompanying laboratory sections may be particularly unlikely to occur at community colleges, where a significant amount of introductory psychology instruction occurs (Goldstein, 2010; Kena et al., 2016). As such, there is strong motivation to identify methods for training scientific reasoning that can easily be incorporated into existing classroom structures.

Although instructors searching for such scientific reasoning activities will find no shortage of possibilities described in the literature, the majority of these activities have not been empirically validated. For example, there are descriptions of activities that address specific aspects of scientific reasoning, such as working with graphs and tables (Holmes, 2008; Lutsky, 2006; Nolan & Heinzen, 2009), analyzing research methodology (Gareis, 1995), or responding to specific written prompts that target aspects of critical thinking (Wade, 1995). In addition, the APA has developed a useful Online Psychology Laboratory, which includes a set of freely available laboratory activities that students can complete even in the absence of a traditional accompanying laboratory section (http://opl.apa.org). However, most of these activities have not been evaluated, leaving it unclear whether implementing them will lead to changes in key student learning outcomes.

Empirically validated activities to train scientific reasoning are less common, although there are exceptions. For example, Adam and Manson (2014) developed an activity for use in a 75-min class session in which students watched a short infomercial containing pseudoscientific claims and then identified and evaluated the evidence used to support the claims. Students in classes that included this activity were able to identify significantly more problems with hypothetical research study descriptions than students in comparison sections of introductory psychology who received only a traditional lecture covering similar content. In another study, Blessing and Blessing (2010) developed a term-long research project, modeled after the TV program MythBusters, in which students found and evaluated research studies to determine whether psychological claims (e.g., "Blind people have unusually sensitive organs of touch") were confirmed, plausible, or "busted." In comparison to students enrolled in a different section of introductory psychology, students in the course including the research project showed larger improvements on a test of scientific reasoning that required students to consider a psychological claim (e.g., "Blondes have more fun") before answering a series of four scientific reasoning questions. Sample questions included describing an experiment to examine the claim or speculating on how possible results of the experiment might support the claim.

In our own prior research (Stevens & Witkow, 2014), we have examined the effects of an in-class activity on students' scientific reasoning outcomes. In this research, we developed a single, 45-min class activity in which students were asked to read a short description of a published research study comparing the effects of Prozac to saffron for the treatment of depression. Similar to MCAT-style passage sets, the summary included a data figure and table. Students worked in small groups to answer a series of scientific reasoning questions, including interpreting patterns of data from the graph and table, drawing appropriate conclusions based on aspects of the study design (e.g., lack of a no-treatment control group), and proposing follow-up experiments to test issues unresolved by the initial study. Relative to students in comparison sections of introductory psychology not receiving the modules, students receiving the modules scored higher on scientific reasoning questions from an MCAT passage set whereas there were no differences between groups on content-oriented questions.

Although these examples of research-validated, modular scientific reasoning activities are promising, a striking limitation remains—namely that all prior work, including our own, has been limited to testing by instructors who developed the materials being implemented. The potential risk of failing to test instructional materials and activities outside of the original creator is twofold. First, although the original creator may have success with the teaching materials, it may not be practical or simple for other instructors to interpret and use the materials. Second, improvements in scientific reasoning might be attributable to other aspects of the instructor and not the activity per se. In addition, a natural consequence of this evaluation strategy is that materials are typically only evaluated at a single institution, leaving it unknown whether the materials can be effective in different contexts with different student populations. There is a diverse range of course structures for the introductory course (e.g., large lecture or online sections), but a first step toward broader assessment could be to implement instructional materials in courses with a similar size and structure (e.g., smaller sized, face-to-face classes) but at different types of institutions.

The present study is designed to address these limitations while also advancing the range of research-validated materials available to instructors wishing to embed scientific reasoning training into existing introductory psychology courses. Here we report on the development of a set of eight scientific reasoning modules for introductory psychology and the evaluation of these materials using instructors uninvolved in their development. Building on our prior research (Stevens & Witkow, 2014), we developed a larger set of eight scientific reasoning modules. Each module required 30–45 min of class time, and instructors could choose how many and which modules to implement. Evaluation took place at a 4-year baccalaureate liberal arts college (Study 1) and a community college (Study 2). Gains in students' scientific reasoning, assessed using MCAT-style multiple-choice passage set questions, were compared in classes in which the instructors implemented at least one module relative to comparison sections of introductory psychology taught as usual. It was predicted that in the liberal arts and community college settings, students in classes receiving the modular scientific reasoning activities would show greater gains in scientific reasoning relative to students in comparison sections.

General Method

Overview

Two studies were conducted to examine the effect of incorporating scientific reasoning modules into introductory psychology. Study 1 was conducted at a selective, 4-year baccalaureate liberal arts college (Willamette University) whereas Study 2 was conducted at a community college (Chemeketa Community College). At each institution, a subset of instructors teaching introductory psychology received a set of eight teaching modules from the study authors. The instructors had no involvement in creating the materials and could select how many and which modules to implement during the term. Using a pretest/posttest assessment focusing on specific scientific reasoning skills, growth in scientific reasoning was compared between students in sections of introductory psychology in which the instructor implemented modules as compared with sections offered at the same institutions but taught as usual. All study procedures were approved by the institutional review board at Willamette University.

Measures

Scientific reasoning modules. A set of eight scientific reasoning modules was developed by the study authors. The modules aligned with scientific reasoning outcomes as defined by the APA's undergraduate guidelines for training in scientific inquiry and critical thinking (Learning Outcome 2; APA, 2013) as well as the American Association of Medical College's MCAT skills of data-based and statistical reasoning (Skill 4) and reasoning about the design and execution of research (Skill 3; Association of American Medical Colleges, 2012). These modules were developed following the same standards as a single successful scientific reasoning module, which focused on an empirical study examining the benefits of saffron as a treatment for depression (Stevens & Witkow, 2014).

Each module was designed to take 30–45 min of class time and addressed one of eight general topics commonly covered in introductory psychology, allowing for integration of the modules across many topics and in any order throughout the term. The modules aligned with the following eight topics: clinical/abnormal, developmental, biological, sensation/perception, learning, memory, personality, and social psychology. Each module centered on an empirical seed article that met two conditions. First, the article needed to contain data relevant to a potential application of psychology to everyday life, such that materials would not become quickly dated (e.g., how the presence of peers affects teen driving behaviors). Second, the article needed to include data that could be presented as simple tables or graphs as well as design features that would allow students to engage in scientific reasoning (e.g., identifying study limitations, designing possible follow-up studies).

Each module included a lesson plan for the instructor as well as a student handout, a PowerPoint slideshow with pertinent visuals, or both. The student handout and/or PowerPoint slideshow provided a concise summary of the empirical article, similar to what would be provided in a passage set question (note that neither the instructor nor the students read the seed article, and the original empirical article was not included in the materials provided to instructors). The instructor lesson plan provided a suggested structure for the class discussion, including notes about possible student misconceptions for different scientific reasoning questions.

The typical structure of a module began with a general class discussion of the topic at hand (e.g., teen driving) to establish interest and common ground. After this general discussion, students received either the study handout or a verbal description of the study accompanied by PowerPoint slides. Students then answered a series of scientific reasoning questions, typically in small groups. For example, students might need to identify the dependent variable(s) of a study, interpret study results from graphs or tables, determine what conclusions might be drawn from a particular study, or design follow-up studies. During this time, the instructor could circulate among groups, addressing common misconceptions in students' scientific reasoning. All modules concluded with a full-class discussion of the scientific reasoning questions, including possible applications of the findings to real-world issues. A complete description of one of the eight modules is available in previously published work (Stevens & Witkow, 2014), and copies of all modules are available from the authors on request.

Scientific reasoning assessment. A 14-question, multiple-choice scientific reasoning assessment was developed to assess changes in scientific reasoning. Each question presented four alternatives. No questions required students to draw from a base of content knowledge, and all could be answered by applying scientific reasoning skills to a passage-based scenario or data tables and graphs.

The assessment included one passage set and related scientific reasoning questions from the first edition of the MCAT preview guide (Association of American Medical Colleges, 2011) and one MCAT-style passage set and related scientific reasoning questions from a textbook test bank (Schacter et al., 2014; this textbook was not used in any of the classes included in the present study). A total of eight questions were drawn from these resources (two questions, concerning identifying the dependent and independent variable in the study, were derived from a single original test bank question). The questions drawn from the MCAT preview guide were classified as tapping MCAT Skill 3 or 4 (reasoning about design and execution of research; data-based and statistical reasoning) and have been used in our previous research (Stevens & Witkow, 2014). The questions drawn from the textbook test bank were classified within the test bank as assessing the APA Guidelines Learning Outcome 2 (scientific inquiry and critical thinking) and were determined by the study authors to also address aspects of scientific reasoning. An additional six questions were written by the authors. Two of these questions required reference to the test bank passage set and four required reference to an isolated table and/or scatterplot created by the authors that did not include accompanying text. The distractors for all additional questions were designed to capture possible student misconceptions.

The 14 questions on the assessment were grouped into four different subskills capturing more specific aspects of scientific thinking addressed in one or more of the scientific reasoning modules: defining variables (2 questions; e.g., *What was the dependent variable in this study?*), interpreting data from graphs and tables (3 questions; e.g., *Which of the following conclusions is NOT supported by Figure 3*?), reasoning about research design and appropriate conclusions (6 questions; e.g., *Based on the study description, which of the following limits the generalizability of findings?*), and reasoning specifically about correlational research design and results (3 questions; e.g., *In psychological research, which of the following is most appropriate for identifying cause and effect?*).

General Procedure

Before the start of the course, instructors of the treatment sections were provided with a zip file containing a table of contents listing the available modules and eight separate folders with the materials for each module. Instructors were told that they could select which modules to implement, and when, during the term, and that they could feel free to modify the modules in any way that would suit their classroom needs (all lesson plans and student handouts were provided as pdf and editable docx files, and the PowerPoint slides were provided as editable pptx files).

Students in all participating introductory psychology sections (treatment and comparison) completed the scientific reasoning preassessment on the first day of class. Students did not sign consent forms because the project was approved as exempt by the institutional review board, falling under the exemption category of normal educational practices. The assessment was included as an in-class activity, but it did not contribute points to students' grades. Instructors were asked not to provide feedback or discuss the goals and purpose of the assessment with students. Students were not aware of any changes being implemented in different sections or to planned comparisons between course sections, and they did not know that a posttest would occur later in the semester. Instructors were asked to pass the completed assessment on to the study authors without reviewing student answers. The exact same assessment was administered a second time, at the end of the term, together with the final exam (Study 1, although not contributing to students' exam scores) or on the final day of class (Study 2, again without contributing points to students' final grades). Students who did not complete the pre- and postassessment were excluded from analysis. Although no strict time limit was imposed, approximately 20 min were allocated to the assessment, and instructors reported that all students completed the assessment comfortably within the available time provided.

Data Analysis

Student scores on the scientific reasoning assessment were converted to percentage correct (number correct/14 possible questions). The same transformation was applied to each subskill (number correct/number of items indexing that subskill). Note that this linear transformation of student scores into percentage correct has no effect on statistical significance but renders the results more directly interpretable when comparing across groups or subskills. Change scores were also calculated for each student by subtracting the percentage correct at pretest from the percentage correct at posttest for the main assessment and each individual subskill. For example, a student who scored 6 of 14 correct at pretest (43%) and 8 of 14 correct at posttest (57%) would show a change score of +14 percentage points.

Study 1

Method

Participants. A total of 195 undergraduate students from 10 different sections of introductory psychology at Willamette University participated in Study 1. Willamette University is a selective, 4-year, baccalaureate liberal arts college located in Salem, Oregon enrolling approximately 1,900 undergraduate students each year. Of these, 34% are from ethnic minority backgrounds and 23% receive Pell Grants. At Willamette, introductory psychology is taught as a single-semester course, with instructors given discretion to create their own assignments, activities, and exams. During course registration, students select the section of their choice (each semester 60% of seats are reserved for first-year students, with first-semester freshmen enrolled in sections that fit their course schedule through the Dean's Office). All sections were offered during the 2015–2016 academic year, with half of the sections in the fall semester and half in the spring semester. Each section is capped at 25 students. Student demographic information was not collected as part of the study assessment.

Five sections were designated the treatment sections because the instructors of these sections received and chose to use the scientific reasoning modules (n = 97 students, three fall semester sections and two spring semester sections). The remaining five sections were designated the comparison sections because the instructors of these sections taught their courses as usual, without inclusion of the scientific reasoning modules (n = 98 students, two fall semester sections and three spring semester sections). None of the instructors of either the treatment or comparison sections were involved in the development of the scientific reasoning modules.

Procedure. As earlier described, at the start of the semester, instructors of the treatment sections were provided with a zip file containing separate folders with the teaching materials for each of the eight modules. One instructor, who taught three treatment sections, implemented three modules during the semester, including the modules on depression (clinical unit), teen driving (developmental unit), and memory in the context of eyewitness testimony (memory unit). A second instructor, who taught two treatment sections, implemented five modules during the semester, including the modules on depression (clinical unit), the effect of school start times on sleep (biological unit), memory in the context of eyewitness testimony (memory unit), the correlation between extraversion and social interaction (personality unit), and the influence of social networks on cold susceptibility (social unit). Instructors for the comparison sections taught their class as they normally would and made no changes for the semester on the basis of the modules.

Results

Table 1 separately presents the percentage correct on the assessment at pre- and posttest for the treatment and comparison groups. Means are presented for the entire assessment ("all skills") as well as each of the four scientific reasoning subskills. Gain scores indicate change from pre- to posttest for each group, along with results of the independent-sample t test comparing gain scores across groups.

As shown in Table 1, at pretest the treatment and comparison groups scored relatively poorly (<50% correct) on the overall assessment, with

Table 1
Percentage Correct at Pre- and Posttest From Study 1 (Liberal Arts College Setting), Separately for the Treatment and Comparison Groups for the Entire Assessment ("All Skills") As Well As the Four Subskills

Outcome measure	Treatment *M* (*SD*)	Control *M* (*SD*)	*t*(193)	*p*	*d*
All skills					
Pretest	42 (15)	44 (16)			
Posttest	56 (17)	47 (18)			
Gain	15 (16)	3 (19)	4.61	.00	0.66*
Definitional					
Pretest	70 (43)	58 (46)			
Posttest	81 (38)	69 (43)			
Gain	12 (49)	11 (49)	0.09	.93	0.02
Correlational reasoning					
Pretest	37 (26)	43 (26)			
Posttest	56 (29)	45 (28)			
Gain	18 (37)	2 (29)	3.46	.00	0.50*
Research design					
Pretest	38 (20)	45 (20)			
Posttest	52 (20)	47 (22)			
Gain	14 (23)	2 (25)	3.43	.00	0.50*
Graph interpretation					
Pretest	35 (30)	36 (29)			
Posttest	48 (28)	36 (30)			
Gain	13 (35)	0 (35)	2.62	.01	0.38*

Note. Gain scores indicate change from pre- to posttest for each group, along with results of independent-sample *t* tests comparing gain scores across groups. Positive values of Cohen's *d* indicate greater gains for the treatment group, with asterisks indicating a significant difference between groups at the .05 significance level.

the pattern of means suggesting that students scored best on the subskill of defining variables (~64% correct) and worst on the subskill of graph interpretation (~35% correct). The pattern of means indicates that gain scores for students in the comparison classes were relatively flat across the overall assessment (3 percentage point increase), with the largest gain in the definitional subskill, in which students in the comparison class improved by 11 percentage points. In contrast, students in the treatment group showed an average gain of 15 percentage points on the overall assessment and 12–18 percentage points on each of the four subskills.

Direct comparison of gain scores across the treatment and comparison groups supported the study predictions. As shown in Table 1 and Figure 1, students in treatment sections made significantly greater gains overall on the 14-item scientific reasoning assessment than students in the comparison sections ($p < .01$, $d = +0.66$). To examine whether this pattern was specific to particular subskills of scientific reasoning, additional analyses separately compared gain scores between the treatment and comparison classes for each subskill. As shown in Table 1 and Figure 1, students in the treatment classes made significantly greater gains than students in the comparison classes in three of the four scientific reasoning subskills: reasoning about research design ($p < .01$, $d = +0.50$), graph interpretation ($p < .05$, $d = +0.38$), and correlational reasoning ($p < .01$, $d = +0.50$). The groups did not differ significantly in gains on the defining variables subskill ($p = .92$, $d = +0.02$).

The analyses previously reported focused on a comparison of gain scores, which has the advantage of being easily interpretable and providing a direct test of the research question. One disadvantage of this approach is that it does not statistically control for possible group differences in performance at pretest. Furthermore, this approach can have reduced statistical power because variance in the outcome accounted for by pretest score is considered unexplained in the model. Thus, supplementary regression analyses were conducted that controlled for pretest score in addition to direct statistical comparisons of the groups at pretest. The direct statistical comparison of pretest scores between groups indicated one significant difference, with

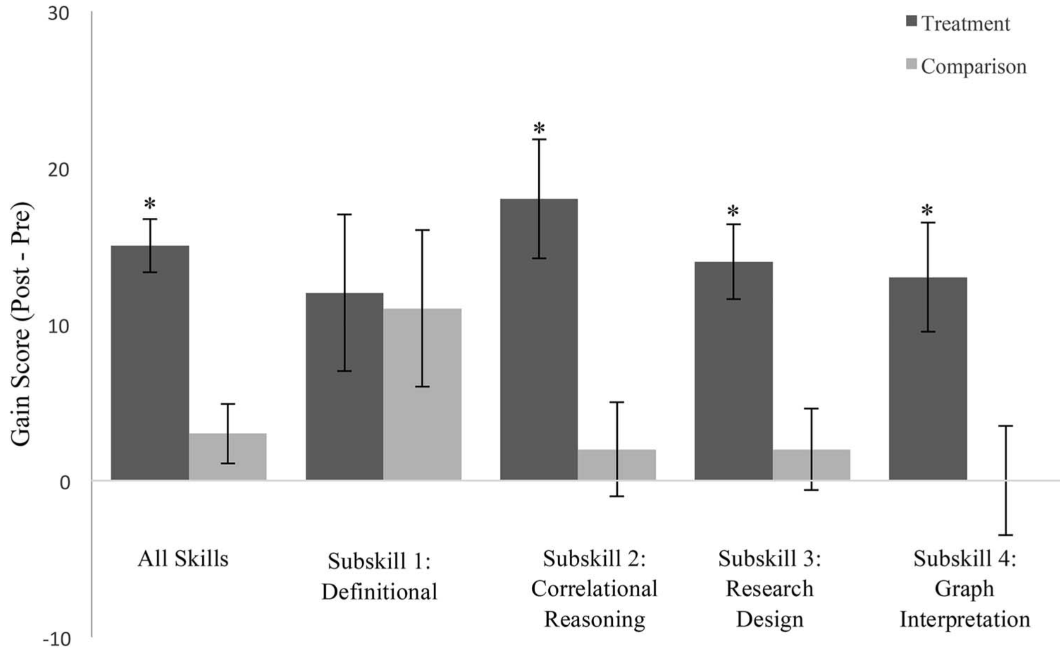

Figure 1. Gain score (posttest – pretest) from Study 1 (liberal arts college setting), separately for the treatment and comparison groups for the entire assessment ("all skills") as well as the four subskills. Error bars represent *SEM*, with asterisks indicating a significant difference in gain scores between groups at the .05 significance level.

the control group starting the semester with significantly higher scores than the treatment group on the research design subskill, $t(193)$, = −1.93, $p < .05$. No other differences between groups were significant at pretest (all $p > .05$). However, for consistency, the supplementary regressions were conducted for all outcomes.

Each supplementary regression predicted gain score from treatment condition (0 = *control*, 1 = *treatment*) while including pretest score on that outcome in the model. Not surprisingly, in all of these analyses pretest scores negatively and significantly predicted growth (i.e., those scoring lowest at pretest, who had the greatest room for improvement, also showed the largest gains). However, critical to the hypothesis of the study, this secondary set of analyses also confirmed all of the main pattern of results, with the effect of treatment remaining significant when controlling for pretest score on the overall assessment ($\beta = 0.28$, $p < .001$) and the three subskills showing significant effects in the primary analysis: research design subskill ($\beta = 0.15$, $p < .01$), graph interpretation subskill ($\beta = 0.18$, $p < .01$), and correlational reasoning subskill ($\beta = 0.19$, $p < .01$). The effect of treatment remained nonsignificant for the definitional subskill ($\beta = 0.08$, $p = .12$).

Discussion

The results of Study 1 indicate that modular activities can be used to improve the scientific reasoning skills of introductory psychology students in a liberal arts college setting, even when the instructors using the activities do not create the curriculum. Instructors using the modules implemented only three to five lessons across the semester, replacing less than 4 hours of traditional class content with activities dedicated to scientific reasoning. The effect size of this change was large ($d = +0.66$), indicating that the modules improved scientific reasoning by two thirds of a standard deviation relative to courses taught as usual. Moreover, the benefits were observed across a range of scientific reasoning subskills, including interpreting data from graphs and tables, reasoning about research design and appropriate conclusions, and reasoning specifically about correlational research designs and results.

The one subskill not showing significantly larger gains in the liberal arts treatment group was defining variables, which requires students to identify the dependent and independent variables in a study. The reason for this is unclear, although one possibility is that because students began the semester with relatively high scores on this subskill, there was less room for students to show improvement. Because this subskill is tightly connected to vocabulary relatively specific to science courses, it is possible that some students entered their college-level psychology course already familiar with the terms *independent* and *dependent variable* from prior coursework in high school. It is also noteworthy that within the comparison sections, defining variables was the subskill showing the largest change (increase in 11 percentage points). As such, it is possible that this aspect of scientific reasoning is more likely to be covered in introductory psychology courses in a liberal arts setting, even without the inclusion of specific scientific reasoning modules. Another possibility is that students may understand the concepts of dependent and independent variables, and even be able to reason scientifically about these concepts, but still not show growth in the use of specific terminology tied to these concepts.

Although Study 1 provided strong support for the benefit of scientific reasoning modules, one limitation was its focus on a single liberal arts college. Study 2 was designed to address this limitation by testing materials in a community college setting.

Study 2

Method

Participants. A total of 94 undergraduate students from three sections of introductory psychology at Chemeketa Community College participated in Study 2. Chemeketa is a 2-year public community college with seven locations. Chemeketa's main campus, located in Salem, Oregon, serves approximately 10,000 students each year, with 45% from ethnic minority backgrounds and 47% receiving Pell Grants. All testing occurred at the Salem campus. At Chemeketa Community College, introductory psychology is offered as a two-quarter sequence (PSY201/202), and all participating students were currently enrolled in a section of PSY201. PSY201 and PSY202 include methodology training. However, PSY201 includes units on biological foundations, development, sensation, perception, consciousness, learning, and memory, whereas PSY202 includes units on language, thinking, emotion, motivation, intelligence, personality, health, abnormal behavior, therapy, and social thinking. Students can take PSY201 and PSY202 in any order. Instructors have discretion to create their own assignments, activities, and exams. During course registration, students select the section of their choice. All sections were offered during the 2015–2016 academic year as face-to-face courses (Chemeketa also offers online and hybrid PSY201/202 sections). Each section was capped at 35 students. Student demographic information was not collected as part of the study assessment.

One section was designated the treatment section; the instructor of this section received the set of scientific reasoning modules and chose to implement one of the modules during the quarter ($n = 31$ students, offered spring quarter). The remaining two sections were designated the comparison sections; the instructors of these sections taught their courses as usual, without inclusion of the scientific reasoning modules ($n = 63$ students, offered winter or spring quarter). This included one section that was taught by a different instructor in the same term as the treatment section ($n = 28$ students, offered spring quarter) and serving as the primary comparison class in the analyses that follow. The remaining comparison section was taught by the same instructor who implemented the modules, but in the term before receiving the modules ($n = 35$ students, offered winter quarter). This permitted a secondary within-instructor comparison of student performance in sections receiving versus not receiving the modules. Neither of the instructors in Study 2 were involved in the development of the scientific reasoning modules.

Procedure. As in Study 1, at the start of the term, the instructor of the treatment section was provided with a zip file containing separate folders with the teaching materials for each of the eight modules. Although all eight modules were provided, the instructor could choose how many and which modules to implement. The instructor chose to implement a single module on depression (clinical unit). Instructors for the comparison sections taught their classes as they normally would and made no changes for the quarter on the basis of the modules.

Results

Between-instructor comparison. In the primary comparison, which paralleled the design of Study 1, we compared gains in scientific reasoning for students in a treatment section versus a comparison section that was taught by a different instructor in the same term. Table 2 presents the percentage correct on the assessment at pre- and posttest for the between-instructor comparison in Study 2, for the entire assessment ("all skills") and for each of the four scientific reasoning subskills. Gain scores indicate change from pre- to posttest for each group along with results of the independent-sample t test comparing gain scores across groups.

Similar to the liberal arts college students in Study 1, students in the community college classes scored relatively poorly on the assessment at pretest, with the treatment and comparison groups well below 50% correct. Although prescores appeared lower overall relative to students in Study 1, students in the community college showed a similar relative pattern of performance across subskills, scoring best on the subskill of defining variables (~47% correct across groups) and worst on the subskill of graph interpretation (~28% correct across groups). The pattern of means indicates that gain scores for students in the comparison class were relatively flat across the overall assessment (2 percentage point decrease) and each subskill. In contrast, students in the treatment group showed an average gain of 14 percentage points on the overall assessment. Gains across subskills in the treatment group varied considerably, with increases of 10–52 percentage points on three of the four subskills but a relatively flat 1 percentage point increase on the correlational reasoning subskill.

Direct comparison of gain scores across the community college treatment and comparison groups supported the study predictions. As shown in Table 2 and Figure 2, students in the treatment section made significantly greater gains overall on the 14-item scientific reasoning assessment than those in the comparison section ($p < .01$, $d = +1.06$). To examine whether this

Table 2
Percentage Correct at Pre- and Posttest From Study 2 (Community College Setting), Separately for the Treatment and Comparison Groups for the Entire Assessment ("All Skills") As Well As the Four Subskills

Outcome measure	Treatment *M* (*SD*)	Control *M* (*SD*)	*t*(57)	*p*	*d*
All skills					
Pretest	35 (09)	40 (14)			
Posttest	49 (12)	39 (16)			
Gain	14 (14)	−2 (17)	3.99	.00	1.06*
Definitional					
Pretest	39 (44)	54 (45)			
Posttest	90 (27)	46 (47)			
Gain	52 (49)	−7 (40)	4.99	.00	1.32*
Correlational reasoning					
Pretest	35 (27)	44 (27)			
Posttest	37 (20)	46 (28)			
Gain	1 (30)	2 (41)	−0.14	.89	−0.04
Research design					
Pretest	35 (15)	42 (17)			
Posttest	46 (17)	39 (20)			
Gain	11 (19)	−2 (22)	2.48	.02	0.66*
Graph interpretation					
Pretest	30 (23)	26 (26)			
Posttest	40 (23)	25 (25)			
Gain	10 (34)	−1 (32)	1.27	.21	0.34

Note. Gain scores indicate change from pre- to posttest for each group, along with results of independent-sample t tests comparing gain scores across groups. Positive values of Cohen's d indicate greater gains for the treatment group, with asterisks indicating a significant difference between groups at the .05 significance level.

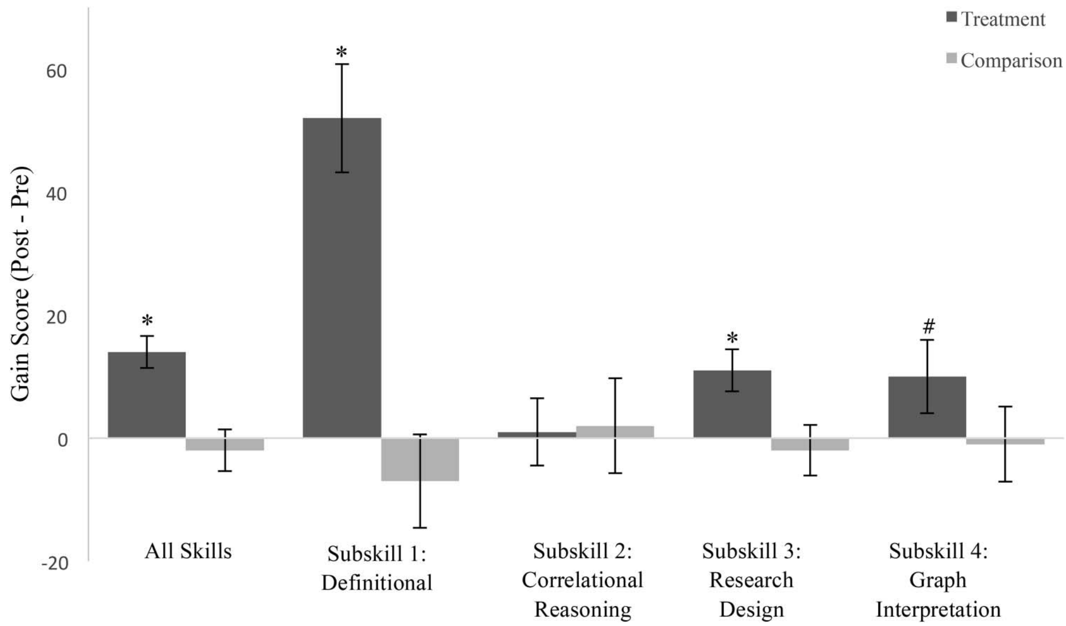

Figure 2. Gain score (posttest – pretest) from the between-instructor comparison in Study 2 (community college setting), separately for the treatment and comparison groups for the entire assessment ("all skills") as well as the four subskills. Error bars represent *SEM*, with asterisks indicating a significant difference in gain scores between groups at the .05 significance level and # indicating a trend toward differences in gain scores between groups at the .10 trend level.

pattern was specific to particular subskills of scientific reasoning, additional analyses separately compared gain scores between the treatment and comparison classes for each subskill. As shown in Table 2 and Figure 2, students in the treatment class made significantly greater gains than students in the comparison class in two of the four scientific reasoning subskills: defining variables ($p < .01$, $d = +1.32$) and reasoning about research design ($p < .05$, $d = +0.66$). For the subskill of graph interpretation, the pattern of means favored the treatment group with a similar effect size to Study 1, but with the smaller sample size this difference was not statistically significant ($p = .21$, $d = +0.34$). There were no significant differences between groups in the correlational reasoning subskill ($p = .89$, $d = -0.04$), with neither the treatment nor comparison group showing much movement from pre- to posttest (mean change close to zero for both groups, as shown in Table 2).

As in Study 1, supplementary regression analyses were conducted that controlled for pretest score in addition to direct statistical comparisons of the groups at pretest. The direct statistical comparison of pretest scores indicated no significant differences between groups on either the overall assessment or any of the four subskills (all $p > .05$). The supplementary regressions predicted gain score on each outcome from treatment condition (0 = *control*, 1 = *treatment*) while including pretest score on that outcome in the model. As in Study 1, in all of these analyses pretest scores negatively and significantly predicted growth in each outcome (i.e., those scoring lowest at pretest, who had the greatest room for improvement, also showed the largest gains). With respect to the effect of treatment condition on gain scores, the regression analyses largely paralleled those of the main analysis, with students in treatment classes making greater gains than students in the comparison classes on the overall assessment ($\beta = 0.36$, $p = .001$) and definitional subskill ($\beta = 0.46$, $p < .001$) but no significant effect of treatment condition on the correlational reasoning subskill ($\beta = 0.08$, $p = .12$). However, when controlling for pretest scores, the effect of treatment became statistically significant for the graph interpretation subskill ($\beta = 0.22$, $p < .05$) and went from significant to marginally significant for the research design subskill (β = 0.22, p = .056).

Within-instructor comparison. In addition to the between-instructor comparison, we were able to compare performance of students in the treatment section to students in a comparison section taught by the same instructor but in the term before receiving the modules. This additional, within-instructor comparison replicated the pattern of results in the between-instructor comparison. Relative to students taught by the same instructor, but before the instructor receiving the modules, students in the treatment section made greater gains on the overall scientific reasoning assessment, $t(64) = 4.17$, $p < .001$, $d = +1.05$. As in the main between-instructor comparison, these gains were significant for the definitional subskill, $t(64) = 4.88$, $p < .001$, $d = +1.24$, and for the research design subskill, $t(64) = 2.73$, $p < .01$, $d = +0.71$. No significant differences were observed between groups for either the graph interpretation subskill, $t(64) = 1.37$, $p = .18$, $d = +0.35$, or the correlational reasoning subskill ($t(64) < 1$, $p = .79$, $d = +0.07$).

Supplementary regression analyses were conducted that controlled for pretest score in addition to direct statistical comparisons of the groups at pretest. At pretest, there was only one difference between groups—in overall score on the scientific reasoning assessment—in which the control group (M = 41% correct) scored significantly higher than the treatment group (M = 35% correct), $t(64)$, = -2.17, $p < .05$. No other differences between groups were significant at pretest (all $p > .05$). However, the overall pattern of group differences favoring the treatment group remained the same if regression analyses were used to predict gain scores from treatment status (0 = *control*, 1 = *treatment*) while controlling for pretest score: overall assessment ($\beta = 0.29$, $p < .001$), definitional subskill ($\beta = 0.38$, $p < .001$), research design subskill ($\beta = 0.21$, $p < .05$), graph interpretation subskill ($\beta = 0.16$, $p = .12$), and correlational reasoning subskill ($\beta = -0.04$, $p = .67$).

Discussion

The results of Study 2 extend our findings on the value of scientific reasoning modules to a community college setting, in which class sizes were slightly larger and students' incoming scientific reasoning skills slightly lower. Students in a community college classroom in which a single scientific reasoning module was implemented showed greater gains in scientific reasoning than students in a comparison section, and the effect size of this difference was more than a standard deviation in magnitude. These gains were observed in a between-instructor comparison and a supplementary within-instructor comparison.

In the between-instructor comparison, community college students in the treatment section showed significantly larger gains than students in the comparison section in the subskills of defining variables ($d = +1.32$) and reasoning about research design and appropriate conclusions ($d = +0.66$). The treatment group also showed larger gains in interpreting data from graphs and tables ($d = +0.34$), but this difference was only statistically significant in the comparison that controlled for pretest scores. However, it is important to note that the effect size for the graph interpretation subskill was similar in magnitude to that observed in Study 1, in which effects were statistically significant with a larger sample size. Finally, contrary to Study 1, there was no evidence that students in the treatment group made greater gains on the subskill related to specifically reasoning about correlational research designs and results, with little evidence for change in this skill in either the treatment or comparison group. Results of the within-instructor comparison largely paralleled this pattern of results, indicating that results were robust even when students taught by the same instructor were compared in terms before versus after implementing a scientific reasoning module.

Summary and Concluding Discussion

Taken together, the two studies reported here provide strong evidence that targeted activities can be used to improve scientific reasoning skills in the introductory psychology course. Through the inclusion of one or more scientific reasoning modules, requiring only 30–45 min of class time each, introductory psychology instructors at a liberal arts college and a community college produced large gains in students' scientific reasoning skills relative to students in comparison sections taught as usual. Although the number of modules implemented by different instructors varied, data from the community college indicate that even a single exposure was associated with significantly greater gains in scientific reasoning compared with courses taught as usual. Across the two implementation sites, student gains were most robust in the subskills of reasoning about research design and appropriate conclusions and interpreting data from graphs and tables. Gains in the subskills of defining variables and reasoning specifically about correlational research design and results were observed only in one implementation site or the other, suggesting the need for further research on whether gains vary as a function of the number and type of modules used or incoming student or classroom characteristics.

It is unclear why a different pattern of subskill development was observed across the two studies. For the definitional subskill, greater gains in the treatment group were observed only in the community college sample. As noted earlier, the lack of effect for the liberal arts setting (Study 1) may be attributed to the relatively high pretest performance on this skill and/or the gains on this skill in the comparison sections. It is possible that a combination of higher incoming scores and coverage of this topic in traditional courses renders little added benefit to the modules in a liberal arts setting whereas students in the community college treatment classes showed large benefit in this subskill. For the correlational reasoning subskill, greater gains in the treatment group were observed only in the liberal arts sample. It is possible that this difference can be attributed to the number and type of modules implemented in each setting. Students in the community college treatment sample were exposed only to a single module, and our post hoc review of that module indicates that it does not specifically address correlational reasoning. However, we cannot rule out the possibility that this skill is more difficult to improve or more difficult to improve in a community college setting. Further research, using additional classes and testing a wider range of the modules, will be needed to help identify the pattern of gains across subskills and why that pattern may be uneven in different contexts.

The present study builds on previous research showing that targeted lessons or student projects can be used to improve aspects of scientific reasoning (Adam & Manson, 2014; Blessing & Blessing, 2010; Stevens & Witkow, 2014). However, the present study takes a novel approach by distributing the modules to instructors not involved in the development of the teaching materials and assessing materials in a liberal arts and community college setting. This suggests that the scientific reasoning activities are flexible enough to be adapted by different instructors, teaching at different types of institutions. Indeed, a key feature of the activities was their modular nature, with instructors able to choose how many and which modules to implement during the term. Each module was designed to stand alone, but several were available so that instructors would have the ability to embed scientific reasoning training across different course content.

The inclusion of several modules, which could be integrated throughout the term, may have been an important component of the curriculum. This may be particularly useful given that a recent review of introductory syllabi indicated that less than 5% of lecture time is devoted specifically to research methods despite two thirds of syllabi listing research methods as a key student learning outcome for the course (Homa et al., 2013). Because repeated exposure to skill-based concepts is recognized to be critical to effective training of higher-order and scientific reasoning skills (Halpern, 1999, 2001; Willingham, 2007), these modules provide a means for instructors to ensure scientific reasoning is not treated as a "one-off" lesson as part of the course. Indeed, the modules provided a means for instructors to provide more time on research methods and distributed practice throughout the term, either or both of which might be key to the favorable results found here. Furthermore, because the activities all include applications of psychology, the modules provide an opportunity to meet APA task force recommendations for including cross-cutting themes, such as the application of psychology, across different content domains (APA, 2014; Gurung et al., 2016). Some of the specific questions students addressed as parts of the modules, such as critiquing claims, evaluating a research study, and designing possible follow-up studies, are also well aligned with key tasks recommended in the APA Summit on National Assessment of Psychology for addressing learning goals related to scientific inquiry and critical thinking (Mueller et al., 2016).

Introductory psychology remains one of the most heavily subscribed courses at colleges and universities nationwide, second only to general composition courses in enrollment (Goldstein, 2010). Although students who major in psychology typically take one or more dedicated research methods courses as part of their degree curriculum (Norcross et al., 2016), for the vast majority of introductory students who do not major in psychology, introductory psychology may serve as their first and only exposure to critical thinking about research methods. As such, any efforts in the introductory course to improve training in scientific reasoning can have an important impact on goals for improving general scientific literacy skills in the broader population (American Association for the Advancement of Science, 2009; Mueller et al., 2016).

Limitations and Future Directions

The present study used an assessment focused on specific scientific reasoning skills, drawing largely from passage-based, multiple-choice scientific reasoning questions from the MCAT preview guide (Association of American Medical Colleges, 2011) and a textbook test bank (Schacter et al., 2014). This allowed us to create a brief assessment requiring minimal intrusion on class time, focused on the specific aspects of scientific reasoning targeted in the modules, while retaining strong face validity and objective scoring. However, other assessments of critical thinking are available, some of which the APA has noted could be used to assess learning outcomes related to scientific inquiry and critical thinking (see Table 2 in APA, 2013). At the same time, most of these other assessments are broader in scope than the assessment used in the present study, and many require fees and/or extensive time to administer and score because of the open-ended nature of questions. However, future studies could explore whether the modules used here also lead to improvements on other measures of scientific reasoning.

Another limitation of the present research is that instructors were not randomly assigned to the treatment and comparison conditions. Thus, it is difficult to rule out the possibility that there is something about instructors who volunteer to try new curricula that leads to greater gains in scientific reasoning for their students rather than the activities themselves. However, it is noteworthy that in the community college setting we were also able to demonstrate gains in scientific reasoning when data were compared within the same instructor (i.e., when comparing student performance on scientific reasoning questions in a term before vs. after implementing the scientific reasoning modules). This suggests that the modules have added value, even for instructors who are willing to implement new curricula. In addition, in prior data presented at a national conference, we have shown similar within-instructor benefits of adding the scientific reasoning modules for instructors at a liberal arts college using a smaller, posttest-only scientific reasoning assessment (Stevens, Witkow, Laughlin, & Yankelevitz, 2016). This suggests that, at least for those instructors willing to implement the materials, the modules described here present a means for improving student learning outcomes in areas of scientific reasoning. However, because only one educational innovation was tested, we cannot assess the degree to which the gains were due to this specific curriculum as opposed to general gains that might be seen whenever an instructor tries something new.

An interesting question, which could not be addressed in the present study, is the possibility of a dose–response function, or that student gains are larger as more modules are implemented during the term. In future research involving more instructors and classes, this question might become tractable. However, in the present study, the number of modules implemented is confounded by institution type and instructor such that this analysis was not attempted. It will also be important to test additional modules in community college settings because only one module was implemented in the community college treatment class.

Finally, although the present study included two diverse settings (community college and liberal arts college), it is unclear the extent to which these activities might be adaptable or generalizable to other environments in which the introductory course is typically taught, including large lecture halls or online environments. The modular activities described here are designed for use in smaller-sized classes where small group and whole class discussion are possible, but they will need to be adapted if they are to be useful in other types of introductory psychology contexts. In ongoing work, we are working to create adaptations of the materials for large lecture and online environments and to test their effectiveness in these settings.

Conclusion

Training scientific reasoning, in introductory psychology and more generally in first-year science courses, remains an important goal of higher education (American Association for the Advancement of Science, 2009; APA, 2013; National Research Council, 2011, 2012). The present study provides a rigorous assessment of materials and demonstrates that these skills can be improved in students when specific, targeted activities are included in the introductory psychology course. The ability of instructors not involved in curriculum development to use the materials, and to realize gains in scientific thinking for their students, provides a promising start for ongoing efforts to improve the training of scientific thinking in introductory psychology courses.

References

Adam, A., & Manson, T. M. (2014). Using a pseudoscience activity to teach critical thinking. *Teaching of Psychology, 41,* 130–134. http://dx.doi.org/10.1177/0098628314530343

American Association for the Advancement of Science. (2009). *Vision and change in education biology education: A call to action*. Washington, DC: Author.

American Psychological Association. (2011). *APA principles for quality undergraduate education in psychology*. Washington, DC: Author.

American Psychological Association. (2013). *APA guidelines for the undergraduate psychology major: Version 2.0*. Retrieved from http://www.apa.org/ed/precollege/about/psymajor-guidelines.pdf

American Psychological Association. (2014). *Strengthening the common core of the Introductory Psychology course*. Washington, DC: American Psychological Association, Board of Educational Affairs. Retrieved from https://www.apa.org/ed/governance/bea/intro-psych-report.pdf

American Psychological Association. (2016). Guidelines for the undergraduate psychology major: Version 2.0. *American Psychologist, 71,* 102–111. http://dx.doi.org/10.1037/a0037562

Association of American Medical Colleges. (2011). *MCAT 2015: A better test for tomorrow's doctors. Preview guide for MCAT 2015*. Washington, DC: Author.

Association of American Medical Colleges. (2012). *MCAT2015: A better test for tomorrow's doctors*. Preview guide for the MCAT2015 exam (2nd ed.). Washington, DC: Author.

Blessing, S., & Blessing, J. (2010). PsychBusters: A means of fostering critical thinking in the introductory course. *Teaching of Psychology, 37,* 178–182. http://dx.doi.org/10.1080/00986283.2010.488540

Frazer, N. B., & Twohig, M. P. (2012). The new social and behavioral requirements: Inspired innovation, missed opportunity, or both? *Peer Review: Emerging Trends and Key Debates in Undergraduate Education, 14,* 14–17.

Gareis, K. C. (1995). Critiquing articles cited in the introductory textbook: A writing assignment. *Teaching of Psychology, 22,* 233–235. http://dx.doi.org/10.1207/s15328023top2204_4

Goldstein, R. (2010). Major developments in undergraduate psychology. *APS Observer, 23,* 1–4.

Gurung, R. A. R., Hackathorn, J., Enns, C., Frantz, S., Cacioppo, J. T., Loop, T., & Freeman, J. E. (2016). Strengthening introductory psychology: A new model for teaching the introductory course. *American Psychologist, 71,* 112–124. http://dx.doi.org/10.1037/a0040012

Halpern, D. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning, 1999,* 69–74. http://dx.doi.org/10.1002/tl.8005

Halpern, D. (2001). Assessing the effectiveness of critical thinking instruction. *The Journal of General Education, 50,* 270–286. http://dx.doi.org/10.1353/jge.2001.0024

Holmes, J. D. (2008). An introductory exercise for promoting critical thinking about psychological measurement. In D. S. Dunn, J. S. Halonen, & R. A. Smith (Eds.), *Teaching critical thinking in psychology* (pp. 277–280). Malden, MA: Wiley-Blackwell. http://dx.doi.org/10.1002/9781444305173.ch28

Homa, N., Hackathorn, J., Brown, C. M., Garczynski, A., Solomon, E. D., Tennial, R., . . . Gurung, R. A. R. (2013). An analysis of learning objectives and content coverage in Introductory Psychology syllabi. *Teaching of Psychology, 40,* 169–174. http://dx.doi.org/10.1177/0098628313487456

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., . . . Dunlop Velez, E. (2016). *The condition of education 2016* (NCES 2016–144). Washington, DC. Retrieved from http://nces.ed.gov/pubsearch

Lutsky, N. (2006). Teaching quantitative reasoning: How to make psychology statistically significant. *American Psychological Society Observer, 19,* 35–38.

Mitchell, K., Lewis, R. S., Satterfield, J., & Hong, B. A. (2016). The new Medical College Admission Test: Implications for teaching psychology. *American Psychologist, 71,* 125–135. http://dx.doi.org/10.1037/a0039975

Mueller, J., Kelly, K., Taylor, H., Brakke, K., Levine, G., Drysdale, M., & Ronquillo-Adachi, J. (2016). *Executive summary goal 2: Scientific inquiry and critical thinking*. Green Bay, WI: APA Summit on National Assessment in Psychology. Retrieved from http://www.apa.org/ed/precollege/undergrad/summary-scientific-inquiry.pdf

National Research Council. (2011). *Promising practices in undergraduate science, technology, engineering, and mathematics education. Summary of two workshops*. Washington DC: Author.

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

Nolan, S. A., & Heinzen, T. E. (2009). Graphing literacy in the psychology major: Florence Nightingale and the creation of a beautiful display of data. *American Psychological Society Observer, 22,* 33–36.

Norcross, J. C., Hailstorks, R., Aiken, L. S., Pfund, R. A., Stamm, K. E., & Christidis, P. (2016). Undergraduate study in psychology: Curriculum and assessment. *American Psychologist, 71,* 89–101. http://dx.doi.org/10.1037/a0040095

Penningroth, S., Despain, L., & Gray, M. (2007). A course designed to improve psychological critical thinking. *Teaching of Psychology, 34,* 153–157. http://dx.doi.org/10.1080/00986280701498509

Schacter, D., Gilbert, D., Wegner, D., & Nock, M. (2014). *Psychology* (3rd ed.). New York, NY: MacMillan.

Stevens, C., & Witkow, M. R. (2014). Training scientific thinking skills: Evidence from an MCAT[2015] aligned classroom module. *Teaching of Psychology, 41,* 115–121. http://dx.doi.org/10.1177/0098628314530341

Stevens, C., Witkow, M., Laughlin, R., & Yankelevitz, R. L. (2016). *A within-instructor evaluation of three scientific thinking modules: Description of activities and evaluation*. Poster presented at the annual meeting of the National Institute for the Teaching of Psychology, St. Pete Beach, FL.

Thieman, T. J., Clary, E. G., Olson, A. M., Dauner, R. C., & Ring, E. E. (2009). Introducing students to psychological research: General psychology as a laboratory course. *Teaching of Psychology, 36,* 160–168. http://dx.doi.org/10.1080/00986280902959994

Wade, C. (1995). Using writing to develop and assess critical thinking. *Teaching of Psychology, 22,* 24–28. http://dx.doi.org/10.1207/s15328023top2201_8

Willingham, D. (2007, Summer). Critical thinking: Why is it so hard to teach? *American Educator,* 9–19.

Received July 29, 2016
Revision received October 2, 2016
Accepted October 7, 2016 ■