Research Experience in Recitation Sections Within Introductory Psychology Courses: The Effectiveness of a Semester-Long Laboratory Component[₸]

Jared G Branch,* and Eric F Dubow

Department of Psychology, Social Sciences & Behavioral Sciences Division

*Corresponding author. Email address: branchj@bgsu.edu. (801) 645-1877

[₸]Accepted for publication and forthcoming in Scholarship of Teaching and Learning in Psychology

Abstract

Undergraduate psychology students struggle to think critically, are hesitant to endorse psychology as a scientific field, and believe scientifically disproven psychological myths. We asked if a semester-long laboratory component (small weekly recitation groups), embedded within a large introductory psychology course, could improve students' critical thinking, perceptions of psychology as a science, and ability to identify psychological myths. In the recitation groups (N = 124), students researched a topic, designed a study, collected data, and presented findings. We compared this format to two control groups employing similar activities: one with a traditional recitation component of discussion of material presented within the large lecture (N = 36), and one with a traditional recitation component coupled with a behavior modification project that students designed halfway through the semester (N = 133). The three large courses were taught by separate instructors, and recitation groups were led by graduate teaching assistants, but otherwise the sections were structured to maximize equivalency. We administered a pretest at the beginning of the semester and a posttest at the end of the semester, and found that our intervention was successful in increasing critical thinking ($\eta_p^2 = .065$); however, across all three conditions, at the posttest students were less likely to view psychology as a science ($\eta_p^2 = .035$).

Keywords: critical thinking, introductory psychology, research experience, lab, psychological myths

Research Experience in Recitation Sections Within Introductory Psychology Courses: The Effectiveness of a Semester-Long Laboratory Component

The general public tends to be skeptical of the view that psychology constitutes a science (Lilienfeld, 2012). Student regard for the scientific nature of psychology is likewise low and does not improve with progression through the undergraduate career (Holmes & Beins, 2009). At the same time, endorsement of psychological myths is high (Hughes, Lyddy, & Lambe, 2013), and students struggle to think critically (Wentworth & Whitmarsh, 2017).

In this study, we asked if a semester-long intervention embedded within a large introductory psychology course, in the form of a student-initiated research component facilitated by graduate student teaching assistants, could improve students' ability to think critically and their perception of psychology as a science, and decrease their endorsement of psychological myths. We hypothesized that if students were able to critically engage with psychological research as psychological scientists, they would show demonstrable improvements in their ability to critically evaluate scientific findings and claims, and better understand the scientific nature of psychology. With the popularity of introductory psychology among undergraduate students and the number of students enrolled annually, the course offers an opportunity for widespread improvement in critical thinking and belief in psychology as a science, and reduction in endorsement of psychological myths.

The APA Guidelines for the undergraduate major (APA, 2013) align with the goals of our study. The Guidelines stress the scientific nature of psychology and hold as a major goal the ability to think critically and use psychological principles in scientific inquiry. The Guidelines further distinguish between *foundation* knowledge, obtained by completing lower-level courses (e.g., introductory psychology), and *baccalaureate* knowledge (i.e., obtained after completing several psychology courses). At the *foundation* level of knowledge, students should adopt a "psychological worldview on how they think about behavior" (p. 3); but students have not yet progressed to the level of applying "scientific principles more systematically to describe, explain, and predict behavior" (p. 4), which is evidently learned at the *baccalaureate* level.

Indeed, as students advance through the major, they show an increased belief that psychology is a science, greater psychological knowledge, and fewer endorsements of psychological myths and misconceptions (Landrum, Gurung, & Amsel, 2019). But what about those students who do not declare psychology as a major? They are at a significant disadvantage to their psychology-major counterparts, who receive *baccalaureate* training in thinking about the world from a scientific standpoint. If introductory psychology courses do not address the scientific nature of psychology, after having completed the course students may continue to struggle to think scientifically and critically about psychological findings.

Two major overhauls to how introductory psychology is taught have been proposed (for an alternative approach to teaching a general education science course, see also Rowe et al. 2015). Bernstein (2017) suggested renaming the course "Myths and Illusions about Human Behavior in Everyday Life," and focusing on rectifying the high rate of student endorsement of psychological myths through active learning techniques. Gurung et al. (2016) suggested, among other changes, covering research methods and including research experiences, but note that more research is required to determine what specific types of research should be offered (see also Gurung & Hackathorn, 2018).

A limited number of studies have examined the appropriate level of research experience and instructor scaffolding to improve learning outcomes for students enrolled in an introductory psychology course, with mixed results. For instance, Lewis (2015) compared a virtual rat laboratory with human demonstrations and found no differences on a class exam. Students in an introductory psychology course who observed public behaviors to test hypotheses (e.g., high-BMI people would eat more at a buffet) showed an overall decrease in their understanding of research methods at the end of the semester relative to the beginning (Downey, 2013). However, Koschmann and Wasp (2001), who also used a dining facility to allow students to conduct observational research, found students had an increased understanding of the scientific method.

In one notable study, Thieman, Clary, Olson, Dauner, and Ring (2009) implemented a semester-long laboratory component within their introductory psychology courses. Students completed five disparate research activities: conducting a literature review through PsycINFO and summarizing abstracts, conducting a structured observation of the behaviors of videotaped children, testing of memory phenomena such as the serial position effect, conducting correlational analyses of heart and respiration rates and scores on a self-report stress inventory, and conducting exploratory correlational analyses on a set of archival data. At the end of the semester, students had significantly improved in their ability to think critically.

Laboratory and other types of active-learning, hands-on experiences have been observed to increase critical thinking skills among students enrolled in courses in biology (Reinbold, 2018) and molecular biology (Felzien, 2016); nursing (Goodstone et al., 2013); chemistry (Quattrucci, 2018) and organic chemistry (Weaver, Samoshin, Lewis, & Gainer, 2016); nutrition (Iwaoka, Li, & Rhee, 2010); and dentistry (Rowland & Joy, 2015).

In spite of APA recommendations (APA, 2013) and improved learning outcomes, only 5% of introductory psychology courses implement a laboratory component, instead relying almost exclusively on lecturing (Peterson & Sesma, 2017). The introductory psychology program offered at our university is a 4-credit hour class, with three hours spent weekly in a large auditorium-style lecture commonly enrolling 150 – 250 students, and one hour spent in weekly breakout recitation sessions usually capped at 30 students. These recitation sessions are led by graduate student teaching assistants, each of whom are assigned to a particular session of a class for the entirety of the semester (a given graduate student conducts 2 - 3 recitation sections), who generally conduct class by discussing lecture material in greater detail and providing one-on-one opportunities for student learning.

We asked if instituting a semester-long intervention in the form of a laboratory component, in lieu of the discussion-based recitation component, would improve student outcomes. We hypothesized that, in comparison to a traditional recitation, and in comparison to one that was part-traditional and part-experiential (i.e., conducting a behavior modification project), students in the laboratory recitation format would (1) improve more in their ability to think critically, and (2) increase more in their perceptions of psychology as a science.

Critical thinking relates to scientific thinking in that they both require learning the standards for one's field and applying those standards to claims and research findings (Bensley, 2018). The terms are often used interchangeably (e.g., Gurung et al., 2018), as thinking like a psychological scientist is operationalized as improved critical thinking (Wentworth & Whitmarsh, 2018). Thus, we hypothesized that engaging in psychological research would lead to both improvements in critical thinking and greater endorsement of psychology as a science.

We also asked if a laboratory component would decrease endorsement of psychological myths. Critical thinking relates to skepticism in that those students who think critically are less likely to endorse psychological myths and misconceptions (Bensley, 2018; see Hughes, Lyddy, & Lambe, 2013 for a discussion). If a curriculum directly addresses psychological myths, students are less likely to endorse them (LaCaille, 2015; McCarthy & Frantz, 2016). For instance, creating a poster that emphasizes the correct information, rather than the myth, results in the endorsement of fewer myths, both for the students who create the poster and the students who view the poster (LaCaille, 2015). However, exposure to myths may lead to a sense of fluency, thereby increasing myth endorsement (e.g., Schwarz, Sanna, Skurnik, & Yoon, 2007). We note that we did not directly address myths in our recitation groups.

We compared the pretest and posttest data of students enrolled in three introductory psychology courses. In small groups, students in the first author's introductory psychology course completed a semester-long research project which culminated in an APA-style paper and presentation of research findings. Students in the second author's introductory psychology course completed the recitation portion as usual (i.e., as a discussion component), with an added component of a behavior modification project halfway through the course (i.e., using operant conditioning to decrease or increase a personal target behavior). Students in a third introductory psychology course were exposed to the traditional recitation component of review and discussion of in-class material in more depth.

Method

Participants

A total of 293 (117 male, 176 female; mean age 19.08, *SD* = 1.41, range 18 to 31) students enrolled in Bowling Green State University introductory psychology courses in the fall semester of 2018 completed both the pretest and posttest assessment. Of the 148 students enrolled in the laboratory recitation group, 124 completed the study (84%). Of the 216 students enrolled in the traditional plus behavior modification project group, 133 completed the study (62%). Of the 104 students enrolled in the traditional recitation group, 36 completed the study (35%)

Materials

Critical thinking was measured using the Psychological Critical Thinking Exam (PCTE) (Lawson, Jordan-Fleming, & Bodle, 2015). The PCTE includes questions related to (1) an event occurring by chance, (2) the use of a control group, (3) concluding causation based on correlational data, (4) generalizing findings from an unrepresentative sample, (5) biased questioning, (6) the inability to falsify hypotheses, and (7) claims to finding *the* cause of a behavior. Because the PCTE includes 2 questions for each of the 7 domains, we counterbalanced the presentation of questions between participants. Questions were answered as a short, free response. In scoring the PCTE, we utilized the coding scheme developed by Lawson, Jordan-Fleming, and Bodle (2015), with 0 indicating no problem identified by the student, 1 indicating a problem recognized but misidentified, 2 indicating the student identified the main problem but also mentioned less relevant problems, and 3 indicating the student identified only the main problem. As such, the total possible score for any one questionnaire was 21. The PCTE was coded by three research assistants who each coded two conditions. Inter-rater reliability was assessed using intraclass correlation coefficient: Across pairs of raters, the average measure ICC ranged from .788 to .947.

Perception of psychology as a science was measured using the Psychology as a Science (PAS) scale (Friedrich, 1996). The PAS includes items related to conducting research (e.g. "It's just as important for psychology students to do experiments as it is for students in chemistry and biology") and the nature of psychology (e.g. "The study of psychology should be seen primarily as a science") Psychological myths were obtained from Hughes, Lyddy, and Lambe (2013), and included questions about topics commonly covered in introductory psychology courses, including neuroscience (e.g. "People only use 10% of their total brain power"), perception ("We see by emitting rays from our eyes that reflect off objects"), disorders ("People with schizophrenia have multiple personalities"), development ("Playing classical music to infants increases their intelligence"), and memory ("Human memory works like a tape recorder or video camera). Responses to both the PAS and psychological myths were solicited via 7-point Likert scales, with 7 indicating strong agreement, 1 indicating strong disagreement, and 4 neither agreeing nor disagreeing. The PAS is a 15-item questionnaire, as such students could score 105 if they strongly agreed that psychology is a science. Students viewed 10 myths, as such students could score 70 if they strongly endorsed psychological myths.

Both the laboratory and traditional plus behavior modification project classes utilized Psychology: Fourth Edition (Schacter, Gilbert, Nock, & Wegner, 2017), while the traditional group utilized Psychology, 5th Edition (Ciccarelli & White, 2017). Because using different texts across classes could be a confound, we analyzed the Flesch-Kincaid level of a portion of the sample chapter freely available through the publisher's website and found that both textbooks were written at a similar grade level (Schacter et al. (2017) was written at an 11.4 grade level, and Ciccarelli and White (2017) was written at an 11.6 grade level). All instructors taught neuroscience, learning, development, social psychology, motivation, emotion, and stress and health. Instructors in the laboratory condition and traditional plus behavior modification condition also taught methods, memory, sensation and perception, personality, and consciousness. Instructors in the traditional plus behavior modification condition and the traditional condition also taught psychological disorders and treatment/therapies. Overall, the material taught in the traditional plus behavior modification condition overlapped with 83% of the material taught in the laboratory condition, and the material taught in the traditional condition overlapped with 75% of the material taught in the laboratory condition.

Procedures

Students enrolled across the three introductory psychology courses completed a pretest assessment during the first week of the 15-week semester and a posttest assessment during the last week of the semester. The pretest and posttest assessments were identical, except that the PCTE was divided into two forms of 7 questions each and presentation counterbalanced across assessments, such that some students saw form A as a pretest assessment and form B as a posttest assessment, whereas others saw form B as a pretest assessment and form A as a posttest assessment. Students first responded to demographics information before completing the PAS scale, psychological myths, and the PCTE. In order to administer the assessment, a trained research assistant administered the survey using Qualtrics survey software during class time, during which the instructor left the room. Students were periodically surveyed by the research assistant to determine if they had finished, and all participants had done so after approximately 30 minutes. Students were told that they were not required to complete the assessment but received partial fulfillment of class requirements if they elected to do so. (These class requirements could be met by completing other research projects as well.) To allow any student to participate who did not attend class that day, the instructor also posted a link to the survey using a Canvas announcement.

Laboratory Recitation Groups. The recitation session met once a week for a semester. Given the length of our semester as well as school holidays, students successfully completed the laboratory component over 13 class meetings. Within each recitation section of 25 - 30 students, students were assigned to semester-long small groups of 5 to 6 students.

Literature Review. During the first week students were given a short lecture on how to write an APA-style literature review. Next, each individual group identified a topic they wanted to research further. Each student within the group was tasked with searching the PsycINFO database for a relevant article and summarizing the main sections. Students then verbally explained their findings to their group members, and each group created a literature review via a shared Google Doc. Student groups were free to choose their own topics of interest, which were quite diverse and included topics such as the effects of cannabis on memory and how music effects the retention of material. The *literature review* component took 3, 50-minute recitation classes to complete.

Methods. Next, students received instruction on research methods and experimental design. The recitation leader briefly discussed such topics as independent and dependent variables, and operationalization of variables. These topics were also briefly covered in the lecture portion of class. Students were then given a worksheet which tasked them with designing both a correlational and experimental study to empirically explore their group's topic. Students identified their hypothesis, procedure, variables, the population from which they would draw their sample, how their variables would be operationally defined, and what they expected their results to be. After completing this task, students searched the PsycINFO database for a standardized measure to incorporate into their research design. Finally, students completed a mock Institutional Review Board (IRB) form and exchanged their form with a second group. Students were tasked with identifying any ethical problems present in another group's proposed procedure. The *methods* component took 3, 50-minute recitation classes to complete.

Data Collection. Students designed a data collection tool, with many groups opting to use freely available online questionnaire software. The next recitation session was excused so that students could use this time to collect their data using students on campus as participants.

Students were instructed on how to approach participants and obtain informed consent. The *data collection* component took 2, 50-minute recitation classes to complete.

Results & Discussion. Students structured their data appropriately and the recitation leaders analyzed the data using the appropriate statistical techniques and discussed with each group how the statistical test worked and what the results meant using plain language. Students added the results to their APA-style paper, as well as a discussion section including limitations and future directions. The *results & discussion* component took 2, 50-minute classes to complete.

Presentations & Final Papers. Students were instructed on how to present their research question and findings. Students verbally presented their findings, within groups, to their classmates. They also completed an APA style report, including a literature review, methods and results, discussion, and reference list. The *presentations and final papers,* which included finalizing the APA report, took 3, 50-minute classes to complete.

Traditional Plus Behavior Modification Project and Traditional Components. Students enrolled in the traditional plus behavior modification project recitation component and the traditional recitation component completed the recitation component as normal: Recitation leaders commonly lead discussions, give short lectures, administer worksheets, and help students review for exams. However, in the traditional plus behavior modification project, beginning half-way through the semester, just after the "learning" unit was presented in the large lecture, students chose a personal behavior they wanted to improve (e.g., study skills) or reduce in frequency (e.g., smoking cigarettes); identified reinforcement or punishments they could use to achieve the goal; tracked their behavior for a baseline week and for two intervention weeks; and described the results of their intervention. This project was modeled after Grasha (1987) and the course instructor has utilized this behavioral modification since prior to the study in order that students better understand operant conditioning. Instruction for this project was accomplished for approximately 5-10 minutes in each of the last 6 recitation sections; otherwise a traditional format for the remainder of these recitations sections was followed.

Results

Pretest Differences among Groups in Demographic and Major Study Variables

Across the three conditions, students were similar in their age and gender (all *p*s > .05). However, a one-way non-parametric ANOVA showed a significant difference in student's year in school, $\chi^2(2) = .764$, $p = .033$. We conducted a series of non-parametric ANOVAs, utilizing an ordinal data coding scheme such that a "1" corresponded to freshman and "4" corresponded to senior, and found that students in the laboratory group ($M = 1.69$, $SD = .814$) were significantly further in their schooling compared to those in the traditional plus behavior modification group ($M = 1.48$, $SD = .718$; $p = .044$) and the traditional group ($M = 1.37$, $SD = .646$; $p = .031$).

Students within the three conditions were similar to each other at the beginning of the semester in respect to their endorsement of psychological myths, $F(2, 286) = .570$, $p =.566$, and critical thinking, $F(2, 234) = 1.25$, $p = .289$, but not endorsement of psychology as a science, $F(2, 290) = 8.707$, $p < .001$, with post hoc Tukey tests showing that students in the traditional plus behavior modification group ($M = 74.0$, $SD = 6.21$) scored higher than students in the laboratory group ($M = 70.9$, $SD = 7.44$; $p = .001$), and scored higher than students in the traditional group ($M = 69.8$, $SD = 6.73$; $p = .004$).

Group Differences from Pretest to Posttest

To ascertain semester-long improvement, we conducted three repeated-measures analyses of variance (one for each of the three dependent variables) with the pretest and posttest data as within-subjects factors and condition (laboratory group, traditional plus behavior modification group, traditional group) as a between-subjects factor.

Not all students completed all aspects of the questionnaire (i.e. critical thinking, psychology as a science, and psychological myths), and as such were excluded from the respective analysis if they failed to complete either the pretest or posttest. The *critical thinking* measure required a written response, and fewer students completed it compared to the *myths* and *psychology as a science* measures, which employed a Likert response. Of the 141 students in the laboratory group who completed the pretest, 124 completed the posttest (88%). All 124 of these students completed both the *myths* and *psychology as a science* measure, while only 102 completed the *critical thinking* measure. Of the 184 students in the traditional plus behavior modification group who completed the pretest, 133 completed the posttest (72%). All 133 of these students completed the *psychology as a science* measure, 130 completed the *myths* measure, and 114 completed the *critical thinking* measure. Of the 88 students in the traditional group who completed the pretest, 36 completed the posttest (41%). All 36 of these students completed the *psychology as a science* measure, 35 completed the *myths* measure, and 21 completed the *critical thinking* measure. Given this attrition rate, we compared students who completed both the pretest and posttest with students who completed just the pretest. Although we did not collect the majority of the demographic information until the posttest, we found that students who completed the posttest were not significantly different from students who completed only the pretest in terms of year in school, endorsement of myths, and perception of psychology as a science (all ps > .289).

For critical thinking, there was no significant main effect for time, $F(1, 234) = .005$, $p = .944$, or condition, $F(2, 234) = .246$, $p = .782$, but, as predicted, there was a significant interaction, $F(2, 234) = 8.155$, $p < .001$, $\eta_p^2 = .065$. Paired-samples *t* tests showed that scores in the laboratory condition significantly improved (pretest $M = 6.02$, $SD = 3.291$; posttest $M = 7.41$, $SD = 3.941$; $t(101) = -4.019$, $p < .001$), scores in the traditional plus behavior modification condition were unchanged (pretest $M = 6.25$, $SD = 3.22$; posttest $M = 6.80$, $SD = 3.938$; $t(113) = -1.593$, $p = .114$), and scores in the traditional condition significantly declined (pretest $M = 7.24$, $SD = 2.897$; posttest $M = 5.24$, $SD = 2.682$; $t(20) = 2.913$, $p = .009$). Because groups were not equivalent on year in school, we also conducted a repeated-measures ANOVA with year in school as a covariate and obtained similar results: a non-significant main effect of time ($p = .454$) and condition ($p = .943$) and a significant interaction, $F(2, 231) = 8.491$, $p < .001$, $\eta_p^2 = .068$. For a visualization of the results, see Figure 1.

For perceptions of psychology as a science, there was a significant main effect for time, $F(1, 290) = 8.19$, $p = .005$, $\eta_p^2 = .027$, with participants being less likely to endorse psychology as a science at the posttest (M = 70.5, SD = 7.96) relative to the pretest (M = 72.2, SD = 6.99). There was a main effect for condition, $F(2, 290) = 5.20$, $p = .006$, $\eta_p^2 = .035$. Adjustments for multiple comparisons were made using Bonferroni correction, and indicated that collapsed across time, students in the traditional plus behavior modification condition ($M = 72.62$, $SE = .562$) were more likely to endorse psychology as a science relative to the laboratory condition ($M = 70.42$, $SE = .582$; $p = .021$) or traditional condition ($M = 69.56$, $SE = 1.081$; $p = .037$). The interaction of time and condition was not significant, $F(2, 290) = 2.23$, $p = .110$, $\eta_p^2 = .015$. We conducted a series of paired samples *t* tests and found that students in the laboratory condition were unchanged on the posttest ($M = 69.9$, $SD = 8.06$) relative to the pretest ($M = 70.9$, $SD = 7.44$), $p = .107$, as were students in the traditional condition (pretest $M = 69.8$, $SD = 6.73$; posttest $M = 69.3$, $SD = 7.02$; $p = .679$), whereas students in the traditional plus behavioral modification condition showed a significant decrease (pretest $M = 74.0$, $SD = 6.21$; posttest $M = 71.3$, $SD = 8.07$; $p < .001$). Because groups were not equivalent in their pretest means for perceptions of psychology as a science and year in school, we also conducted an ANCOVA with the psychology as a science post-test as the dependent variable, condition as the fixed factor, and the psychology as a science pre-test and year in school as covariates, and found a non-significant effect of condition, $F(2, 282) = .4850$, $p = .616$.

For endorsement of psychological myths, the main effect of time approached significance, with students endorsing fewer myths at the posttest ($M = 37.1$, $SD = 8.59$) compared to the pretest ($M = 37.8$, $SD = 8.03$), $F(1, 286) = 3.590$, $p = .059$, $\eta_p^2 = .012$. There was no significant main effect for condition, $F(2, 286) = .367$, $p = .693$, and there was no significant interaction effect, $F(2, 286) = .191$, $p = .826$. Because groups were not equivalent on year in school, we also conducted an ANCOVA with the endorsement of psychological myths post-test score as the dependent variable, condition as the fixed factor, and the endorsement of psychological myths pre-test and year in school as covariates, and found a non-significant effect of condition, $F(2, 282) = .0377$, $p = .963$.

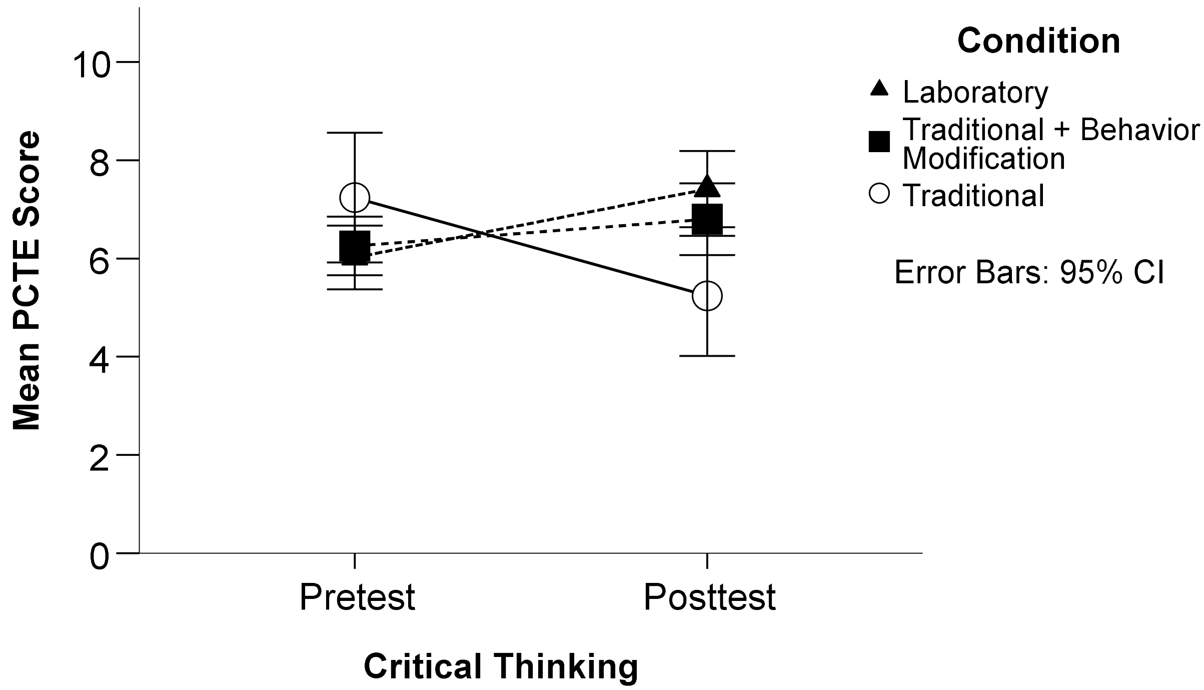

Figure 1. The mean PCTE score on the pretest and posttest, as a function of condition. The total possible score for either the pretest or posttest is 21, indicating that the student correctly identified each problem over 7 questions.

Discussion

We found that a laboratory component recitation section embedded within a large introductory psychology course was more effective in improving critical thinking skills compared to traditional recitation formats. The laboratory intervention focused on teaching critical research tasks, rather than critical thinking skills per se, but was effective in turning psychological *students* into psychological *scientists*, in that students thought more critically about scientific claims.

Students, regardless of condition, also tended to endorse fewer psychological myths at the end of the semester. Across the three courses, none of the instructors directly addressed common psychological myths. Given that skepticism is often conceptualized as a hallmark of a good scientist (Normand, 2008), our results bode positively for introductory psychology courses in general as a means of reducing the belief in psychological myths. However, decreasing endorsement of psychological myths was not a benefit of engaging in psychological research. These findings are consistent with prior research showing that, in order to decrease belief in myths, a curriculum must directly address them, and do so in a way that allows students to think critically about them so as not to paradoxically increase their endorsement (LaCaille, 2015; Schwarz et al., 2007). As well, students further in their undergraduate careers showed a greater ability to think critically. These findings suggest that, in general, higher education effectively increases critical thinking (see also Landrum et al., 2019)

According to the American Psychological Association (APA, 2013; APA, 2014) and others' recommendations (e.g., Gurung et al., 2016), instructors of introductory psychology should consider implementing a laboratory section to improve student outcomes. However, what form that laboratory section should take is still an issue of consideration, given the failures (Downey, 2013; Lewis, 2015) and successes (Koschmann & Wasp, 2001; Thieman et al., 2009) of past interventions.

Our findings provide evidence that students are capable of carrying out research within introductory psychology, and a semester-long project is one such form a laboratory component could take. However, another such form a laboratory component could take is exposing students to several different types of laboratory experiences. Past research has found that this type of intervention is also effective in improving critical thinking (Thieman et al., 2009). Future research could directly compare a semester-long project with multiple, smaller projects, such as those freely available through the Online Psychology Laboratory hosted online by the American Psychological Association (see https://opl.apa.org/).

As well, we allowed students to pick their own projects both in the laboratory and traditional plus behavior modification components. Might we see different outcomes if the instructor were to assign projects to students? For instance, students could attempt to replicate classic psychological findings. Student learning outcomes in contributing to the reproducibility of prior findings is an exciting avenue for future research and might be more appropriate for highly supervised and mentored advanced students.

However, we also found decreases across all three recitation conditions in students' perception of psychology as a science. A good scientist is one who thinks critically, scientifically, and skeptically, yet our intervention increased only critical thinking. That these elements are dissociable implies that simply conducting research in a given field does not make one perceive that field as being more scientific, nor does one think more skeptically about the claims made by that field. Given these findings, one avenue for future research is determining the full extent of the benefit conferred by conducting research.

Our study had several limitations. A major limitation is that we cannot conclude what specifically within our laboratory intervention improved critical thinking. Thus, while we can conclude that conducting a study improves critical thinking, we are unable to conclude which aspects specifically improve critical thinking. Future research should aim to make the process more visible, either by testing aspects of the research process in isolation or by administering more frequent assessments of critical thinking. For instance, students in the laboratory component did not conduct their own statistical analyses because it was deemed too difficult to accomplish in the short amount of time available. Rather, we relied on graduate-level teaching assistants to conduct the statistics for the students and briefly explain to them the meaning of the findings. It may be beneficial to student outcomes if students were tasked with learning basic statistical procedures and analyzing their own data (see Gurung et al., 2016).

We also cannot conclude that any effects could not be attributed to characteristics of the instructors. The three large courses were taught by different instructors, although the material overlapped significantly with the laboratory component (83% with the traditional plus behavioral modification and 75% with the traditional condition) in a lecture format, and the course textbooks were written at similar grade levels. In addition, within each course there were three graduate teaching assistants assigned to the recitation sections. However, across courses, the material covered within the recitations was closely supervised by the instructors to ensure the material was reliably covered across recitation sections within the course. Nevertheless, with only three courses in this study, it is impossible to rule out instructor characteristics as affecting study outcomes.

Our real-world approach was limited by the differences between what was taught in these sections, which may have had an effect on our observed results. Although we consider these differences between conditions to be limitations, we do not consider them to be confounds. The textbooks covered similar content at a similar grade level. As well, at the end of the semester students did not differ in their perception of psychology as a science nor in their endorsement of psychological myths. That they only differed in their ability to critically think suggests that this was an effect of our manipulation, rather than an effect of what was taught in the lecture.

Completing demographic information immediately prior to responding to our variables of interest also may have posed a limitation to our study, in that this may have activated stereotype threat. According to stereotype threat, being reminded of membership in a group that historically underperforms on some task may cause a member of that group to conform to those stereotypes and underperform where they otherwise might not (Spencer, Logel, & Davies, 2016).

That our laboratory intervention was only successful in improving critical thinking and not endorsement of myths or perceptions of psychology as a science may suggest that the time and effort required to implement a laboratory component may not lead to better learning outcomes. In order to assess the full extent of a laboratory component, the exam scores within the large lecture courses should be compared. Unfortunately, the exams given by the instructors between conditions were not consistent, nor were they asked for as part of the consent process, making it impossible to assess this type of learning outcome. A similar limitation in our study is that there are no measures for which we can identify which components of the recitation sections were effective in improving learning outcomes. In order to determine if a laboratory component leads to improved test scores, future research should vary laboratory participation within a single class, such that some enrolled students complete a laboratory component while others complete a control group comparison.

In summary, we found that a semester-long laboratory component, in which students were responsible for the genesis and completion of a chosen research project, increased students' ability to critically think. However, it did not increase students' perceptions of psychology as a science, nor decrease students' endorsement of psychological myths. Although this study is beneficial in informing the pedagogical literature concerning research experience in introductory psychology, the study did raise many avenues for future research.

References

American Psychological Association. (2013). *APA guidelines for the undergraduate psychology major: Version 2.0*. Retrieved from http://www.apa.org/ed/precollege/undergrad/index.aspx

American Psychological Association. (2014). Strengthening the common core of the introductory psychology course. Retrieved from https://www.apa.org/ed/governance/bea/intro-psych-report.pdf

Bensley, D. A. (2018). *Critical thinking in psychology and everyday life: A guide to effective thinking*. Worth Publishers, Macmillan Learning.

Bernstein, D. A. (2017). Bye-bye intro: A proposal for transforming introductory psychology. *Scholarship of Teaching and Learning in Psychology*, *3*(3), 191.

Ciccarelli, S.K., & White, J.N. (2017). *Psychology, 5th Edition*. Pearson.

Downey, C. A. (2013). Student Research in an Introductory Psychology Course: Outcomes of Two Experiential Learning Projects and Implications for Instruction of Human Subjects Research. *Journal of Effective Teaching*, *13*(2), 21-37.

Felzien, L. K. (2016). Integration of a zebrafish research project into a molecular biology course to support critical thinking and course content goals. *Biochemistry and Molecular Biology Education*, *44*(6), 565-573.

Friedrich, J. (1996). Assessing students' perceptions of psychology as a science: Validation of a self-report measure. *Teaching of Psychology*, *23*(1), 6-13.

Goodstone, L., Goodstone, M. S., Cino, K., Glaser, C. A., Kupferman, K., & Dember-Neal, T. (2013). Effect of simulation on the development of critical thinking in associate degree nursing students. *Nursing Education Perspectives*, *34*(3), 159-162.

Grasha, A. F. (1987). *Practical applications of psychology. (3rd ed.)* New York: Harper Collins.

Gurung, R. A., & Hackathorn, J. (2018). Ramp it up: A call for more research in introductory psychology. *Teaching of Psychology*, *45*(4), 302-311.

Gurung, R. A., 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*(2), 112.

Holmes, J. D., & Beins, B. C. (2009). Psychology is a science: At least some students think so. *Teaching of Psychology*, *36*(1), 5-11.

Hughes, S., Lyddy, F., & Lambe, S. (2013). Misconceptions about psychological science: A review. *Psychology Learning & Teaching*, *12*(1), 20-31.

Iwaoka, W. T., Li, Y., & Rhee, W. Y. (2010). Measuring gains in critical thinking in food science and human nutrition courses: The Cornell Critical Thinking Test, problem-based learning activities, and student journal entries. *Journal of Food Science Education*, *9*(3), 68-75.

Koschmann, N., & Wesp, R. (2001). Using a dining facility as an introductory psychology research laboratory. *Teaching of Psychology*, *28*(2), 105-108.

LaCaille, R. A. (2015). Two birds with one myth-debunking campaign: Engaging students to target psychological misconceptions. *Teaching of Psychology*, *42*(4), 323-329.

Landrum, R. E., Gurung, R. A., & Amsel, E. (2019). The importance of taking psychology: A comparison of three levels of exposure. *Teaching of Psychology*, *46*(4), 290-298.

Lawson, T. J., Jordan-Fleming, M. K., & Bodle, J. H. (2015). Measuring psychological critical thinking: An update. *Teaching of Psychology*, *42*(3), 248-253.

Lewis, J. L. (2015). A comparison between two different activities for teaching learning principles: Virtual animal labs versus human demonstrations. *Scholarship of Teaching and Learning in Psychology*, *1*(2), 182.

Lilienfeld, S. O. (2012). Public skepticism of psychology: why many people perceive the study of human behavior as unscientific. *American Psychologist*, *67*(2), 111.

McCarthy, M. A., & Frantz, S. (2016). Challenging the status quo: Evidence that introductory psychology can dispel myths. *Teaching of Psychology*, *43*(3), 211-214.

Normand, M. P. (2008). Science, skepticism, and applied behavior analysis. *Behavior Analysis in Practice*, *1*(2), 42-49.

Peterson, J. J., & Sesma Jr, A. (2017). Introductory Psychology: What's Lab Got to Do With It?. *Teaching of Psychology*, *44*(4), 313-323.

Quattrucci, J. G. (2018). Problem-Based Approach to Teaching Advanced Chemistry Laboratories and Developing Students' Critical Thinking Skills. *Journal of Chemical Education*, *95*(2), 259-266.

Reinbold, S. L. (2018). Critical Thinking Assessment of Students in Nonmajors Biology Classes with Corn or Fly Genetics Laboratory Studies. *Journal of College Science Teaching*, *48*(2).

Rowe, M. P., Gillespie, B. M., Harris, K. R., Koether, S. D., Shannon, L. J. Y., & Rose, L. A. (2015). Redesigning a general education science course to promote critical thinking. *CBE—Life Sciences Education*, *14*(3), 1-12.

Rowland, K. C., & Joy, A. (2015). The gross anatomy laboratory: a novel venue for critical thinking and interdisciplinary teaching in dental education. *Journal of Dental Education*, *79*(3), 295-300.Schacter, D.L., Gilbert, D.T., Nock, M.K., & Wegner, D.M. (2017). *Psychology: Fourth Edition.* Macmillan Learning.

Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology*, *39*, 127-161.Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, *67*, 415-437.

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*(3), 160-168.

Weaver, M. G., Samoshin, A. V., Lewis, R. B., & Gainer, M. J. (2016). Developing students' critical thinking, problem solving, and analysis skills in an inquiry-based synthetic organic laboratory course. *Journal of Chemical Education*, *93*(5), 847-851.

Wentworth, D. K., & Whitmarsh, L. (2017). Thinking Like a Psychologist Introductory Psychology Writing Assignments: Encouraging Critical Thinking and Resisting Plagiarism. *Teaching of Psychology*, *44*(4), 335-341