

TO BE PUBLISHED IN OCT 2013 EDITION OF *TEACHING OF PSYCHOLOGY* JOURNAL

Critical Analysis:

A Comparison of Critical Thinking Changes in Psychology and Philosophy Classes

Brian L. Burke,¹ Sharon R. Sears,¹ Sue Kraus,¹ and Sarah Roberts-Cady²

¹Department of Psychology, Fort Lewis College, Durango, USA

²Department of Philosophy, Fort Lewis College, Durango, USA

Corresponding Author:

Brian L. Burke, Department of Psychology, Fort Lewis College, Durango, CO 81301 USA

Email: burke_b@fortlewis.edu

Abstract

This study compared changes in psychology and philosophy classes in two distinct components of critical thinking: general skills and personal beliefs. Participants were 128 undergraduates enrolled in critical thinking in psychology, other psychology courses, or philosophy courses. Critical thinking and philosophy students significantly reduced beliefs in paranormal phenomena at the end of the semester compared to other psychology students. Only philosophy students improved on the Watson–Glaser Critical Thinking Appraisal (Watson & Glaser, 2006). The Watson-Glaser may not fully measure critical thinking emphasized in psychology, but psychology instructors can still effectively teach students to examine their own beliefs and think differently in their daily lives. Differentiated assessment of critical thinking is important as instructors evaluate student learning against specific goals.

Keywords: critical thinking, teaching, psychology, Watson-Glaser, paranormal beliefs, philosophy

Critical Analysis:

A Comparison of Critical Thinking Changes in Psychology and Philosophy Classes

“We must sooner or later abandon the traditional attempt to teach our fellow citizens *what* to think. Such efforts cannot prepare us for the real world we must, in fact, face. We must concentrate instead on teaching ourselves *how* to think...”

– Paul (1995, p. 16)

Critical thinking (CT) involves the ability to evaluate claims on the basis of evidence so that a sound conclusion can be drawn (Bensley, 1998). CT is typically reported across disciplines as one of the top goals of higher education, with 99.6 percent of U.S. faculty indicating that CT skills are "very important" or "essential" (Wyer, 2009). Further, CT is central to psychology as it lies at the junction of the discipline’s emphasis on scientific methods and its seminal content domain—i.e., the study of how human beings think (Myers, 2009). In fact, CT is the third listed goal in the American Psychological Association guidelines for the undergraduate psychology major, stated as “respect and use critical and creative thinking, skeptical inquiry, and, when possible, the scientific approach to solve problems related to behavior and mental processes” (American Psychological Association, 2007, p. 14). Thus, if we aim to teach our students psychology, we must above all teach them how to think critically about the world within and around them. We do not want them to simply memorize facts, or believe everything they learn from Wikipedia, but rather to develop a structured and effective process for how to think about various issues.

One key issue is to what extent we can assess the fundamental elements of CT development and evaluate whether it can be improved by specific college coursework. Toward this end, previous research has taken several complementary approaches, measuring changes in:

(1) CT skills, both psychology-specific (Lawson, 1999; Wesp & Montgomery, 1998) and general ability (e.g., Sandor, Clark, Campbell, Rains, & Cascio, 1998; Scott, Markert, & Dunn, 1998), and (2) specific beliefs such as endorsement of paranormal phenomena (McLean & Miller, 2010).

Psychology-specific CT is the ability to evaluate claims using psychological research principles—i.e., to examine evidence, detect flaws, or design studies. Measuring psychology-specific CT is a relatively recent undertaking. Our literature search illustrated that, from 1974 (the first full-size journal) to 1998, 26 articles in *Teaching of Psychology* (85% of which appeared in the February 1995 special issue on CT) described various teaching methods to enhance CT, ranging from case studies and technology to cooperative learning and questioning. However, none actually assessed whether these teaching strategies resulted in CT gains for students. In 1998, roughly two thirds of Introductory Psychology textbooks defined and discussed the CT process (Griggs, Jackson, Marek, & Christopher, 1998).

Since that time, psychology's efforts to measure discipline-relevant CT skills have proliferated, showing that psychology courses designed to teach CT lead to improved psychology-specific abilities. These researchers have taken two different approaches to teaching CT: either constructing a stand-alone parapsychology/skeptical inquiry course or infusing CT instruction and content—e.g., via textbooks such as Stanovich, 2004, and Bensley, 1998—into existing psychology courses (or adding a 1-credit seminar). For instance, parapsychology students identified significantly more flaws ($d = .82$) in a brief article describing a claim about a hypothetical scientific discovery compared to students in a psychology class about self-control (Wesp & Montgomery, 1998). Different parapsychology students also generated significantly

more advanced explanations for these flaws ($d = .39$) compared to students in a research methods class (McLean & Miller, 2010).

Furthermore, students in a 1-credit first-year seminar on CT (Penningroth, Despain, & Gray, 2007), a CT-infused human development course (Williams, Oliver, Allin, Winn, & Booher, 2003), or a CT-infused research methods course (Stark, 2012) improved their scores on Lawson's (1999) *Psychological Critical Thinking Test*. These outcomes were significant compared to students in a general psychology class ($d = 1.50$; Penningroth et al., 2007) or from pre- to post-semester ($d = .65$ in Williams et al., 2003, and $d = 1.25$ in Stark, 2012). Lawson's (1999) test assesses how effectively students can find the errors in each of 14 short research study descriptions. A sample description is: "A researcher tested a new drug designed to decrease depression by giving it to 100 clinically depressed patients; she discovered that their scores on a standardized depression inventory declined after 4 months of taking the drug and concluded that the drug reduces depression." Not surprisingly, Lawson (1999) demonstrated that psychology majors scored significantly higher ($d = 1.22$) compared to natural science majors on this test of psychology-specific CT.

Students in a different CT-infused research methods course (Bensley, Crowe, Bernhardt, Buckner, & Allman, 2010) had higher scores compared to students in a control research methods course ($d = 1.50$) on an argument analysis test designed by the researchers, one of three subtests comprising their *Critical Thinking in Psychology Test* (Bensley & Baxter, 2006). The argument analysis test had 15 multiple-choice items describing psychology-related situations or psychological research or clinical practice examples. The test included three items on recognizing kinds of evidence, five items on evaluating different evidence, four items on

determining whether an example is an argument or not, and three items asking participants to find assumptions in examples (Bensley & Baxter, 2006).

Blessing and Blessing (2010) illustrated that even a single class assignment may boost psychology-specific analytic abilities. Students in introductory psychology who completed a “PsychBusters” project in which they produced a group presentation investigating a psychological myth scored significantly higher on a test created for the study than students in different course sections who did not do the project ($d = .68$). The test used as the dependent measure gave students one of two statements—either “Blondes have more fun” or “You can’t teach an old dog new tricks”—and asked them how they would design an experiment to test that claim. It is therefore clear that psychology courses designed to stimulate critical analysis of discipline-specific content—detecting errors, evaluating evidence, or designing studies—have been effective in doing so, with medium to large effect sizes (i.e., d s ranging from .60 to 1.50; Cohen, 1992). What remains unclear is to what extent students improve their CT skills that could apply to non-psychology topics. Will the development of discipline-specific CT skills translate into a general improved ability to think critically outside of the discipline? How is this best measured?

The most widely used evidence-based measure of general CT in post-secondary students has been the Watson-Glaser Critical Thinking Appraisal (WGCTA; Watson & Glaser, 2006), which has good psychometric properties (Gadzella, Stacks, Stephens, & Masten, 2005). Several studies have evaluated CT changes in health profession classes in nursing or medical schools (Brown, Alverson, & Pepa, 2001; Frye, Alfred, & Campbell, 1999; Sandor et al., 1998; Scott et al., 1998). Overall, this research has found that years of problem-based nursing or medical curricula improve students’ WGCTA scores, but the effects are typically in the small range (i.e.,

ds below .50; Cohen, 1992). A meta-analysis on WGCTA outcomes in higher education determined that public speaking, debate, and argumentation (e.g., mock trial) classes produced CT gains with a mean effect size of $d = .32$ ($k = 12$; Allen, Berkowitz, Hunt, & Loudon, 1999). A more recent meta-analysis examined all standardized measures of CT (including the WGCTA) and found a similar mean effect size ($d = .24$) for instructional interventions ranging from elementary school to graduate school ($k = 91$; Abrami et al., 2008), with elementary and secondary students showing significantly higher CT skills gains than undergraduate students overall, perhaps due to a ceiling effect.

Note also that five of the seven previous studies on teaching CT to psychology students published in *Teaching of Psychology* (Bensley et al., 2010; Blessing & Blessing, 2010; Penningroth et al., 2007; Wesp & Montgomery, 1998; Williams et al., 2003) measured only gains in *psychology-specific* CT abilities, such as students' proficiency at analyzing psychological research. Only two studies, to our knowledge, assessed general CT skill gains after specific psychology classes. One found no significant differences between parapsychology students and those in an advanced research methods class on their WGCTA scores (McLean & Miller, 2010). Another recent study using the Cornell Critical Thinking Test (CCTT; Ennis, Millman, & Tomko, 2004), a different measure of general CT ability, also failed to find any improvements in psychology students after a CT-infused research methods course (Stark, 2012).

What may be more amenable to change than general CT abilities are specific unscientific beliefs, such as endorsement of paranormal phenomena ranging from alien encounters and life after death to psychic powers and the Loch Ness monster (e.g., Tobacyk, 2004). Four different studies have shown that teaching interventions can reduce student scores on paranormal belief scales. A 1-week intensive course on parapsychology and skeptical inquiry significantly reduced

paranormal beliefs in participants aged 55+ ($d = .45$; Banziger, 1983). More recently, students in two different parapsychology courses decreased their paranormal beliefs significantly throughout the semester (by about 32% in McLean & Miller, 2010, and 45% in Manza et al., 2010) and compared to advanced research methods students ($d = 1.42$; McLean & Miller, 2010) or statistics students ($d = 1.62$; Manza et al., 2010). Furthermore, students in a CT-infused research methods course also lowered their paranormal beliefs, but only by about 10% ($d = .49$) from pre- to post-semester assessment (Stark, 2012). All of these studies used the 26-item Revised Paranormal Belief Scale (RPBS; Tobacyk, 2004), except for Manza et al. (2010), who used a 20-item Paranormal Belief Scale instead (taken from Sparks, Nelson, & Campbell, 1997).

The present study was the first to our knowledge to directly compare general CT gains in specific psychology and non-psychology undergraduate classes after a semester of coursework. This research was designed to gain information on how well psychology courses develop general CT skills, as compared to courses in other disciplines. Further, this study could help answer the question of what the WGCTA measures and its relevance for psychology departments. We chose philosophy as a comparison discipline because those courses emphasize deductive reasoning and argument analysis, key components of the WGCTA. Moreover, philosophy students generally exhibit strong CT skills as evidenced by their high scores on various standardized reasoning tests. For instance, philosophy majors typically earn the highest scores on the Graduate Record Examination (overall and on the Verbal Reasoning and Analytical Writing sections), whereas psychology majors are in the middle of the pack of the 44 reported undergraduate majors (Educational Testing Services, 2011). Rather than assessing improvements in course-specific skills, which is already well-established in psychology (cf. Lawson, 1999; Bensley et al., 2010), our primary aim was to gauge how disciplinary courses impact progress in students' ability to

apply CT skills outside the discipline, as well as how these courses impact unscientific belief content.

Accordingly, we had two main hypotheses for this study, one concerning changes in beliefs and the other concerning changes in skills:

1. Both the CT and philosophy students would reduce their paranormal beliefs throughout the semester. As described below, each of these classes contained explicit skeptical analysis regarding paranormal claims and encouraged students to reflect on their own beliefs, which has yielded reductions in such beliefs in previous research. We did not expect the comparison (non-CT) psychology classes to show any change in paranormal beliefs.

2. Only philosophy students would show significant changes in general CT skills on the standardized test (WGCTA). Whereas general CT skill components have not yet been studied heavily in undergraduate education and may not improve after a single semester course, the philosophy (but not the psychology) classes in this study specifically targeted logic and argument analysis, which could improve WGCTA scores. We did not predict that any of the psychology classes would improve student scores based on previous research of the WGCTA and other general skills tests in psychology instruction.

Method

Participants

We recruited 128 undergraduate students (82 females and 46 males) at a small, liberal arts college enrolled in one of seven different psychology class sections of Introduction to Psychology, Research Methods, Senior Seminar, or a new Critical Thinking course developed for this study, or one of two different philosophy classes (Introduction to Philosophy or Logic). The study had IRB approval and was conducted during class time; students did not receive inducement for their participation and were allowed to sit quietly and not participate if they so

chose. Although no one declined to participate, only 63% of eligible students provided useable data at both pre- and post-test due to absences and incomplete surveys. The overall sample was comprised of 43% psychology majors, with a mean age of 21.24 ($SD = 4.69$) and a relatively even distribution among class standing (30% first-years, 19% sophomores, 20% juniors, and 31% seniors). The ethnic breakdown was 73% Caucasian, 17% Native American, and 10% other ethnicities. The high proportion of Native American participants reflects the college's unique population comprised of approximately 21% Native American students.

Measures

Watson–Glaser Critical Thinking Appraisal. The Watson–Glaser Critical Thinking Appraisal, Form S (WGCTA–FS; Watson & Glaser, 2006) is a 40-item self-report measure that asks respondents to read and evaluate passages that include problems, statements, arguments, and interpretations covering inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. The WGCTA is best viewed as a measure of general competency, and its five subscales should not be interpreted individually (Bernard et al., 2008). Total scores range from 0 to 40, with higher scores reflecting greater CT abilities. The WGCTA has a documented history of use in educational and organizational settings (e.g., Gadzella et al., 2005). Psychometric properties for the full scale are sound, with good internal consistency (Cronbach's alpha coefficient r ranging from .76 to .85; Watson & Glaser, 2006) and test-retest reliability ($r = .81$; Watson & Glaser, 1994). In addition, there is ample evidence of criterion-related validity, with WGCTA scores consistently showing significant correlations with on-the-job performance (Watson & Glaser, 2006) and decision-making effectiveness (Shin, 1998).

Revised Paranormal Belief Scale. The Revised Paranormal Belief Scale (RPBS; Tobacyk, 2004; Tobacyk & Milford, 1983) is the most widely used instrument for measuring

paranormal beliefs (Goulding & Parker, 2001). The 26-item scale assesses traditional Western religious beliefs, psychic phenomena, witchcraft, superstition, and anomalous natural phenomena such as Big Foot and the Abominable Snowman. Respondents are asked to rate their level of agreement with statements of beliefs on a seven-point Likert scale with higher ratings indicating stronger endorsement. As with the WGCTA above, the RPBS is best analyzed as a full scale rather than as seven separate subscales (Lawrence, 1995). Test-retest reliability for the full scale over a 4-week interval was $r = .92$ (Tobacyk, 2004), and internal consistencies ranged from .86 to .89 (McLean & Miller, 2010; Shiah, Tam, Wu, & Chang, 2010). Construct validity for the RPBS is still emerging, although the scale has shown theoretically sound correlations with specific aspects of religiosity. For instance, the traditional religious belief factor of the RPBS had the highest correlation with the afterlife factor of the Personal Religiosity Scale (Shiah et al., 2010); however, the overall RPBS was largely independent of religious beliefs (Williams, Francis, & Lewis, 2009). In addition, RPBS scores were correlated with reports of paranormal practices such as ‘out-of-body’ experiences (Tobacyk, 2004).

Procedure

During the 2011 academic year, we compared the psychology CT course to two philosophy courses—Introduction and Logic—and several of our core psychology class sections: two sections each of Introduction to Psychology, Research Methods, and Senior Seminar. All courses met for three to four hours weekly throughout a 15-week semester. Participants could not be randomized to class types because the study occurred in the context of existing courses. There were no students who were taking any of those courses simultaneously. Students who chose to participate in the study completed a basic demographic form (with gender, ethnicity, religion, political orientation, age, and parents’ level of education); they were asked to take the WGCTA

and RPBS as an in-class assignment during both the first and last weeks of class, which took about 45-55 minutes for most students to complete.

Psychology CT Course Instruction. In the Winter 2011 semester, we designed a full-semester four-credit course called “Critical Thinking in Psychology” taught by the first author. The course covered topics ranging from psychic powers and UFOs to TV game shows, global warming, and the medical marijuana debate (see Kraus, Sears, & Burke, 2012 in prep, for more details about the course content). Lilienfeld, Lynn, and Lohr’s (2004) *Science & Pseudoscience in Clinical Psychology* was the primary textbook for the course; thus, many of the topics had clinical relevance, such as recovered memories, treatments for trauma such as eye-movement desensitization and reprocessing (EMDR), substance abuse treatment programs, and psychology in the media. The course incorporated strategies, readings, and websites related to skeptical inquiry (see Lilienfeld, Lohr, & Morier, 2001). In addition, the students took their CT skills actively into the community—by giving hands-on demonstrations for children at a local science museum, judging a regional middle school science fair, and attending a nearby UFO symposium as a class.

Each CT class featured an informational presentation and associated class activity (e.g., discussion, debate, writing) about a different controversial topic or issue. Class discussion and analysis of each topic was structured around inductive reasoning according to what we termed the “seven steps to critical thinking.” These were modified from Bernstein’s five steps (2007) as follows (underlines included in version handed out to students):

1. What am I being asked to believe or accept?
2. What evidence is available to support the claim?
3. What alternative ways are there to interpret the evidence?

4. Rate all the evidence/alternatives on a 0-10 scale based on validity/strength.
5. What assumptions or biases came up when doing the above steps?
(e.g., using intuition/emotion, authority, or personal experience rather than science)
6. What additional evidence would help us evaluate the alternatives?
7. What conclusions are most reasonable or likely?

Although students were not explicitly tested on these seven steps, each CT topic increased in complexity and in progression through the steps. For example, the first topic addressed the first two steps only while the fourth topic addressed all seven steps above.

Comparison Psychology Course Instruction. All three different psychology comparison courses had some CT instruction built into the preexisting curriculum. The college's psychology department learning outcomes list CT as one of four overarching goals, and include the following subcomponents: ability to assess the validity and reliability of psychological sources; full development of conclusions and concepts; use of clarity, open-mindedness and skepticism; and ability to identify common flaws in psychological research and practice. Because we could not prohibit instructors from encouraging CT in their classes, these comparison groups thus represent more of a "treatment as usual" model than true control groups. Introduction to Psychology featured a broad overview of the scientific study of behavior and mental processes, including the concepts of learning, development, personality, psychotherapy, cognitive psychology, and the brain. Research Methods introduced students to the basic assumptions, concepts, and methodology of experimental and psychological research, including critical evaluation of published research and a discussion of research ethics. In this course, each student designed, conducted, and reported on an experimental research project. Senior Seminar had a component on careers and graduate school issues (e.g., cover letters, CVs, GREs), but also

involved advanced study in psychology research and selected topics around each student's individual interests, such as a major literature review on the student's chosen subject area.

Philosophy Course Instruction. Both philosophy classes were taught by the same instructor and included explicit CT components. The college's philosophy department learning outcome related to CT states that students will be able to identify arguments, including their underlying assumptions, and critically evaluate them in a rigorous and fair-minded way. Introduction to Philosophy featured discussion and analysis of representative readings from the history of philosophy. Issues considered included ethics, the relation between reason and religious belief, philosophy of art, the relation of knowledge to experience, and the nature of free will. As students explored readings on these topics, they were introduced to skills in the identification and analysis of philosophical arguments (both deductive and inductive), which they were tested on repeatedly throughout the semester. Logic taught a broad range of different methods of evaluating both deductive and inductive arguments. The textbook used in the Logic class states that "the aim of logic is to develop a system of methods and principles that we may use as criteria for evaluating the arguments of others and as guides in constructing arguments of our own" (Hurley, 2011, p. 1). Both philosophy classes focused on real-world applications of logic and philosophy. For instance, students in Introduction to Philosophy regularly discussed personal beliefs that may deeply affect their lives outside the classroom—such as moral and religious beliefs—and engaged in critical analysis of paranormal phenomena. In Logic, students were encouraged to bring in and analyze arguments from the media (such as advertisements), other courses, and even Supreme Court hearings. Further, the Logic textbook (Hurley, 2011) contained a chapter on distinguishing science from superstition, which is directly relevant to

paranormal beliefs. Students in the Logic class took their exam that included this chapter right before taking the RPBS post-test.

Results

Table 1 displays the group means pre- and post-semester of the CT measures. Note that both philosophy classes were combined into a single group, as there were no significant differences between them in any pattern of results. Preliminary analyses using ANOVA suggested that participants in the three different class types—CT class, other psychology classes, and philosophy classes—were not significantly different prior to the intervention in demographics: age, gender, ethnicity, mother's education, father's education, religious affiliation, religiosity, or political affiliation (all $p > .22$). Likewise, participants were not significantly different between groups in their pre-test WGCTA scores or RPBS scores (see Table 1), $F(2, 125) = 1.09, p = .34$ and $F(2, 49) = 1.30, p = .28$ respectively. Thus, although participants could not be randomized to class types, we are reasonably confident that they entered the study with similar relevant attributes across groups.

We used a mixed model 2 (pre/post) X 3 (class type: CT, other psychology, or philosophy) two-way ANOVA to analyze both WGCTA and RPBS scores. All ANOVA results are shown in Table 2, with post-hoc test results (*Bonferroni*) provided in answer to the first two questions below.

Our first hypothesis was that both CT and philosophy students would reduce their paranormal beliefs throughout the semester but that psychology controls would not. As shown in Table 2, our two-way ANOVA yielded a significant interaction effect for RPBS scores, $p = .03$. In addition, there was a significant pre-post effect, $p = .01$, indicating that scores decreased throughout the semester. To determine which classes were driving the score reductions, we

examined each group separately for pre-post changes in paranormal beliefs using paired-sample t tests for repeated measures. As Table 1 displays, students in the CT class, $t(9) = 2.73, p = .02, d = .74$, and philosophy classes, $t(25) = 3.89, p = .01, d = .65$, significantly decreased their paranormal beliefs from pre- to post-semester; whereas, students in the comparison psychology classes did not, $t(13) = .54, p = .60, d = .10$.

Our second hypothesis was that only philosophy students would improve their CT skills as measured by the WGCTA. Our two-way ANOVA (see Table 2) did not yield a significant interaction effect for WGCTA scores, $p = .11$, but there was a significant effect of class type, $p = .03$. To determine what accounted for that effect, we performed one-way ANOVAs for both pre- and post-test WGCTA scores separately. As described above, there were no pre-test differences between classes; however, there was a significant between-group difference in post-test WGCTA scores, $F(2, 126) = 5.31, p = .01$. As predicted, students in the philosophy classes showed significantly increased CT skills at the end of the semester compared to other psychology students, *Bonferroni* $p = .01, d = .70$, and approached significantly higher CT skills than students in the CT class, *Bonferroni* $p = .07, d = .90$. Overall, though, the pre-post gains on the WGCTA in the philosophy classes were small, with repeated measures $t(25) = 2.42, p = .02, d = .45$, an 8.2% boost. As expected, there were no significant differences on post-semester WGCTA scores between students in the CT class and those in the other psychology classes, *Bonferroni* $p = 1.00, d = -.08$.

Discussion

Findings supported our overarching hypothesis that different classes would differentially affect specific elements of undergraduate student CT. A specially designed course on “Critical Thinking in Psychology” reduced students’ specific paranormal beliefs (by 30%) but did not

change CT skills as measured by the WGCTA. By contrast, philosophy classes featuring logic and argument analysis improved students' WGCTA scores and reduced their specific paranormal beliefs (by 20%). These differential findings have important implications for instructors, departments, and institutions as they attempt to define, assess, and develop interventions to increase CT.

We therefore have mixed news for psychology instructors regarding teaching different elements of CT—i.e., skills and beliefs—to students. We will discuss the bad news first. A semester-long CT course on skeptical inquiry using classroom and community-based learning did not improve students' CT skills outside the discipline, as measured by the WGCTA. One plausible interpretation of these results is that the development of psychology-specific CT skills does not necessarily translate into CT skills outside the discipline. Given that one of our most fervent goals as educators—and psychologists—is to teach our students how to think critically and effectively beyond the confines of our discipline, these results may be cause for sobering reflection.

A closer look at the data suggests exercising caution before drawing this conclusion. Although the WGCTA is a well-validated CT measure, it is possible that there are general (non-disciplinary) CT skills that psychology instructors are developing in students that the WGCTA does not measure. The key to understanding the results may be in grasping the distinction between different forms of deductive and inductive reasoning. Deductive reasoning involves reasoning about logically necessary relationships between claims (Hurley, 2011). For example, if one knows that the mean paranormal belief scores of CT students decreased from 90 to 60 in this study, one can deduce that it follows necessarily that, on average, student beliefs decreased by one-third from pre- to post-semester. In contrast, inductive reasoning involves probabilistic

reasoning; that is, it involves examining conclusions that follow from the evidence with probability rather than necessity. Suppose we give students in philosophy and psychology classes the WGCTA at the beginning and at the end of the semester. If we see a significant improvement in the scores of the philosophy students but not the psychology students, this will offer some probabilistic (but not certain) evidence that participation in the philosophy class is the cause of the improved scores.

The WGCTA appears primarily—although not exclusively—to test deductive reasoning. In fact, four of its five subscales mainly test propositional reasoning, a form of deductive reasoning in which the validity of inferences depends upon the relationships between the information content of various statements/premises (Hurley, 2011). Two of its five subscales also appear to test syllogistic or categorical reasoning, a form of deductive reasoning in which a conclusion is inferred from two or more statements/premises (Hurley, 2011), along with inductive reasoning, chiefly causal—i.e., examining evidence for claims that one thing caused another (Hurley, 2011).

Both deductive and inductive reasoning skills are crucial to “reasonable, reflective thinking focused on deciding what to believe or do” (Ennis, 1987, p. 10) and are used regularly in psychology, philosophy, and many other fields. However, psychologists tend to focus on developing in students the specific forms of inductive reasoning that are important to their field—i.e., hypothesis-based, causal, and statistical reasoning. Whereas these skills may have broad, cross-disciplinary application, they are forms of inductive reasoning not measured by the WGCTA. In contrast, philosophy focuses broadly on evaluating both deductive and inductive arguments, many forms of which are measured by the WGCTA; this could explain why philosophy students improved on the WGCTA, but psychology students did not. This conclusion

is supported by the fact that a longitudinal study found that undergraduate training in psychology, and, more generally, the social sciences, significantly boosted inductive (statistical and methodological) reasoning about a wide range of problems; conversely, students majoring in the natural sciences and the humanities (like philosophy) improved significantly more in their deductive (conditional) reasoning during their undergraduate years (Lehman & Nisbett, 1990). Note also that two different studies using the Cornell Critical Thinking Test (CCTT; Ennis, Millman, & Tomko, 2004), another general CT measure, found small but significant improvements in psychology students after psychology courses that infused *generic* (not discipline-specific) CT instruction including deductive reasoning training ($d = .75$ in Solon, 2007, and $d = .46$ in Nieto & Saiz, 2008).

Our CT findings may thus have important implications for higher education in psychology. The American Psychological Association has recently endorsed guidelines that call for college and universities to increase efforts to measure and improve student learning (Chamberlin, 2012). One major suggestion as student assessment expands has been to systematically measure changes in CT (Bensley & Murtagh, 2012). However, the present results suggest that moving forward toward this goal will require attention to two specific issues. First, educators in psychology should reflect on what specific CT skills they would like to teach: some inductive reasoning only or also deductive reasoning skills. If improvement in deductive reasoning is an appropriate outcome for psychology courses, then increased emphasis on teaching such skills may be warranted. Second, educators should choose an assessment tool carefully. Our analysis suggests that psychology classes focused on inductive reasoning may have limited ability to impact the CT skills assessed by WGCTA. Perhaps one could highlight the broad value of psychology by specifically designing and administering tests that assess

progress in general inductive reasoning (Bensley & Murtagh, 2012; Halpern, 2010; Butler et al., 2012).

Now the good news for psychology instructors: A full-semester CT course resulted in significant changes—both from pre- to post-semester and compared to other psychology courses—in belief content. These reductions in paranormal beliefs (a 30% decrease, similar to the 32% decrease found in McLean and Miller, 2010) could have powerful real-world benefits for students. There is a small yet significant inverse correlation between paranormal beliefs and college GPA (Tobacyk, 1984), as well as a significant positive correlation between paranormal beliefs and negative affect (Dudley, 2000). In addition, media consumption has been associated with increased paranormal beliefs (Sparks, Nelson, & Campbell, 1997). It is possible that teaching students how to critically evaluate and be more aware of their own potential biases (paranormal beliefs) could inoculate them against some of the negative effects of media and render their life choices more accurate. In order to do this, however, our psychology classes would also need to increase their motivation to use CT outside the classroom, as previous research has shown (Burke, Sears, & Kraus, 2012). Future studies should investigate to what extent psychology, philosophy, nursing/medical, or other classes differentially improve student motivation to use CT outside of class, which can now be assessed via a new instrument with good psychometric properties (Valenzuela, Nieto, & Saiz, 2011).

Our study had several notable strengths, including the use of comparison classes both within and outside the discipline of psychology, as well as pre- and post-test assessments of two fundamental elements of CT—unscientific belief content and CT skills—using standardized measures. Nonetheless, there were several important limitations of our research design and procedure, in addition to a small sample size. Firstly, we elected not to measure psychology-

specific CT gains, as that seemed to us to be “teaching to the test”—i.e., of course students do better on a psychology-specific test after taking a targeted psychology course. Previous studies have clearly established that CT gains do occur in psychology-specific content, but we were interested in assessing generic CT gains here. However, demand characteristics may still have contributed to the CT score changes in our study; as noted above, the philosophy students had instruction more apropos to the WGCTA and both the CT psychology and philosophy students had specific course content that addressed paranormal beliefs. Thus, whereas we did not explicitly “teach to the test” like psychology-specific CT classes, our students’ responses on the RPBS may have reflected what they thought we wanted.

As mentioned previously, several studies have found that nursing and medical students improve their WGCTA scores (with small to medium effect size gains compared to control groups) as a result of their curriculum (Brown et al., 2001; Frye et al., 1999; Sandor et al., 1998; Scott et al., 1998). One possibility worth exploring is that the teaching strategies used to develop CT abilities may differ. For instance, our CT class students primarily experienced lectures, reading, group work, discussion, and videos, with some problem-based learning (and community engagement) but little explicit grade-related testing of their CT abilities. In contrast, it is likely that nursing/medical students were exposed to more problem-based and case-based learning strategies along with more frequent assessment via tests and real-world situations (i.e., caring for a live patient). The philosophy classes tested in this study, which did produce a significant improvement in WGCTA scores, also included more frequent problem-solving homework and more traditional testing than the CT class. Future research should therefore investigate specific classroom strategies and methods to determine how psychology instructors can teach their

subject matter optimally, especially if improving CT across a wider spectrum of reasoning types remains a trenchant and desirable goal.

References

- Abrami, P. C., Bernard, R. M., Borokhovhovski, E., Wade, A., Surkes, M. A., Tamim, R., et al., (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 4*, 1102–1134.
doi:10.3102/0034654308326084
- Allen, M., Berkowitz, S., Hunt, S., & Louden, A. (1999). A meta-analysis of the impact of forensics and communication education on critical thinking. *Communication Education, 48*, 18-30. doi:10.1080/03634529909379149
- American Psychological Association. (2007). *APA guidelines for the undergraduate psychology major*. Washington, DC: Author. Retrieved from www.apa.org/ed/resources.html on June 14, 2011.
- Banziger, G. (1983). Normalizing the Paranormal: Short-term and long-term change in belief in the paranormal among older learners during a short course. *Teaching of Psychology, 10*, 212-214. doi:10.1207/s15328023top1004_6
- Bensley, D. A. (1998). *Critical thinking in psychology: A unified skills approach*. Pacific Grove, CA: Brooks/Cole.
- Bensley, D. A., & Baxter, C. (2006). *The Critical Thinking in Psychology Test*. Unpublished manuscript, Frostburg State University, Frostburg, MD.
- Bensley, D. A., Crowe, D. S., Bernhardt, P., Buckner, C., & Allman, A. L. (2010). Teaching and assessing critical thinking skills for argument analysis in psychology. *Teaching of Psychology, 37*, 91–96. doi:10.1080/00986281003626656
- Bensley, D. A., & Murtagh, M. P. (2012). Guidelines for a scientific approach to critical thinking assessment. *Teaching of Psychology, 39*, 5-16. doi:10.1177/0098628311430642

Bernard, R. M., Zhang, D., Abrami, P. C., Sicol, F., Borokhovski, E., & Surkes, M. A. (2008).

Exploring the structure of the Watson-Glaser Critical Thinking appraisal: One scale or many subscales? *Thinking Skills and Creativity*, 3, 15-22. doi:10.1016/j.tsc.2007.11.001

Bernstein, D. A. (2007). *Promoting critical thinking and active learning in psychology courses*.

Keynote presentation given at the Mountain States Conference on the Teaching of Psychology. Durango, CO (October 6-7).

Blessing, S. B., & Blessing, J. S. (2010). PsychBusters: A means of fostering critical thinking in the introductory course. *Teaching of Psychology*, 37, 178-182.

doi:10.1080/00986283.2010.488540

Brown, J. M., Alverson, E. M., & Pepa, C. A. (2001). The influence of a baccalaureate program on traditional, RN-BSN, and accelerated students' critical thinking abilities. *Holistic Nursing Practice*, 15, 4-8.

Burke, B. L., Sears, S. R., & Kraus, S. (2012). *I Don't Think, Therefore I Cram: Critical Thinking in Psychology Classrooms*. Poster presentation given at the National Institute for the Teaching of Psychology (NITOP). St. Petersburg Beach, FL (January 3-6).

Butler, H. A., Dwyer, C. P., Hogan, M. J., Franco, A., Rivas, S. F., Saiz, C., & Almeida, L. S. (2012). The Halpern Critical Thinking Assessment and real-world outcomes: Cross-

national applications. *Thinking Skills and Creativity*, 7, 112-121.

doi:10.1016/j.tsc.2012.04.001

Chamberlin, J. (2012). APA endorses higher education guidelines. *Monitor on Psychology*, 43, 22.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. doi:10.1037/0033-2909.112.1.155

- Dudley, R. T. (2000). The relationship between negative affect and paranormal belief. *Personality and Individual Differences, 28*, 315-321.
- Ennis, R. (1987). A taxonomy of critical thinking abilities and dispositions. In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills* (pp. 9-26). New York: W. H. Freeman.
- Ennis, R. H., Millman, J., & Tomko, T. N. (2004). *Cornell Critical Thinking Tests, Level X and Level Z manual. 4th ed.* Seaside, CA. Critical Thinking Company.
- Educational Testing Service. (2011). *GRE Scores by Major* [Data File]. Retrieved from http://www.nmu.edu/sites/DrupalPhilosophy/files/UserFiles/Files/Pre-Drupal/SiteSections/Resources/GRE_Scores_by_Intended_Major.pdf
- Frye, B., Alfred, N., & Campbell, M. (1999). Use of the Watson-Glaser Critical Thinking Appraisal with BSN students. *Nursing and Health Care Perspectives, 20*, 253-255.
- Gadzella, B. M., Stacks, J., Stephens, R. C., & Masten, W. G. (2005). Watson-Glaser Critical Thinking Appraisal, Form-S For Education Majors. *Journal of Instructional Psychology, 32*, 9-12.
- Goulding, A., & Parker, A. (2001). Finding psi in the paranormal: Psychometric measures used in research in paranormal beliefs/experiences and in research on psi-ability. *European Journal of Parapsychology, 16*, 73-101.
- Griggs, R. A., Jackson, S. L., Marek, P., & Christopher, A. N. (1998). Critical thinking in introductory psychology texts and supplements. *Teaching of Psychology, 25*, 254-266. doi:10.1080/00986289809709711
- Halpern, D.F. (2010). Halpern Critical Thinking Assessment. SCHUHFRIED (Vienna Test System). <http://www.schuhfried.com/vienna-test-system-vts/all-tests-from-a-z/test/hcta-halpern-critical-thinking-assessment-1/>

- Hurley, P. J. (2011). *A Concise Introduction to Logic, 11th ed.* Independence, KY: Cengage.
- Kraus, S., Sears, S. R., & Burke, B. L. (2012). *Is Truthiness Enough? Classroom activities for encouraging evidence-based critical thinking.* Manuscript in Preparation.
- Lawrence, T. (1995). How many factors of paranormal belief are there? A critique of the PBS. *Journal of Parapsychology, 59*, 3 – 25.
- Lawson, T. J. (1999). Assessing psychological critical thinking as a learning outcome for psychology majors. *Teaching of Psychology, 26*, 207-209.
- Lehman, D. R., & Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology, 26*, 952-960.
- Lilienfeld, S. O., Lohr, J. M., & Morier, D. (2001). The teaching of courses in the science and pseudoscience of psychology: Useful resources. *Teaching of Psychology, 28*, 182-191.
- Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2004). *Science and Pseudoscience in Clinical Psychology.* New York, NY: Guilford Press.
- Manza, L., Hilperts, K., Hindley, L., Marco, C., Santana, A., & Hawk, M. V. (2010). Exposure to science is not enough: The influence of classroom experiences on belief in paranormal phenomena. *Teaching of Psychology, 37*, 165-171. doi:10.1080/00986283.2010.488554
- McLean, C. P., & Miller, N. A. (2010). Changes in critical thinking skills following a course on science and pseudoscience: A quasi-experimental study. *Teaching of Psychology, 37*, 85-90. doi:10.1080/00986281003626714
- Myers, D. G. (2009). *Psychology: Ninth Edition.* New York, NY: Worth.
- Nieto, A. M., & Saiz, C. (2008). Evaluation of Halpern's "Structural Component" for improving Critical Thinking. *The Spanish Journal of Psychology, 11*, 266-274.

Paul, R. (1995). *Critical thinking: How to prepare students for a rapidly changing world.*

Rohnert Park, CA: Foundation for Critical Thinking.

Penningroth, S. L., Despain, L. H., & Gray, M. J. (2007). A course designed to improve psychological critical thinking. *Teaching of Psychology, 34*, 153-157.

doi:10.1080/00986280701498509

Sandor, M. K., Clark, M., Campbell, D., Rains, A. P., & Cascio, R. (1998). Evaluating critical thinking skills in a scenario-based community health course. *Journal of Community Health Nursing, 15*, 21-29. doi:10.1207/s15327655jchn1501_3

Scott, J. N., Markert, R. J., & Dunn, M. M. (1998). Critical Thinking: Change during medical school and relationship to performance in clinical clerkships. *Medical Education, 32*, 14-18. doi:10.1046/j.1365-2923.1998.00701.x

Shiah, Y., Tam, W. C., Wu, M., & Chang, F. (2010). Paranormal beliefs and religiosity: Chinese version of the Revised Paranormal Belief Scales. *Psychological Reports, 107*, 367-382.

Shin, K. R. (1998). Critical thinking ability and clinical decision-making skills among senior nursing students in associate and baccalaureate programmes in Korea. *Journal of Advanced Nursing, 27*, 414-418. doi:10.1046/j.1365-2648.1998.00499.x

Solon, T. (2007). Generic critical thinking infusion and course content learning in Introductory Psychology. *Journal of Instructional Psychology, 34*, 95-104.

Sparks, G. G., Nelson, C. L., & Campbell, R. G. (1997). The relationship between exposure to televised messages about paranormal phenomena and paranormal beliefs. *Journal of Broadcasting & Electronic Media, 41*, 345-359. doi:10.1080/08838159709364412

Stanovich, K. E. (2004). *How to think straight about psychology.* Boston, MA: Pearson Higher Education.

- Stark, E. (2012). Enhancing and assessing critical thinking in a Psychological Research Methods course. *Teaching of Psychology, 39*, 107-112. doi:10.1177/0098628312437725
- Tobacyk, J. (1984). Paranormal belief and college grade point average. *Psychological Reports, 54*, 217-218. doi:10.2466/pr0.1984.54.1.217
- Tobacyk, J. (2004). A revised paranormal belief scale. *The International Journal of Transpersonal Studies, 23*, 94-98.
- Tobacyk, J., & Milford, G. (1983). Belief in paranormal phenomena: Assessment instrument development and implications for personality functioning. *Journal of Personality and Social Psychology, 44*, 1029-1037. doi:10.1037//0022-3514.44.5.1029
- Valenzuela, J., Nieto, A. M., & Saiz, C. (2011). Critical Thinking Motivational Scale: a contribution to the study of relationship between critical thinking and motivation. *Electronic Journal of Research in Educational Psychology, 9*, 823-848.
- Watson, G., & Glaser, E. M. (1994). *Watson-Glaser Critical Thinking Appraisal: Form S manual*. San Antonio, TX: The Psychological Corporation.
- Watson, G., & Glaser, E. M. (2006). *Watson-Glaser Critical Thinking Appraisal: Short form manual*. San Antonio, TX: Pearson.
- Williams, E., Francis, L., & Lewis, C. A. (2009). Introducing the modified paranormal belief scale: Distinguishing between classic paranormal beliefs, religious paranormal beliefs and conventional religiosity among undergraduates in Northern Ireland and Wales. *Archive for the Psychology of Religions, 31*, 345-356.
- Williams, R. L., Oliver, R., Allin, J. L., Winn, B., & Booher, C. S. (2003). Psychological critical thinking as a course predictor and outcome variable. *Teaching of Psychology, 30*, 220-223. doi:10.1163/008467209X12499946199605

Wesp, R., & Montgomery, K. (1998). Developing critical thinking through the study of paranormal phenomena. *Teaching of Psychology, 25*, 275-278.

doi:10.1080/00986289809709714

Wyer, K. (2009). *U.S. faculty: Civic engagement, diversity important goals for undergraduate education*. Retrieved from <http://www.heri.ucla.edu/PDFs/press/pr030509-08faculty.pdf> on June 14, 2011.

Table 1

Course Differences in CT Belief Content & Skills

Class Type	<i>n</i>	RPBS		WGCTA	
		PRE <i>M (SD)</i>	POST <i>M (SD)</i>	PRE <i>M (SD)</i>	POST <i>M (SD)</i>
CT	12	89.70 (38.09)	63.90 (21.04)	25.17 (4.76)	24.33 (5.88)
Other Psych	90	^a 86.21 (21.15)	^a 84.29 (17.74)	24.94 (5.15)	24.97 (5.70)
Philosophy	26	76.27 (25.99)	60.62 (20.87)	26.62 (4.98)	28.81 (4.56)
<i>Total</i>	<i>128</i>	<i>81.74 (27.65)</i>	<i>67.90 (22.25)</i>	<i>25.30 (5.09)</i>	<i>25.68 (5.71)</i>

Note. ^aOnly 14 students (out of 90) in the other psychology classes took the RPBS at pre- and post-test due to a data collection error. CT = Critical Thinking; RPBS = Revised Paranormal Belief Scale; WGCTA = Watson-Glaser Short Form. Possible scores on the RPBS ranged from 26-182, and WGCTA from 0-40. CT Class = semester-long course dedicated to critical thinking in psychology; Other Psych(ology) Classes = Introduction to Psychology, Research Methods, Senior Seminar; Philosophy Classes = Logic, Introduction to Philosophy.

Table 2

ANOVAs for Effects of Class Types

Dependent Variable	<i>df</i>	<i>F</i>	<i>p</i>	<i>Partial eta squared</i>
RPBS				
Pre-post * Class Type	2, 47	3.94	.026*	.143
Pre-post	1, 47	20.16	.001***	.300
Class Type	2, 47	2.78	.072	.106
WGCTA				
Pre-post * Class Type	2, 125	2.28	.107	.035
Pre-post	1, 125	0.56	.457	.004
Class Type	2, 125	3.68	.028*	.056

Note. RPBS = Revised Paranormal Belief Scale; WGCTA = Watson-Glaser Short Form. Pre-post * Class Type represents the interaction between pre- and post-semester scores and class type (Critical Thinking in Psychology class, other psychology classes, philosophy classes). Pre-post alone represents main effect differences between before-and-after assessments aggregated across class types. Class type alone represents main effect differences in groups aggregated across pre-post assessments.

* $p < .05$. *** $p < .001$.