Learning and Individual Differences 36 (2014) 9–18

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

A new measure of psychological misconceptions: Relations with academic background, critical thinking, and acceptance of paranormal and pseudoscientific claims

D. Alan Bensley [a,*], Scott O. Lilienfeld [b,1], Lauren A. Powell [a,2]

[a] Department of Psychology, Frostburg State University, Frostburg, MD 21532 USA
[b] Department of Psychology, Emory University, 473 Psychology Building, 36 Eagle Row, Atlanta, GA 30322 USA

ARTICLE INFO

Article history:
Received 4 June 2013
Received in revised form 29 April 2014
Accepted 18 July 2014
Available online xxxx

Keywords:
Misconceptions
Intuition
Critical thinking
Knowledge of psychology
Paranormal belief

ABSTRACT

Many studies of psychological misconceptions have used tests with methodological and psychometric shortcomings, creating problems for interpreting individual differences related to misconceptions. To address these problems, we developed the Test of Psychological Knowledge and Misconceptions (TOPKAM), administering it to two samples of psychology students. Results from the first study ($N = 162$) supported the TOPKAM's internal consistency and showed that the number correct on the TOPKAM was significantly predicted by measures of paranormal belief, faith in intuition, the ability to distinguish scientific fields and practices from pseudoscientific ones, and SAT scores. Also, scores on a measure of critical thinking dispositions in psychology predicted TOPKAM scores. A second study ($N = 178$) supported the TOPKAM's test–retest reliability at four weeks and showed that TOPKAM scores were significantly predicted by the same critical thinking dispositions measure and also by scores on a test of critical thinking, argument analysis skill.

1. Introduction

1.1. Overview of psychological misconceptions

The study of misconceptions has become an important and frequently researched topic, partly because of the hope that science education can contribute to the rejection of incorrect but popular ideas. Several studies have shown that misconceptions regarding scientific issues are prevalent (e.g., Crowe & Miura, 1995; Swami et al., 2012). Of particular interest are the many studies suggesting that students are highly susceptible to psychological misconceptions (e.g., Brown, 1983; Kowalski & Taylor, 2009; Lamal, 1979; McKeachie, 1960; Standing & Huber, 2003; Vaughan, 1977). For example, students often believe incorrectly that people with schizophrenia have split personalities and that opposites tend to attract in romantic relationships. Because misconceptions are often resistant to traditional instruction (Best, 1982; Gardner & Dalsing, 1986; McKeachie, 1960; Vaughan, 1977), they are potentially an important obstacle to effective science teaching. Yet, the actual frequency of misconceptions and our understanding of them are limited because most studies assessing misconceptions have used tests with methodological and psychometric shortcomings.

The purpose of the present investigation is to report on the development and initial validation of a new psychological misconceptions test designed to remedy some of these problems. As part of its development, we investigated its relationship to several measures expected to be related to individual differences in learning that might further inform us about the nature of psychological misconceptions.

Taylor and Kowalski (2004, p. 15) defined misconceptions as "beliefs that are held contrary to known evidence." In the case of psychological misconceptions, the relevant known evidence is high quality research that supports well-established data and theories about human behavior and mental processes. As such, psychological misconceptions are widely-held beliefs, contrary to the well-replicated findings of psychological science. For example, a recent book discusses many misconceptions based on commonsense psychology including but not limited to such paranormal claims as extrasensory perception, the claim that the mind leaves the body during an out-of-body experience, and other false beliefs commonly associated with pseudoscience (Lilienfeld, Lynn, Ruscio, & Beyerstein, 2010).

Failure to reject these incorrect ideas may be due to a lack of (a) knowledge, (b) skills, or both needed to think scientifically about such questions. An alternative hypothesis is that individuals possess thinking styles and other enduring dispositions that dispose them to endorse poorly-supported claims. They may lack the interest or willingness

* Corresponding author. Tel.: +1 301 687 4195.
E-mail addresses: abensley@frostburg.edu (D.A. Bensley), slilien@emory.edu (S.O. Lilienfeld).
[1] Tel.: +1 404 727 1125.
[2] Tel.: +1 301 687 4195.

http://dx.doi.org/10.1016/j.lindif.2014.07.009
1041-6080/

to engage in the effortful processing and open-minded thinking needed to revise their incorrect beliefs. Or, they may be less willing than other individuals to rely upon a rational, scientific approach to evidence.

A third hypothesis is that both critical thinking (CT) knowledge/skills and thinking style/dispositions are related to endorsement of misconceptions. This view is consistent with cognitive-experiential self-theory (CEST), a dual-process theory proposed by Epstein (2008; Pacini & Epstein, 1999). According to CEST, people have an intuitive-experiential system that automatically learns from experience and is largely unconscious, and a second rational-analytic system for engaging in verbal reasoning that is conscious, deliberate, and analytic. The knowledge acquired through the intuitive-experiential system is tacit and more resistant to change than the knowledge acquired through rational-analytic thinking. Some dual-process theories associate intuitive thinking with processing in a heuristic-driven cognitive system called "System 1" and reflective thinking with an analytic system called "System 2" (Stanovich & West, 2000), see also Evans (2010), Evans and Stanovich (2013), and Kahneman (2011).

From the perspective of CEST, we might expect people who endorse unsubstantiated claims to be more intuitively-oriented, acquiring their misconceptions through experience and relying more on their tacit knowledge. They may also be less interested in seeking out new information that could disconfirm their experience-based knowledge and less inclined to analyze and reflect upon their misconceptions.

The differences in intuitive-experiential thinking and rational-analytic thinking seem to parallel the origins of misconceptions versus scientifically-supported beliefs. Misconceptions typically originate from such informal knowledge sources as everyday conversation, the media, works of fiction, and rumors (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012) and in other cases derive from misinterpretations of personal experience (Hughes, Lyddy, & Lambe, 2013). This information is seldom supported by high-quality evidence and is tacitly accepted because it seems familiar or intuitively true. In contrast, claims that achieve the status of scientific knowledge usually develop through careful analysis of systematically-collected observations, passing the effortful, deliberate scrutiny of researchers.

Indeed, some research shows that people who hold beliefs that lack empirical support tend to adopt an intuitive approach in their thinking. Saher and Lindeman (2005) found that people who endorsed greater belief in complementary and alternative medicine (CAM), the paranormal, and in magical food and health-related practices showed more faith in intuition. In contrast, those with a more rational thinking style showed less belief in the paranormal and in magical food- and health-related practices, but not less belief in CAM. These findings are consistent with a dual-process explanation, but no study has examined whether such explanations also apply to psychological misconceptions.

Nevertheless, a full understanding of psychological misconceptions is not possible without a reliable and valid test that is free from problematic response biases (see the next section). To this end, we report on the development and preliminary validation of a new measure called the Test of Psychological Knowledge and Misconceptions (TOPKAM), designed to avoid some of the shortcomings of previous tests. We also investigate individual differences in CT skills and dispositions, belief in pseudoscientific and unsubstantiated claims, as well as academic background variables potentially related to belief in psychological misconceptions.

1.2. Review of misconceptions tests and their problems

Since the seminal psychological misconceptions test of Nixon (1925), most tests have employed a true–false (T/F) response format (e.g., Brown, 1983; Gardner & Dalsing, 1986; Griggs & Ransdell, 1987; Gutman, 1979; Kuhle, Barber, & Bristol, 2009; Lamal, 1979; McKeachie, 1960; Taylor & Kowalski, 2004; Vaughan, 1977). Many using the T/F format have used the Test of Common Beliefs (TCB) of Vaughan (1977) or items from it to assess introductory psychology students' psychological misconceptions (e.g., Gardner & Dalsing, 1986; Griggs & Ransdell, 1987; Gutman, 1979; Kuhle et al., 2009; Landau & Bavaria, 2003). Each of the 80 T/F items on the TCB is scored as correct when answered false.

The use of T/F response format in misconceptions tests, especially in which true responses are scored as misconceptions, can create problems when interpreting scores. For example, a yea-saying response style (acquiescence) could lead to inflated estimates of their susceptibility to misconceptions; whereas, nay-saying (counteracquiescence) could deflate estimates. Conversely, negatively keyed items could induce a response set in which some respondents who are biased in their responding to appear more positive or agreeable would produce inaccurate estimates of knowledge. In addition, T/F format with correct items always keyed false could make it easier to guess correctly when respondents discerned the pattern of correct answers in the format of the test.

Other researchers have criticized misconception items with the T/F format on the grounds that they constrain responses to be completely true or completely false, a position that does not accurately capture the difference between most misconceptions and scientifically-supported ideas in psychology. For example, Brown (1984) provided several examples of misconception items written in language that allowed them to be interpreted as at least partly true. Ruble (1986) argued that because some items are too ambiguous to be answered as completely true or false, qualifiers should sometimes be used. Supporting this objection, Hughes, Lyddy, and Kaplan (2013) found that the language and response format of items in a misconceptions test affected the level of endorsement of misconceptions, with ambiguously phrased items yielding higher levels of misconceptions than non-ambiguously-phrased items. Moreover, the T/F format used in many misconceptions tests is inconsistent with the provisional status of knowledge in science. Specifically, the inductive and informal reasoning used to build scientific theories is defeasible, often resulting in conclusions that are only tentative and qualified. Indeed, many psychological misconceptions contain a kernel of truth (Hughes, Lyddy, & Lambe, 2013; Lilienfeld et al., 2010). For example, although the claim that some people are exclusively "left-brained" and others "right-brained" is false, it is at least partly true that the brain's two hemispheres subserve somewhat different functions.

Yet another criticism of most T/F format tests is that they do not allow respondents to indicate that they do not know an answer. To control for this limitation, Gardner and Dalsing (1986) administered a 60-item version of the TCB to 531 college students in T/F format but added a third option of "don't know/no opinion." They found that students chose this option 12.2% of the time. After discarding these responses and calculating misconceptions only from the remaining responses, they found that this change reduced the level of misconceptions by 8% on 14 common items. Although this strategy may control for guessing, it produces total test scores that are based on an unequal number of responses to items. Moreover, judging that one does not know an answer or has no opinion about a question is not necessarily equivalent to the more continuously varying judgment of one's ability to provide a correct answer. The ability to accurately assess the veracity of one's own knowledge is better viewed as a metacognitive dimension in which respondents judge the certainty of the correctness of their answers. Another potential problem is that responding with "no opinion" about a question might indicate a lack of motivation to answer the question. This ambiguity suggests the need to separate the assessment of a knowledge dimension underlying misconceptions from the metacognitive dimension reflected by confidence or certainty in a knowledge response.

One study, conducted by Landau and Bavaria (2003), has assessed confidence on a continuous scale, asking respondents to rate their confidence after answering each question using a 5-point Likert scale. They found that respondents were significantly more confident on incorrect items (misconceptions) than on items they got correct, consistent with the hypothesis that most people are not aware that they are endorsing misconceptions.

Few studies have dealt with the problems of the T/F format when misconceptions are always associated with a true response. In one study, Brown (1983) reworded 18 of 37 false items obtained from lists of misconceptions in instructional and other materials. He found that only 19 of the 37 (both true and false) items were missed by at least 50% of the students and concluded that misconceptions may be less frequent than supposed. In another study, Kowalski and Taylor (2009) developed a true–false instrument designed to measure adherence to psychological misconceptions along with knowledge of psychology. About half of their test items were false when correct, and misconceptions were intermixed with more conventional general psychology questions. Although their new test showed clear improvements over previous T/F misconception tests, Kowalski and Taylor did not report the reliability and validity of their new instrument, and did not assess guessing or other metacognitive aspects of response.

Exploring another alternative to T/F misconception tests, McCutcheon (1991) developed a 62-item, multiple-choice test with response options that presented both factual and incorrect psychological information. Although the multiple-choice response format may have lowered the probability that respondents would guess half of the items correctly, wording of the response options was sometimes inconsistent and seemed to target different aspects of a psychological construct within the same item.

Support for the validity of this misconceptions test (see also McCutcheon, 1991) came when McCutcheon, Apperson, Hanson, and Wynn (1992) found that performance on the Watson-Glaser Critical Thinking Test and GPA predicted performance on this test. Taylor and Kowalski (2004) similarly found that performance on their misconception test was positively correlated with six items from the Scottsdale Critical Thinking Test. The results of these studies support the hypothesis that endorsing misconceptions is associated with poorer CT skills; but they do not address other aspects of CT, such as thinking dispositions and metacognition (Bensley, 2011a; Halpern, 1998).

In a more recent attempt to rectify the problems of T/F tests, Gardner and Brown (2013) developed a new test of psychological misconceptions based on the *50 Great Myths of Popular Psychology* of Lilienfeld et al. (2010). They stated some misconceptions in T format and others in F format to examine the effect of wording the truth value of items. To take into account the fact that misconceptions are not completely false, the test employs a Likert-type scale assessing endorsement of misconceptions on a scale ranging from "completely false" to "completely true." Furthermore, to take guessing into account, the test allows respondents to report that they did not know the answer, using the "don't know/no opinion" option of Gardner and Dalsing (1986). Although their test showed good internal consistency, problems remain with regard to interpreting "don't know/no opinion" responses.

1.3. Development of a new test of psychological misconceptions

We developed the Test of Psychological Knowledge and Misconceptions (TOPKAM) to address the limitations of earlier misconception tests (Bensley & Lilienfeld, 2010). To reduce potential bias in responding associated with the T/F format, we constructed the test in a forced-choice format in which an evidence-based response option representing factual knowledge from psychology is counterposed against an alternative corresponding to the misconception. Each item presents a common misconception paired with an evidence-based response option contradicting the misconception based on literature reviews found in the Lilienfeld et al. (2010) book and other sources. For example, a common misconception discussed in the Lilienfeld et al. (2010) book is that venting of anger is a good way to control it. This was expressed in the TOPKAM false option as "It is better to express your anger or 'blow off steam' than to hold it in." Contrary to this false option, we constructed the correct option, "It is better to control the expression of your anger" consistent with the research (Bushman, Baumeister, & Strack, 1999).

To address the fact that psychological knowledge is tentative and rarely is completely true or false, TOPKAM's general instructions ask test takers to answer questions by judging which option is "best" in each question. Likewise, question stems ask them to select the option that is "most true."

To address the problem of guessing and the unequal number of responses comprising test scores when "don't know/no opinion" leads to eliminating responses, the TOPKAM treats guessing as part of a separate dimension of certainty (analyses of correlations of scores on this dimension are reported in separate manuscripts). Specifically, respondents rate the certainty of the correctness of their answer following each question.

To evaluate the psychometric quality of the TOPKAM, we used two samples to assess its reliability and validity. We examined its internal consistency, reliability, and concurrent validity with measures of individual difference variables thought to be related to misconceptions. In the first study, we examined TOPKAM's concurrent validity by administering it with measures of knowledge of science, pseudoscience, paranormal belief, and CT dispositions. We also examined its relation to academic background variables, such as GPA, SAT, and number of course credits earned, all of which would ostensibly be related to performance on a knowledge-based test.

2. Study 1

2.1. Method

2.1.1. Participants

We assessed 205 students enrolled in psychology courses during the 2010–2011 academic year at a small, mid-Atlantic university, retaining data from 162 participants who completed all TOPKAM items and corresponding certainty ratings. This final sample included 36 males and 126 females with ages ranging from 18 to 47 years ($M = 22.20$; $SD = 4.49$). Of those reporting ethnic origin, 76.3% reported white (non-Hispanic), 20.6% reported African American, 1% reported Latino or Hispanic, and 2.1% reported "other". Overall, the 81 beginning majors included 8 first years, 46 sophomores, and 27 juniors (excluding seniors) from beginning psychology classes, as well as 49 seniors from senior capstone courses, 11 students beginning a graduate program in counseling psychology, 20 second- and third-year graduate students from the same program who had finished the academic part of their program, and 1 not reporting class rank. Except for graduate students and some senior internship students, all students received course credit for their participation.

2.1.2. Measures

The TOPKAM contains 40 questions presented in a forced-choice, two-response format for assessing both factual knowledge of psychology and susceptibility to common psychological misconceptions (Bensley & Lilienfeld, 2010). For questions, we selected 40 of the misconceptions discussed in the 50 major essays reviewing literature on widespread psychological misconceptions in Lilienfeld et al. (2010). To ensure content validity, the 40 items were selected to represent each of the 11 general categories of psychological myths in the Lilienfeld et al. book. The domains comprised myths about the brain and perception, development and aging, memory, intelligence and learning, consciousness, emotion and motivation, interpersonal behavior, personality, mental illness, psychology and the law, and psychological treatments. Scores on the TOPKAM range from 0 to 40 correct with higher scores reflecting more accurate knowledge (i.e., greater rejection of the alternate misconception options).

Before the 40 items, general instructions asked respondents, "Please use what you know about psychology to select the best answer for each question below" and then further requested that they rate their certainty of the correctness of their answer for each question. They were instructed to use a 5-point Likert scale ranging from 1 = *not at all*

certain to 5 = *completely certain* that appeared at the end of instructions and before the first item. Instructions asked students to answer each test question by bubbling in their answer on a Scantron form and then to write the certainty rating next to the number of the question after answering it. After the TOPKAM, other metacognitive questions (reported elsewhere) asked students to postdict the accuracy of their answer. Postdiction involves providing an immediate, retrospective measure of metacognitive monitoring (Hacker, Bol, & Keener, 2008).

TOPKAM items described misconceptions in language that combined everyday language with scientific language while avoiding jargon. Each item began with "Which is most true…". For example, one question asked, "Which is most true about the Rorschach (inkblot) Test?" followed by two response options: "a. It is like a 'psychological X-ray' because it can penetrate the unconscious mind and tell a great deal about personality" versus "b. It can detect marked thinking disturbances but is not effective in detecting depression or anxiety disorders." The "b" option is correct based on many studies showing that many or most indices of the Rorschach (Inkblot) Test lack adequate scoring reliability for most psychiatric diagnoses (Wood, Nezworski, Lilienfeld, & Garb, 2003). A second item, asked, "Which is most true about how much of the brain people typically use?" followed by "a. People use most of their brains, but not all at once." and "b. Most people only use about 10% of their brains." The answer to this question is "a" because research using various methods of study clearly show that virtually all parts of the brain are active but at different times depending on the task and behavior (Beyerstein, 1999; Lilienfeld et al., 2010). All TOPKAM items were randomized with an equal number of "a" and "b" response options correct.

To assess the ability to distinguish poorly-supported and pseudoscientific treatments and practices from scientifically-supported (evidence-based) treatments and practices, we administered the Test of Evidence-based Theories and Practices—Revised (TEBTP) of Bensley, Edwards, and Murtagh (Bensley, 2011b). The 40-item TEBTP contains one item for each of 40 different therapies, treatments, or practices with one 20-item scale designed to assess knowledge of poorly-supported and pseudoscientific practices and another 20-item scale for well-supported ones. Respondents rated how well each claim is supported by high quality scientific research on a 5-point Likert scale: 1 = *not at all well supported* to 5 = *very well supported*. The 20-item evidence-based scale contains statements about evidence-based treatments and therapies supported by high-quality outcome research (Chambless & Ollendick, 2001). The second 20-item, non-evidence-based scale includes poorly-supported and pseudoscientific treatments and practices as drawn from the literature on pseudoscientific and poorly supported practices (Lilienfeld, Lynn, & Lohr, 2003) and a Delphi study of clinicians rating poorly supported treatments and practices (Norcross, Koocher, & Garafalo, 2006). Scores on each 20-item scale range from 20 to 100. Bensley and Crowe (2010) reported that the four-week, test–retest reliability of the TEBTP on a sample of undergraduate and graduate psychology students was $r(135) = .77$, and the internal consistency of both subscales was *alpha* = .82; the *alpha* for the total score was .84. Supporting its convergent validity, both seniors and graduate students were significantly more accurate in rating support than were beginning majors. Scores on the evidence-based scale were positively and significantly correlated with the ITDP-R and the Scientist-Practitioner Inventory practitioner scale.

To assess the ability to distinguish pseudoscientific from scientific fields, participants completed the 14-item Knowledge of Fields Inventory (KOF) on which they rated how scientific each of the 14 fields are on a 5-point Likert scale: 1 = *not at all scientific* to 5 = *very scientific* (Bensley, 2011b). Possible scores on each of the two, 7-item scales range from 7 to 35. Respondents rated the seven scientific fields and approaches (astronomy, chemistry, cognitive behavior therapy, evolutionary biology, neuroscience, physics, and psychology) as significantly more scientific than the pseudosciences (alchemy, astrology, creation science-intelligent design, Freudian psychoanalysis, parapsychology, phrenology, and Scientology; see Bensley and Crowe, 2010). When scores are reversed on the pseudoscience scale and added to the science scale, the full-scale KOF indicates the ability of respondents to distinguish pseudosciences from sciences. Bensley and Crowe (2010) found that KOF pseudoscience scores were significantly correlated with scores on the TEBTP poorly-supported practices scale, but little is known about the reliability of the KOF.

To assess paranormal belief, participants completed the 26-item Revised Paranormal Beliefs scale (RPBS) of Tobacyk (2004) with each item describing a different paranormal phenomenon rated on 7-point Likert scale: 1 = *strongly disagree* to 7 = *strongly agree*. Revision of the original PBS increased the four-week, test–retest reliabilities of its subscales to levels ranging from $r = .71$ to .95 and its full-scale reliability to .95. Tobacyk (2004) cited numerous studies supporting its validity.

To assess individual differences in thinking style, we used four measures. The first two were versions of the 20-item scales, the intuitive-experiential and the rational-analytic scales, adapted from the Rational-Experiential Inventory of Pacini and Epstein (1999). Each of the 20 items on the respective scales was rated on a 5-point Likert scale: 1 = *not at all true of self* to 5 = *very true of self*, yielding possible scores of 20 to 100. The rational scale was based on the Need for Cognition scale of Cacioppo and Petty (1982). This scale has been found to be reliable and to be correlated with cognitive innovativeness, openness to experience, the need to evaluate information, and scores on the ACT (Cacioppo, Petty, Feinstein, & Jarvis, 1996). The intuitive-experiential scale also called the "faith in intuition" (FI) scale has high internal consistency (*alpha* = .88) and is essentially uncorrelated with the rational-analytic scale (Pacini & Epstein, 1999). Other studies have found that FI is associated with scores on the RPBS (Irwin & Young, 2002; Saher & Lindeman, 2005).

We also administered the Objectivism scale (OBJ) of Leary, Shepperd, McNeill, Jenkins, and Barnes (1986) to measure the dispositions to adopt an objective, rational approach to information. The OBJ contains 11 items rated on a 5-point Likert scale, with possible scores ranging from 11 to 55. Research demonstrates that it possesses adequate internal consistency (*alpha* = .80) and exhibits significant positive correlations with NFC.

Finally, to measure thinking dispositions more related to psychology, we administered the Inventory of Thinking Dispositions in Psychology—Revised (ITDP-R), a 24-item, self-report inventory designed to assess active, open-minded interest in psychology and the disposition to take a rational, scientific approach to psychology (Bensley, 2011b). Respondents rate each of its statements on a 5-point Likert scale: 1 = *not at all true of me* to 5 = *very true of me* with possible scores ranging from 24 to 120. Research on the ITDP-R has shown its test–retest reliability at four weeks to be $r = .74$ and its internal consistency (*alpha*) to be .84. Significant positive correlations with OBJ and NFC provided support for its concurrent validity (Parsons, Powell, Bensley, & Crowe, 2010).

To measure more content-related interest in psychology, we administered the Scientist-Practitioner Interest Inventory-Brief Version B (SPI-20), a short (20-item) form of the Scientist-Practitioner Inventory of Leong and Zachar (1993) that contains 10 items each assessing interest in the Scientist role and Practitioner role, respectively. The SPI-20 demonstrated good test–retest reliabilities at six months of .80 and .93 for the scientist and practitioner roles respectively (Leong & Zachar, 1993). Evidence of its concurrent validity comes from a significant, positive correlation between the scientist scale and the Holland Vocational Preference Inventory Investigative scale. The practitioner scale, in contrast, was negatively correlated with the Investigative scale and was positively correlated with the Social scale.

To assess students' test-taking motivation, participants completed the 10-item Student Opinion scale (SOS) of Sundre (2007). Each item of the SOS is reported on a scale with five different categorical response options ranging from A = *strongly disagree* to E = *strongly agree*. The SOS contains two factors, effort expended and assessment importance, with reliabilities ranging from .80 to .89 in a sample of over 15,000 students (Sundre & Moore, 2002).

Finally, to better understand the possible contributions of academic ability, achievement, and experience, we obtained academic background

information from the university's institutional research office. This information included psychology GPA, overall GPA, SAT scores, and the number of credits completed both in psychology and overall prior to the semester in which students were assessed. All measures, including the collection of academic background information, were approved by the local IRB.

2.2. Procedure

During their regularly scheduled classes, the first author administered the TOPKAM and the other measures to beginning majors early in the fall semester and to senior majors in the middle of the spring semester. He distributed the assessment forms in a booklet with the consent form on top followed by the ITDP-R, a demographic form, the TEBTP, the SPI, the KOF, the Rational scale, the Intuitive scale, the OBJ, and the RPBS. On a separate form, students received the TOPKAM, along with a Scantron form and finally the SOS, and a form for estimating their total number of correct answers. He discussed the consent form and instructed all students to carefully read the instructions, do their best, report honestly, and complete all forms in the order they were presented. Almost all students completed the booklet in the 50 minute class period, but were given an additional few minutes to finish if they needed it.

2.3. Results and discussion: study 1

After reversing appropriate items on the ITDP-R, Rational scale, Intuitive scale, OBJ, SPI, KOF-Pseudoscience scale, TEBTP not evidence-based scale, and the RPBS, we summed scores to create total scores for the respective scales. We used the total number correct on the TOPKAM as a measure of accuracy of psychological knowledge in all analyses (the total number of misconceptions endorsed can be readily calculated by subtracting the number correct from 40). We reversed the TEBTP not evidence-based items and the KOF pseudoscience items, so that higher scores indicated that these items lacked good scientific support and were not scientific, allowing them to go in the same direction as the TEBTP evidence-based items and KOF sciences, respectively. We converted the unnumbered categorical responses of the SOS to numeric values corresponding *strongly disagree* to 1, *disagree* to 2, *neutral* to 3, *agree* to 4, and *strongly agree* to 5. We reverse-scored four of the SOS items and then summed scores to create SOS interest and importance scales.

We eliminated a participant's score on the TOPKAM if the respondent was missing any value on the measure ($n = 16$) and those who did not complete the TOPKAM certainty ratings ($n = 24$). Completing the certainty ratings was considered an integral part of answering the questions and following instructions, and missing any misconception item would not allow us to accurately estimate the total TOPKAM score, as each question was independent and not duplicated. We compared the TOPKAM scores of the 24 students who did not complete the certainty ratings to the 162 who did and found no significant difference in the scores, $t(182) = 1.03$, $p = .31$ although the mean of those completing the certainty ratings ($M = 21.44$, $SD = 5.24$) was slightly higher than the mean of those who did not ($M = 20.29$, $SD = 3.95$). The results of analyses of the certainty data are reported elsewhere (Bensley & Lilienfeld, 2014).

Next, comparisons of the TOPKAM scores of male and female students revealed no significant sex differences on either the TOPKAM number correct or the sum of certainty scores. Consequently, sex of participant was not included in any further analyses. In contrast, we found that age was significantly correlated with the number correct on the TOPKAM, $r(157) = .23$, $p < .01$. A t-test on age, $t(33) = 3.78$, $p = .001$ (equal variances not assumed), showed that graduate students ($M = 25.97$, $SD = 6.55$) were significantly older than undergraduates ($M = 21.32$, $SD = 3.33$); this difference was large in magnitude (Cohen's $d = .89$). Likewise, a second t-test, $t(157) = 5.57$, $p < .001$, showed the mean TOPKAM score of graduate students ($M = 25.10$, $SD = 4.49$) was significantly higher than the undergraduate mean ($M = 20.57$, $SD = 5.04$); this difference was again large in magnitude ($d = .90$). Reanalyzing the correlation between age and TOPKAM scores of only the undergraduates showed that it was no longer significant without graduate students, $r(127) = .14$, $p = .12$. Because our intent was to include psychology students at various educational levels in our sample, the main analyses that follow report results from the entire sample unless otherwise noted, but we also report results that changed when graduate students were omitted from such analyses.

The internal consistency of the TOPKAM across the entire sample using the Kuder–Richardson coefficient was KR-20 = .74, suggesting adequate, although not extremely high, internal consistency. The internal consistency of the TOPKAM for undergraduates was KR-20 = .71, and for graduate students, KR-20 = .70.

To examine the concurrent validity of the TOPKAM with measures related to knowledge of science, pseudoscience, and paranormal belief, we calculated Pearson correlations between TOPKAM and the scores from the relevant measures. As seen in Table 1, the number correct on the TOPKAM was significantly and positively correlated with KOF and TEBTP full scale scores. These positive correlations showed that as students' knowledge of psychology increased, their ability to distinguish sciences from pseudosciences and evidence-based from poorly supported practices increased. The negative correlation between TOPKAM correct and RPBS showed that better knowledge of psychology on the TOPKAM was associated with less paranormal belief. Analyses of the undergraduate data using the same measures showed the same pattern of significant correlations as in the entire sample. Table 1 also shows the mean number correct on the TOPKAM and the other measures. The mean number of misconceptions (40—TOPKAM correct) was 21.44 or 53.6% of the items missed, a rather high percentage and significantly greater than 20 (expected by chance), $t(161) = 3.50$, $p = .001$.

To test the significance of the contribution of the variables significantly correlated with the TOPKAM, we conducted a standard (simultaneous) multiple regression, regressing KOF, TEBTP, and RPBS onto the TOPKAM scores from the entire sample. The results were significant, $F(3,126) = 19.27$, $p < .001$, with the model accounting for 31.4% of the variance, $R^2 = .31$, adjusted $R^2 = .30$. Knowledge of Fields scores significantly predicted the number correct on the TOPKAM, $B = .30$, $t(126) = 3.32$, $p = .001$, as did scores on the Test of Evidence-based Theories and Practices, $B = .27$, $t(126) = 3.03$, $p < .01$ and scores on the Revised Paranormal Belief scale, $B = -.20$, $t(126) = -2.71$, $p < .01$.

To more closely examine the basis of these correlations, we examined the correlations between TOPKAM scores and KOF and TEBTP subscale scores. KOF pseudosciences (reversed scores) were significantly correlated with TOPKAM correct, $r(152) = .50$, $p < .001$, showing that

Table 1
Descriptive statistics and bivariate correlations for the test of psychological knowledge and misconceptions and other knowledge measures.

Measure	1	2	3	4
1. TOPKAM	–	.49***	.46***	−.22**
N =		155	152	140
2. KOF (full scale)		–	.61***	−.12
N =			148	137
3. TEBTP (full scale)			–	−.06
N =				134
4. Paranormal belief				–
M	21.44	53.95	138.60	81.60
SD	5.24	6.74	13.73	24.73

Note. TOPKAM = Test of Psychological Knowledge and Misconceptions of Bensley and Lilienfeld (2010); KOF = Knowledge of Fields from Bensley (2011b); TEBTP = Test of Evidence-based Theories and Practices from Bensley (2011b); Paranormal Belief = Revised Paranormal Belief scale of Tobacyk (2004).

* p < .05.
** p < .01.
*** p < .001.

as participants scored better on the TOPKAM they rated pseudosciences as less scientific; however, no significant correlation was found with KOF sciences. Similarly, TEBTP poorly-supported practice reversed scores were significantly correlated with the TOPKAM, $r(150) = .51$, $p < .001$, showing that as participants scored better on the TOPKAM they rated poorly-supported practices as having less scientific support; however, no significant correlation was found with TEBTP evidence-based practices. A standard multiple regression analysis regressing KOF-pseudosciences and TEBTP poorly-supported practices and RPBS onto TOPKAM scores was significant, $F(3,127) = 18.80$, $p < .001$, with the model accounting for 30.8% of the variance, $R^2 = .31$, adjusted $R^2 = .29$. KOF pseudoscience scores significantly predicted the number correct on the TOPKAM, $B = .21$, $t(126) = 2.04$, $p < .05$, as did scores on the TEBTP poorly-supported practices, $B = .34$, $t(126) = 3.26$, $p = .001$; however, the RPBS was not a significant predictor of TOPKAM scores as it had been in the multiple regression with full-scale KOF and TEBTP scores as predictors.

Next, we calculated correlations between TOPKAM correct and four measures of thinking dispositions in the entire sample. Examination of ITDP-R scores showed positive kurtosis and a rightward skew. We used a square root transformation of the original ITDP-R scores to normalize them (see Tabachnick & Fidell, 2007) and repeated all analyses on the transformed scores. As shown in Table 2, the TOPKAM exhibited a significant, positive correlation with the transformed ITDP-R scores and a significant, negative correlation with FI (intuition scores). Although both NFC and OBJ were positively correlated with ITDP-R, supporting its validity, neither was significantly correlated with TOPKAM scores. Analyses of the undergraduate data using the same measures showed a similar pattern of correlations as in the entire sample; however, only FI was significantly correlated with the TOPKAM, $r(129) = -.18$, $p < .05$, and the transformed ITDP-R scores were no longer significantly correlated with it, $r(129) = .10$, $p = .26$.

A standard multiple regression analysis regressing ITDP-R and FI scores onto TOPKAM scores from the entire sample was significant, $F(2,155) = 5.35$, $p < .01$, with the model accounting for 6.5% of the variance, $R^2 = .07$, adjusted $R^2 = .05$. Transformed ITDP-R scores significantly predicted the number correct on the TOPKAM, $B = .18$, $t(154) = 2.37$, $p < .02$, as did scores on the FI, $B = -.19$, $t(154) = -2.38$, $p < .02$.

Further analyses showed that FI was positively correlated with RPBS, $r(137) = .18$, $p = .03$ and negatively correlated with OBJ, $r(156) = -.18$, $p = .02$. Thus, as students tended to trust their intuition more, they believed paranormal claims more and relied less upon objective evidence. Although NFC did not correlate significantly with TOPKAM scores, it was positively correlated with scores on the KOF recoded pseudoscience scale, $r(152) = .17$, $p = .03$, showing that as participants reported being more intellectually-engaged their recoded ratings of pseudosciences indicated that they recognized them to be less scientific. In a complementary fashion, the positive correlation of NFC with scores on the KOF Sciences scale, $r(152) = .23$, $p < .01$ showed that the more intellectually-engaged students rated the sciences as more scientific. As in earlier research with the REI (Pacini & Epstein, 1999), NFC and FI were not significantly correlated, suggesting that they measure different thinking styles.

Other analyses indicated that TOPKAM scores were correlated with some kinds of interest but not others. ITDP-R scores, which correlated with TOPKAM scores, also showed significant positive correlations with SPI interest in the scientist role, $r(150) = .28$, $p < .001$, SPI interest in the practitioner role, $r(150) = .43$, $p < .001$, and with SOS Importance (a scale containing two interest items), $r(153) = .27$, $p = .001$. Despite the positive correlation between ITDP-R and TOPKAM, none of these three other interest measures correlated significantly with TOPKAM scores. This may be because the ITDP-R measures actively open-minded interest in psychology and the endorsement of a rational, scientific approach to psychology; whereas, the SPI scales measure interest in professional roles and the SOS measures interest and engagement in the assessment activity. This pattern of correlations between TOPKAM and actively open-minded interest in psychology, but not with interest in professional roles or in the assessment activity, provides evidence of its discriminant validity.

To further examine the concurrent validity of TOPKAM with measures of academic achievement and aptitude, we correlated TOPKAM scores with SAT, overall GPA, psychology GPA, total psychology credits earned, and total non-psychology credits earned by the undergraduates in the sample. Because two of these variables, overall GPA and Psychology GPA, showed positive kurtosis and negative skewness, we transformed the scores by first reflecting them on each variable and then calculating the square root of those values (see Tabachnick & Fidell, 2007). TOPKAM scores were significantly correlated with SAT, $r(87) = .45$, $p < .001$ and with total psychology credits earned, $r(89) = .27$, $p < .05$, but not with the transformed scores of either overall GPA or psychology GPA. Because psychology GPA was positively correlated with number of psychology credits, $r(93) = .21$, $p < .05$, we controlled statistically for the contribution of psychology GPA to TOPKAM scores. When we calculated the partial correlation between TOPKAM and the number of psychology credits, partialling out psychology GPA, the correlation remained significant, $r(82) = .26$, $p < .05$; but the partial correlation between TOPKAM and psychology GPA, controlling for number of psychology credits, was not significant. This finding suggests that course experience but not grades in psychology accounted for the TOPKAM scores.

The results of the first study suggest that our attempt to create a new psychological misconceptions test, addressing psychometric shortcomings of previous tests, has yielded a measure with adequate internal consistency and concurrent validity. In particular, we found the ability to distinguish sciences from pseudosciences and evidence-based from poorly-supported practices predicted scores on the TOPKAM while low scores on the Revised Paranormal Belief scale predicted higher scores on the TOPKAM.

Other new results included the finding that those who scored better on the TOPKAM showed differences in thinking style from those who held more misconceptions. Scores on the ITDP-R, a measure of actively open-minded interest in psychology and the tendency to take a rational, scientific approach to psychology significantly predicted higher TOPKAM scores; whereas, higher scores on the FI or the tendency to take a more intuitive approach predicted lower scores. Taken together, these findings are consistent with dual-process theory in that a more actively open-minded, rational scientific approach and less trust in intuition predicted better scientifically-based knowledge of psychology and a concomitant greater rejection of misconceptions.

Table 2
Descriptive statistics and bivariate correlations for the test of psychological knowledge and misconceptions and measures of thinking dispositions.

Measure	1	2	3	4	5
1. TOPKAM	–	.18*	.10	.14	−.17*
N =		159	160	159	159
2. ITDP-R		–	.53***	.36***	.05
N =			159	158	158
3. Rational			–	.34***	.01
N =				159	159
4. OBJ				–	−.18*
N =					158
5. Intuitive					–
M	21.44	102.73	77.23	40.29	67.92
SD	5.24	9.15	9.50	5.19	12.93

Note. TOPKAM = Test of Psychological Knowledge and Misconceptions from Bensley and Lilienfeld (2010); ITDP-R = Inventory of Thinking Dispositions in Psychology—Revised from Bensley (2011b); Rational = Rational-Analytic scale adapted from the Rational-Experiential Inventory of Pacini and Epstein (1999); OBJ = Objectivism scale of Leary et al. (1986); Intuitive = Intuitive-Experiential scale adapted from the Rational-Experiential Inventory of Pacini and Epstein (1999).

* $p < .05$.
** $p < .01$.
*** $p < .001$.

The finding that students who hold more psychological misconceptions differ in thinking style from those who possess more accurate knowledge further supports a connection with critical thinking that has been demonstrated in other studies (McCutcheon et al., 1992; Taylor & Kowalski, 2004). Those scoring higher on the TOPKAM also tended to have higher SAT scores and more psychology course experience. CT skill has often been associated with better SAT test performance (Erwin & Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, 2000). The results of the first study imply that there is a relationship between TOPKAM performance and both the skills and dispositions for thinking critically in psychology.

Accordingly, in the second study we investigated performance on the TOPKAM in relation to scores on the ITDP-R, the measure of CT disposition used in the first study, and also in relation to performance on a test of CT knowledge and skill in psychology. To further examine the reliability of the TOPKAM, we administered it and the other CT measures again four weeks after the first administration. We expected that the test–retest reliability of the TOPKAM would be adequate and that all three measures would be intercorrelated at both administrations.

3. Study 2

3.1. Method

3.1.1. Participants

The second sample included undergraduate psychology majors assessed during the 2011–2012 academic year at the same university as in the first study and who had not been tested on the TOPKAM and who had completed all TOPKAM items and certainty ratings. The students came from a beginning critical thinking course ($n = 52$), a beginning introduction to the profession of psychology course ($n = 66$) with 11.4% first years, 30.3% sophomores, 24.0% juniors and students completing the major including 34.3% seniors from various senior psychology capstone courses ($n = 60$). In the second study, no students were excluded from the first TOPKAM administration for omitting certainty ratings, but seven were excluded for not answering all TOPKAM questions. Also excluded were five students from the CT course who reported on a post-experimental questionnaire that they had read portions of the Lilienfeld et al. (2010) misconceptions book before the first day of class when assessment was conducted. The final sample of 183 students included 39 males, 134 females, and five not reporting their sex, ranging in age from 18 to 54 years ($M = 20.68$; $SD = 3.51$). Of those reporting ethnic origin, 66.1% reported white (non-Hispanic), 28.6% reported African American, 2.4% reported Latino/a, and .6% reported Asian American and 2.4% "other."

3.1.2. Measures

Participants completed the TOPKAM, the ITDP-R, the SOS, and the same demographic form as in the first study. Because we were interested in psychological misconceptions, we used a measure of CT knowledge and skill in psychology, a 20-item, multiple-choice, argument analysis test called Analyzing Psychological Statements (APS) developed by Bensley (2011a). Little is known about the psychometrics of the APS, but a slightly shorter version showed sensitivity to change following CT instruction (Bensley, Crowe, Bernhardt, Buckner, & Allman, 2010). The revised APS included seven items describing everyday, psychology-related situations and 13 describing psychological research or clinical practice examples. Of the 20 items, three were designed to test the ability to distinguish arguments from non-arguments, seven the ability to identify and evaluate kinds of evidence, three on finding assumptions, three on drawing appropriate conclusions, and four on identifying problems in reasoning about psychology-related questions. The consent form, TOPKAM, APS, ITDP-R, SOS, and demographic form were assembled in that order into a single booklet. All measures were again approved by the local IRB.

3.1.3. Procedure

During their regularly scheduled classes, the first author assessed participants following a similar general instructional procedure as in the first study, passing out a single booklet of forms containing the TOPKAM and new measures along with two Scantron forms. He first assessed students in the CT course on the first day of class and students in the introduction to the field of psychology course in the third week of classes. Most of the students in the senior seminars were tested in the early to middle part of the semester. He distributed the assessment booklet and two Scantron forms for completing the TOPKAM and the APS, respectively. After explaining the consent form, he read the TOPKAM test instructions to make sure students knew how to complete the certainty ratings on the Scantron form. Students in the classes were tested again four weeks after their first testing, following the same procedure as at the first testing and giving them new copies of the same booklet and Scantrons. During each test session, most students completed the booklet within 50 min, but if they needed more time, they were given a few minutes to finish.

3.2. Results and discussion

We scored the ITDP-R and TOPKAM as in Study 1. The APS was scored as the number correct out of 20. As in the first study, comparisons of male and female students revealed no significant sex differences on the TOPKAM number correct in either the first administration, $t(166) = 1.55$, $p = .12$ or the second administration, $t(118) = 1.29$, $p = .20$; nor were there differences on certainty at either the first administration, $t(162) = 1.77$, $p = .08$ or the second administration, $t(119) = 1.48$, $p = .14$. Consequently, the following analyses do not include sex as a variable. Nor did we find any significant correlations between age and the TOPKAM at the first administration; however, age did show a low, but significant, positive correlation with the TOPKAM in the second administration, $r(118) = .20$, $p = .03$.

Table 3 shows the test–retest reliability of the TOPKAM at four weeks, excluding the students in the CT class who received instruction in analyzing arguments and recognizing psychological misconceptions. The test–retest reliability was acceptable for the TOPKAM but higher for the APS and ITDP-R. The magnitude of the test–retest reliability in the senior capstone, $r(59) = .86$, $p < .001$ was somewhat higher than the test–retest reliability for beginning students in the CT class, $r(51) = .64$, $p < .001$, and in the introduction to the profession course, $r(65) = .56$, $p < .001$. These results suggest that the TOPKAM showed generally adequate test–retest reliability overall, but varied somewhat across groups.

The *KR*-20 across the overall sample was .70 on the first administration rising to .75 on the second administration. The *KR*-20 for the CT class was .66 on the first administration rising to .83 on the second administration. The substantial increase for the CT class was probably due to more consistent responding after instruction about specific psychological misconceptions. These results suggest that the internal consistency was adequate in the samples, but may also be sensitive to instruction.

Concurrent validity was evaluated by examining the correlations between TOPKAM and the two CT measures. In the entire sample, the TOPKAM was positively correlated with scores on the APS, the CT measure of argument analysis knowledge and skill, at the first administration, $r(158) = .44$, $p < .001$, and the second, $r(110) = .51$, $p < .001$. Table 3 shows a similar pattern of correlations in analyses of the sample that excluded students from the CT class. These results are consistent with those of studies by McCutcheon et al. (1992) and Taylor and Kowalski (2004) in showing a positive correlation between their measures of psychological misconceptions and CT.

Extending this previous connection with CT knowledge and skill, TOPKAM scores in the entire sample showed a moderate, but significant and positive correlation with ITDP-R scores at both the first administration, $r(171) = .29$, $p < .001$, and the second, $r(138) = .25$, $p = .003$, replicating the positive correlation found in the first study.

Table 3
Descriptive statistics and bivariate correlations for the first and second administrations of the test of psychological knowledge and misconceptions and critical thinking measures.

Measure	1	2	3	4	5	6
1. TOPKAM1	–	.43***	.30**	.69***	.39***	.23*
N =		113	126	91	90	101
2. APS1		–	.17	.45***	.73***	.14
N =			113	79	90	89
3. ITDP-R1			–	.20	.13	.80***
N =				91	90	101
4 TOPKAM2				–	.39***	.19
N =					78	89
5. APS2					–	.15
N =						88
6. ITDP-R2						–
M	19.98	9.13	99.33	20.59	8.82	99.37
SD	4.90	3.59	10.09	4.82	3.94	10.62

Note. TOPKAM = Test of Psychological Knowledge and Misconceptions of Bensley and Lilienfeld (2010); APS = Analyzing Psychological Statements of Bensley (2011b); ITDP-R = Inventory of Thinking Dispositions in Psychology—Revised from Bensley (2011b).
* $p < .05$.
** $p < .01$.
*** $p < .001$.

These findings suggests that, besides CT skills, CT dispositions are related to holding misconceptions and to an actively open-minded interest in psychology and a tendency to adopt a rational–scientific approach to psychology.

To test the significance of the contribution of both CT skill and CT disposition in psychology, we conducted a standard multiple regression of scores from the first administration, regressing APS and ITDP-R scores onto the TOPKAM scores. The results were significant, $F(2,155) = 26.17$, $p < .001$, with the model accounting for 25.2% of the variance, $R^2 = .25$, adjusted $R^2 = .24$. CT skill for argument analysis significantly predicted the number correct on the TOPKAM, $B = .44$, $t(152) = 6.17$, $p < .001$, as did CT disposition in psychology, $B = .17$, $t(152) = 2.32$, $p < .05$.

Table 3 also shows the mean number correct on the TOPKAM and the other CT measures at time 1 and 2. The mean number of misconceptions (40—TOPKAM correct) was 20.36 or 50.9% of the items missed on the first administration of the TOPKAM. Although this percentage is rather high, it was not significantly greater than 20 (expected by chance), $t(172) = 0.97$, $p = .34$.

4. General discussion

In this article, we documented the development of the Test of Psychological Knowledge and Misconceptions, a new test designed to address several shortcomings of previous tests of psychological misconceptions. We used a forced-choice format to avoid potential response biases that can arise in T/F instruments in which true responses are always scored as misconceptions. The correct response for each TOPKAM item was based on recent literature reviews on psychological misconceptions (Lilienfeld et al., 2010). To address the problem of the inherently provisional basis of most psychological knowledge, TOPKAM questions asked participants to choose which of the two response options was most true. The results from both samples supported the TOPKAM's internal consistency, and retesting of participants at four weeks in the second study supported its test–retest reliability.

Correlations between the TOPKAM and both new and established measures supported its concurrent validity and showed its relations with individual difference correlates. Following up on significant correlations, multiple regression analyses showed that the ability to distinguish sciences from pseudosciences and evidence-based from poorly-supported practices significantly predicted the number correct on the TOPKAM; whereas, paranormal belief was a significant, negatively weighted predictor of TOPKAM scores. This finding indicated that the more items participants got correct on the TOPKAM (viz., the fewer their misconceptions) the better they were able to distinguish sciences from pseudosciences and evidence-based from poorly-supported practices. Moreover, those scoring higher on the TOPKAM were less accepting of paranormal claims on the RPBS. Finally, all four measures were significantly intercorrelated, suggesting commonality in what they measure. This pattern of results further suggests that less knowledge of psychology is related to a tendency to accept unsubstantiated claims including those regarding pseudoscience, poorly-supported practices, and the paranormal.

At the same time, TOPKAM scores were positively correlated with the ITDP-R, a measure of CT disposition, and negatively correlated with FI, a measure of the tendency to trust one's intuition. The complementary pattern of these correlations with the TOPKAM and follow-up multiple regression analyses suggest that those least susceptible to psychological misconceptions are more disposed to take an active, open-minded interest in psychology and a rational–scientific approach while trusting their intuition less. The results from the second study further suggest that CT skill for argumentation in psychology and the disposition to think critically in psychology both predict susceptibility to psychological misconceptions. These findings replicate the association between misconceptions and poorer critical thinking found in previous studies (McCutcheon et al., 1992; Taylor & Kowalski, 2004) but also extend them to indicate the involvement of CT disposition. Being more disposed to reflect on psychological claims in combination with greater CT skill predict more successful rejection of psychological misconceptions.

In general, the results were consistent with dual-process theories, such as cognitive–experiential self-theory, which posit individual differences in thinking styles. In people who tend to endorse psychological misconceptions, the intuitive–experiential process may be dominant; whereas, in those who hold fewer misconceptions, the rational–analytic process may be dominant. The results of our first study suggest that those who have more psychological misconceptions tend to accept paranormal claims more and trust their intuition more. This finding is supported by studies showing a positive relationship between FI and paranormal belief (Irwin & Young, 2002; Lindeman and Aarnio, 2006; Saher & Lindeman, 2005). The greater acceptance of these scientifically unsupported ideas may be associated with a tendency to not reflect on such claims and a tacit acceptance of them, consistent with an intuitive thinking style.

The fact that performance on the CT argument analysis test predicted TOPKAM scores suggests that the effortful cognitive processing and reasoning skills needed for thinking critically are related to the reflective thinking involved in rejection of misconceptions. Future research should examine this relationship with respect to measures of global intelligence and cognitive ability. In particular, recent work on dual-process theory has emphasized the contribution of working memory to System 2 (sometimes called Type 2) thinking (Evans, 2010; Evans & Stanovich, 2013), and future studies should examine the possible contribution of individual differences in working memory to rejection of misconceptions with respect to CT.

Supporting the involvement of a rational, analytic thinking style was the finding in our second study that those who scored higher on the TOPKAM reported higher levels of CT disposition on the ITDP-R at both administrations. Students with better psychological knowledge may rely more on a rational–analytic style and use their CT knowledge and skill to reflect on misconceptions. It is not clear, however, why neither the TOPKAM number correct nor the RPBS was significantly correlated with the rational scale (NFC) although the ITDP-R, also measuring a rational thinking style, was positively correlated with both NFC and OBJ. Moreover, NFC was positively correlated with the scientific ratings on the KOF Science scale and negatively correlated with scores on the KOF Pseudoscience scale.

Some studies have found that NFC is negatively correlated with unsubstantiated belief. Lindeman and Aarnio (2006) found that paranormal belief was negatively correlated with NFC and positively correlated with

FI supporting a dual-process conception of superstitious, magical, and paranormal beliefs. In an earlier study, Epstein, Pacini, Denes-Raj and Heier (1996) found that NFC was negatively correlated with personal superstitious beliefs. Likewise, Saher and Lindeman (2005) found that NFC scores were negatively correlated with the RPBS and with a measure of magical food- and health beliefs but were not correlated with beliefs about CAM. The present study extended the dual-process account to endorsement of psychological misconceptions and many pseudoscientific fields and practices besides paranormal belief.

Taken together, our results and those of other studies suggest that rejection of a variety of unsupported claims such as misconceptions, paranormal and pseudoscientific claims, and other unsubstantiated claims may be associated with a rational, reflective thinking style; whereas, paranormal, superstitious, and pseudoscientific beliefs tend to be associated with an intuitive thinking style. Other claims that are often unsubstantiated, such as the claims regarding CAM, may be related differentially to the two thinking styles. Endorsement of unsubstantiated claims may depend on their content and the thinking errors that believers of unsubstantiated claims tend to make. Future research on the TOPKAM in relation to other measures of specific kinds of unsupported beliefs may clarify these relationships, a direction originally proposal by Lindeman and Svedholm (2012).

Further evidence for the concurrent validity of the TOPKAM comes from its significant correlations with academic ability. The positive correlation of TOPKAM scores with SAT in the first study suggests that students with better academic aptitude tended to have more accurate knowledge of psychology, accepting misconceptions less. The TOPKAM and SAT may both assess CT ability, given that SAT is often correlated with CT test performance (Erwin & Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, 2000) and that misconception test performance is correlated with CT test performance (McCutcheon et al., 1992). This interpretation is further supported by the positive correlations between TOPKAM and ITDP-R and the APS. The significant correlation of TOPKAM with the number of psychology course credits earned shows that the rejection of misconceptions increases with psychology course experience, as found previously (Gardner & Dalsing, 1986; Griggs & Ransdell, 1987). This correlation is weak, however; and a multiple regression analysis revealed that SAT, not psychology course credits, significantly predicted TOPKAM correct.

In general, our initial results with the TOPKAM are encouraging. They provide evidence for its internal consistency and test–retest reliability. Correlational and multiple regression analyses provided evidence for its concurrent validity although the results with the less well established measures, such as the TEBTP and KOF, invite further study. Moreover, our results suggest that the TOPKAM may serve as a viable alternative to T/F misconception tests in which misconceptions were always true responses and to tests that have not measured metacognitive aspects of response. The relationship of TOPKAM with variables related to CT offers a wealth of opportunities for future research. Because CT is a multi-dimensional construct involved in belief revision (Bensley, 2011b; Ennis, 1987), the TOPKAM may provide another means to study the relationships among belief revision, metacognition, CT dispositions, and CT skills needed to advance learning outcomes assessment research (Bensley & Murtagh, 2012).

In addition, future research should examine the TOPKAM's relation to established measures to clarify its discriminant validity. Our preliminary findings showed that performance on the TOPKAM was significantly correlated with active open-minded interest in psychology, but not interest in roles in psychology or interest in assessment. Future research should examine whether TOPKAM scores can be differentiated from measures of theoretically separable constructs such as global intelligence. Moreover, it would be useful to examine whether TOPKAM misconceptions decrease less than non-misconception items when both are tested at the beginning and after instruction in an introductory psychology course. Misconceptions should be more resistant to traditional psychology instruction than common factual information covered in such a course if misconceptions are persistent, default responses.

Our results have other implications for instruction, too. The dual-process account suggests that helping students correct their misconceptions may be facilitated when instructors explicitly encourage students to adopt a critical thinking approach to psychological claims, especially those students who adopt an experiential–intuitive approach. Our results imply that these students not only tend to lack CT skills but also the necessary CT disposition for questioning tacit beliefs. Our results may also help to explain why refutational approaches to correcting misconceptions are often successful (e.g., Braasch, Golden, & Wiley, 2013; Guzzetti, 2000; Kowalski & Taylor, 2009). Refutation encourages students to explicitly confront their misconceptions and critically evaluate them, creating the kind of tension that makes it difficult to continue to accept familiar, tacit knowledge in the presence of evidence-based alternatives that might otherwise not be considered. Given the tendency of many students to entertain various unsupported claims, instructors may find it useful to help them adopt a CT approach that encourages refutation of not only psychological misconceptions, but also other unsubstantiated claims.

Potential research and instructional applications of our results show promise, but certain limitations of our findings should be borne in mind. The samples included psychology students at various levels in their programs at one university, and so the TOPKAM should be used at other universities to determine whether misconceptions occur at the same high levels as in our samples. However, psychological misconceptions were at high levels at a different university using a test with similar item content as the TOPKAM (Gardner & Brown, 2013).

Another question is whether the students in our samples may have actually deflated their endorsement of misconceptions due to experimenter bias and their wanting to provide correct answers that were counter to their intuitions. The data reported in this study cannot conclusively answer this question, but other analyses of our data (Bensley & Lilienfeld, 2014) argue against this interpretation. We found that students were more certain of the correctness of misconceptions they endorsed more often than of misconceptions they endorsed less often, suggesting that they lack accurate knowledge of the correctness of their responses and so are not in a position to answer with correct answers to please experimenters. However, further research examining the relation between TOPKAM scores and self-report measures of social desirability and other motives for responding may shed more light on this question.

5. Conclusion

The results provided evidence for the internal consistency and test–retest reliability of the Test of Psychological Knowledge and Misconceptions, a new test designed to eliminate shortcomings of previous misconceptions tests. Supporting its concurrent validity, higher TOPKAM scores were predicted by (a) a measure of the ability to distinguish science from pseudoscience, (b) a measure of the ability to distinguish evidence-based from poorly-supported psychological practices, and (c) lower levels of paranormal belief. Less faith in intuition and greater CT disposition and skill, as well as higher SAT scores, also predicted higher TOPKAM scores.

Besides providing preliminary support for the construct validity of the TOPKAM, this pattern of findings suggests that the rejection of psychological misconceptions is related to the rejection of other unsupported paranormal and pseudoscientific claims, better academic and critical thinking ability, and to a more active, open-minded interest in psychology and less reliance on intuition. In general, our results are consistent with dual-process theories of thinking and offer a new avenue for examining individual differences in learning in relation to psychological misconceptions.

Acknowledgments

We thank two anonymous reviewers and the editor of this journal for their helpful comments on the manuscript. We also wish to acknowledge Drs. Barney Beins of Ithaca College, Tom Heinzen and Kate Makarec of William Patterson University, and Paul Bernhardt, Erica Kennedy, Chris Masciocchi, and Michael Murtagh of Frostburg State University for helping revise Analyzing Psychological Statements. Thanks to Stephanie Kuehne for assistance with the manuscript, Selina Smith of the Frostburg State University Office of Institutional Research for providing academic background information on student participants, and to the students, themselves, who consented to let us use their assessment data.

References

Bensley, D. A. (2011a). Rules for reasoning revisited: Toward a scientific conception of critical thinking. In C. Horvath, & J. M. Forte (Eds.), *Critical thinking* (pp. 1–36). Hauppauge, NY: Nova Science Publishers.

Bensley, D. A. (2011b). The critical thinking in psychology test battery. Unpublished manuscript.

Bensley, D. A., & Crowe, D. S. (2010, May). Reliability and validity of the Test of Evidence-based Theories And Practices. *Poster presented at the annual meeting of the Association for Psychological Science, Boston, MA.*

Bensley, D. A., & Lilienfeld, S. O. (2010). *The test of psychological knowledge and misconceptions.* Unpublished instrument.

Bensley, D. A., & Murtagh, M. P. (2012). Taking a scientific approach to critical thinking assessment: Guidelines for psychology programs. *Teaching of Psychology, 37*, 5–16.

Bensley, D. A., & Lilienfeld, S. O. (2014). *What is a psychological misconception? Moving towards an empirical answer.* (submitted for publication).

Bensley, D. A., Crowe, D. S., Bernhardt, P. C., Buckner, C., & Allman, A. (2010). Teaching and assessing critical thinking skills for argument analysis. *Teaching of Psychology, 37*, 91–96.

Best, J. B. (1982). Misconceptions about psychology among students who perform highly. *Psychological Reports, 51*, 239–244.

Beyerstein, B.L. (1999). Whence cometh the myth that we only use ten percent of our brains? In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 1–24). Chichester: John Wiley & Sons.

Braasch, J. L., Golden, S. R., & Wiley, J. (2013). The influence of text and reader characteristics on learning from refutations in science texts. *Journal of Educational Psychology, 105*, 561–578.

Brown, L. T. (1983). Some more misconceptions about psychology among introductory psychology students. *Teaching of Psychology, 10*, 207–210.

Brown, L. T. (1984). Misconceptions about psychology are not always what they seem. *Teaching of Psychology, 11*, 75–78.

Bushman, B. J., Baumeister, R. F., & Strack, A. D. (1999). Catharsis, aggression, and persuasive influence: Self-fulfilling or self-defeating prophesies. *Journal of Personality and Social Psychology, 76*, 367–376.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197–253.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*, 685–716.

Crowe, R. A., & Miura, C. K. (1995). Challenging pseudoscientific beliefs: Surveying evidence for exotic claims. *Psychological Reports, 77*, 1263–1282.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron, & R. J. Sternberg (Eds.), *Teaching thinking skills* (pp. 9–37). New York: W. H. Freeman.

Epstein, S. (2008). Intuition from the perspective of cognitive–experiential self-theory. In H. Plessner, C. Betsch, & T. Betsch (Eds.), *Intuition in judgment and decision making* (pp. 23–37). New York: Lawrence Erlbaum.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology, 71*, 390–405.

Evans, J. St. B. T. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry, 21*, 313–326.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241.

Erwin, T. D., & Student Outcomes Pilot Working Group: Cognitive and Intellectual Development (2000). *NPEC sourcebook on assessment: Definitions and assessment methods for critical thinking, problem solving, and writing. Vol. 1*, Washington, DC: US Government Printing Office (Retrieved from www.nces.ed.gov/pubs20000/2000195.pdf).

Gardner, R. M., & Brown, D. L. (2013). A test of contemporary misconceptions in psychology. *Learning and Individual Differences, 24*, 211–215.

Gardner, R. M., & Dalsing, S. (1986). Misconceptions about psychology among college students. *Teaching of Psychology, 13*, 32–34.

Griggs, R. A., & Ransdell, S. E. (1987). Misconceptions tests or misconceived tests? *Teaching of Psychology, 6*, 210–214.

Gutman, A. (1979). Misconceptions of psychology and performance in the introductory course. *Teaching of Psychology, 6*, 159–161.

Guzzetti, B. J. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading and Writing Quarterly, 16*, 89–98.

Hacker, D. J., Bol, L., & Kenner, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky, & R. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 429–455). New York, NY: Psychology Press.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*, 449–455.

Hughes, S., Lyddy, F., & Kaplan, R. (2013). The impact of language and response format on student endorsement of psychological misconceptions. *Teaching of Psychology, 40*, 31–37.

Hughes, S., Lyddy, F., & Lambe, S. (2013). Misconceptions about psychological science. *Psychology Learning and Teaching, 12*, 20–31.

Irwin, H. J., & Young, J. M. (2002). Intuitive versus reflective processes in formation of paranormal beliefs. *European Journal of Parapsychology, 17*, 45–53.

Kahneman, D. (2011). *Thinking fast and slow.* New York: Farrar, Straus, & Giroux.

Kowalski, P., & Taylor, A. K. (2009). The effect of refuting misconceptions in the introductory psychology class. *Teaching of Psychology, 36*, 153–159.

Kuhle, B. X., Barber, J. M., & Bristol, A. S. (2009). Predicting students' performance inintroductory psychology from their psychology misconceptions. *Journal of Instructional Psychology, 36*, 119–124.

Lamal, P. A. (1979). College students' common beliefs about psychology. *Teaching of Psychology, 6*, 155–158.

Landau, J.D., & Bavaria, A. J. (2003). Does deliberate source monitoring reduce students' misconceptions about psychology? *Teaching of Psychology, 30*, 311–314.

Leary, M. R., Shepperd, J. A., McNeill, M. S., Jenkins, B., & Barnes, B. (1986). Objectivism in information utilization: Theory and measurement. *Journal of Personality Assessment, 50*, 32–42.

Leong, F. T. L., & Zachar, P. (1993). Presenting two brief versions of the Scientist Practitioner Inventory. *Journal of Career Assessment, 1*, 162–170.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*, 106–131.

Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (Eds.). (2003). *Science and pseudoscience in clinical practice.* New York: Guilford Press.

Lilienfeld, S., Lynn, S., Ruscio, J., & Beyerstein, B.L. (2010). *50 great myths of popular psychology: Shattering widespread misconceptions about human behavior.* West Sussex, United Kingdom: Wiley-Blackwell.

Lindeman, M., & Aarnio, K. (2006). Paranormal beliefs: Their dimensionality and correlates. *European Journal of Personality, 20*, 585–602.

Lindeman, M., & Svedholm, A.M. (2012). What's in a term? Paranormal, superstitious, magical, and supernatural beliefs by any other name would mean he same. *Review of General Psychology, 16*, 241–255.

McCutcheon, L. E. (1991). A new test of misconceptions about psychology. *Psychological Reports, 68*, 647–653.

McCutcheon, L. E., Apperson, J. M., Hanson, E., & Wynn, V. (1992). Relationships among critical thinking skills, academic achievement, and misconceptions about psychology. *Psychological Reports, 71*, 635–639.

McKeachie, W. J. (1960). Changes in scores on the northwestern misconceptions test in six elementary psychology courses. *Journal of Educational Psychology, 51*, 240–244.

Nixon, H. K. (1925). Popular answers to some psychological questions. *The American Journal of Psychology, 36*, 418–423.

Norcross, J. C., Koocher, G. P., & Garafalo, A. (2006). Discredited psychological treatments and tests: A Delphi poll. *Professional Psychology: Research and Practice, 37*, 515–522.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*, 972–987.

Parsons, E. A., Powell, L., Bensley, D. A., & Crowe, D. S. (2010, March). Reliability and validity of the inventory of thinking dispositions in psychology. *Poster presented at the 81st annual meeting of the Eastern Psychological Association, Brooklyn, NY.*

Ruble, R. (1986). Ambiguous psychological misconceptions. *Teaching of Psychology, 13*, 34–36.

Saher, M., & Lindeman, M. (2005). Alternative medicine: A psychological perspective. *Personality and Individual Differences, 39*, 1169–1178.

Standing, L., & Huber, H. (2003). Do psychology courses reduce belief in psychological myths? *Social Behavior and Personality: An International Journal, 31*, 585–591.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*, 645–726.

Sundre, D. L. (2007). *The student opinion scale (SOS): A measure of examinee motivation test manual.* Harrisonburg, VA: James Madison University Center for Assessment & Research Studies.

Sundre, D. L., & Moore, D. L. (2002). The student opinion scale: A measure of examinee motivation. *Assessment Update, 14*, 8–9.

Swami, V., Stieger, S., Pietschnig, Nader, I. W., & Voracek, M. (2012). Using more than 10% of our brains: Examining belief in science-related myths from an individual difference perspective. *Learning and Individual Differences, 22*, 404–408.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson/Allyn & Bacon.

Taylor, A. K., & Kowalski, P. (2004). Naïve psychological science: The prevalence, strength, and sources of misconceptions. *The Psychological Record, 54*, 15–25.

Tobacyk, J. L. (2004). A revised paranormal belief scale. *The International Journal of Transpersonal Studies, 23*, 94–98.

Vaughan, E. (1977). Misconceptions about psychology among introductory psychology students. *Teaching of Psychology, 4*, 138–141.

Wood, J. M., Nezworski, M. T., Lilienfeld, S. O., & Garb, H. N. (2003). *What's wrong with the Rorschach? Science confronts the controversial inkblot test.* San Francisco: Jossey-Bass.