reboot
elevating critical thinking

A Scalable, Versatile Approach for Improving Critical Thinking Skills

Ben Motz
Emily Fyfe
Helen Lee Bouygues
Taylor Guba

Indiana University

Introduction

Critical thinking is often seen as a vexing educational challenge. On one hand, it holds high status as an urgent outcome of academic programs, stemming in part from a need to counteract growing volumes of online disinformation (Machete & Turpin, 2020). On the other hand, it is seen as a uniquely difficult outcome to achieve, with a history of contention about how to define, measure, and improve this skill (Abrami et al., 2015). In the current study, we address both challenges, testing an easy-to-implement approach for improving learners' ability to identify fallacious claims.

Although contention remains (Johnson & Hamby, 2017), a growing consensus of education researchers view critical thinking as involving the ability to detect and identify fallacies in peoples' claims (e.g., Lawson, 1999; Lawson, Jordan-Fleming, & Bodle, 2015; Schmaltz & Lilienfeld, 2014; Schmaltz et al., 2017). Essentially, when presented with a statement about the world, someone who is skilled at critical thinking should be able to discern whether the claim is justified based on the information provided, and furthermore, should be able to describe the specific issues present when claims are unjustified.

Conventional educational efforts to build a learner's disciplinary knowledge base might be assumed to achieve such critical thinking goals for "free," but this is not the case. Improving a student's understanding of domain knowledge (e.g., neuroscience) does not necessarily improve a student's ability to identify fallacies (e.g., neuromyths; Im, Cho, Dubinsky, & Varma, 2018). Instead, recent research on instructional strategies to improve critical thinking emphasize the importance of practice with critical thinking (Heijltjes et al., 2014; Morewedge et al., 2015), particularly with scenarios (Self & Self, 2017) for improving critical thinking outcomes. However, the standard instructional approaches that can be beneficially applied at large scales remain uncharted.

In the current study, we examine a specific, scalable, and versatile approach for improving critical thinking: having learners practice categorizing scenarios according to fallacies. Learners are presented with a scenario where an individual makes a claim based on some evidence or observations, and in a multiple-choice response format, the learner marks which fallacy, if any, the individual in the scenario was committing.
We measure the benefits of these multiple-choice practice problems using a validated,open-form critical thinking assessment, implemented both pre- and post-practice. The materials in the current study were designed for college-level students in Introductory Psychology, but they could easily be adapted for other domains. Introductory Psychology is one of the most popular college courses, taken by roughly 1.5 million students each year in the United States (Gurung et al., 2016), with programs placing heavy emphasis on the importance of critical thinking (Homa et al., 2013), and where critical thinking outcomes similarly include the ability to recognize flaws in explanations (American Psychological Association, 2013).

Method

All materials, data, and analysis scripts are publicly available at https://osf.io/vcgzf. The study protocol was approved by the Indiana University Institutional Review Board (IRB). Prior to data collection, we publicly registered all study methods at https://osf.io/9j583, which included fully-specified analysis scripts based on simulated data. The experiment, as described below, contained no deviations from this pre-registration.

Participants

Study participants, based in the US and at least 18 years old, were recruited on MTurk (https://mturk.com), among those who had an MTurk approval rating above 50%. The study's sample size was determined by a financial constraint — we continued to recruit participants in batches until funding was exhausted. Initially, 601 potential participants completed the screening session, but following exclusions and attrition over the four-part, multi-day, online study, a total of 253 participants completed the full study. All results described hereafter are limited to these 253 participants with complete datasets. According to responses provided in the initial screening session, 68.4% of these participants had completed a bachelor's degree or higher, and 85.0% agreed or strongly agreed with the statement "I am skilled at critical thinking."

Procedure

There were four sessions in the current study: (1) Screener, (2) Pre-test and Training, (3) Intervention, and (4) Post-test (see Figure 1), all conducted online using Qualtrics (Provo, UT). When running a batch of participants, each day we monitored who had satisfactorily completed each session, and sent invitations for subsequent sessions accordingly (using MTurk's messaging system). Thus, the average time between each session was about 1.4 days.

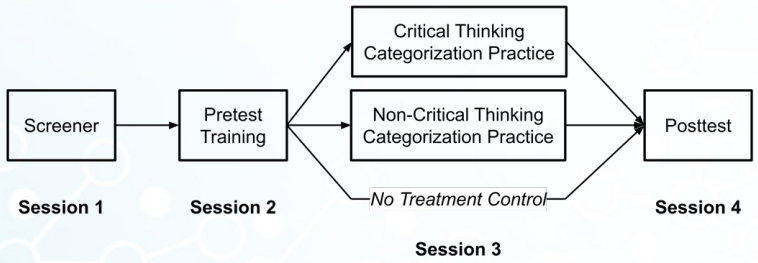

Figure 1. Overview of the study.

Session 1: Screener.

The purpose of the Screener was to identify human participants (not automated bots) who could follow instructions and provide coherent responses to open-ended questions. Participants were paid $0.20 this session, and the median completion time for participants in the final sample was 3.2 minutes.

Session 2: Pre-test and Training.

This session included a pre-test assessment of participants' critical thinking, immediately followed by training about critical thinking fallacies. The pretest was half the items on the Psychology Critical Thinking Exam (P CTE; Lawson et al., 2015), which was counterbalanced: half the participants got odd items, and the other half got even items for pre-test. We used a slightly reduced version of the PCTE, with 12 items total (6 item pretest, 6 item posttest), available at https://osf.io/pc37b.

Each item presents a scenario describing a person's fallacious interpretation of some observations, and respondents are prompted to "State whether or not there is a problem with the person's conclusions and explain the problem (if there is one)." There are 6 such fallacies, labeled and described in the table below:

Fallacy	Description
Random chance	Inferring systematic relationships from chance occurrences
Lack of control	Inferring that improvement was due to an experimental intervention without comparison to a control group
Correlation is not causation	Inferring a causal relationship on the basis of correlational data
Overgeneralization	Inferring that findings generalize to a larger group based on a biased sample
Experimenter bias	Inferences based on questions that were biased, loaded, or leading
Confirmation bias	Inferences based only on positive evidence and ignoring disconfirming evidence

Training followed the pretest, and included two phases: a study phase and a test phase. During the study phase, participants were asked to read and study instructional text about the six fallacies described above. During the subsequent test phase, participants were asked to respond to seven multiple choice item s, corresponding to each of the 6 fallacies plus an additional item where there was no fallacy, available at https://osf.io/n2hu8. Each item included a scenario similar to the PCTE items, and participants were asked to mark which fallacy (if any) the person committed. The training text was included on the same page as each test item, so participants could refer back to definitions of these fallacies when selecting their responses. Participants received feedback indicating the correct answer after each selection.

Participants were paid $2 for completing session 2, and the median completion time for participants in the final sample was 20.5 minutes.

Session 3: Intervention.

Participants were randomly assigned to one of three intervention conditions: (1) critical thinking categorization practice (n=81); (2) non-critical thinking categorization practice (n=91); or (3) no intervention control (n=81). Participants in the no intervention control group were not invited to complete session three; after completing session two they were invited directly to session four. Participants in the other two conditions were paid $3 for completing session three.

For the critical thinking practice group, we composed four practice sets, each containing seven multiple-choice critical thinking scenarios, available at https://osf.io/h8fpk. Like those in the training session, each item asked participants to mark the fallacy the person committed (if any), full descriptions of each fallacy were shown alongside the questions for reference, and validation feedback (correct/incorrect) was shown after each response. The sets were organized around typical content units in an Introductory Psychology course (neuroscience, sensation and perception, memory, and learning). Participants had to get at least 5 of 7 questions correct in each set; if they got fewer than five items correct, they had to repeat the set a maximum of three times. The median completion time was 36.6 minutes for the final sample of participants in the critical thinking practice group.

For the non-critical thinking practice group, we modified existing multiple-choice Introductory Psychology exam questions corresponding to each content unit (neuroscience, sensation and perception, memory, and learning), so that they were length-matched with the critical thinking practice items, and so that there were seven response options, available at https://osf.io/bq5a8. Each item asked participants to categorize a scenario according to a structure or concept in psychology, with no mention of logical fallacies. Like the critical thinking questions, these non-critical thinking questions were shown alongside full definitions of all the relevant structures and concepts, validation feedback (correct/incorrect) was shown after each response, and participants had to get 5 of 7 correct to proceed to the next set. The median completion time for participants in the final sample was 32.6 minutes for final sample participants in the non-critical thinking practice group.

Critical Thinking Categorization Practice

Directions: Read the scenario below and select which fallacy (if any) the person made. You can use the descriptions of the fallacies in the table to help. After you have made a selection, click the arrow to proceed. You will get feedback on your answer, which you can look over.

Many people have suggested that some of the symptoms of ADHD might be caused by hyperactivity of the mirror neuron system. In order to treat these symptoms, Douglas has designed a therapy that should improve cognitive control of mirror neurons. This therapy, which involves watching cartoons of animals while keeping perfectly still, was administered to one hundred 12-year-olds who had previously been diagnosed with ADHD. After 6 months of therapy, these patients showed an overall decline in symptoms of ADHD. Douglas concluded that the therapy reduces symptoms of ADHD.

- ○ Random chance
- ● Lack of control
- ○ Correlation is not causation
- ○ Overgeneralization
- ○ Confirmation bias
- ○ Experimenter bias
- ○ There is no fallacy in this conclusion

Non-Critical Thinking Categorization Practice

Directions: Read the scenario below and select which concept applies to that scenario. You can use the definitions of the concepts in the table to help. After you have made a selection, click the arrow to proceed. You will get feedback on your answer, which you can look over.

A group of people suffering from depression recently began a new experimental drug therapy trial. Like most pharmacological treatments for depression, this experimental drug affects the action of a certain neurotransmitter. However, this new drug is designed to avoid some of the adverse effects of previous drug therapies affecting this neurotransmitter, such as insomnia, drowsiness, and food cravings. Therapy for people suffering from depression often involves drugs that affect the action of which neurotransmitter?

- ○ Cerebellum
- ○ Substantia nigra
- ● Serotonin
- ○ Prefrontal cortex
- ○ Temporal lobe
- ○ Amygdala
- ○ Hippocampus

Session 4: Post-test.

In this session, participants completed the remaining half of the PCTE items. No information about the fallacies were displayed during the post-test. Participants were paid $5 for completing session four, and the median completion time was 11.1 minutes.

Data Analysis

Responses to PCTE items were scored by a trained coder who was not able to see participants' condition assignments, nor whether the responses were from a pre-test or post-test. Scoring used the PCTE's original coding scheme: 0 (no problem identified), 1 (a problem recognized but misidentified), 2 (identified main problem, but also mentioned less relevant problems), and 3 (identified only the main problem). Our primary analytical goal is to assess whether improvement from pre-test to post-test differs between the three intervention conditions: (1) critical thinking categorization practice; (2) non-critical thinking categorization practice; or (3) no intervention control.(scored 3-out-of-3) on the pretest and 3.1 items correct out of 6 on the post-test, corresponding to percent scores of 40.4% and 52.3% respectively. It makes sense that we see modest improvement (0.72 more items correct on posttest) across the full study sample, considering that all participants

We calculated the number of test items that participants got precisely correct (scored 3-out-of-3) on the post-test and on the pre-test, and subtracted the pre-test from post-test, creating an improvement score . We estimated the tendency of this improvement score between different intervention conditions using a robust, hierarchical, Bayesian version of the t-test, described in Kruschke (2013). Priors for group-level estimates for each intervention condition were the empirical mean and standard deviation of the full study sample. We estimated model parameters with 100 adaptation steps, 500 burn-in steps, and 100,000 samples thinned to every 5th step (20,000 saved samples) across 4 MCMC chains, using JAGS (Plummer, 2003) and the runjags package (Denwood, 2016) for R, and the full registered model specification is available at https://osf.io/vju8w/. The effective sample size (ESS) was at least 20,000 for all model estimates, well above the 10,000 recommended by Kruschke (2014).

Results

Averaging across all treatment conditions, participants got 2.4 items precisely correct out of 6 (scored 3-out-of-3) on the pretest and 3.1 items correct out of 6 on the post-test, corresponding to percent scores of 40.4% and 52.3% respectively. It makes sense that we see modest improvement (0.72 more items correct on posttest) across the full study sample, considering that all participants received some training on critical thinking fallacies immediately after taking the pretest.

We estimated improvement scores (post-test-pre-test) for each treatment condition at the group-level. The modal improvement estimate in the critical thinking practice condition was 1.33 items (95% HDI: 0.96 to 1.67). In the non-critical thinking practice condition was 0.46 items (95% HDI: 0.13 to 0.81), and in the no treatment control was 0.41 items (95% HDI: 0.13 to 0.70).

Improvement in the critical thinking practice condition was credibly larger than the non-critical thinking practice condition (difference mode = 0.82; 95% HDI: 0.33 to 1.32), and larger than the no treatment control (difference mode = 0.89; 95% HDI: 0.44 to 1.35). There was no evidence of any difference between the non-critical thinking practice condition and the no treatment control condition (difference mode = 0.03; 95% HDI: -0.40 to 0.49). As shown in Figure 1, scores are larger in the critical thinking practice condition than in the other two conditions.

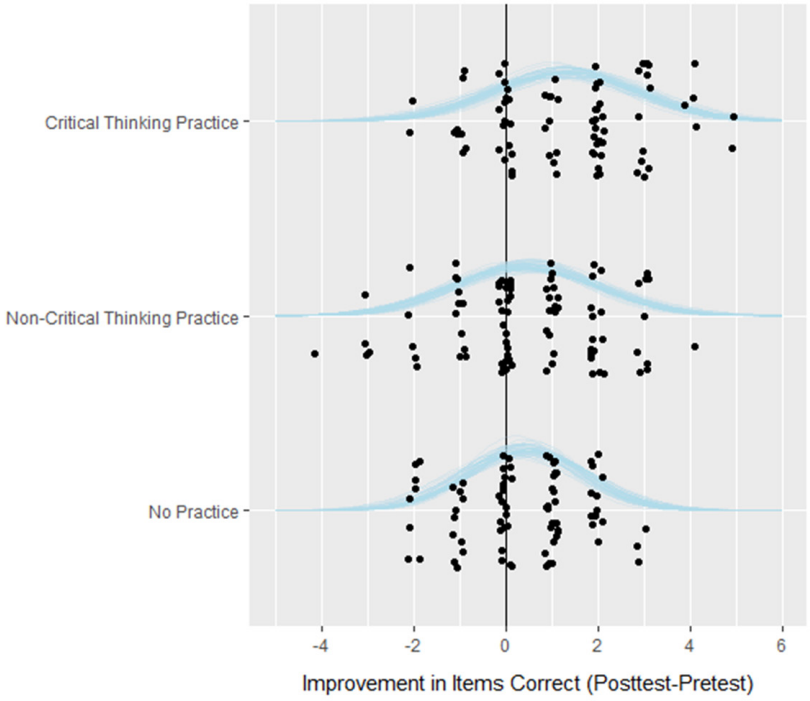

Figure 2. Improvement in the number of items correct between pretest and posttest. Each dot illustrates a single participant's improvement. Semi-transparent blue lines show credible estimates of the mean and standard deviation, modeled as a normal distribution at the group level. Results indicate larger improvements in the critical thinking practice condition (top row) compared with the other two conditions.

Discussion

In the present study, participants in all three conditions received training about logical fallacies, and in all three conditions, participants demonstrated improved performance on an open-ended critical thinking assessment. However, participants who had a separate session practicing categorizing scenarios according to these logical fallacies had significantly higher gains.

Even though the critical thinking practice session involved simple, auto-graded multiple choice problems, participants assigned to this condition had roughly three times larger improvement pre-to-post (1.33 items) than participants who practiced categorizing basic concepts (0.46 items) and participants who had no practice (0.41 items).

Whereas past research conflated the benefits of training and practice (Bensley & Spero, 2014), the current research separates the two, and finds that substantial critical thinking gains depend on the learner's ability to apply the "to-be-learned" critical thinking concepts during deliberate practice after initial learning (see also Heijltjes et al., 2014). Practice categorizing non-critical thinking concepts had no credible benefits for critical thinking performance compared with no practice at all.

Thus the current study demonstrates a simple scalable multiple-choice practice intervention that caused significant improvements in performance on an open-form critical thinking assessment instrument. In this regard our research design was opposite Renaud & Murray's (2008); rather than designing an intervention involving open-form questions, and evaluating critical thinking improvement using a multiple-choice assessment instrument (Watson-Glaser Critical Thinking Appraisal; WGCTA), we created an intervention using multiple-choice practice questions, and assessed improvement on an open-form critical thinking thinking assessment. Thus we believe that our proposed intervention is more portable and easy to implement at scale, addressing calls for incorporating modular critical thinking practice throughout a course (Stevens, Witkow, & Smelt, 2016).