Teaching Critical Thinking with the Personalized System of Instruction

Penultimate draft

1. Introduction

Many instructors want to improve the critical thinking skills of their students. But it's often unclear how to do this, and whether we're succeeding. For one thing, some evidence suggests that students make only slight gains in critical thinking over the course of their undergraduate careers (Arum & Roksa, 2011; Liu et al., 2021; Pascarella et al., 2011). Consider also the evidence from critical thinking courses. Surveys of the evidence find that traditional, stand-alone critical thinking courses produce only small improvements in students' critical thinking skills (Hitchcock, 2017: 505-6; Ortiz, 2007: 732).

Can we do better? In this essay, I describe an approach to teaching critical thinking that seems effective. My critical thinking course uses the Personalized System of Instruction (PSI). In PSI courses, the material is divided into units, students must pass a test on each unit before advancing to the next unit, and students have unlimited attempts on tests. PSI courses are self-paced, there's no group-level instruction, and class attendance is not required. A large body of evidence finds that PSI reliably improves student learning. But, to my knowledge, researchers haven't studied the effectiveness of PSI in critical thinking classes. I'll present evidence from pre- and post-tests that indicates that students substantially improved their reasoning skills in a class that used PSI to teach critical thinking.

I've organized this essay as follows. In section 2, I clarify which critical thinking skills that I aimed to promote. In section 3, I discuss the nature of PSI, the evidence for its effectiveness, and other studies that have used elements of PSI to enhance critical thinking skills. In section 4, I describe the structure and content of my course. In section 5, I present evidence

that this course was effective at promoting critical thinking skills. In section 6, I consider the problems and costs of teaching critical thinking with PSI. Section 7 concludes the paper.

2. Which Critical Thinking Skills?

Before I begin, let me say more about the critical thinking skills that I'd like to promote. First, I want my students to be able to recognize, analyze, and evaluate arguments in a precise and thoughtful way. Second, I would like my students to be able to construct credible arguments and to clearly communicate these arguments through prose. These strike me as core critical thinking skills.

While scholars disagree about how to define critical thinking (van der Brugge, 2018: 21-58), most would agree that the skills that I've highlighted are important. For example, Deana Kuhn claims that "argumentative reasoning skills are in fact fundamental to what educators call 'critical thinking" (Kuhn, 1991: 5). Robert Ennis (2015) lists argument analysis and evaluations among the ideal critical thinking abilities. Kevin Possin describes the major tests of critical thinking and notes that these tests focus on the analysis and evaluation of arguments and reasoning (Possin, 2008: 204). David Hitchcock surveys many different conceptions of critical thinking and finds that scholars of critical thinking agree that the following skills are important: the ability to clarify meaning, analyze arguments, evaluate evidence, judge whether a conclusion follows, and draw warranted conclusions (Hitchcock, 2017: 482). Hitchock's list overlaps substantially with mine, although I'd also include the ability to construct and communicate arguments.

So, it's uncontroversial that the ability to analyze, evaluate, and construct arguments are valuable critical thinking skills, although I'll refrain from claiming that these are the *only* critical

thinking abilities that universities should promote. Other skills are valuable too, such as the ability to do formal logic, to calculate probabilities, to understand cognitive biases, and so on. But Charles Twardy (2004) notes that instructors have reasons to focus on honing foundational argumentative skills before teaching more advanced ones: "Teaching extra flourishes does not mean they are being learned, and every such addition takes away time that could be used to help students master the fundamental critical thinking skill of argument analysis. Indeed, most other critical thinking skills centrally depend upon argument analysis: a student must be able to identify claims and lines of reasoning before they can possibly engage them critically, and then must be able to assess evidence and support 'in the wild' of informal arguments" (101).

3. The Personalized System of Instruction

I'll now describe PSI, which is the major feature of my critical thinking course. PSI is a method of instruction that was pioneered by the psychologist Fred Keller in the 1960s and 70s (people also refer to PSI as "the Keller Plan").

PSI has the following features (Keller & Sherman, 1974):

A. *Course Structure*. PSI courses are typically divided into units and student grades depend on how many units they complete before the end of the semester. PSI course are often divided into about 20 units.

B. *Self-Pacing*. Students move through the course at their own pace and class attendance is not required.

C. *Content Delivery*. PSI courses deliver instructional material through textbooks, articles, and videos. Some instructors lecture for motivational purposes, although

instructors refrain from delivering important course content in this way. PSI courses have little or no whole-group instruction.

D. *Unit Mastery*. To advance in the course, students must demonstrate mastery of the material. This means that students need to achieve a certain score on unit tests. PSI courses usually require students to score 90 percent or above on these tests before they can move to the next unit. Students can take unit tests an unlimited number of times, although instructors may set limits on how many attempts students can make during a class session.

E. *Proctors and Immediate Feedback*. When students complete unit tests, they submit these tests to a proctor. The proctor then provides these students with rapid feedback. If a student has passed the unit test, then the proctor notifies the student of that fact. If a student has yet to pass, then the proctor explains the student's mistakes and suggests areas where the student can improve.

PSI departs radically from standard methods of instruction. So, why would an instructor use PSI? A sizable amount of evidence shows that PSI reliably improves student learning. Michael Dunkin and Jennifer Barnes review the evidence on PSI (or, as they refer to it, the Keller plan) and conclude:

> The single most significant conclusion to be reached from research on innovative teaching methods in higher education is that the Keller Plan is clearly superior to other methods with which it has been compared. Indeed, the Keller Plan has been so consistently found superior that it must rank as the method with the greatest research support in the history of research on teaching. (Dunkin & Barnes, 1986: 756)

In a more recent review of the evidence, Eric Fox says:

> Few educational innovations have been subjected to the empirical scrutiny PSI has, and fewer still have emerged so unscathed….It has been estimated that over 2,000 PSI research studies have been conducted, and the vast majority of these have shown that students in PSI courses learn the course content better, remember it longer, and like the experience more than students in traditional classes. (J. Fox, 2004: 206)

The most comprehensive meta-analysis on PSI finds that this instructional approach improved final test scores and retention of the material when compared to more standard methods of instruction (Kulik et al., 1979).

Instructors have used PSI in a wide variety of courses, such as psychology, calculus, sociology, and physics (Clark, 1974; Klopfenstein, 1977; Protopapas, 1974). Have any instructors implemented PSI in critical thinking courses? Perhaps the most relevant source here is Ann Cahill and Stephen Bloch-Schulman's report (2012) on their experience with teaching argumentation. This report is relevant because their course resembles PSI in many ways. Cahill and Bloch-Schulman divided their course into ten steps, the course was self-paced, and students needed to demonstrate mastery on quizzes before advancing in the course. The learning goal for the course was "argumentative fluency." More specifically, the course required students to analyze, evaluate, and construct arguments. Cahill and Bloch-Shulman suggest that this "mastery learning" approach to critical thinking is more effective than more standard methods of instruction.

Cahill and Bloch-Schulman's experiment is intriguing. In fact, their experience has influenced the design of my critical thinking course. Yet their approach has certain limitations. First, it's a little unclear how closely Cahill and Bloch-Schulman's approach resembles PSI. Here's one difference. PSI emphasizes written materials, such as textbooks, to convey the course

material. But Cahill and Bloch-Schulman say that "new material is only taught face-to-face directly by the instructor" (54). In other words, Cahill and Bloch-Schulman convey the course material through in-person lectures to small groups. There may be other differences between Cahill and Bloch-Schulman's approach and PSI as well.[1] Second, while Cahill and Bloch-Schulman hold that their instructional approach improved learning, this conclusion is based only on their informal observations. They do not provide evidence from, say, pre-tests and post-tests to demonstrate better learning outcomes. One contribution of my paper is to provide evidence along these lines.

In an unpublished study, another group of critical thinking instructors attempted to implement aspects of PSI into their courses (Thomason et al., 2013). These courses focused on argumentation and used argument mapping – a technique for visualizing reasoning – to improve critical thinking skills (I'll discuss argument mapping in greater detail below). Furthermore, the researchers in this study designed mastery learning quizzes that students needed to complete before advancing in the class. Data from pre-tests and post-tests suggest that the courses improved critical thinking skills, but the effect size of these courses on test scores varied widely depending on the test. Although the courses in this study did use aspects of mastery learning, they were not PSI courses. Some differences include: there was whole-group instruction, the course was not self-paced, and the mastery learning quizzes were automated and did not involve proctors. So, the structure of these critical thinking courses differs substantially from PSI.

To sum up, the evidence indicates that PSI is an effective way of promoting learning. However, we lack evidence that critical thinking courses that use PSI as the primary method of instruction significantly improve argumentative reasoning skills.

4. PSI in a Critical Thinking Course

I'll now describe the structure and content of my critical thinking course. First, some quick background. I teach at a selective liberal arts college that emphasizes small class sizes (most of my classes enroll about 18 to 24 students). The course in question is entitled "Critical Thinking and Methods of Inquiry" and it's a required course for undergraduate students in the [information omitted for blind review] at the [information omitted for blind review]. This is an interdisciplinary school and faculty have diverse disciplinary backgrounds, ranging from literature and philosophy to economics and psychology. Most students have minimal familiarity with the content of the class before they enroll.

Course Structure and Content

Here's how my course is structured. I've divided the course into seven major steps. These are:

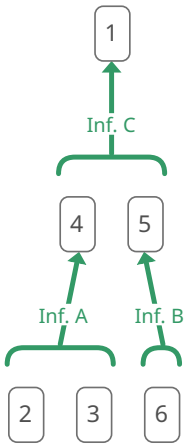
Step	**Learning Goals: in this step, you'll learn …**
Step 1: Premise/Ultimate Conclusion Arguments	How to recognize, analyze, and evaluate simple arguments; identify indicator expressions.
Step 2: Sub-Conclusions	How to recognize, analyze, and evaluate arguments with sub-conclusions.
Step 3: Dependent Reasons	How to analyze and evaluate arguments with dependent reasons - reasons that are linked with other premises in an argument; how to identify missing premises.

Step 4: Independent Reasons	How to identify and evaluate arguments with independent reasons (reasons that don't depend on other claims); understand the logic of conditional "if, then" statements; use rules of logical inference, such as modus ponens and modus tollens.
Step 5: Implicit Conclusions	How to identify missing ultimate conclusions and sub-conclusions; use logically valid forms such as dilemmas, hypothetical syllogism, and more.
Step 6: Objections	How to analyze and evaluate objections; how to distinguish genuine objections to an argument from fallacies and pseudo-objections.
Step 7: Argument Construction	How to construct and communicate your own argument; what qualifies as good evidence and how to find it.

The material is highly scaffolded. The course proceeds from simple arguments, such as arguments with just one premise and conclusion, to complex arguments that contain several different argumentative patterns. Each step requires students to practice their skills at recognizing, analyzing, and evaluating arguments. In the final step, students defend their own argument.

Much of the course material relies on a technique called argument mapping. Argument mapping is a way of visually representing the logical structure of an argument. Here's an example:

> One idea that's growing in popularity is the universal basic income, the idea that governments should simply give cash to all citizens of a country. What do you think about the idea of a universal basic income? I personally believe that [1] the universal basic income is a bad idea. This is so because, [2]if all citizens of a country received a universal basic income, then many of them would stop working. But, [3]if citizens stopped working, then economic growth would decline. Thus, [4]a universal basic income would lead to less economic growth. Obviously, though, [5]policies that bring about less economic growth are bad ideas since [6]the goal of public policy is to promote economic prosperity.



The conclusion of the argument goes at the top of the map. Claims that are lower on the map work to support claims that higher up. For example, claims 2 and 3 work together to support claim 4, and 4 and 5 in turn support the conclusion of the argument, claim 1. The arrows represent the inferences between claims. When we construct an argument map of an argument that we find in prose, we ignore non-argumentative material, such as stage-setting and other kinds of "filler." The map only includes the material that is part of the structure of the argument. For primers on argument mapping, see the references in this endnote.[2]

The course uses argument mapping as a vehicle for teaching argument analysis, evaluation, and construction. Students learn to map arguments with pencil and paper and use

these maps to analyze and evaluate arguments. Several studies find that courses in which students often practice mapping arguments effectively promote critical thinking skills (Dwyer et al., 2012; Gelder, 2015; Harrell, 2005). There are likely several reasons why argument mapping benefits students. Argument mapping encourages students to understand the logical structure of reasoning and read more carefully. Argument mapping also familiarizes students with critical thinking concepts and externalize reasoning in a way that allows students to evaluate it more precisely (Thomason et al., 2013: 13-4). I suspect that one reason that argument mapping can improve critical thinking skills is that argument mapping makes it possible for instructors to give students specific and rapid feedback on their work. An experienced instructor can review a student's map and quickly determine whether this map accurately describes the relevant argument. And, if the map is flawed, an instructor can show the student exactly where she went wrong.

My course teaches argument mapping through a textbook and videos. Students use a free critical thinking textbook that the philosopher Dona Warren has written that makes extensive use of argument mapping.[3] Furthermore, I create and upload videos in which I explain the content in more detail or demonstrate how to solve problems. While much of the material focuses on argument analysis, the textbook and videos also introduce other critical thinking concepts along the way, such as criteria for evaluating different kinds of arguments, rules of logical inference, certain fallacies, and guidelines for finding and evaluating empirical evidence.

Keys and Mastery Checks

To advance in the course, students must demonstrate mastery of the material by passing "mastery checks." Most mastery checks are short quizzes that students can complete in about 5

to 20 minutes. Some of these are multiple choice, some require students to map arguments, and others ask students to write one or two paragraphs evaluating arguments. There are several mastery checks in each step, which test students on different aspects of the material in that step, and there's a total of 28 mastery checks in the course.[4] Grades in the course depend on how many mastery checks students complete. Students need to pass every mastery check in order to receive an "A," students who pass most but not all mastery checks receive a "B," and so on.

Before students can attempt mastery checks, they must complete assignments that I refer to as "keys." There are two kinds of keys: reading keys and quiz keys. Reading keys require students to read part of the textbook or watch relevant videos while quiz keys ask students to complete a quiz that tests their understanding of the material. Both reading and quiz keys are automated. Students use Perusall, an online social annotation platform, to do the readings and watch the videos. One major advantage of Perusall is that it automatically assigns students a passing grade when they've finished the reading. Students then complete quiz keys on the course's learning management system. The quizzes cover material that is similar to the content of the mastery checks. Students can take these quizzes an unlimited number of times, but they're required to score 90 percent or higher on each quiz key before they pass. For any given mastery check, students must usually pass both reading and quiz keys.

The key system helps prepare students for mastery checks. If students weren't required to complete the reading and practice problems before they attempted mastery checks, then some students would refrain from doing so. These students might then attempt mastery checks anyway, despite the fact that they're unprepared. The likely outcome is that students would do badly on mastery checks, which can frustrate them and lead to an inefficient use of the instructor's time. The quiz and reading keys make this inefficient outcome less likely. I also provide additional

practice exercises to students and encourage them to finish them before they attempt mastery checks. But these additional practice exercises are optional.

Let's suppose that a student has completed the necessary keys and would like to take a mastery check. I place folders throughout the classroom that contain mastery checks and each folder is clearly labeled. When a student is ready to take a mastery check, she takes them out of the folder and go back to her desks and works on the assignment. Once this student is done, she brings the mastery check to me. I then grade the mastery check in front of her. If this student scores 90 percent or higher on the mastery check, then she passes and can advance to the next part of the course. If the student has yet to pass, I point out her mistakes and suggest ways to improve on the next attempt. When she's ready, this student tries again and takes a different version of the mastery check, and the process repeats itself.

While students can attempt mastery checks an unlimited number of times, I only allow three attempts on the same mastery check per class session. For example, suppose that a student is working on mastery check 2.2, which requires this student to map arguments with intermediate conclusions. And let's assume that this student takes mastery check 2.2 three times and has yet to score 90 percent or higher. The student must wait until the next class session to try again. This rule discourages students from repeatedly taking mastery checks on the off chance that they'll pass, even though they haven't actually mastered the material. If students can only take three of the same mastery check per session, then this is another incentive to adequately prepare for the check.

Constructing and Communicating Arguments

The final step in the class focuses on the construction and communication of arguments. This step differs from other ones in the class. For one thing, several of the mastery checks in the final step are not quizzes *per se*. Rather, they are steps in the process of constructing an argument.

For the first mastery check in the final step, a student must research a topic of interest and formulate a thesis that reasonable and well-informed people disagree with. Many students want to select a topic that is uncontroversial among well-informed observers ("racism is unfair" or "climate change is real"). But I encourage them to pick a conclusion that confronts serious opposition and that's also appropriate for an academic assignment. Next, students map their own arguments for their conclusions. This forces students to clarify their argument and allows me to give them preliminary feedback on the cogency of their reasoning. Students then complete a module on identifying good empirical evidence. For the final mastery check in the step, students submit a paper that's about 3,000 words long. This is the only mastery check that students can submit outside of class – students need to submit every other mastery check during a class session or office hours.[5] And, given that it is time-consuming to write comments on a paper of this size, I only allow students to submit their papers once every week. It often takes students several tries before they pass this final mastery check (to pass, papers must satisfy all of the criteria on a detailed rubric). Once students complete this final mastery check, they're done with the course.

Course Policies and the Class Environment

The major policies of the course follow standard PSI practices. I refrain from giving any whole group instruction, with an exception for the first week of classes. There's no attendance

requirement. Students come and go as they please. Students often arrive after class has started or leave well before class is over. When I teach multiple sections of the course, some students come to every section in order to make rapid progress in the class. Others decide to take a break from the course for a week or two. What are class sessions like? Some students quietly work on keys or mastery checks, some students chat or do practice problems together, and others discuss their mastery checks or practice problems with me. During any given class session, almost every student in class is actively engaged with the material because this engagement is necessary to advance in the class.

5. Evidence

In this section, I'll explore whether teaching critical thinking with PSI significantly improves the critical thinking skills of students.

Evidence from Pre-Tests and Post-Tests

A total of seventy-three students (sixteen men and fifty-seven women) enrolled in my sections of Critical Thinking during Fall 2021 and Spring 2022. I administered a pre-test and post-test at the beginning and end of the semester. This test consists of twenty-five questions from the logical reasoning section of the LSAT. Ten questions covered argument analysis, such as the relation between claims in the argument, eight questions asked about the identification of missing premises and conclusions, and seven questions involved argument evaluation. I used question from the logical reasoning portion of the LSAT because this is widely regarded as a reliable and psychometrically valid test of reasoning skills (Anthony et al., 2016; Walzer et al.,

2019). My own judgment is that the logical reasoning section of the LSAT is a challenging but accurate measure of one's skills at argument analysis and evaluation.

While I took all of the questions on the pre- and post-test from official versions of the LSAT, I didn't use the official forms of the test. The reason is the pre-tests and post-tests were part of students' grades. Thus, I felt an obligation to align the content of these tests with the material that students learned in the class, at least in a general way. Yet some questions in the official forms of the LSAT ask about material that students never learned in much detail, such as arguments by analogy and inference to the best explanation.[6] So, I excised material that was not related to the content of my course and only used questions that were connected, if only in a loose way, to the course material. Furthermore, I administered the same test as the pre- and post-test. But I never revealed the answers to the questions on the test and I didn't allow students to study or keep the test. I also administered the post-test fourteen weeks after I administered the pre-test. While it's difficult to be sure, it seems unlikely that familiarity with the test significantly influenced performance on the post-test.[7]

Students performed significantly better on the post-test (M = 17.53, SD = 3.23) than they did on the pre-test (M = 14.31, SD = 3.56), $t(72) = 7.812$, $p < .0001$, $d = 0.94$, 95%, CI: [0.661, 1.227]. Students answered 3.2 additional questions correctly on the post-test and 80% of the class improved. I assessed the effect size of the course using Cohen's *d*. Cohen's *d* is a measure of effect size based on the differences between two means – in this case, the mean scores on the pre-test and the post-test. A common convention is to describe an effect size of 0.2 as small, 0.5 as medium, and 0.8 as large. Thus, the effect size of my course ($d = 0.94$) indicates a large effect. These results suggest that participating in the course improved students' argumentative reasoning skills.[8]

Now, you might worry that the course involves "teaching to the test" and that this explains the above gains. There's a grain of truth to this worry. Both the course material and the logical reasoning of the LSAT focus on argument analysis and evaluation. Thus, it stands to reason that participating in the course would fosters skills that also help on the LSAT. Yet this is hardly an objection. After all, the goal of the class is to teach these skills. Moreover, the problems on the LSAT are different from the problems that students encounter in the class. For example, no question on the LSAT asks students to map arguments and the techniques for evaluating arguments and identifying missing claims that the course teaches have only partial applicability to the LSAT. Despite my attempts to align the pre- and post-test with the course, some students complained on the course evaluations that the pre- and post-tests were not closely related to the course material and were unfair for that reason! So, the fact that the course improved student performance on the logical reasoning section of the LSAT suggests that students did learn reasoning skills that transfer, at least somewhat, to new material.[9]

Nonetheless, I lack a true control group. Therefore, it's unclear to what extent teaching critical thinking with PSI *caused* the improvement on the post-test. Maybe students would have improved just as much if I taught the class using a more traditional format. While I can't rule out this possibility, this seems unlikely. To explain why, I'll now compare the above results with evidence from earlier versions of the course.

Comparison with an Earlier Course

Before I started teaching critical thinking with PSI, I taught the course in a more standard format. The course was not self-paced and it used a conventional grading scheme. Class activities included lectures, whole group discussion, and assignments that students completed in

small groups. Argument mapping was also a central feature of the course. Students would often work with partners to map out arguments using a computer program. I adopted this approach because the past research on argument mapping used a similar format. I was inspired in particular by Simon Cullen's seminars on argument mapping and philosophical analysis at Princeton University. In these seminars, the instructors gave students passages from complex philosophical texts and students worked in pairs to map the arguments in these texts. The instructor would then give students written feedback on their maps and provide model answers to the exercises. Cullen et al. (2018) report that participating in this seminar had large positive effects on the reasoning skills of students. I tried out a similar approach in my course. Students would be required to map out arguments that I adapted from philosophical texts and newspaper editorials.

To determine whether my course was effective, I administered a pre-test at the beginning and a post-test at the end of the course. This was also the logical reasoning version of the LSAT. A total of thirty-five students (twelve men and twenty-three women) completed both the pre-test and the post-test in the spring of 2019. To my surprise (and, to be honest, disappointment), the students did not perform better on the on the post-test (M = 13.57, SD = 3.94) than they did on the pre-test (M = 13.91, SD = 3.65), $t(34) = -0.576$, $p = 0.569$, $d = -0.09$, 95%, CI: [-0.397, 0.218]. In fact, students did slightly worse on the post-test, although the effect was not statistically significant. So, judging from the post-test, my class failed to improve the argumentative reasoning skills of students.

What explains the difference between the PSI course and the earlier, more traditional course? Several differences between these courses could have influenced the results. First, in the earlier course, the only incentive for students to take the pre-test and post-test was extra credit,

which was equivalent to about 1-2 percent of students' final grades. In contrast, students in the PSI classes were required to complete these tests and performance on them was worth 10 percent of the final grade.[10] Given the smaller stakes, students in the earlier course may have been less motivated to do well. Second, much of the material in the earlier class was different from the material in the later version of the course. We used a different textbook and many of the course exercises and assignments were different. As I mentioned, students in the earlier course focused more on mapping longer and more complex arguments. Perhaps this difference in materials and emphasis were a less effective preparation for the logical reasoning section of the LSAT.

Nonetheless, the contrast between my earlier, more traditional course and the later version that was built around PSI is suggestive. A natural interpretation of the results is that PSI improved learning and this improved learning explains, at least in part, the difference between the classes. For what it's worth, my informal observation is that students learned the material much better in the PSI version of the course than in the more traditional version.

Comparison with Other Research

How do my results compare to the effect sizes that are reported in the literature on critical thinking instruction? This question is difficult to answer because I used a non-standard test of reasoning skills. This makes it fraught to compare my results with the results from studies that use common critical thinking tests. Nevertheless, I'm aware of two studies that used the logical reasoning section of the LSAT to study gains in critical thinking skills. It may be useful to compare the results from these studies with my findings.

First, there's the study by Simon Cullen, et al. (2018) that I mentioned above. The authors used the logical reasoning section of the LSAT to measure reasoning skills and they report an

effect size of $d = 0.77$ for students who were enrolled in an freshmen seminar dedicated to philosophical analysis and 0.11 for a control group of Princeton undergraduates who weren't enrolled in the course. The second relevant study comes from Thomason et al (2013). This group of researchers used the logical reasoning section of the LSAT as a pre-test and post-test to study the impact of critical thinking courses that focus on argument mapping, and report an effect size of 0.307. Notably, Thomason et al. report effect sizes from different standardized tests of critical thinking and find that improvements on the LSAT were much smaller than improvements on these other tests, which the authors attribute to the greater difficulty of the LSAT and the higher literacy level that it requires.

Although we should take these comparisons with a grain of salt[11], the effect size from critical thinking courses with PSI is as large or larger than others reported in the literature.

6. Problems, Challenges, and (Some) Solutions

In this section, I'll discuss some challenges that I've encountered while teaching critical thinking with PSI. I'll also explain how I've addressed some of them.[12]

Instructor Workload

Despite evidence of the superior effectiveness of PSI over traditional instruction, few instructors now use PSI. Why? Observers have given different reasons. But one common explanation is that PSI increases instructor workload (Buskist et al., 1991). Instructors who use PSI must create an enormous number of tests, create study guides and practice exercises, and hire and manage proctors. Many instructors who have tried PSI observe that it takes more time and effort than other pedagogical approaches (A. Tyree, 2013). And this point also applies to

teaching critical thinking with PSI. Here's my rule of thumb: it's usually necessary to have at least six versions of each mastery check. Thus, my course requires about 160 tests and a large volume of practice exercises. Furthermore, most of the materials for the course must be finished before the semester starts since it's often infeasible to create the necessary number of tests and practice problems during the semester (Keller & Sherman, 1974: 48). So, PSI can sharply increase instructor workload.

In my experience, creating mastery checks is the aspect of PSI that takes the most time. It's very challenging to create dozens of high-quality tests that are well-aligned with the material. But there's a partial solution to this problem. Instructors can share material with one another. This is what allowed me to start using PSI. Dona Warren, the author of the textbook that my class uses, shared her quizzes and tests with me. I then adapted these materials into mastery checks. While I still had to create about 40 additional mastery checks, I doubt I could have started using PSI without Warren's generous assistance. I'd also be pleased to share the extensive set of materials that I've created with any other instructors who are interested in using PSI to teach critical thinking.[13] If instructors share materials in this way, the cost of transitioning to PSI drop considerably.

Once an instructor has sufficient tests and practice exercises, a PSI course probably takes about as much time as a conventional one. I grade most of the mastery checks during class and, thus, I don't need to devote much time to grading outside of class. Instead of preparing lectures and in-class activities or grading work, I spend more time tutoring students, organizing the materials for the class, and keeping track of student progress.

Less Fun

A PSI instructor abandons the role of sage on the stage and becomes more of a tutor and manager of learning. Some instructors might find the less central role of the professor in PSI courses to be unattractive. William Buskist et al. (1991: 230) put the worry as follows: "In a PSI-based course, the student is the star of the show, not the teacher. For many teachers this is not a desirable state of affairs. Many teachers like to lecture, they enjoy the stardom, and are reluctant to give it up…. To be a manger of learning just doesn't have the same ring to it."

It might be tempting to dismiss this resistance to PSI as narcissism. But I think that's a mistake. If instructors stop enjoying their courses, then this is a problem. For one thing, it's unlikely that a method of instruction is sustainable if it makes instructors miserable to use it, regardless of its superior efficacy. And I can report that, in my experience, using PSI can be less fun than teaching in a more traditional format. When you use PSI, you must give up the high of a well-received lecture or a vibrant class discussion. This is a genuine cost. Nonetheless, PSI has some distinctive joys of its own. First, I believe that students learn better through PSI. And it's satisfying to see students learn well. Second, I often get to know students better when using PSI because I often interact with them on a one-to-one basis while evaluating their mastery checks and providing them feedback. Whether these benefits of PSI outweigh the costs depends, of course, on the personality, interests, and situation of particular instructors.

Procrastination

One concern about PSI is student procrastination. Observers worry that self-pacing will cause students to procrastinate and this will cause them to fail to pass the class. Or perhaps students will submit most of their work near the end of the semester, which may place a

considerable strain on the instructor. Are these concerns well-founded? And, if so, how should instructors who use PSI respond?

In a meta-analysis of PSI, James Kulik et al. (1979) find that PSI courses don't have lower completion rates than conventional classes. This is true to my experience. Only a small handful of students have failed to pass my class, and I doubt that I would reduce this number by using a more conventional method of instruction. This falls short of showing that procrastination isn't a problem, though. PSI instructors often report that it is (J. Fox, 2004).

But I've found ways of reducing student procrastination. First, I've created what I call a "Pacing Guide" that gives students example paths in the class, depending on their goals, and I share this guide with students on the first day of class. For example, I'll provide a sample schedule for a student who wants to receive an "A" in the course, another schedule for students who want to earn a "B," and so on. These sample schedules tell students approximately what they should complete each week in order to achieve their goals in the course. Second, I send out weekly emails to the class reminding students which mastery checks they should finish during the week to stay on their preferred track. I'll say things like: "if you're on the B Track, then you should pass mastery check 3.3 and 3.4 this week…." These simple interventions have dramatically decreased procrastination in my classes.

Struggling Students

Another problem with PSI is struggling students who are unable to pass a mastery check. Struggling students can become demoralized and resentful, and helping these students is a major challenge. Here's my advice. First, the best cure is prevention. The primary reason that students struggle is that they invest an inadequate amount of time and effort in preparing for mastery

checks. They skip the reading and neglect to do practice problems. For this reason, the key system is vital. If a mastery check is difficult, then the keys for this mastery check must require students to do a considerable amount of practice on tasks that resemble the mastery checks. So, once a student completes the keys, then she'll be in a decent position to pass the mastery check.

Second, you should be prepared to tutor struggling students and even to permit them to take mastery checks during office hours. My rule is that students who are falling behind the "A-Track" (the schedule to receive an "A" in the course) on the pacing guide are eligible to take mastery checks in office hours. If students are struggling, I insist that they come to office hours so that I can provide them with detailed feedback on their performance. These measures – keys that requires considerable practice and individualized tutoring – are sufficient to resolve the problems that students are having in the vast majority of cases.

Providing Rapid Feedback

PSI depends on rapid feedback. So, I aim to provide students with feedback on their mastery checks within a few minutes of submission. Yet this can be challenging when I have, say, twenty-four students who are submitting mastery checks during the same class session. How can instructors reduce the backlog of mastery checks?

One simple intervention is answer keys. I created a book of answer keys to all of the mastery checks in the course with index tabs that made it easy to find the answers to specific mastery checks. This book of answer keys helps me to swiftly grade mastery checks.[14] Another solution to the backlog problem is costlier and, unfortunately, won't be available to every instructor. I hire a student tutor. When students have submitted a large number of mastery checks, the student tutor helps me to grade them. The tutors use the book of answer keys to

review mastery checks and, if a student's answers are different from those in the book of answer keys, then the tutor marks these answers as incorrect. The tutor then hands me this mastery check and I give it another quick review. I then call up the student who submitted the mastery check and give him or her feedback. This system allows me to provide feedback on most mastery checks within about five minutes of submission, even on the busiest days. However, some of the mastery checks in the advanced steps of the course are more time-consuming to evaluate since they may ask students to analyze complex, real-world arguments. For these mastery checks, I'll often provide students with preliminary feedback on their work in class and then send them additional feedback after I've had the chance to review their work in greater detail.

Other instructors who teach PSI courses use automated mastery checks on computers in order to provide students with rapid feedback (Bergmann, 2023; A. L. Tyree & Rawson, 1993). But this isn't fully possible in my courses since computers can't yet grade argument maps. At least, I haven't discovered a way for them to do this.

Student Evaluations

Kulik's et al.'s (1979) meta-analysis of PSI finds that students gave PSI courses higher ratings than lectures. But I'm sorry to say that PSI has not made me more popular with students. Students rate my Critical Thinking courses significantly lower than other courses that I teach and my evaluations have declined since I adopted PSI. Common complaints include: the class is too hard, there's too much work, some students dislike self-pacing, some students want discussion or whole-group lectures, and there's less community in my classes compared to more conventional ones. A majority of students report that they like the course and feel that they learned a great deal. But a non-trivial minority of students rate the class poorly.

Unfortunately, a course that improves learning can also decrease student satisfaction. Meta-analyses and large-sample studies find that student ratings are not correlated with learning (Boring et al., 2016; Kornell & Hausman, 2016; Uttl et al., 2017). Furthermore, students may give more demanding courses worse ratings (Stroebe, 2016). The few randomized controlled trials that exist on this subject find that there's a negative correlation between student ratings in a course and students' performance in follow-on courses, which indicates that students give lower ratings to more effective teachers (Braga et al., 2014; Carrell & West, 2010).

One finding from the literature on student evaluations seems especially relevant to PSI. Some studies suggest that students believe that they've learned more when they passively listen to a lecture instead of engaging in active learning tasks, even though they actually learn more from active learning (Carpenter et al., 2020; Deslauriers et al., 2019). Why? One reason is that many students have poor metacognition. They fail to understand how much, or how little, they've learned. Active learning in which students must struggle to solve problems makes students aware of their lack of understanding. This leads students to lower their estimates of their learning. This contrast with a passive lecture where students can avoid testing their understanding of the material. I suspect something like this process is at work in my classes. My course requires that students constantly demonstrate how much they've learned. If students have poor metacognition, then this will cause students to lower their estimates of their learning. Yet other courses that require less testing will refrain from upsetting students' overly optimistic estimates of how much they've learned. So, students prefer those courses.

Regardless of whether this is the right explanation or not, teaching critical thinking with PSI might pose a tradeoff. There's more learning but less student satisfaction. Although I've tried different strategies to increase student satisfaction, my efforts have largely failed. And it's

possible that the tradeoff between learning and student satisfaction would generalize to other instructors, if they were to try teaching critical thinking with PSI. But universities tend to reward or punish instructors on the basis of student ratings rather than learning. Thus, if administrators and faculty prioritize high student ratings and the tradeoff between student satisfaction and learning generalizes, then teaching critical thinking with PSI is a risk for instructors, especially non-tenured faculty who are more vulnerable to administrative interference and sanction.

7. Conclusion

In their review of the history of PSI, Buskist et al. (1991) note that one underexplored issue with PSI is the relationship between this method of instruction and "higher-level" thinking skills:

> Another relatively unaddressed issue is developing (wider spread) applications of PSI to teaching "thinking" and other cognitive skills. Many educators are aware that PSI works well in lower-level college courses, but complain that it cannot be applied to upper-division courses involving more "thinking" (298).

The evidence presented in this paper suggests that PSI can, in fact, promote key cognitive abilities, such as critical thinking skills. PSI may also have downsides, such as an increased instructor workload and lower student satisfaction. But, if teaching critical thinking with PSI is as effective as the preliminary evidence indicates, then this method of instruction is worth further consideration and study.

Works Cited

Anthony, L. C., Dalessandro, S. P., & Trierweiler, T. J. (2016). *Predictive Validity of the LSAT: A National Summary of the 2013 and 2014 LSAT Correlation Studies* (LSAT Technical Report Series). Law School Admission Council, Inc.

Arum, R., & Roksa, J. (2011). *Academically Adrift: Limited Learning on College Campuses*. University of Chicago Press.

Bergmann, J. (2023). *The Mastery Learning Handbook: A Competency-Based Approach to Student Achievement*. ASCD.

Boring, A., Ottoboni, K., & Stark, P. (2016). Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. *ScienceOpen Research*. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Boucher, S., Barnett, A., & Thomason, N. (2015). *Improving Your Reasoning*. Melbourne Critical Thinking Project.

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, *41*, 71–88. https://doi.org/10.1016/j.econedurev.2014.04.002

Buskist, W., Cush, D., & DeGrandpre, R. J. (1991). The Life and Times of PSI. *Journal of Behavioral Education*, *1*(2), 215–234. JSTOR.

Cahill, A., & Bloch-Schulman, S. (2012). Argumentation Step-By-Step: Learning Critical Thinking through Deliberate Practice. *Teaching Philosophy*, *35*(1), 41–62.

Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On Students' (Mis)judgments of Learning and Teaching Effectiveness. *Journal of Applied Research in Memory and Cognition*, *9*(2), 137–151.

Carrell, S. E., & West, J. E. (2010). Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors. *Journal of Political Economy*, *118*(3), 409–432. https://doi.org/10.1086/653808

Clark, S. G. (1974). An Innovation for Introductory Sociology: Personalized System of Instruction. *Teaching Sociology*, *1*(2), 131–142. JSTOR. https://doi.org/10.2307/1317148

Cullen, S., Fan, J., van der Brugge, E., & Elga, A. (2018). Improving analytical reasoning and argument understanding: A quasi-experimental field study of argument visualization. *Npj Science of Learning*, *3*(1), 21.

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, *116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

Dunkin, M., & Barnes, J. (1986). Research on Teaching in Higher Education. *Handbook of Research on Teaching*, *3*, 754–777.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning*, *7*(3), 219–244. https://doi.org/10.1007/s11409-012-9092-1

Ennis, R. H. (2015). Critical Thinking: A Streamlined Conception. In M. Davies & R. Barnett (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education* (pp. 31–47). Palgrave Macmillan US. https://doi.org/10.1057/9781137378057_2

Gelder, T. van. (2015). Using Argument Mapping to Improve Critical Thinking Skills. In *The Palgrave Handbook of Critical Thinking in Higher Education* (pp. 183–192). Palgrave Macmillan, New York. https://doi.org/10.1057/9781137378057_12

Harrell, M. (2005). Using Argument Diagramming Software in the Classroom. *Teaching Philosophy*, *28*. https://doi.org/10.5840/teachphil200528222

Harrell, M. (2016). *What Is the Argument?: An Introduction to Philosophical Argument and Analysis*. The MIT Press.

Haskell, R. E. (2001). *Transfer of learning: Cognition, instruction, and reasoning* (pp. xx, 241). Academic Press. https://doi.org/10.1016/B978-012330595-4/50003-2

Hitchcock, D. (2017). *On Reasoning and Argument: Essays in Informal Logic and on Critical Thinking* (1st ed. 2017 edition). Springer.

J. Fox, E. (2004). Chapter 12—The Personalized System of Instruction: A Flexible and Effective Approach to Mastery Learning. In D. J. Moran & R. W. Malott (Eds.), *Evidence-Based Educational Methods* (pp. 201–221). Academic Press. https://doi.org/10.1016/B978-012506041-7/50013-9

Keller, F. S., & Sherman, J. G. (1974). *Psi, the Keller Plan Handbook: Essays on a Personalized System of Instruction*. W. A. Benjamin Advanced Bk Program.

Klopfenstein, K. F. (1977). The Personalized System of Instruction in Introductory Calculus. *The American Mathematical Monthly*, *84*(2), 120–124.

Kornell, N., & Hausman, H. (2016). Do the Best Teachers Get the Best Ratings? *Frontiers in Psychology*, *7*, 570–570. PubMed. https://doi.org/10.3389/fpsyg.2016.00570

Kuhn, D. (1991). *The Skills of Argument*. Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9780511571350

Kulik, J. A., Kulik, C. C., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist*, *34*(4), 307–318. https://doi.org/10.1037/0003-066X.34.4.307

Liu, O. L., Roohr, K. C., Seybert, J. M., & Fishtein, D. (2021). Are Fourth-Year College Students Better Critical Thinkers than Their First-Year Peers? Not So Much, and College Major and Ethnicity Matter. *Educational Measurement: Issues and Practice*, *n/a*(n/a). https://doi.org/10.1111/emip.12430

Ortiz, C. (2007). *Does philosophy improve critical thinking skills?* [The University of Melbourne, Victoria, Australia.].

Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How Robust Are the Findings of Academically Adrift? *Change: The Magazine of Higher Learning*, *43*(3), 20–24.

Possin, K. (2008). A Field Guide to Critical-Thinking Assessment. *Teaching Philosophy*, *31*(3), 201–228.

Protopapas, P. N. (1974). The Keller Plan Implementation of the Personalized System of Instruction in a Freshman Biology Course. *The Science Teacher*, *41*(5), 44–46. JSTOR.

Stroebe, W. (2016). Why Good Teaching Evaluations May Reward Bad Teaching: On Grade Inflation and Other Unintended Consequences of Student Evaluations. *Perspectives on Psychological Science*, *11*(6), 800–816. https://doi.org/10.1177/1745691616650284

Thomason, N., Adajian, T., Barnett, A., Boucher, S., Van der Brugge, E., & Campbell, J. (2013). *A012. Critical Thinking Final Report* (N66001-12-C–2004; Critical Thinking and Argument Mapping Project, pp. 1–139). University of Melbourne.

Twardy, C. (2004). Argument Maps Improve Critical Thinking. *Teaching Philosophy*, *27*(2), 95–116.

Tyree, A. (2013). *The Keller Plan at Law School*. http://www2.austlii.edu.au/~alan/j_leged.html

Tyree, A. L., & Rawson, S. L. (1993). The Keller plan at Sydney law school. *The Law Teacher*, *27*(1), 58–68.

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22–42. https://doi.org/10.1016/j.stueduc.2016.08.007

van der Brugge, E. (2018). *The use of argument mapping in improving critical thinking* [University of Melbourne]. http://hdl.handle.net/11343/214519

Walzer, M., Dustman, K., Langer, D., Hegarty, M., Camilli, G., & Gallagher, A. (2019). *LSAC 2018 Skills Analysis Study: Content Validity of the LSAT*. Law School Admission Council, Inc.

[1] Cahill and Bloch-Schulman (2012: fn. 22) suggest that their approach in fact does differ from PSI: "One crucial difference is in our reconceptualizing the content of the logic course, and thus, the way the step-by-step method leads to deliberate practice and deep learning. In Keller's plan, the units remain traditional (he describes a class having thirty units over the course of the semester), and the content remains the same (in kind and amount) as in a traditional psychology class." Setting aside the question of whether Cahill and Bloch-Schulman's course qualifies as a version of PSI, there are also several other differences between their course and the course that I will describe later in this paper. Among other differences, students in Cahill and Bloch-Schulman's course must request permission from the instructor before taking mastery checks and the final step in their course involves a group project. In contrast, my course doesn't require students to receive permission before taking mastery checks and there are no group projects.

[2] There are several textbooks that focus on argument mapping (Boucher et al., 2015; Harrell, 2016). The non-profit organization ThinkerAnalytix has an online course that teaches argument mapping, which can be located at the following address: https://course.thinkeranalytix.org/

[3] You can access a version of the textbook at the following address: https://www.critical-thinking-resources.org/critical-thinking-textbook

[4] When I started using PSI, I initially assigned just one mastery check for each step. Thus, there were only seven mastery checks in the course. Yet this caused problems. Since the mastery checks were long and covered all of the material in a step, students felt intimidated and overwhelmed by the mastery checks. So, a concerning number of students procrastinated and fell behind in the course. To address this problem, I broke each mastery check into multiple, shorter mastery checks that test only one component of the material in each step. Now, there are about three or four mastery checks per step. This change had the effect of making each mastery check less intimidating to students and it helped reduce procrastination.

[5] One benefit of this policy is that most of the course is ChatGPT-proof. The majority of mastery checks must be completed with pencil-and-paper and in-person. Students must also put away their computers and phones before taking a mastery check. But the final paper is an exception to these policies. Since students can work on the paper outside of class, it's possible that they will use large language models (LLMs) to write the paper. Yet I find that

popular LLMs such as ChatGPT are currently unable to reliably mimic the kind of writing that I assign for the final mastery check. This is so in part because the final paper requires extensive research and LLMs continue to hallucinate citations and empirical findings. In addition, my experience is that LLMs have trouble defending an argument over thousands of words in a sustained way. Nonetheless, like other instructors, I'm also grappling with how to assign and evaluate out-of-class writing in light of the growing capabilities of LLMs and I may need to reconsider the final paper in the future.

[6] You might object that a critical thinking course should teach argumentative patterns such as inference to the best explanation and argument by analogy, and that it's a defect of my approach that my course neglects to cover this material in much detail. Perhaps this objection is correct. But my course aims to hone the foundational argumentative skills of my students and I'm wary about adding additional material into the course that may detract from this aim. That said, it's difficult to determine whether I've made the right tradeoff between honing fundamental skills and covering important critical thinking topics. I want to thank an anonymous referee for raising this concern.

[7] I'm happy to share both the tests that I've used and the all of the data that I report upon request.

[8] I also administered the same pre-test to three sections of Critical Thinking during the spring of 2020. However, the COVID pandemic prevented me from administering the post-test to the whole class. But I did offer students the opportunity to take the post-test for extra credit. Eight students ultimately took the post-test. On average, these students performed significantly better on the post-test (M = 17.5, SD = 2.67) than they did on the pre-test (M = 15.625, SD = 2.55), $t(7) = 1.488$, $p = 0.18$, $d = 0.716$, 95%, CI: [13.48, 15.26]. Since the conditions under which students took the post-test during the COVID pandemic differed substantially from the other semesters in which I administered these tests, I've decided not to include this data in the sample that I discuss in the main text.

[9] But you might object that the questions on the logical reasoning section of the LSAT are highly artificial and curated. So, even if students do better on this test, this fails to show that students' argumentative reasoning skills transfer to real-world arguments, which are often more ambiguous and complex than the arguments on the LSAT. I concede that my students' skills at argument analysis and evaluation might not transfer in this way. But here are two considerations that are worth keeping in mind. First, my course does require students to analyze and evaluate some real-world arguments, and it's possible that this task helps students to transfer their skills (see note 14 for further details). Second, a concern about transfer of learning applies to a wide variety of interventions that aim to promote learning, including interventions that seek to improve critical thinking skills. Consider the research on argument mapping. When researchers study the effects of argument mapping, they typically use standardized tests of critical thinking skills to measure learning gains. Yet it's often uncertain whether the benefits of argument mapping transfer beyond the standardized tests that researchers use to study learning. The challenge of measuring transfer of learning is a difficult one and I'm unable to satisfactorily address this issue in this paper. I want to thank an anonymous reviewer for raising these concerns. For further discussion of transfer of learning, including interventions that aim to improve reasoning abilities, see Haskell, (2001).

[10] In the PSI classes, I dropped the lowest score on the pre-test and post-test and only counted the highest score across both tests.

[11] There are several differences between these studies and mine that could have affected the outcome. One difference is that I was using a modified version of the logical reasoning portion of the LSAT, not an official version. These other studies used official versions of the test.

[12] My students differ from students at other universities in important ways. Since my university is selective, the students that I teach are more academically prepared than most college students on average. While instructors have used PSI successfully at a range of different institutions, I can only draw on my own experience. I'll avoid claiming that my advice can apply to every educational context.

[13] If you'd like access to the materials for my course, please email me at [omitted for blind review].

[14] There are sometimes multiple correct maps of an argument and, if that's the case, a student will pass the mastery check if he or she produces one of the correct maps. But it becomes difficult to rapidly evaluate students' argument maps if there are many different, but equally correct, maps of the argument in question. For this reason, I curate the arguments on the mastery checks to reduce their ambiguity, which constrains the number of correct solutions. Yet this strategy generates a tradeoff. On the one hand, I can provide students with rapid feedback on their work if there's only or two viable interpretations of an argument. On the other hand, arguments in the real world are often ambiguous and there are sometimes many plausible interpretations of them. So, if students only analyze arguments that are heavily curated, then their skills of argument analysis will be less applicable to arguments "in the wild." Here's how I've addressed this tradeoff. The earlier steps in the class use highly curated arguments that have only one or two solutions. I find that these relatively simple exercises are useful for teaching fundamental skills of argument analysis and evaluation. But more advanced steps sometimes require students to analyze and evaluate messy, real-world arguments. For example, I use the following assignment in the sixth step of the course: students

must have a conversation with another person, such as another student or family member, about an issue over which they disagree, the student then maps the main arguments for his or her interlocutor, and writes an evaluation of this interlocutor's arguments. This assignment may help students to extend their skills at argument analysis and evaluation to real-world arguments. I'm grateful to an anonymous reviewer for pressing me on these issues.