Journal of the American Philosophical Association (2024) 1–22 © The Author(s), 2024. Published by Cambridge University Press on behalf of The American Philosophical Association
DOI:10.1017/apa.2023.30

Does Studying Philosophy Make People Better Thinkers?

ABSTRACT: *Philosophers often claim that doing philosophy makes people better thinkers. But what evidence is there for this empirical claim? This paper reviews extant evidence and presents some novel findings. We discuss standardized testing scores, review research on Philosophy for Children and critical thinking skills among college students, and present new empirical findings. On average, philosophers are better at logical reasoning, more reflective, and more open-minded than non-philosophers. However, there is an absence of evidence for the claim that studying philosophy led to these differences. We present some preliminary and suggestive evidence that although some of these differences may be attributable to philosophical training, others appear to be selection effects. The key takeaway is that more data are needed. We conclude by urging philosophers and interdisciplinary collaborators to gather more data to test the claim that studying philosophy makes people better thinkers.*

KEYWORDS: Philosophy for Children, critical thinking, intellectual virtue, intellectual humility, teaching philosophy

Does Studying Philosophy Make People Better Thinkers?

Many philosophers think that, in addition to any intrinsic value that philosophy may have, the discipline is also instrumentally valuable insofar as it makes people better thinkers. Philosophers often claim that doing philosophy encourages people to question ideas or assumptions that others take for granted, to reflect more deeply, and to reason more carefully. Studying philosophy is also said to help people recognize more acutely the limits of their own understanding, opening their minds and awakening them from dogmatic slumbers. To illustrate, the website of the American Philosophical Association (APA 2017) states that:

> The study of philosophy enhances, in a way no other activity does, one's problem-solving capacities. It helps students to analyze concepts, definitions, arguments, and problems. It contributes to students' capacity to organize ideas and issues, to deal with questions of value, and to extract what is essential from masses of information. It helps students distinguish fine differences between views and discover common ground between opposing positions.

In short, it is often claimed that the study of philosophy is distinctively well-suited to cultivating intellectual virtue.

We find this idea interesting in its own right. But there are also pragmatic reasons to investigate whether or not it is true. If we had compelling evidence that philosophy is instrumentally valuable in this way, this might encourage students to enroll in philosophy courses and encourage academic administrators to devote funds to philosophy programs. So, have these claims been empirically tested?

In recent years, numerous studies have empirically assessed philosophers and students of philosophy, often comparing them with non-philosophers (Kilov and Hendy 2022; Livengood et al. 2010; Schwitzgebel and Cushman 2012; Yaden and Anderson 2021). This article follows in that tradition. We review evidence from past research and also present some new findings of our own. Our aim is not to advance a specific view about what it means to produce better thinkers, but to determine what we can infer from existing studies using widely accepted measures. For this reason, we will not offer a precise definition of 'better thinker', but will remain open to diverse conceptions. There are numerous intellectual skills or virtues that might be cultivated by studying philosophy, such as critical thinking, logical reasoning, open-mindedness, or intellectual humility. By remaining open in this way, we can review a broader range of existing research and data and assess evidence relevant to various possible conceptions.

As shall become clear, evidence in favor of the view that studying philosophy improves thinking is weak, mixed, and ultimately inconclusive. There is a fundamental problem with much of the extant data—namely, they cannot differentiate between treatment effects and selection effects. A 'treatment effect' is a difference in outcomes that results from an external intervention. In this case, the 'treatment' is philosophy education. Hence, a treatment effect would simply mean that studying philosophy does make people better thinkers. By contrast, a 'selection effect' is a difference in outcomes that results from the way in which people's choices place them into different groups. In this case, people usually decide for themselves whether to study philosophy, and people who choose this are likely different from those who do not. Hence, simply comparing philosophers and non-philosophers might reveal differences. But it would not reveal whether such differences result from studying philosophy. Some of the evidence reviewed here will be able to differentiate at least partially between selection and treatment effects. However, much of it cannot. Hence, a key takeaway is that in order to answer this article's titular question, we need more data that enable tests for treatment effects specifically.

We will begin by reviewing existing evidence, including the oft-discussed topic of standardized testing scores (which we argue is very poor evidence indeed), studies on pre-college philosophy programs, and studies on critical thinking skills among college students. We then present some new empirical findings of our own. We first compare the intellectual traits of philosophers and non-philosophers at various levels of education. Then, we compare students at the beginning of a Philosophy 101 course with the general population, and examine some longitudinal data (i.e., data collected across multiple points in time). Throughout, we discuss the results of statistical tests in colloquial terms. However, detailed results from these analyses, along with the data, and R scripts used to run the analyses, are available online: https://osf.io/mbvpr/. Finally, we conclude the

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

article with a call for philosophers to begin working with others to collect more, and more illuminating, data.

Standardized Test Scores

Since the 1980s (Hoekema 1986), philosophers have observed that students who major in philosophy tend to score remarkably well on post-college standardized tests such as the Graduate Record Examination (GRE), Law School Admission Test (LSAT), and Graduate Management Admission Test (GMAT). Naturally, the rankings change somewhat from year to year, but it is common for philosophy to be one of the top-ranked majors, especially on the LSAT (APA 2019) and verbal reasoning portion of the GRE (APA 2014). These kinds of statistics are flattering and widely advertised by philosophy departments in the hopes of attracting greater enrollment. But are they compelling evidence for the claim that studying philosophy makes people better thinkers?

The obvious limitation of this form of evidence is that it does not differentiate between selection and treatment effects. That is, students who majored in philosophy may have high test scores because studying philosophy improved their thinking. But it is equally possible that people likely to get good test scores are more interested in philosophy to begin with. For example, people who are already academically talented or resourced (e.g., for reasons related to sociodemographics) may be disproportionately interested in studying philosophy or likely to stick with it. Perhaps the most likely explanation is that both selection and treatment effects are present. But we simply do not know how much of these differences in test scores can be attributed to treatment versus selection effects.

One way to address this issue would be to look at test scores for people who are interested in studying philosophy but have not yet taken any philosophy classes. That is, in addition to post-college tests like the GRE, one could look at pre-college tests such as the SAT and ACT (Metcalf 2021). If intended philosophy majors already score remarkably well on tests, then this would suggest that the high scores on post-college tests reflect a selection effect. However, if intended philosophy majors do not score especially well on the SAT and ACT, then this might be at least some evidence for a treatment effect.

Using data from the National Center for Education Statistics and the company that administers the ACT, we examined SAT scores from 2017 to 2021 and the ACT scores from 2013 to 2021. (For figures, see https://osf.io/dw3me.) Unfortunately, the testing companies do not treat philosophy as a distinct major but instead group philosophy with religious studies and theology. (We return to this problem below.) Intended philosophy and religious studies majors rank 10th out of 38 on the SAT and 9th out of 20 on the ACT. Considering specifically the reading and writing portion of the SAT, intended philosophy and religious studies majors rank 7th out of 38. In other words, even before going to college, these students are in the top half of the distribution.

Nonetheless, Thomas Metcalf (2021) has argued that philosophy majors' post-college test scores are *even better* than one would predict based on these pre-college scores. His idea is that if philosophy students move up in the rankings

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

(i.e., if their average percentile on post-college tests is higher than their average percentile on pre-college tests), then this improvement in relative position could be attributed to studying philosophy. Indeed, he finds that the average post-college percentile tends to be higher than the average pre-college percentile.

We argue that standardized test scores are not a good form of evidence because average scores on pre- and post-college tests do not enable meaningful comparisons. First, one problem is that philosophy is grouped with religious studies and theology for the SAT and ACT, but not for the GRE, LSAT, etc. This means that part of the first group is not in the second group. We do not know what proportion of that pre-college group is constituted by students intending to study religion and theology. It could be a small fraction, or it could be nearly the entire group.

Second, between the time when students take the pre-college tests and the time they start college, some students are likely to change their minds and pursue other fields of study. What proportion of the *intended* major group actually does major in philosophy? Again, we do not know.

Third, there are changes that occur during the college years. Because pre-college philosophy programs are so rare (at least in places like the United States), it is common for undergraduates to decide to major in philosophy only after taking their first philosophy class in college. Many students who major in philosophy do so after transferring from community colleges and so never take the SAT or ACT in the first place. We do not know what fraction of the post-college group falls into this category. Similarly, we do not know what fraction of the students who enter college with plans to study philosophy end up dropping the major sometime later.

Fourth and finally, only some college students decide to pursue further education and hence take post-college standardized tests. The subset that chooses to do this probably varies from discipline to discipline. This matters because the only way to compare pre- and post-college test scores is with a relative ranking (i.e., percentile). To illustrate why this is a problem, suppose that only the brightest and best philosophy majors take post-college standardized tests, whereas many of the more middling students from other disciplines do so. If that were the case, then the philosophy majors would place particularly well in the post-college ranking. But, again, that would be a selection effect.

In short, although people who study philosophy score very well on standardized tests, we do not know whether studying philosophy had any effect on this outcome. Even if there were no treatment effects whatsoever, changes in the composition of the groups being compared could easily explain differences in scores on pre- and post-college tests. Thus, standardized testing statistics do not provide evidence for the claim that studying philosophy makes people better thinkers.

In any case, there seems to be growing dissatisfaction with standardized testing at all levels of education in the United States. According to recent estimates, about 80 percent of colleges and universities in the United States either do not require or do not consider test scores for admission (FairTest 2022). As admissions committees and universities move away from considering and collecting these scores, we should perhaps also take this opportunity to look elsewhere when considering whether philosophy produces better thinkers.

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

Research on Philosophy for Children

Although most philosophy education takes place at the college level, some takes place in primary schools. These kinds of programs tend to be far more studied than their collegiate counterparts. The most well-studied program is Philosophy for Children (P4C). Early meta-analyses of studies on the impact of P4C found that the program led to small to medium-sized improvements in students' academic abilities (García-Moriyón, Rebollo, and Colom 2005; Trickey and Topping 2004). Those (older) studies tended to have quite small sample sizes. But one study, including over 2,000 primary school students in the United Kingdom, also found small but positive impacts on reading, mathematical, and reasoning abilities (Gorard, Siddiqui, and See 2015). However, the statistical analyses used in that study were harshly criticized (Inglis 2015; Thornton 2015), and a larger, more rigorous study intended to replicate those findings found no significant effects (Lord et al. 2021). Hence, although P4C is a comparatively well-studied program, it remains unclear whether it improves children's academic abilities.

Both the standardized testing scores and the metrics employed in prior P4C research are focused on general academic ability. Yet, is it really plausible that studying philosophy—or any particular subject, for that matter—increases a person's general academic ability or overall intelligence? If we are going to find effects of a philosophical education, we might do better to focus on specific intellectual skills or virtues. For example, at the beginning of this article, we articulated what we take to be relatively common claims about how philosophy opens one's mind, prompts one to think more clearly and more deeply, and so on. Hence, we might do better by focusing on outcomes like critical thinking and logical reasoning or open-mindedness and intellectual humility.

Research on Critical Thinking

One way in which studying philosophy might make people better thinkers is by improving their critical thinking skills. Many universities have implemented critical thinking courses, both within philosophy departments and without, as part of their general education curricula. But do philosophy courses typically improve students' critical thinking? And if so, how do those improvements compare with the improvements students could expect from taking other kinds of courses? According to the most recent and comprehensive meta-analysis (Huber and Kuncel 2016), students across majors show small to moderate increases in critical thinking skills while in college. Is there any evidence, then, that philosophy courses are special in this regard?

Empirical research on this question stretches back several decades. Some studies have found evidence that students in philosophy courses show greater improvements in critical thinking skills than do students in non-philosophy courses (Ross and Semb 1981). However, other studies have not shown this (Annis and Annis 1979). For example, one comparatively recent study (Burke et al. 2014) found evidence of a selection effect (i.e., students in a philosophy class showed better critical thinking skills than psychology students on both pre- and

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

post-tests), but no evidence of a treatment effect (i.e., neither the philosophy nor psychology students showed significant increases over the course of a semester). Because of these inconsistencies, it is valuable to look across many studies of college students' critical thinking skills.

A meta-analysis of fifty-two studies specifically investigated whether students in philosophy courses show greater gains in critical thinking than students in non-philosophy courses (Ortiz 2007). The results did not support that conclusion. However, the meta-analysis did find that a specific technique called 'argument mapping' (Harrell 2004) leads to substantially greater increases in critical thinking skills. Argument mapping is a technique for visualizing arguments by drawing hierarchical diagrams that illustrate the logical relations among propositions.

Since this meta-analysis was conducted, a number of further studies have tested for effects of training in argument mapping. One study compared students in a philosophy course that focused on argument mapping with control students who had expressed interest in the course but who were not able to enroll (Cullen et al. 2018). The results indicated that over the course of a semester the students who learned argument mapping showed substantially greater improvement in their performance on logical reasoning puzzles. Independent judges also rated the students' final papers for clarity, structure, and understanding of relevant arguments. Students in the argument mapping course received significantly higher scores. However, another study yielded less positive results (Dwyer, Hogan, and Stewart 2015). It found that university students with poor critical thinking skills at the start of a critical thinking course appeared to benefit from training in argument mapping. Yet, students who began the course with strong critical thinking skills actually showed *declines* over the course of the semester. Hence, the technique may be useful for helping students who are struggling with critical thinking, but it may also hold back students who are not.

There are at least two other reasons for tempered enthusiasm about argument mapping. First, argument mapping instruction usually involves having students work in groups. Other research has found that people reason better in groups (Dutilh Novaes 2020; Moshman and Geil 1998). Hence, it may be that observed benefits of argument mapping instruction come not from learning to map arguments, but from reasoning with other people. Second, given that studies failing to find effects of this technique are less likely to be published, the published evidence regarding argument mapping instruction may be skewed.

In sum, extant research does not show that philosophy courses are better suited to fostering critical thinking skills than other kinds of courses. However, it does point to a way in which philosophy courses might be made more effective in this regard—at least for some students. More studies are needed to confirm when and why argument mapping is beneficial.

Are Students of Philosophy More Intellectually Virtuous?

If we focus on whether studying philosophy cultivates specific intellectual skills or virtues then, apart from critical thinking skills, which should we focus on? Which skills or virtues is a philosophical education most likely to cultivate?

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

Doing philosophy often involves offering, dissecting, and reformulating arguments, and may therefore strengthen a person's ability to reason logically. Because logical reasoning tends to be slow and methodical, doing philosophy may also cultivate reflectiveness. When considering a question, some people tend to simply endorse the first idea that comes to mind, whereas others are inclined to stop and reflect on the question further. It seems plausible that studying philosophy might incline people toward the latter response. Through this process of reasoning and reflection, it is common to discover that one knows far less than one thought and that reasonable people hold many different views on fundamental questions. Thus, studying philosophy might foster the virtues of intellectual humility (i.e., an acute awareness of the limits of one's understanding; Whitcomb et al. 2017) and/or open-mindedness (i.e., a willingness to take new or unfamiliar ideas seriously; Montmarquet 1992).

In this section, we present empirical evidence that, on average, philosophers are more skilled at logical reasoning, more reflective, and more open-minded than non-philosophers. We also present some preliminary evidence about whether this is pure selection or whether there might also be a treatment effect.

Comparing Philosophers with Non-Philosophers

Nick Byrd (2022) recently conducted a study that found that certain intellectual traits correlate with the philosophical views that people hold. Because his data are publicly available (https://osf.io/a98ck/), we were also able to reanalyze them in order to investigate questions that he did not. Specifically, we tested whether people who have studied philosophy score differently on his measures than people who have not studied philosophy.

The participants in this study were recruited through ads on blogs such as *Leiter Reports* and *Daily Nous* and separately through an online research platform. The complete sample included 705 adults. However, we were unable to classify 27 of these as either having or not having studied philosophy. Accordingly, in these analyses, we examine the remaining $N = 678$ participants. Ages ranged from 19 to and 78 years ($M = 36.73$, $SD = 11.03$), 158 (23%) identified as female, 512 (76%) as male, 8 (1%) as other or declined to state; 64 (9%) identified as Asian, 16 (2%) as Black or African American, 526 (78%) as White, 29 (4%) as other, 41 (6%) as mixed race or ethnicity. Participants were asked whether they had or were a candidate for a PhD in philosophy. To this, 279 participants responded with 'Yes'. The other 399 were then asked to indicate the highest level of education that they had received and their primary subject of study. Of these, $n = 187$ indicated that they had studied philosophy.

The study included a 7-question logical reasoning measure, a multiple choice test asking participants what could be inferred from pairs of premises. To illustrate: 'All laloobays are rich. Sandy is a laloobay. If these two statements are true, can we conclude from them that Sandy is rich?'; 'In a box, some red things are square, and some square things are large. What can we conclude?'

The survey also included the Cognitive Reflection Test (CRT; Frederick 2005), which includes three questions designed to lure people into giving an intuitive

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

(non-reflective) but incorrect answer. For example, one question is, 'If it takes 5 machines 5 minutes to make 5 widgets, how long would it take for 100 machines to make 100 widgets?' For many people, '100 minutes' initially jumps out as the obvious answer. However, a moment of reflection will reveal that the correct answer is '5 minutes'. Along with the three questions from the original CRT, the reflectiveness measure in this study included fourteen additional questions of a similar form, each with a text-entry answer format. This measure can be scored in two ways: by summing the number of correct answers or by summing the number of 'lured' answers. The resulting scores are very highly correlated because when people do not give the correct answers, they *usually* give the lured answers. But this is not always the case. Hence, we examined both the number of correct or reflective answers and the number of lured or intuitive answers.

Finally, the study included the Actively Open-Minded Thinking Scale (Baron 2019). Unlike the previous measures, which are *tests,* this last measure is a self-report measure of open-mindedness. It asks participants how strongly they agree or disagree with ten statements such as, 'People should revise their beliefs in response to new information or evidence'. Participants responded on Likert scales ranging from 'Completely disagree' to 'Completely agree'. We took the average of the ten items.

Figure 1 shows the mean scores for each measure, grouped by education level and philosophical training. The results are striking. For people with doctorate or professional degrees (i.e., PhDs, MDs, and JDs), there are no differences between those who have studied philosophy and those who have not. In other words, philosophy PhDs do not seem to be any more reflective, skilled in logical reasoning, or open-minded than physicians, lawyers, and the like. Both groups basically max out the scales. However, at lower levels of educational attainment, people who have studied philosophy tend to score *substantially* higher than those

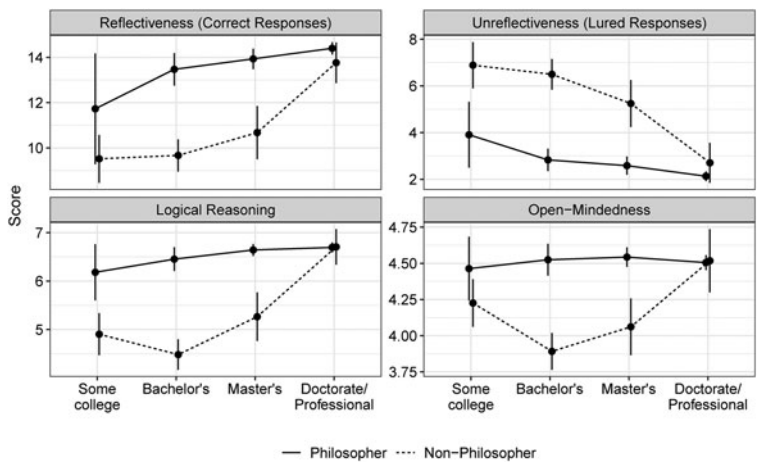

Figure 1. Logical reasoning ability, reflectiveness, and open-minded thinking, grouped by philosophical training. Points and vertical lines indicate means and 95 percent confidence intervals.

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

who have not. Replicating a finding from previous research (Livengood et al. 2010), we found that philosophers scored significantly higher on reflectiveness than non-philosophers at every level of education apart from doctorate/professional. (The pattern of results for correct versus lured answers was identical, only inverted.) This same basic pattern of results emerged for logical reasoning. However, for open-mindedness, the differences are statistically significant only for those with bachelor's and master's degrees. All the statistically significant differences between philosophers and non-philosophers would, by standard conventions, be considered 'large' or 'very large'.

Overall, higher levels of education generally come with more logical reasoning ability, reflectiveness, and open-mindedness. Curiously, however, for people who have *not* studied philosophy, those holding a bachelor's degree were significantly less open-minded than those who have not completed a college degree. (There may be a similar difference in logical reasoning ability. But in these data the difference was not statistically significant.) We are unsure what to make of this. One possibility is that upon receiving a college degree, the non-philosophers tended to become more dogmatic because—seeing that they are now '*educated*'—they need revise their beliefs no longer. Another possibility is that this difference arises from the fact that the group indicating 'some college' education encompasses both people who dropped out of college and people who are currently in college. There may be a mind-opening effect of the college context that wears off after one graduates.

Although these data are revealing, they share the same basic limitation as the standardized testing statistics. That is, they cannot differentiate between selection and treatment effects. It can be easy, when looking at an image like Figure 1, to forget that the variable on the *y*-axis is itself a potential and indeed *likely* cause of the variable on the *x*-axis. In this case, a person who is more logical, reflective, and open-minded is likely to pursue additional education and perhaps philosophical education specifically. Hence, although we see striking differences between groups, we do not know how many of these differences result from studying philosophy.

Comparing Philosophy 101 Students with US Adults

One strategy for addressing the question of treatment versus selection effects would involve assessing students as they start their first philosophy course. If students at the beginning of their philosophical education are no different from their peers, or from the population at large, then the differences we have observed between philosophers and non-philosophers might be due to a treatment effect. On the other hand, if those students already score substantially higher than others, then the differences we have observed are likely due primarily, if not entirely, to selection effects.

To address this question, we administered a survey with measures of several intellectual virtues to students during the first week of a Philosophy 101 class at the University of North Carolina at Chapel Hill. Given the paucity of pre-college philosophy education in North Carolina, it is safe to assume that most if not all of these students had no prior experience with philosophy. We received $N = 157$

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

complete responses. Ages ranged from 18 to 24 years ($M = 19.5$, $SD = 1.15$); 82 (52%) of these students identified as men, 63 (40%) as women, and 12 (8%) declined to state a gender; 33 (21%) identified as Asian, 9 (6%) as Black or African American, 6 (4%) as Hispanic or Latinx, 85 (54%) as White, 8 (5%) as mixed, 3 (2%) as other, and 13 (8%) declined to state.

The survey included four common, psychometrically validated measures. (Table 1 presents the full text of all questions from these measures.) One was the CRT-2 (Thomson and Oppenheimer 2016), a four-item variation on the original CRT that was designed to be less mathematical and less familiar to participants who may have previously seen the original questions. The other measures were self-reports. These included the General Intellectual Humility Scale (Leary et al. 2017), Open-Minded Cognition Scale (Price et al. 2015), and Situated Wise Reasoning Scale (SWIS; Brienza et al. 2018).

The SWIS differs from the other self-reports in that it does not ask respondents about what they are like *in general*. Instead, it asks respondents to think about, and mentally relive, the last time they had an interpersonal conflict and then answer twenty-one questions about that specific occasion. 'Situated' measures like this are thought to be less subject to certain kinds of self-report biases (Kahneman et al. 2004), such as 'social desirability bias' (where people tend to answer in ways that they think will make them look good). Each SWIS item is introduced with the following prompt: 'While this situation was unfolding, I did the following . . .'. Respondents then respond to each statement as a description of their behavior on that specific occasion. The twenty-one items are divided into five subscales: *Other's Perspective* is about the degree to which respondents attempted to understand the point of view of their interlocutors; *Multiple Outcomes* is about the degree to which they considered the various ways in which the situation could play out; *Intellectual Humility* is about the degree to which they recognized that they might not have all the relevant information and their interlocutors might have known things that they did not; *Search for Compromise* is about the degree to which they searched for a mutually beneficial resolution; and *Outside Vantage Point* is about the degree to which they tried to understand how an impartial third party might interpret the situation.

Because these are widely used measures, we were able to find a large amount of publicly available data from studies that have used them (data sources are documented online: https://osf.io/mbvpr/). Hence, in addition to the student data we collected, we have 9,014 observations from adults across the United States. Some of these people may have studied philosophy themselves, but this is likely only a small minority. Figure 2 shows the average score for each measure, grouped by educational attainment, with Week 1 Philosophy 101 students on the right-hand side.

We found no statistically significant differences on the General Intellectual Humility Scale or Open-Minded Cognition Scale. That is, there appears to be no association between educational attainment and either of these intellectual virtues. Crucially, the Week 1 Philosophy 101 students did not differ from the other groups. Hence, differences in these traits between philosophers and non-philosophers found in Figure 1 above may result from studying philosophy as

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

Table 1. Measures and items from Philosophy 101 study

Measure	Items
OMCS	1. I have no patience for arguments I disagree with. 2. I often 'tune out' messages I disagree with. 3. I believe it is a waste of time to pay attention to certain ideas. 4. I try to reserve judgment until I have a chance to hear arguments from both sides of an issue. 5. I am open to considering other viewpoints. 6. When thinking about an issue, I consider as many different opinions as possible.
GIHS	1. I question my own opinions, positions, and viewpoints because they could be wrong. 2. I reconsider my opinions when presented with new evidence. 3. I recognize the value in opinions that are different from my own. 4. I accept that my beliefs and attitudes may be wrong. 5. In the face of conflicting evidence, I am open to changing my opinions. 6. I like finding out new information that differs from what I already think is true.
CRT-2	1. If you're running a race and you pass the person in second place, what place are you in? (answer: second) 2. A farmer had 15 sheep and all but 8 died. How many are left? (answer: 8) 3. Emily's father has three daughters. The first two are named April and May. What is the third daughter's name? (answer: Emily) 4. How many cubic feet of dirt are there in a hole that is 3' deep x 3' wide x 3' long? (answer: 0)
SWIS	1. Put myself in the other person's shoes 2. Tried to communicate with the other person what we might have in common 3. Made an effort to take the other person's perspective 4. Took time to get the other person's opinions on the matter before coming to a conclusion 5. Looked for different solutions as the situation evolved 6. Considered alternative solutions as the situation evolved 7. Believed the situation could lead to a number of different outcomes 8. Thought the situation could unfold in many different ways 9. Double-checked whether my opinion on the situation might be incorrect 10. Double-checked whether the other person's opinions might be correct 11. Looked for any extraordinary circumstances before forming my opinion 12. Behaved as if there may be some information to which I did not have access 13. Tried my best to find a way to accommodate both of us 14. Though it may not have been possible, I searched for a solution that could result in both of us being satisfied

(*Continued*)

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

Table 1. Continued.

Measure	Items
	15. Considered first whether a compromise was possible in resolving the situation
	16. Viewed it as very important that we resolve the situation
	17. Tried to anticipate how the conflict might be resolved
	18. Wondered what I would think if I was somebody else watching the situation
	19. Tried to see the conflict from the point of view of an uninvolved person
	20. Asked myself what other people might think or feel if they were watching the conflict
	21. Thought about whether an outside person might have a different opinion from mine about the situation

Note. GIHS indicates the General Intellectual Humility Scale. OMCS indicates the Open-Minded Cognition Scale. CRT-2 indicates the Cognitive Reflection Test – 2. SWIS indicates the Situated Wise Reasoning Scale (items 1-4 are Others' Perspectives; 5-8 are Multiple Outcomes; 9-12 are Intellectual Humility; 13-17 are Search for Compromise; 18-21 are Outside Vantage Point).

opposed to preexisting differences between those interested in philosophy and those not interested in it.

By contrast, we observed striking differences for reflectiveness. More education tends to come with more correct answers and fewer lured answers on the reflection test. Nevertheless, during their first week, Philosophy 101 students gave significantly more correct answers than all other groups—including those with graduate and professional degrees—and fewer lured answers than people with some college education or less. Although the mean number of lured answers was lower for the Week 1 Philosophy 101 students than for all other groups, the difference was not statistically significant when comparing them to people with bachelor's degrees or graduate/professional degrees. In other words, the Week 1 Philosophy 101 students give more correct answers than all other groups but only give fewer lured answers than people with comparable levels of education. It is not entirely clear what to make of this difference. One possibility is that ancillary cognitive abilities (e.g., intelligence or numeracy) play more of a role in determining the number of correct answers than they do in determining the number of lured answers. If so, this would imply that some part of the difference we observe between Philosophy 101 students and US adults is attributable to differences in such abilities.

It is conceivable that these differences in reflectiveness are due to a priming effect —that is, being in a philosophy classroom cues students to be more reflective than they otherwise would be. Although we cannot rule out this possibility, it does not strike us as especially plausible given that these students took the reflection test during the very first week of the class, before they had spent much time doing philosophy in that room. Moreover, a recent meta-analysis found that contextual priming effects tend to be very small (Dai et al. 2023). Hence, the *large* difference we observed is likely not explained by contextual priming.

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

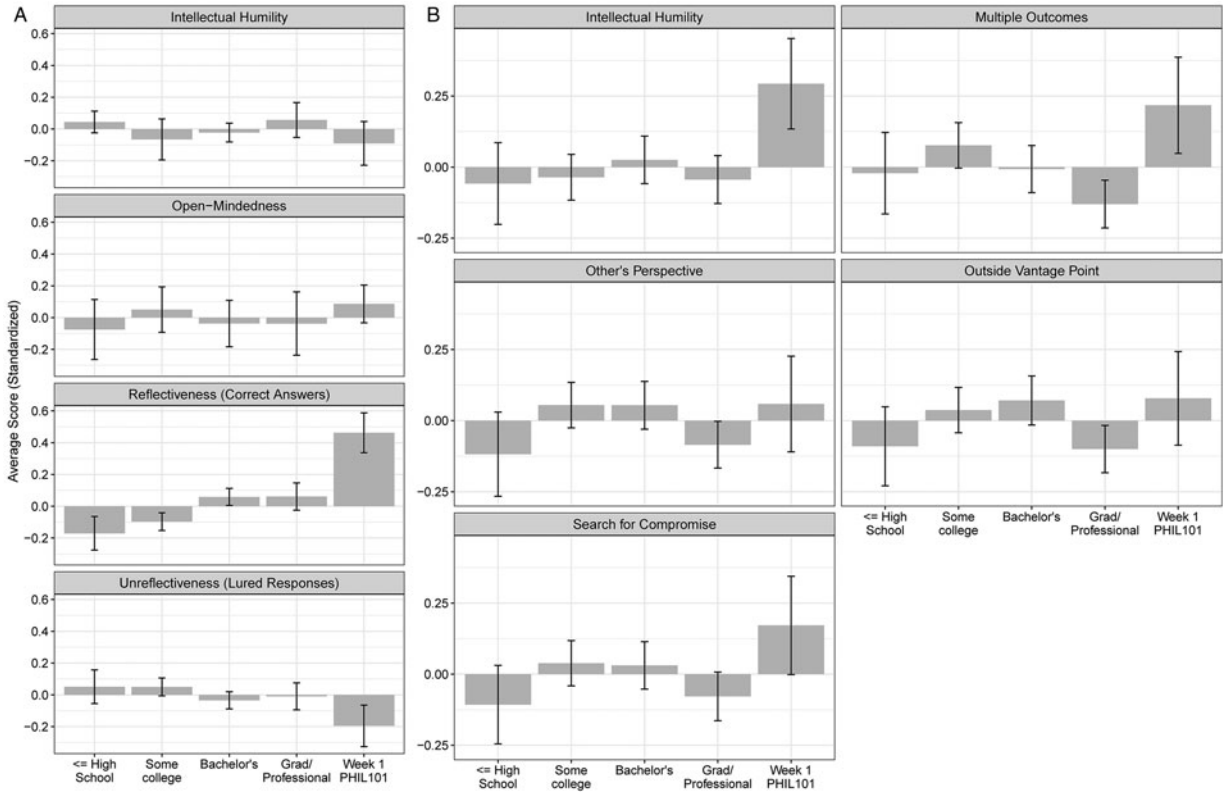

Figure 2. Intellectual virtues across levels of educational attainment, and comparison with Philosophy 101 students. Panel A presents the results for the General Intellectual Humility Scale, Open-Minded Cognition Scale, and the Cognitive Reflection Test – 2. Panel B presents the results for the subscales of the Situated Wise Reasoning Scale. Points and vertical lines indicate means and 95 percent confidence intervals. The *y*-axis has been standardized for ease of comparisons across measures.

Considering the SWIS, we found somewhat different results for each of the five subscales. For the Others' Perspectives and Outside Vantage Point subscales, the Week 1 Philosophy 101 students did not differ significantly from any of the other groups, indicating that, compared with US adults in general, philosophy students are no more likely to try to see things from another's point of view or to imagine what an impartial third party might think about their situation. For Multiple Outcomes and Search for Compromise, the Week 1 Philosophy 101 students scored higher than some of the other groups, but not all of them. Specifically, for Multiple Outcomes, there appear to be declines from the 'some college' group to the group with graduate and professional degrees. The Week 1 Philosophy 101 students scored significantly higher than the people with graduate and professional degrees, and marginally significantly higher than those with bachelor's and high school diplomas or less, but did not score differently from those with 'some college'. This suggests that philosophy students may be somewhat more likely than others to try to consider many different possibilities for how an interpersonal situation could play out. But they did not differ from the most comparable group, namely, those with some college education. Finally, we found clear evidence of a selection effect when considering the Intellectual Humility subscale. The Week 1 Philosophy 101 students scored significantly higher than all other groups. These differences would be considered 'small', but it is striking how the Week 1 Philosophy 101 group stands out from all the rest, which do not differ from each other.

This last finding appears to be inconsistent with our findings on the General Intellectual Humility Scale, which revealed no differences between Week 1 Philosophy 101 students and US adults. Because the SWIS is less affected by social desirability bias (Brienza et al. 2018), one interpretation of this inconsistency is that although philosophy students do not *aspire* to intellectual humility any more than others (hence there is no difference on the more abstract measure), they do *display* it more often (hence there is a difference on the situated measure). Another interpretation could be that these two measures, though they both purport to assess the same thing, are actually tapping into slightly different psychological phenomena. If so, then it will be important for future studies to attend to the questions of which phenomenon is of most interest and how that phenomenon is most effectively measured.

Overall, these results provide preliminary, suggestive evidence that some, but not all, of the observed differences between philosophers and non-philosophers may result purely from selection effects. For reflectiveness specifically, we suspect that the observed difference between philosophers and non-philosophers is probably due primarily or entirely to selection effects. However, for open-mindedness, our findings suggest that the previously observed difference between philosophers and non-philosophers might *not* be due solely to selection effects. Similarly, for intellectual humility, we found evidence of a selection effect using a 'situated' measure, but not with a more general measure. Hence, it is possible that studying philosophy has some effect on intellectual virtues like these. Of course, this is merely evidence of the *possibility* of a treatment effect, not direct evidence of such an effect.

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

Testing for Change over Time

The empirical results that we have presented thus far have all involved between-person comparisons. That is, we have compared people who have studied philosophy with people who have not, and we have compared students at the start of their first philosophy class with US adults at various levels of education. However, a more compelling sort of evidence would monitor people over time to observe the effects of philosophical education as they unfold.

In a recent paper, Kerem Oktar and colleagues (2023) attempted to do just this. They had students in an introductory ethics course ($n = 137$) and a control group of psychology students ($n = 62$) report their views on twelve controversial ethical questions at the start and end of a semester. This sort of study design is sometimes referred to as 'quasi-experimental', in that it is controlled but not randomized. In this sort of study, baseline differences between groups are indicative of selection effects. However, if one group shows changes over time that the other group does not, then this growth might be explained by the differences in coursework. This kind of study design does not rule out selection effects entirely because it is possible that students' *trajectories* were shaped by preexisting influences. But it does rule out the most obvious kind of selection effect—namely, that the groups already differ at the start.

In this study, the mean age for the sample was 19.73 ($SD = 1.4$). Of these, 98 (49%) participants identified as male, 94 (47%) as female, and 7 (4%) declined to state. The twelve controversial ethical questions included ones such as whether it is morally permissible to eat meat, restrict immigration, or have an abortion. Besides indicating their views on each question, participants also answered a pair of questions about how they arrived at their views: 'Is your judgment based on intuition/emotion?' and 'Is your judgment based on deliberation/analysis?' (Response scales ranged from 'Not at all' to 'Entirely'.)

The results indicated that the philosophy students significantly changed their views on these ethical questions and more so than the psychology students. Additionally, the researchers found that the philosophy students, but not the psychology students, showed a reduced tendency to base their ethical views on intuition and emotion versus deliberation and analysis. Among philosophy students specifically, the degree to which students reduced their reliance on intuition and emotion predicted the degree to which they changed their ethical beliefs.

When it comes to moral beliefs—unlike, for example, beliefs about how chemicals interact under various conditions—people tend to dig in their heels and dogmatically maintain preexisting views (Haidt 2001; Heinzelmann, Höltgen, and Tran 2021; Skitka 2010). Yet, these results suggest that philosophy courses can influence the way people think about controversial ethical questions. The results also suggest that the mechanism behind these changes is a reduced tendency to trust one's gut instead of carefully considering reasons and arguments. The evidence discussed in the previous sections suggests that philosophers are distinctive in this regard. Although we found that introductory philosophy students are already especially reflective and less intuitive in their thinking than others, this study provides some initial evidence that philosophical training might amplify this tendency.

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

In addition to students' views on specific ethical questions, Oktar and colleagues (2023) assessed several intellectual traits. The measures included five self-report scales: the Actively Open-Minded Thinking about Evidence Scale (AOT-E; Pennycook et al. 2020), which asks participants for their agreement with eight statements such as, 'Beliefs should always be revised in response to new information or evidence'. The researchers also included abridged versions of the Moralized Rationality Scale (example item: 'It is morally wrong to trust your intuitions without rationally examining them'), the Importance of Rationality Scale (example item: 'It is important to me personally to examine traditionally held beliefs using logic and evidence'; Ståhl, Zaal, and Skitka 2016), and the Unified Scale to Assess Individual Differences in Intuition and Deliberation (example items: 'When I make a decision, it is more important for me to feel the decision is right than to have a rational reason for it' and 'I study every problem until I understand the underlying logic'; Pachur and Spaar 2015). Additional details regarding these measures can be found in Oktar et al. (2023).

The researchers did not report the results for the trait measures. However, because the data are publicly available (https://osf.io/y5tdu/), we analyzed them to investigate whether the students showed changes in these traits and whether such changes differed between the philosophy and psychology students. Figure 3 shows the average scores for both groups of students at the start and at the end of the semester.

For Moralization of Rationality and Preference for Deliberation, the philosophy students increased over the course of the semester, while the psychology students did not. For Preference for Intuition, the philosophy students increased, while the psychology students decreased. Hence, for these three outcomes, we do actually have some evidence of treatment effects. However, considering Open-Mindedness about Evidence and Importance of Rationality, there were no statistically significant differences between groups and no changes over time for either group.

Overall, the results of this study offer some preliminary evidence that philosophy courses can change the way students think. Although we found no evidence of an effect on open-mindedness, we did find evidence that philosophy courses might increase the tendency to deliberate, which is plausibly a desirable change—especially when paired with a balanced reliance on intuition and emotion. A greater tendency to be *moralistic* about rationality is less obviously valuable. Indeed, this question has long been debated (Clifford 1879; James 1896). But perhaps most important, this study shows what can be done when philosophers and psychologists collaborate. Future studies could follow a similar model, assessing other outcomes and examining a wider range of philosophy courses.

Implications and Conclusion

We have covered a lot of ground in this article. To review, we have seen very clear evidence that people who have studied philosophy tend to have remarkably strong academic abilities in general and strong verbal and logical reasoning skills in particular. They also tend to be highly reflective and open-minded compared with people who have not studied philosophy. We have also seen clear evidence that at least some of these differences result from selection effects—that is, preexisting

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

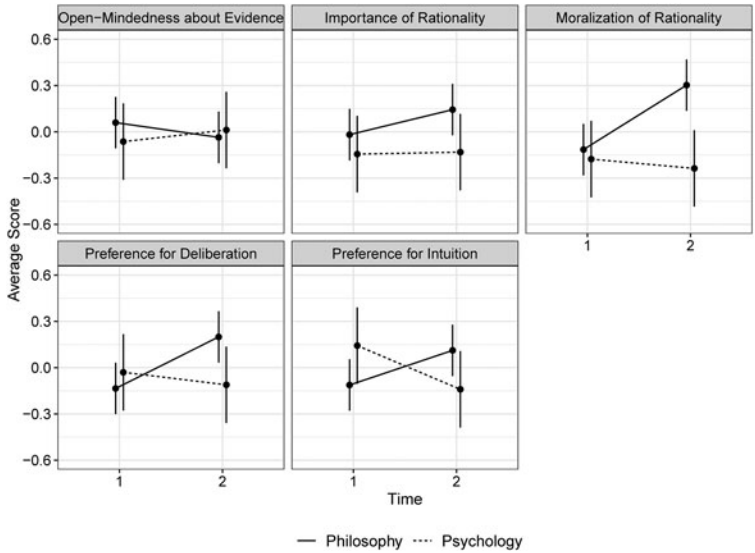

Figure 3. Longitudinal changes in intellectual traits for students taking introductory courses in philosophy and psychology. Points and vertical lines indicate means and 95 percent confidence intervals. The *y*-axis has been standardized for ease of comparisons across measures.

differences between those who choose to study philosophy and those who do not. For example, students at the very beginning of their philosophical education (during the first week of Philosophy 101) are already far more reflective than most people. Of course, the presence of strong selection effects does not rule out the presence of treatment effects. After all, one could think of a philosophical education as amplifying a strength, cultivating potentials that are revealed by preexisting interests and inclinations.

When we turn to treatment effects, however, the evidence is far less clear. Some studies find that philosophy education improves reading, writing, and mathematical abilities in young children, whereas others find no such effects. Additionally, although college students tend to improve in their critical thinking skills over time, there is no clear evidence that philosophy courses are especially effective at teaching critical thinking. That said, a specific technique called 'argument mapping', which is sometimes taught in philosophy courses, does hold promise for teaching critical thinking. Finally, although there is only a very limited amount of longitudinal data, we have seen that, relative to their peers, philosophy students show greater changes in their thinking about the specific topics covered in philosophy courses and in certain general attitudes (e.g., the degree to which they moralize rationality). However, we do not have clear evidence that students in these classes become, for instance, more open-minded or intellectually humble. Naturally, this lack of evidence does not demonstrate the absence of any such effects. Rather, it shows that more data are needed.

In some ways, our findings fit a larger pattern in education research, stretching back many decades, which is that learning often does not 'transfer' (Barnett and

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

Ceci 2002; Cormier and Hagman 1987). That is, people often do not generalize the ideas, techniques, or skills that they learn in one area or context and apply them in other areas or contexts. Teaching people to solve math problems makes them better at solving math problems, and teaching them historical facts makes them better at recalling historical facts. But one should not assume that when people learn math or history, they will improve in their more general analytic reasoning skills or abilities to remember. The learning tends not to 'transfer' in that way. Similarly, teaching students to wrestle with philosophical problems could simply make them better at wrestling with philosophical problems and not improve their domain-general abilities in logical reasoning, critical thinking, and so on.

Philosophy instructors would do well to consider the insights coming out of research on how to facilitate the transfer of learning (Fiorella and Mayer 2016; van Peppen et al. 2022)—for example, by prompting students to interpret and actively apply course content to questions from other domains or from their own lives. Nonetheless, there are some possible reasons for thinking that philosophical learning might be more transferable than some other kinds of learning.

Whereas some disciplines are characterized by a particular body of knowledge, philosophy is characterized by a distinctive kind of *activity*. Philosophy is often understood as a particular way of thinking about abstract questions, a style of thinking that is characterized by critical scrutiny but also by openness to unusual ideas (Edmonds and Warburton 2010; Priest 2006). This style of thinking could plausibly be applied to a wide range of topics beyond those that would normally be considered 'philosophical questions'. Indeed, for nearly any *X* there can be a philosophy of *X*. For example, there can be the philosophy of food, philosophy of dating, philosophy of most anything that one might do in ordinary life. Hence, the skills acquired from philosophical education may be more transferable to ordinary life than the skills acquired from mathematical education, for example. This distinctly philosophical way of thinking may, to some degree, be captured by measures like the CRT and the Actively Open-Minded Thinking Scale. But such measures were certainly not *designed* to assess this philosophical style of thinking. Perhaps, in future work, researchers might create measures designed specifically for this purpose though this would undoubtedly be a very difficult task and one plagued by a specter of parochialism. Although philosophers may value the distinctly philosophical way of thinking, others may not.

We conclude that we do not have strong evidence one way or the other about whether studying philosophy makes people better thinkers. The primary takeaway from our review of empirical evidence, therefore, is that there simply is not enough of it. If we want better evidence, it seems likely that we will need to gather it ourselves. Of course, many philosophers lack the requisite training in empirical and statistical methods. In such cases, we urge them to connect with collaborators in other disciplines (such as psychology) to design rigorous studies testing for effects of philosophical training. Naturally, the ideal would be a randomized controlled trial (RCT). Random assignment to groups rules out selection effects, and this means that any differences subsequently observed between groups can be attributed to the treatment. For a variety of reasons, however, RCTs are often not feasible in educational contexts. Hence, quasi-experiments like the study by Oktar

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

and colleagues are a valuable alternative. A philosopher might, for example, administer pre- and post-tests, at the start and end of a semester or academic year and compare their own students with students not enrolled in any philosophy classes.

Future research may find stronger and clearer evidence that studying philosophy makes people better thinkers. But if we find that a philosophical education has no such effects, what implications would this have? First, we do not think that the value of philosophy can be understood in purely instrumental terms, and such findings would in no way undermine the claim that philosophy is *intrinsically* valuable. Second, we would consider such a finding cause for greater reflection on our teaching. Courses might be designed with greater emphasis on intellectual virtue (e.g., Battaly 2006; Lamb et al. 2022). And empirical research on what does and does not influence intellectual virtues would be invaluable for informing such efforts. In any case, we will not know how philosophy affects its students until we go out and look.

MICHAEL PRINZING
YALE UNIVERSITY
michael@prinzing.net

MICHAEL VAZQUEZ
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
michael.vazquez@unc.edu

References

American Philosophical Association. (2014) 'Philosophy Student Performance on the Graduate Record Examinations'. American Philosophical Association (blog). https://cdn.ymaws.com/www.apaonline.org/resource/resmgr/Data_on_Profession/2014_Philosophy_Performance_.pdf.

American Philosophical Association. (2017) 'Philosophy: A Brief Guide for Undergraduates.' American Philosophical Association (blog). https://www.apaonline.org/page/undergraduates.

American Philosophical Association. (2019) 'Philosophy Student Performance on the Law School Admissions Test'. American Philosophical Association (blog). https://cdn.ymaws.com/www.apaonline.org/resource/resmgr/Data_on_Profession/Philosophy_performance_on_LS.pdf.

Annis, Linda F., and David B. Annis. (1979) 'The Impact of Philosophy on Students' Critical Thinking Ability'. *Contemporary Educational Psychology,* 4, 219–26. https://doi.org/10.1016/0361-476X(79)90042-0.

Barnett, Susan M., and Stephen J. Ceci. (2002) 'When and Where Do We Apply What We Learn?: A Taxonomy for Far Transfer'. *Psychological Bulletin,* 128, 612–37. https://doi.org/10.1037/0033-2909.128.4.612.

Baron, Jonathan. (2019) 'Actively Open-Minded Thinking in Politics'. *Cognition, The Cognitive Science of Political Thought,* 188, 8–18. https://doi.org/10.1016/j.cognition.2018.10.004.

Battaly, Heather. (2006) 'Teaching Intellectual Virtues: Applying Virtue Epistemology in the Classroom'. *Teaching Philosophy,* 29, 191–222. https://doi.org/10.5840/teachphil200629333.

Brienza, Justin P., Franki Y. H. Kung, Henri C. Santos, D. Ramona Bobocel, and Igor Grossmann. (2018) 'Wisdom, Bias, and Balance: Toward a Process-Sensitive Measurement of Wisdom-Related Cognition'. *Journal of Personality and Social Psychology,* 115, 1093–1126.

Burke, Brian L., Sharon R. Sears, Sue Kraus, and Sarah Roberts-Cady. (2014) 'Critical Analysis: A Comparison of Critical Thinking Changes in Psychology and Philosophy Classes'. *Teaching of Psychology,* 41, 28–36. https://doi.org/10.1177/0098628313514175.

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

Byrd, Nick. (2022) 'Great Minds Do Not Think Alike: Philosophers' Views Predicted by Reflection, Education, Personality, and Other Demographic Differences'. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-022-00628-y.
Clifford, William K. (1879) 'The Ethics of Belief'. In *Lectures and Essays, II*, 177–211. MacMillan.
Cormier, Stephen M., and Joseph D. Hagman. (1987) *Transfer of Learning: Contemporary Research and Applications*. Academic Press.
Cullen, Simon, Judith Fan, Eva van der Brugge, and Adam Elga. (2018) 'Improving Analytical Reasoning and Argument Understanding: A Quasi-Experimental Field Study of Argument Visualization'. *Npj Science of Learning*, 3, 1–6. https://doi.org/10.1038/s41539-018-0038-5.
Dai, Wenhao, Tianshu Yang, Benjamin X. White, Ryan Palmer, Emily K. Sanders, Jack A. McDonald, Melody Leung, and Dolores Albarracín. (2023) 'Priming Behavior: A Meta-Analysis of the Effects of Behavioral and Nonbehavioral Primes on Overt Behavioral Outcomes'. *Psychological Bulletin*, 149, 67–98. https://doi.org/10.1037/bul0000374.
Dutilh Novaes, Catarina. (2020) *The Dialogical Roots of Deduction: Historical, Cognitive, and Philosophical Perspectives on Reasoning*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108800792.
Dwyer, Christopher P., Michael J. Hogan, and Ian Stewart. (2015) 'The Effects of Argument Mapping-Infused Critical Thinking Instruction on Reflective Judgement Performance'. *Thinking Skills and Creativity*, 16, 11–26. https://doi.org/10.1016/j.tsc.2014.12.002.
Edmonds, David, and Nigel Warburton. (2010) *Philosophy Bites*. Oxford: Oxford University Press.
FairTest. (2022) 'Record 1,835+ Schools Are Test Optional or Test Free for Admissions'. FairTest. https://fairtest.org/record-1835-schools-are-test-optional-or-test-free-for-admissions/.
Fiorella, Logan, and Richard E. Mayer. (2016) 'Eight Ways to Promote Generative Learning'. *Educational Psychology Review*, 28, 717–41. https://doi.org/10.1007/s10648-015-9348-9.
Frederick, Shane. (2005) 'Cognitive Reflection and Decision Making'. *Journal of Economic Perspectives*, 19, 25–42. https://doi.org/10.1257/089533005775196732.
García-Moriyón, Felix, Irene Rebollo, and Roberto Colom. (2005) 'Evaluating Philosophy for Children: A Meta-Analysis'. *Thinking: The Journal of Philosophy for Children*, 17, 14–22. https://doi.org/10.5840/thinking20051743.
Gorard, Stephen, Nadia Siddiqui, and Beng Huat See. (2015) 'Philosophy for Children Evaluation Report and Executive Summary'. The Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/documents/projects/EEF_Project_Report_PhilosophyForChildren.pdf?v=1681896334.
Haidt, Jonathan. (2001) 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment'. *Psychological Review*, 108, 814–34.
Harrell, Maralee. (2004) 'The Improvement of Critical Thinking Skills in What Philosophy'. CMU-PHIL-158. Pittsburgh, Pennsylvania: Carnegie Mellon University.
Heinzelmann, Nora, Benedikt T. A. Höltgen, and Viet Tran. (2021) 'Moral Discourse Boosts Confidence in Moral Judgments'. *Philosophical Psychology*, 34, 1192–1216. https://doi.org/10.1080/09515089.2021.1959026.
Hoekema, David A. (1986) 'Why Major in Philosophy?' *Proceedings and Addresses of the American Philosophical Association*, 59, 601–6. https://doi.org/10.2307/3131573.
Huber, Christopher R., and Nathan R. Kuncel. (2016) 'Does College Teach Critical Thinking? A Meta-Analysis'. *Review of Educational Research*, 86, 431–68. https://doi.org/10.3102/0034654315605917.
Inglis, Matthew. (2015) 'Does Teaching Philosophy to Children Improve Their Reading, Writing and Mathematics Achievement?' *Scenes From The Battleground* (blog). July 14, 2015. https://teachingbattleground.wordpress.com/2015/07/14/does-teaching-philosophy-to-children-improve-their-reading-writing-and-mathematics-achievement-guest-post-by-mjinglis/.
James, William. (1896) 'The Will to Believe'. *New World*, 5, 327–47.
Kahneman, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone. (2004) 'A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method'. *Science*, 306, 1776–80. https://doi.org/10.1126/science.1103572.

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

Kilov, Daniel, and Caroline Hendy. (2022) 'Pundits and Possibilities: Philosophers Are Not Modal Experts'. *Australasian Journal of Philosophy,* 75, 824–43. https://doi.org/10.1080/00048402.2022.2058034.

Lamb, Michael, Elise M. Dykhuis, Sara E. Mendonça, and Eranda Jayawickreme. (2022) 'Commencing Character: A Case Study of Character Development in College'. *Journal of Moral Education,* 51, 238–60. https://doi.org/10.1080/03057240.2021.1953451.

Leary, Mark R., Kate J. Diebels, Erin K. Davisson, Katrina P. Jongman-Sereno, Jennifer C. Isherwood, Kaitlin T. Raimi, Samantha A. Deffler, and Rick H. Hoyle. (2017) 'Cognitive and Interpersonal Features of Intellectual Humility'. *Personality and Social Psychology Bulletin,* 43, 793–813. https://doi.org/10.1177/0146167217697695.

Livengood, Jonathan, Justin Sytsma, Adam Feltz, Richard Scheines, and Edouard Machery. (2010) 'Philosophical Temperament'. *Philosophical Psychology,* 23, 313–30. https://doi.org/10.1080/09515089.2010.490941.

Lord, Pippa, Afrah Dirie, Kelly Kettlewell, and Ben Styles. (2021) 'Evaluation of Philosophy for Children: An Effectiveness Trial'. Evaluation Report. Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/documents/projects/Philosophy_for_Children_report_-_final_-_pdf.pdf?v=1681843322.

Metcalf, Thomas. (2021) 'Philosophy Majors and High Standardized Test Scores: Not Just Correlation'. *Daily Nous* (blog). July 14. https://dailynous.com/2021/07/14/philosophy-majors-high-standardized-test-scores/.

Montmarquet, James. (1992) 'Epistemic Virtue and Doxastic Responsibility'. *American Philosophical Quarterly,* 29, 331–41.

Moshman, David, and Molly Geil. (1998) 'Collaborative Reasoning: Evidence for Collective Rationality'. *Thinking & Reasoning,* 4, 231–48. https://doi.org/10.1080/135467898394148.

Oktar, Kerem, Adam Lerner, Maya Malaviya, and Tania Lombrozo. (2023) 'Philosophy Instruction Changes Views on Moral Controversies by Decreasing Reliance on Intuition'. *Cognition,* 236, 105434.https://doi.org/10.1016/j.cognition.2023.105434.

Ortiz, Claudia María Álvarez. (2007) 'Does Philosophy Improve Critical Thinking Skills?' Master's Thesis, Melbourne, Australia: University of Melbourne. https://www.reasoninglab.com/wp-content/uploads/2015/10/Alvarez-Final_Version.pdf.

Pachur, Thorsten, and Melanie Spaar. (2015) 'Domain-Specific Preferences for Intuition and Deliberation in Decision Making'. *Journal of Applied Research in Memory and Cognition,* 4, 303–11. https://doi.org/10.1016/j.jarmac.2015.07.006.

Pennycook, Gordon, James Allan Cheyne, Derek J. Koehler, and Jonathan A. Fugelsang. (2020) 'On the Belief That Beliefs Should Change According to Evidence: Implications for Conspiratorial, Moral, Paranormal, Political, Religious, and Science Beliefs'. *Judgment and Decision Making,* 15, 476–98. https://doi.org/10.1017/S1930297500007439.

Peppen, Lara M. van, Tamara van Gog, Peter P. J. L. Verkoeijen, and Patricia A. Alexander. (2022) 'Identifying Obstacles to Transfer of Critical Thinking Skills'. *Journal of Cognitive Psychology,* 34, 261–88. https://doi.org/10.1080/20445911.2021.1990302.

Price, Erika, Victor Ottati, Chase Wilson, and Soyeon Kim. (2015) 'Open-Minded Cognition'. *Personality and Social Psychology Bulletin,* 41, 1488–1504. https://doi.org/10.1177/0146167215600528.

Priest, Graham. (2006) 'What Is Philosophy?' *Philosophy*, 81, 189. https://doi.org/10.1017/S0031819106316026.

Ross, Gregory A., and George Semb. (1981) 'Philosophy Can Teach Critical Thinking Skills'. *Teaching Philosophy,* 4, 111–22. https://doi.org/10.5840/teachphil19814233.

Schwitzgebel, Eric, and Fiery Cushman. (2012) 'Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers'. *Mind & Language,* 27, 135–53. https://doi.org/10.1111/j.1468-0017.2012.01438.x.

Skitka, Linda J. (2010) 'The Psychology of Moral Conviction'. *Social and Personality Psychology Compass,* 4, 267–81. https://doi.org/10.1111/j.1751-9004.2010.00254.x.

Ståhl, Tomas, Maarten P. Zaal, and Linda J. Skitka. (2016) 'Moralized Rationality: Relying on Logic and Evidence in the Formation and Evaluation of Belief Can Be Seen as a Moral Issue'. *PLOS ONE,* 11, e0166332. https://doi.org/10.1371/journal.pone.0166332.

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press

Thomson, Keela S, and Daniel M Oppenheimer. (2016) 'Investigating an Alternate Form of the Cognitive Reflection Test'. *Judgment and Decision Making,* 11, 15.

Thornton, Jim. (2015) 'Teaching Philosophy in Primary Schools'. *Ripe-Tomato.Org* (blog). July 14, 2015. https://ripe-tomato.org/2015/07/14/teaching-philosophy-in-primary-schools/.

Trickey, S., and K. J. Topping. (2004) ''Philosophy for Children': A Systematic Review'. *Research Papers in Education,* 19, 365–80. https://doi.org/10.1080/0267152042000248016.

Whitcomb, Dennis, Heather Battaly, Jason Baehr, and Daniel Howard-Snyder. (2017) 'Intellectual Humility: Owning Our Limitations'. *Philosophy and Phenomenological Research,* 94, 509–39. https://doi.org/10.1111/phpr.12228.

Yaden, David B., and Derek E. Anderson. (2021) 'The Psychology of Philosophy: Associating Philosophical Views with Psychological Traits in Professional Philosophers'. *Philosophical Psychology,* 34, 721–55. https://doi.org/10.1080/09515089.2021.1915972.

https://doi.org/10.1017/apa.2023.30 Published online by Cambridge University Press