

Studying Philosophy Does Make People Better Thinkers

ABSTRACT: *Many philosophers think that doing philosophy cultivates valuable intellectual abilities and dispositions. Indeed this is a premise in a venerable argument for philosophy's value. Unfortunately, empirical support for this premise has heretofore been lacking. We provide evidence that philosophical study has such effects. Using a large dataset (including records from over half a million undergraduates at hundreds of institutions across the United States), we investigate philosophy students' performance on verbal and logical reasoning tests, as well as measures of valuable intellectual dispositions. Results indicate that students with stronger verbal abilities, and who are more curious, open-minded, and intellectually rigorous, are more likely to study philosophy. Nonetheless, after accounting for such baseline differences, philosophy majors outperform all other majors on tests of verbal and logical reasoning and on a measure of valuable habits of mind. This offers the strongest evidence to date that studying philosophy does indeed make people better thinkers.*

KEYWORDS: causal inference, empirically-engaged philosophy, intellectual virtue, value of philosophy

1. Introduction

Philosophy has a reputation for making people better thinkers. Its students learn to dissect arguments and untangle complex problems with clarity and precision. They are taught to question assumptions and consider a variety of possible answers to any particular question, even answers that might initially seem strange or unconventional. The methods and practice of philosophy, in other words, seem especially well-suited to sharpening one's intellectual faculties and inculcating good habits of mind.

This idea has deep historical roots. Plato and Aristotle, for instance, regarded a philosophical education as essential for cultivating the rational faculties that underpin both individual and societal flourishing (*Republic* V, 473c-d; *Nicomachean Ethics* 1178b-c). In more recent history, Bertrand Russell (1912) argued that much of the value of philosophy lies in its ability to “enlarge our conception of what is possible, enrich our intellectual imagination and diminish the dogmatic assurance which closes the mind” (p. 131). And contemporary philosophers continue this tradition. For example, Jennifer Morton (2019) has claimed that “philosophy teaches you to think and write logically and clearly” and is an “antidote to the uncritical acceptance of the world and ourselves as we are” (pp. 103–104).



Such claims constitute a crucial premise in arguments for the value of philosophical study. There are many variants of this argument, depending on the precise details of how one thinks that philosophical study makes people better thinkers. But the basic structure of the argument goes something like this:

- A. If an activity cultivates valuable intellectual abilities and dispositions, then that activity is valuable.
- B. Philosophical study cultivates valuable intellectual abilities and dispositions.
- C. Therefore, philosophical study is valuable.

We find premise A extremely plausible, but we will not offer a defense of it here. Instead, our intent is to scrutinize premise B. Naturally, this premise is an empirical claim. Whether studying philosophy causes improvements in, e.g., logical reasoning or facility with language, or whether it fosters dispositions like open-mindedness or curiosity, is not a matter of *a priori* or conceptual truth. Such claims are not to be verified or rebutted through armchair reflection, but rather through careful observation. Thus, if we are to put forth such claims, we should seek to support them with rigorous analyses of empirical data.

In a recent article (Prinzing and Vazquez 2024), we reviewed extant empirical findings relevant to premise B and presented some new findings of our own. Unfortunately, however, our conclusion was that “we do not have strong evidence one way or the other about whether studying philosophy makes people better thinkers” (p. 18). The aim of this article, therefore, is to provide evidence that would allow for a more definite conclusion. We accessed a very large dataset—including records from over half a million undergraduate students who attended over 800 colleges and universities across the United States—and applied causal inference techniques to test whether studying philosophy might cultivate valuable intellectual abilities and dispositions.

The remainder of the article proceeds as follows. In Section 2 we explain why extant empirical evidence does not support causal conclusions and discuss the kind of evidence that could do so. Then, in Section 3, we consider the merits of different approaches to quantifying intellectual abilities and dispositions. Sections 4 and 5 are the critical ones, in which we describe the data and measures and present the results of our analyses. Briefly, our results indicate that students with higher levels of valuable intellectual abilities and dispositions are more likely to study philosophy. However, even after accounting for these baseline differences, philosophy majors still outperform students in other majors. In fact, philosophy majors top the charts on three out of the five outcomes that we will examine. Finally, in Section 6, we discuss the implications of these findings. We argue that they constitute probably the strongest available evidence that philosophy fosters valuable intellectual abilities and dispositions, while recognizing that more and different evidence is needed to determine whether it fosters genuine intellectual *virtue*. That is, we find clear support for (at least certain popular versions of) premise B and, accordingly, conclude that the above argument for the value of philosophical study is sound. We wrap up by considering the implications, connecting these ideas with debates over the value and

proper aims of education (including what educators should strive to cultivate in students) and with arguments from political philosophers about how education can sustain democratic norms and institutions.

2. Why Extant Evidence Does Not Support Causal Conclusions

Since the 1980s (Hoekema 1986), philosophers have observed that undergraduates who major in philosophy tend to score remarkably well on standardized tests like the Graduate Record Examination (GRE) and the Law School Admission Test (LSAT). Indeed, this fact is widely advertised by philosophy departments at many institutions, and by the American Philosophical Association (APA 2014; 2019). Moreover, numerous studies have found that people who have studied philosophy tend to be more reflective, open-minded, and skilled in logical reasoning than those who have not done so (for a review, see Prinzing and Vazquez 2024).

Unfortunately, although these results establish some striking differences between people who have and have not studied philosophy, they do not support claims about the effects of a philosophical education. This is because group differences of this kind can be explained by self-selection as well as by effects of philosophical study. That is, philosophers may be more reflective, open-minded, and logical because studying philosophy cultivated these dispositions and abilities. But it's also possible that people who are already more reflective, open-minded, and logical are more likely to study philosophy in the first place. Of course, these are not exclusive possibilities. It may be that people who are more skilled with language and logic, more open-minded, and so on are more drawn to philosophy, but then studying philosophy further increases these abilities and dispositions.

This is not simply a problem “in theory.” Recent empirical work has found clear evidence of such self-selection. To illustrate, probably the most well-established difference between those who have and have not studied philosophy is that the former tend to be far more reflective (Byrd 2021; Livengood et al. 2010). “Reflectiveness” is typically measured using the Cognitive Reflection Test (Frederick 2005) or similar measures with questions that lure people into giving intuitive but incorrect responses. One such question asks, “If it takes 5 machines 5 minutes to make 5 widgets, how long will it take 100 machines to make 100 widgets?” For many people, “100 minutes” jumps to mind. Yet, a moment of reflection reveals that the correct answer is “5 minutes.” People who have studied philosophy are significantly more likely to correctly answer questions like this one, compared with those who haven’t. However, in prior work, we found that students in the first week of a Philosophy 101 course also scored dramatically higher than the population average on a version of the Cognitive Reflection Test (Prinzing and Vazquez 2024). That is, these students who were at the very beginning of their philosophical education were already considerably more reflective than most people. Findings like these highlight the risk that observed differences between philosophers and non-philosophers might result largely or even entirely from self-selection.

Naturally, the ideal way to test for effects of philosophical study on intellectual abilities and dispositions would be to use a randomized, controlled experiment. With a large sample, randomization to treatment and control groups ensures that, on

average, baseline characteristics do not differ across groups. This means that subsequent differences in the groups' outcomes are highly unlikely to result from pre-existing differences and are highly likely to instead result from the treatment itself. Unfortunately, although a short-term and small-scale randomized experiment might be feasible, anything at a large-scale or over a long-term does not seem likely ever to take place. For instance, convincing a large group of college freshmen to allow a coin flip to decide whether they will major in philosophy is both practically and ethically fraught.

When it is not possible, not ethical, or simply not practicable to conduct a randomized experiment, some people will throw up their hands and abandon hope for knowledge of causation. But this reaction is premature. After all, there are many sciences, such as epidemiology, economics, and environmental science, in which randomized experiments are rare and, in many cases, impossible or imprudent. For this reason, an entire field of research, spanning numerous disciplines, has developed methods for supporting causal conclusions with non-experimental data (Pearl 2015). Within this field of causal inference research, scholars have developed a range of techniques that can be used to rule out selection effects or other sources of "confounding" (i.e., where two variables are correlated because they share a common cause, known as a confound).¹

One particularly straightforward causal inference technique, called "covariate adjustment," involves statistically controlling for confounds. The obvious limitation of this approach is that one can never know whether one has measured and controlled for all confounds. But there is an elegant way of effectively controlling for a very broad range of confounds, even without knowing exactly what they are. This is by controlling for baseline differences in the outcome of interest. To illustrate, suppose we wanted to test whether majoring in philosophy improves logical reasoning abilities. No doubt, a multitude of factors besides students' majors will influence their scores on logic tests, anything from students' educational opportunities in high school, to their parents' levels of education, their household incomes, or other aspects of their socio-demographic backgrounds, etc. The impact of these factors will very likely have occurred by the time that the students arrive at college. Thus, if we measured students' logical reasoning abilities both at the start of college and again at the end, then by statistically controlling for freshman year scores when examining differences in senior year scores, we can remove the influence of these confounds. In other words, if philosophy students' performance on measures of intellectual abilities or dispositions were due to self-selection, then statistically controlling for baseline differences should make philosophy majors look unremarkable. Conversely, if

¹ To our knowledge, apart from ourselves, only two other researchers have ever applied causal inference techniques to investigate the effects of studying philosophy. Specifically, Farieta and Delprato (2024) used a method called "propensity score matching" (Austin 2011) with a sample of Columbian students training to be schoolteachers. Their findings suggest that studying philosophy improved scores on a test designed to assess the ability to critically analyze texts, reconstructing and evaluating their arguments. In the Appendix, we explain propensity scores further and report supplemental analyses of our data that use propensity scores. However, because the results of those analyses are identical to what we report in Section 5, we do not discuss them further here.

philosophy students continue to score well on a variety of measures, even after controlling for baseline differences, then this would be at least some evidence that studying philosophy improved their scores.

3. How to Measure Good Thinking

One centrally important question for a study of this kind concerns the kinds of measures that could aptly gauge the potential effects of a philosophical education. Naturally, there are many different abilities and dispositions that can make a person a good thinker. Some of these might be well-captured by standardized tests, such as the aforementioned LSAT and GRE. After all, these are fairly comprehensive indexes of verbal, logical, and quantitative reasoning abilities. In fact, the LSAT is primarily a logic test (though it also includes a section that assesses reading comprehension). The GRE has a Verbal Reasoning section that assesses a person's general facility with language, as well as a Quantitative Reasoning section, assessing mathematical ability. Since the study of philosophy is often claimed to teach students to use language skillfully, to make subtle distinctions and cogent arguments, and to recognize the logical relations among propositions and reason carefully from them, the LSAT and GRE Verbal each seem like fitting empirical metrics.

One potential objection to a focus on standardized tests (especially, but not exclusively the SAT) relates to accusations of racial and/or socio-economic biases (Eberle and Peltier 1989). It was on the basis of such concerns that, around the time of the COVID-19 pandemic, numerous colleges and universities dropped standardized tests from their admissions processes. In the years since, many of these institutions have reversed course, and are now requiring applicants to submit test scores again. This is because performance on standardized tests reliably predicts academic success (Friedman et al. 2025; Leonhardt 2024), and removing them from consideration in admissions actually exacerbated the underrepresentation of minority groups (see, e.g., Schmill 2022). Thus, recognizing that all empirical measures are flawed, limited, and subject to measurement errors of various kinds, standardized tests appear to be some of the best available tools for measuring a wide range of valuable intellectual abilities (Currid-Halkett 2020).

Yet, philosophers also often claim that studying philosophy cultivates certain intellectual virtues like curiosity, intellectual humility, or open-mindedness. Alongside intellectual *abilities*, these sorts of dispositions also seem quite important for being a good thinker (King 2021). After all, high test scores show that a person can reason well when prompted and incentivized to do so. But that leaves open the question of whether the person is generally disposed to use their abilities to earnestly and thoroughly pursue truth—to think carefully, critically, and with openness and humility. Naturally, such intellectual dispositions can be very difficult to quantify. Empirical measures of, e.g., curiosity, intellectual humility, and open-mindedness are generally going to be self-reports (e.g., Hoyle et al. 2016; Price et al. 2015). And, in contrast to standardized tests, self-reports are subject to biases, such as “self-enhancement” bias (Weiner and Guenther 2020), where individuals may try to present themselves in overly flattering ways.

In short, these two different kinds of measures have complementary strengths and weaknesses. Standardized tests are “objective” in the sense that they are immune to reporting biases. They also capture a broad range of important abilities but might be thought to reflect a relatively thin conception of good thinking. On the other hand, self-reports can capture dispositions like curiosity and open-mindedness, that seem to be important aspects of good thinking. But these are less “objective” in the aforementioned sense. Given these relative advantages and disadvantages, we would ideally like to see converging evidence from both kinds of measures. That is, although either result would be interesting in its own right, evidence that studying philosophy improves both test scores *and* self-reported intellectual dispositions would provide particularly strong evidence that the discipline makes people better thinkers.

4. Data and Measures for the Present Study

We analyzed data collected between 1990 and 2019 by the Higher Education Research Institute (HERI; <https://heri.ucla.edu/>) and Cooperative Institutional Research Program, based at the University of California, Los Angeles. Data from students graduating between 1994 and 2008 are publicly accessible in the HERI’s Data Archive (<https://heri.ucla.edu/heri-data-archive/>). To access the more recent data, we had to apply and pay a fee.

Participating students completed surveys at the start of their freshman year and end of their senior year, in which they reported on their academic majors, standardized test scores, sociodemographic backgrounds, and more. Incoming freshmen were asked to report their SAT scores and, prior to 2005, graduating seniors were also asked to report their GRE and LSAT scores (assuming they had taken these tests). Starting with the 2010 cohort, the surveys also included self-report measures called the “Habits of Mind” and “Pluralistic Orientation” scales. The individual questions in these scales are presented in Table 1. In their online documentation, the HERI describes these, respectively, as assessing “the behaviors and traits associated with academic success... [and] lifelong learning” and the “skills and dispositions appropriate for living and working in a diverse society” (<https://www.heri.ucla.edu/PDFs/constructs/Appendix2017.pdf>). But just looking at the survey questions themselves, the Habits of Mind scale seems to assess traits like curiosity (see, e.g., items 3, 7, 8), intellectual rigor (items 2, 9), and to some extent intellectual humility (item 5) and open-mindedness (item 10). The Pluralistic Orientation scale seems to be, in effect, a measure of open-mindedness.

The HERI computes composite scores from each set of items. (Specifically, these are factor scores from item response theory models. Further details can be found online: <https://heri.ucla.edu/cirp-constructs/>.) For ease of interpretation, we standardized these composite scores, meaning that we scaled them to have a mean of 0 and a standard deviation of 1. Hence, a student with a score of 1 is one standard deviation above average, whereas someone with a score of -0.5 is half a standard deviation below average.

The full sample includes $N = 649,511$ students (including $n = 4,843$ philosophy majors), attending 804 colleges and universities around the United States. Of these,

Table 1. Items from self-report measures

Habits of Mind

How often in the past year did you:

(Response options: Not at all, Occasionally, Frequently)

1. Seek solutions to problems and explain them to others
2. Support your opinions with a logical argument
3. Look up scientific research articles and resources
4. Take a risk because you felt you had more to gain
5. Accept mistakes as part of the learning process
6. Take on a challenge that scares you
7. Ask questions in class
8. Explore topics on your own, even though it was not required for a class
9. Evaluate the quality or reliability of information you received
10. Seek alternative solutions to a problem

Pluralistic Orientation

How would you rate yourself in the following areas:

(Response options: A Major Weakness, Somewhat Weak, Average, Somewhat Strong, A Major Strength)

1. Ability to see the world from someone else's perspective
2. Tolerance of others with different beliefs
3. Openness to having my own views challenged
4. Ability to discuss and negotiate controversial issues
5. Ability to work cooperatively with diverse people

59% identified as female, 35% as male, and 5% did not indicate a sex; 72% identified as White, 5% as Asian or Pacific Islander, 5% as Black or African American, 4% as Latino/a/x, 5% as mixed and 2% as another race or ethnicity, and 7% did not indicate a race or ethnicity. Because GRE and LSAT scores were only collected from students graduating between 1994 and 2004 ($n = 392,858$), and the Habits of Mind and Pluralistic Orientation scales were only administered to students graduating in 2010 or later ($n = 122,352$), some of our analyses use a subset of the full sample.

5. Empirical Findings

We first looked for evidence of selection effects by testing whether students who score higher on the SAT and the self-report measures of intellectual dispositions in the freshman year survey are more likely to major in philosophy. Second, we tested whether, at senior year, philosophy majors score higher than non-philosophy majors on the GRE, LSAT, Habits of Mind, and Pluralistic Orientation, even after controlling for baseline differences. Third, we compared the baseline-adjusted averages for specific majors, to see where philosophy places in the rankings. Finally, we looked for evidence of a distinct kind of selection effect related to the kinds of students who take the GRE and LSAT (explained further below) that could potentially have biased the standardized testing results.

Our analyses used mixed-effects regression models (logistic regressions for dichotomous outcomes) with random intercepts for institutions (i.e., the colleges

and universities that students attended). We fit these models using the *lme4* and *lmerTest* packages in R (Bates et al. 2020; Kuznetsova, Brockhoff, and Christensen R. H. B. 2017), computed estimated marginal means using the *emmeans* package (Lenth et al. 2018), and used multiple imputation to accommodate missing data with the *mice* package (Buuren and Groothuis-Oudshoorn 2011). The code used in these analyses is available online (<https://osf.io/4s693>). Below, we describe the results in colloquial English. More detailed results of the statistical models are given in Tables A1 and A2 in the Appendix.

5.1. Intellectual Abilities and Dispositions Predict Whether Students Major in Philosophy

We looked for evidence of selection effects by testing whether students' scores on the SAT Verbal, SAT Math, and freshman year Habits of Mind and Pluralistic Orientation scales could predict whether they would end up majoring in philosophy. Results indicated that, apart from SAT Math, each of these predictors was statistically significant. Scores on the Verbal section of the SAT were the strongest predictor. More precisely, a one standard deviation increase in SAT Verbal is associated with 57% greater odds of majoring in philosophy. One standard deviation increases in the Habits of Mind and Pluralistic Orientation scales are associated, respectively, with 34% and 13% greater odds of majoring in philosophy. (Again, SAT Math was not statistically significant.)

In short, these results indicate that students with better verbal reasoning ability (though not mathematical ability), and who are more curious, open-minded, and intellectually rigorous than their peers are more likely than their peers to major in philosophy. These results continue to indicate that people who have studied philosophy score well on measures of intellectual abilities and valuable intellectual dispositions at least partly because philosophy attracts people who independently score well on such measures. The question remains, however, does studying philosophy itself foster these intellectual abilities and dispositions?

5.2. Adjusting for Baseline Differences, Philosophers Still Outperform Their Peers

Next, we compared philosophy and non-philosophy majors' scores on senior year standardized tests and self-report measures while adjusting for baseline differences, observed at the start of the freshman year. The idea, again, is that students' intellectual abilities and dispositions are shaped by many factors beyond their academic majors, such as prior educational opportunities and socio-demographic backgrounds. These factors will also affect students' scores at the start of college. So, by controlling for freshman year scores, we should be able to remove the influence of pre-college confounds, thereby giving a more accurate estimate of the treatment effects. Figure 1 plots the results.

Starting with the standardized tests, we ran three separate models for scores on the GRE Verbal, GRE Quantitative, and LSAT. Unsurprisingly, SAT scores were significantly, positively associated with GRE and LSAT scores across the board.

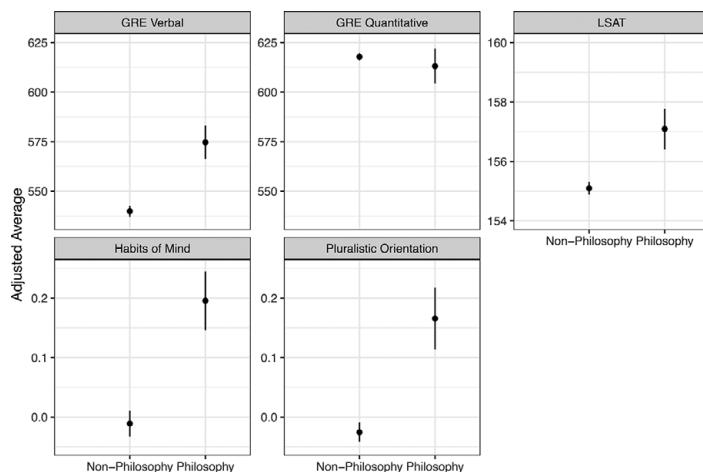


Figure 1. Baseline-adjusted average scores for philosophy and non-philosophy majors. Points and error bars indicate estimated marginal means with 95% confidence intervals, derived from mixed-effects regression models. For GRE Verbal, GRE Quantitative, and LSAT, means are adjusted for SAT scores. For Habits of Mind and Pluralistic Orientation, means are adjusted for scores in the freshman year survey.

But crucially, after accounting for SAT scores, philosophy majors scored significantly higher than non-philosophy majors on the GRE Verbal and LSAT. On the GRE Quantitative, by contrast, there was no significant difference between philosophy and non-philosophy majors. Turning to the self-report measures, we tested whether philosophy majors scored higher on the Habits of Mind and Pluralistic Orientation scales during senior year while adjusting for their scores during freshman year. Unsurprisingly, the freshman year scores were significant predictors in both cases. But, crucially, after accounting for these baseline differences, philosophy majors still scored significantly higher than non-philosophy majors on both Habits of Mind and Pluralistic Orientation.

These results indicate that, even after accounting for differences in pre-college verbal and quantitative reasoning abilities, philosophy majors display stronger verbal and logical (though not quantitative) reasoning abilities than their peers. More precisely, the SAT-adjusted average score for philosophy majors on the GRE Verbal (i.e., 573 out of a possible 800) is about 33 points higher than that of non-philosophy majors (i.e., 540). This amounts to a standardized mean difference (i.e., the difference between groups in terms of standard deviations) of 0.48, which, according to a common convention, would be considered a “medium-sized” effect (Cohen 1988). On the LSAT, the SAT-adjusted average score for philosophy majors (i.e., 157 out of a possible 180) is about 2 points higher than that of non-philosophy majors (i.e., 155). This is a standardized mean difference of 0.35, or a “small to medium-sized” effect. Similarly, senior philosophy majors reported greater levels of positive intellectual dispositions than their peers—encompassing curiosity, intellectual rigor, intellectual humility, and open-mindedness—even after controlling for reports of such dispositions in freshman year. These effects would conventionally be termed “small” effects

(standardized mean differences of .24 for Habits of Mind and .21 for Pluralistic Orientation).

Now, although contrasting philosophy majors with non-philosophy majors enables a simple, statistically powerful comparison, it also conceals a great deal of diversity within the “non-philosophy” group. For all we have shown so far, there may be other majors that substantially outperform philosophy. Hence, to test the claim that there is something truly special about the study of philosophy, making it distinctively well-suited to cultivating good thinking, we compared philosophy majors with other specific majors, again while controlling for baseline levels of the outcomes. The sample includes students with 90 different majors, but some of these majors are represented by only small numbers of students. Hence, in these analyses, we excluded majors with fewer than 200 students. This ensures that each major’s average stably reflects characteristics of that major, rather than sampling variability. After these exclusions, the analyses involving the standardized tests included 57 majors and the analyses involving the self-reports included 62 majors.

Comparing the baseline-adjusted average scores for all of these different majors revealed something truly striking. Although philosophy majors are unremarkable when it comes to the GRE Quantitative (placing 30th out of 57), they rank first on the GRE Verbal, first on the LSAT, first on the Habits of Mind scale, and sixth on the Pluralistic Orientation scale. These rankings are presented in [Figures 2-3](#).

5.3. Do Only the Best and Brightest Philosophy Students Take the GRE or LSAT?

Thus far, we have focused on trying to rule out the possibility that philosophy majors’ impressive intellectual abilities and dispositions result from the fact that the major attracts students who already have such abilities and dispositions. But there is another kind of selection bias that might pose a problem for our standardized testing results. This has to do with the fact that only a small proportion of undergraduates actually take the GRE or LSAT, and the characteristics of those who choose to do so probably varies from one discipline to the next. Hence, it could be that only the best and brightest philosophy majors decide to go to graduate school and so take the GRE or LSAT, whereas many of the more middling students from other disciplines do so. If so, then the average score for philosophy majors might be especially high, not because of philosophy majors’ impressive intellectual abilities, but merely because only the top philosophy students actually take the tests.

If this speculation were correct, then we should see a stronger association between SAT scores and the odds of taking the GRE or LSAT among philosophy majors than among non-philosophy majors. It may be that students with higher SAT scores are, in general, more likely to take subsequent standardized tests. But, if there were a selection effect of this second kind, then the association between SAT scores and whether students take subsequent tests should be *especially* strong among philosophy majors compared with non-philosophy majors. We looked for evidence of this second kind of selection effect by testing whether students’ SAT scores and whether students’ major in philosophy predict whether they take the GRE or LSAT, and also whether there is an interaction between these two predictors.

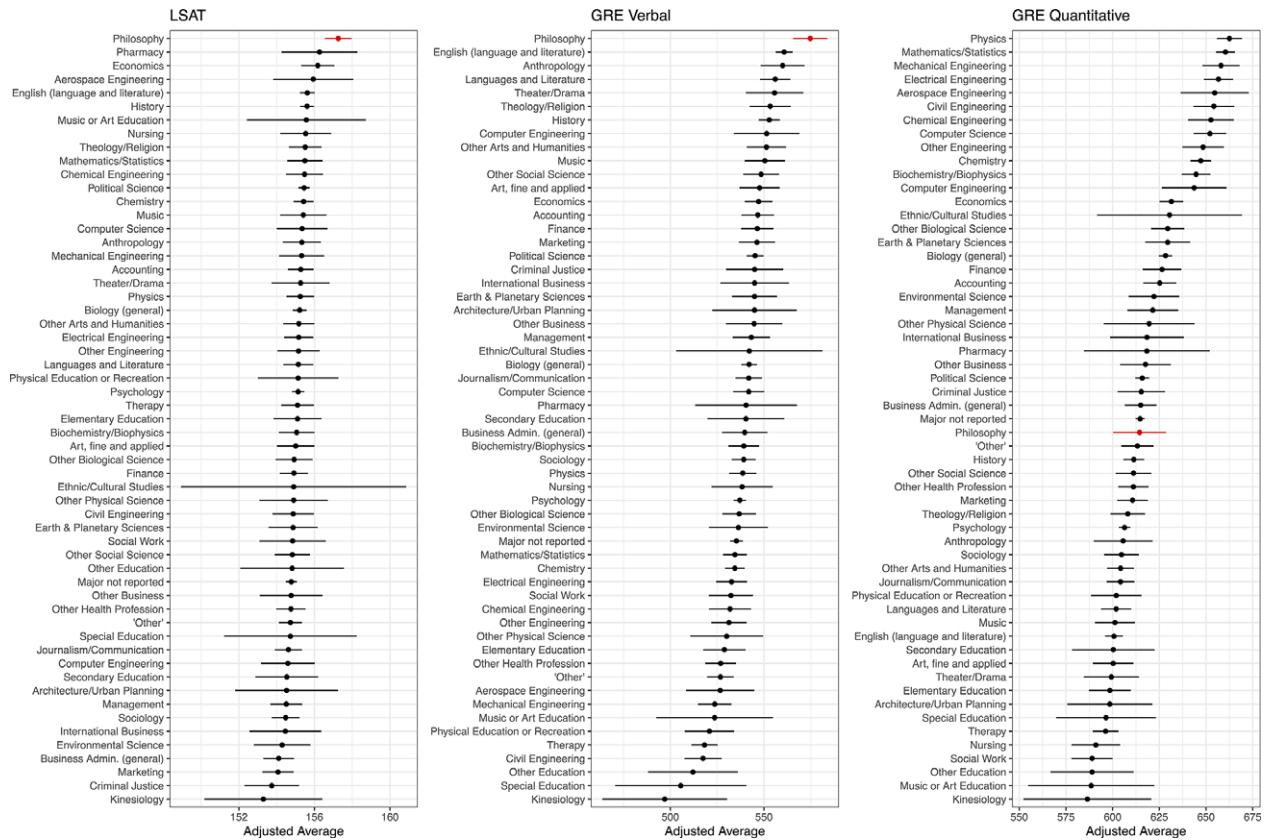


Figure 2. SAT-adjusted average scores on standardized tests for specific majors. Points and error bars indicate estimated marginal means with 95% confidence intervals derived from mixed-effects regression models. Philosophy is highlighted with red.

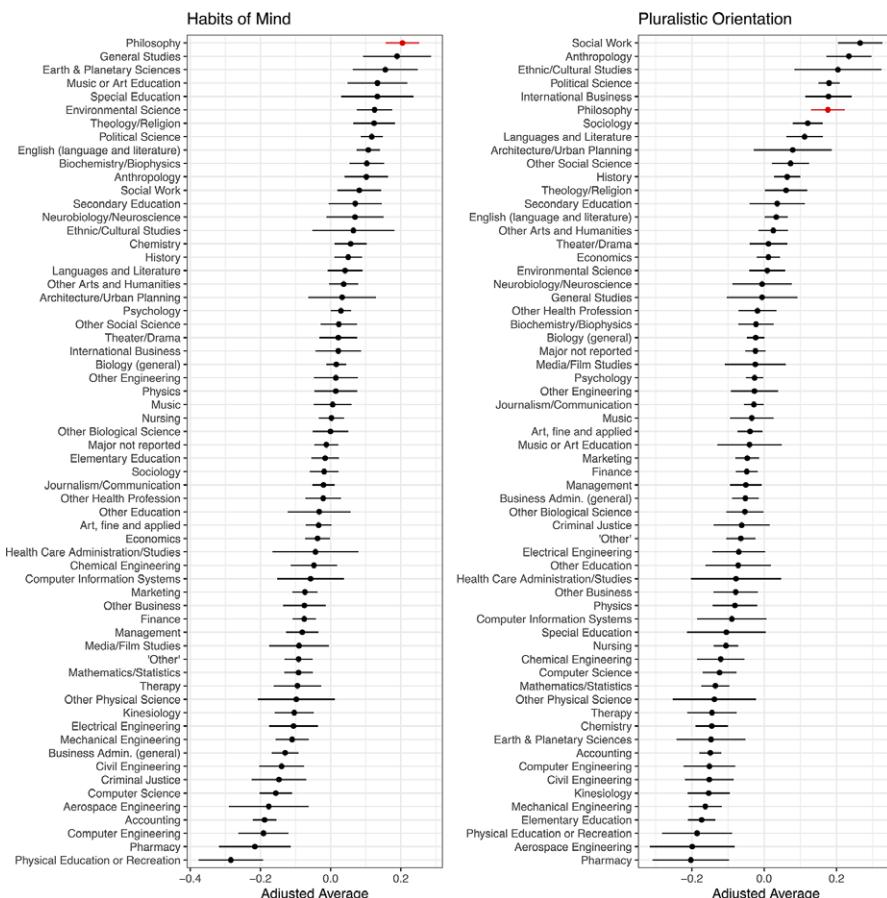


Figure 3. Baseline-adjusted average scores on self-report measures for specific majors. Points and error bars indicate estimated marginal means with 95% confidence intervals derived from mixed-effects regression models. Philosophy is highlighted with red.

Results indicated that students with higher SAT scores are indeed more likely to take the GRE and that philosophy majors are more likely to take the GRE than non-philosophy majors. Crucially, however, we found no interaction—that is, no evidence that the association between SAT scores and the odds of taking the GRE differ between philosophy and non-philosophy majors. Turning to the LSAT, we found a similar pattern of results. Unsurprisingly, again, students with higher SAT scores are more likely to take the LSAT, and philosophy majors are more likely (about three and a half times more likely) to take the LSAT than non-philosophy majors. Crucially, we again found no interaction. Thus, we found no evidence of this second kind of potential selection effect, whereby philosophy majors' impressive performance on the GRE and LSAT results from a tendency for only the best and brightest philosophy majors to take these tests in the first place.

6. Implications and Conclusion

These findings support a popular and venerable argument for the value of philosophy. That argument, as discussed in Section 1, rests on a crucial, empirical premise—namely, that studying philosophy makes people better thinkers. In a recent review, we concluded that there was an absence of evidence concerning this premise (Prinzing and Vazquez 2024). Here, we brought to bear a far larger body of data than any prior study and conducted more rigorous analyses. The data included records from over half a million undergraduate students, attending over 800 institutions across the United States. Moreover, because we have data from these students both at the start and end of their time in higher education, we were able to account for pre-existing differences when comparing philosophy majors with their peers. In this way, we controlled for a host of potential confounds when estimating the effects of philosophical study.

Our results indicated that students with better verbal reasoning abilities and more curiosity, intellectual rigor, and open-mindedness are more likely to major in philosophy. They also indicated that, after adjusting for baseline differences, philosophy majors outperform other students on these measures. In fact, on average, philosophy majors score higher than all other majors on the GRE Verbal and LSAT, as well as a self-report measure designed to assess good habits of mind. Short of a randomized experiment—which for various ethical and practical reasons is unlikely ever to take place—these findings arguably constitute the clearest and strongest kind of empirical support that we will find for the claim that studying philosophy makes students better thinkers.

One important caveat comes from the possibility that certain aspects of the intellectual abilities assessed by the GRE or LSAT are not well assessed by the SAT. For example, unlike the LSAT, the SAT does not include a section dedicated specifically to logical reasoning. Thus, it may be that controlling for SAT scores does not fully account for pre-college differences in logical reasoning abilities. One way to overcome this limitation would be to have students complete the exact same tests (which could be based on the LSAT or other tests, such as the California Critical Thinking Skills Test) at the start and end of their education. Future work may also be able to extend these findings by exploring whether specific forms of philosophical study might affect students differently. For example, do students who focus on ethics show different effects from those who focus on metaphysics? What about students in more analytic versus continental departments? Another interesting avenue would be to look for a “dose-response” relationship—for example by testing whether philosophy majors differ from philosophy minors (who might be thought to have received a smaller “dose” of philosophy).

One limitation of these findings is that, although the Habits of Mind and Pluralistic Orientation items capture some of the behaviors and motivations characteristic of intellectual virtues, they do not speak to the objects, occasions, and means of their exercise. Thus, they do not capture the full profile of intellectual virtue as philosophers traditionally conceive of it (King 2021). To determine whether students of philosophy use their intellectual abilities for the right reasons, with the right means, on the right occasions, and directed at the right objects would require further, and potentially a very different kind of, evidence.

Relatedly, there are many intellectual virtues that one could examine, and the measures used here only touched on a few. For example, one of the items from the Habits of Mind scale could be seen as tapping into intellectual humility (item 5, see Table 1). But there are other, more effective measures of intellectual humility that one could use (see, e.g., Hoyle et al. 2016). And, naturally, one could also consider whether studying philosophy supports virtues like intellectual autonomy, fair-mindedness, intellectual courage, or others (King 2021).

Finally, there are numerous other kinds of claims about the value of philosophy, besides its ability to cultivate intellectual abilities or dispositions. Some would claim that philosophical study supports students' holistic formation as persons (Standish and Saito 2023), or that it fosters autonomy (Standish 1999), or bestows them with "powerful knowledge" (Young and Muller 2013). These are plausibly laudable educational goals, but our findings do not speak to these claims. Perhaps future work could investigate these sorts of outcomes—though, there is room for reasonable skepticism about the prospects for empirically assessing them.

Nonetheless, the intellectual abilities and dispositions assessed in this study seem both useful and admirable. Skill with logic enables people to make better inferences, which should plausibly lead them to form more true and fewer false beliefs. Facility with language enables people to articulate their beliefs more clearly to others. The dispositions critical for lifelong learning and open, humble thinking are plausibly useful for people in all walks of life. (Though, the degree to which the various measures used here predict specific outcomes later in life could itself be tested empirically.) Hence, if philosophical study cultivates these abilities and dispositions, as our findings suggest, then this is good news indeed for philosophy. In the highly technocratic and bureaucratized world of the 21st century academy, the ability to point to such measurable outcomes is often necessary to maintain institutional support for departments and programs. Hence, our findings may have some utility for those advocating for the discipline.

Finally, although we have focused specifically on intellectual outcomes, these same sorts of abilities and dispositions are thought to have substantial implications for the political life of pluralistic societies. For decades, political philosophers have argued that dispositions that facilitate rigorous and autonomous reflection are essential not only for individual flourishing but also for sustaining democratic norms and institutions (Brighouse 2006; Gutmann 1995). According to this view, the kinds of intellectual capacities that we have examined support, for example, the ability to tolerate and engage charitably with diverse perspectives and to find common ground within pluralistic societies, and that these are, in turn, essential for the health of democracy (Lynch 2019; Nussbaum 2024; Samaržija and Cassam 2023).

Although our study does not directly assess these sorts of civic outcomes, it is noteworthy that philosophy majors ranked highly on a measure like the Pluralistic Orientation scale, which was designed specifically to assess "skills and dispositions appropriate for living and working in a diverse society" and is regarded as an important indicator of the success of a liberal education (Hurtado and DeAngelo 2012). It would be worthwhile, in future work, to explore these broader civic and moral effects of philosophical study more directly. After all, it is one thing to form

sharp, analytical thinkers, and quite another to cultivate intellectually virtuous citizens inclined to use their minds responsibly in service of the common good.

MICHAEL PRINZING 

¹WAKE FOREST UNIVERSITY, UNITED STATES

prinzim@wfu.edu

MICHAEL VAZQUEZ 

²THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL, UNITED STATES

michael.vazquez@unc.edu

References

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

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

Austin, Peter C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46 (3): 399–424. <https://doi.org/10.1080/00273171.2011.568786>.

Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, R. H. B. Christensen, Henrik Singmann, Bin Dai, et al. 2020. "Package 'Lme4'." R Package. 2020. <https://cran.r-project.org/web/packages/lme4/lme4.pdf>.

Brighouse, Harry. 2006. *On Education*. London: Routledge. <https://doi.org/10.4324/9780203390740>.

Buuren, Stef van, and Karin Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45 (December):1–67. <https://doi.org/10.18637/jss.v045.i03>.

Byrd, Nick. 2021. "Reflective Reasoning & Philosophy." *Philosophy Compass* 16 (11). <https://doi.org/10.1111/phc3.12786>.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates.

Currid-Halkett, Elizabeth. 2020. "A Pandemic Isn't a Reason to Abolish the SAT." *The New York Times*, May 1, 2020, sec. Opinion. <https://www.nytimes.com/2020/05/01/opinion/coronavirus-test-optional-sat.html>.

Eberle, Joan L., and Gary L. Peltier. 1989. "Is the SAT Biased? A Review of Research." *American Secondary Education* 18 (1): 17–24.

Farieta, Alejandro, and Marcos Delprato. 2024. "The Effect of Philosophy on Critical Reading: Evidence from Initial Teacher Education in Colombia." *International Journal of Educational Development* 104 (January):102974. <https://doi.org/10.1016/j.ijedudev.2023.102974>.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19 (4): 25–42. <https://doi.org/10.1257/089533005775196732>.

Friedman, John N., Bruce Sacerdote, Douglas O Staiger, and Michele Tine. 2025. "Standardized Test Scores and Academic Performance at Ivy-plus Colleges." Working Paper 33570. National Bureau of Economic Research. <http://www.nber.org/papers/w33570>.

Gutmann, Amy. 1995. "Civic Education and Social Diversity." *Ethics* 105 (3): 557–79.

Hoekema, David A. 1986. "Why Major in Philosophy?" *Proceedings and Addresses of the American Philosophical Association* 59 (4): 601–6. <https://doi.org/10.2307/3131573>.

Hoyle, Rick H., Erin K. Davisson, Kate J. Diebels, and Mark R. Leary. 2016. "Holding Specific Views with Humility: Conceptualization and Measurement of Specific Intellectual Humility." *Personality and Individual Differences* 97 (July):165–72. <https://doi.org/10.1016/j.paid.2016.03.043>.

Hurtado, Sylvia, and Linda DeAngelo. 2012. "Linking Diversity and Civic Minded Practices with Student Outcomes." *Liberal Education* 2:14–23.

King, Nathan L. 2021. *The Excellent Mind: Intellectual Virtues for Everyday Life*. Oxford University Press.

Kuznetsova, A., P. B. Brockhoff, and Christensen R. H. B. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82 (13): 1–26. <https://doi.org/10.18637/jss.v082.i13>.

Leite, Walter. 2017. *Practical Propensity Score Methods Using R*. SAGE Publications, Inc. <https://doi.org/10.4135/9781071802854>.

Lenth, R., H. Singmann, J. Love, P. Buerkner, and M. Herve. 2018. "Package 'Emmeans'." <https://cran.microsoft.com/snapshot/2018-01-13/web/packages/emmeans/emmeans.pdf>.

Leonhardt, David. 2024. "The Misguided War on the SAT." *The New York Times*, January 7, 2024, sec. Briefing. <https://www.nytimes.com/2024/01/07/briefing/the-misguided-war-on-the-sat.html>.

Livengood, Jonathan, Justin Sysma, Adam Feltz, Richard Scheines, and Edouard Machery. 2010. "Philosophical Temperament." *Philosophical Psychology* 23 (3): 313–30. <https://doi.org/10.1080/09515089.2010.490941>.

Lynch, Michael P. 2019. *Know-It-All Society: Truth and Arrogance in Political Culture*. New York: Liveright.

Morton, Jennifer. 2019. "Philosophy as an Antidote to Injustice." *The Philosophers' Magazine*, 2019.

Nussbaum, Martha. 2024. *Not for Profit: Why Democracy Needs the Humanities*. Princeton, N.J.: Princeton University Press.

Pearl, Judea. 2015. *An Introduction to Causal Inference*. CreateSpace Independent Publishing Platform.

Price, Erika, Victor Ottati, Chase Wilson, and Soyeon Kim. 2015. "Open-Minded Cognition." *Personality and Social Psychology Bulletin* 41 (11): 1488–504. <https://doi.org/10.1177/0146167215600528>.

Prinzing, Michael M., and Michael Vazquez. 2024. "Does Studying Philosophy Make People Better Thinkers?" *Journal of the American Philosophical Association*, March, 1–22. <https://doi.org/10.1017/apa.2023.30>.

Russell, Bertrand. 1912. *The Problems of Philosophy*. Oxford University Press.

Samaržija, Hana, and Quassim Cassam, eds. 2023. *The Epistemology of Democracy*. New York: Routledge. <https://doi.org/10.4324/9781003311003>.

Schmill, Stuart. 2022. "We Are Reinstating Our SAT/ACT Requirement for Future Admissions Cycles." *MIT Admissions* (blog). March 28, 2022. <https://mitadmissions.org/blogs/entry/we-are-reinstating-our-sat-act-requirement-for-future-admissions-cycles>.

Standish, Paul. 1999. "Education Without Aims?" In *The Aims of Education*, edited by Roger Marples, 35–49. Oxford, UNITED KINGDOM: Taylor & Francis Group. <http://ebookcentral.proquest.com/lib/wfu/detail.action?docID=168793>.

Standish, Paul, and Naoko Saito. 2023. "Learning and Human Development." In *Philosophical Foundations of Education*, edited by Winston C. Thompson, 101–24. Bloomsbury Publishing.

Weiner, Drew S., and Corey L. Guenther. 2020. "Self-Enhancement Bias." In *Encyclopedia of Personality and Individual Differences*, edited by Virgil Zeigler-Hill and Todd K. Shackelford. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-24612-3>.

Young, Michael, and Johan Muller. 2013. "On the Powers of Powerful Knowledge." *Review of Education* 1 (3): 229–50. <https://doi.org/10.1002/rev3.3017>.

Appendix

Detailed Results of Statistical Models

Table A1. Results of Mixed-Effects Logistic Regressions

Independent variable	b	95% CI	OR	p
<i>Philosophy Major</i> (0 = student did not major in philosophy, 1 = student did major in philosophy)				
Intercept	-5.79	[-5.91, -5.67]	0.00	< .001
SAT Verbal	0.45	[0.38, 0.53]	1.57	< .001
SAT Math	-0.01	[-0.10, 0.08]	0.99	.872
Freshman Year Habits of Mind	0.29	[0.22, 0.35]	1.34	< .001
Freshman Year Pluralistic Orientation	0.12	[0.06, 0.19]	1.13	.008
<i>Take GRE</i> (0 = student did not take the GRE, 1 = student did take the GRE)				
Intercept	-3.16	[-3.25, -3.08]	0.04	< .001
SAT (total)	0.27	[0.26, 0.29]	1.31	< .001
Philosophy Major	0.25	[0.02, 0.49]	1.29	.039
SAT × Philosophy Major	0.03	[-0.10, 0.15]	1.03	.669
<i>Take LSAT</i> (0 = student did not take the LSAT, 1 = student did take the LSAT)				
Intercept	-4.53	[-4.65, -4.41]	0.01	< .001
SAT (total)	0.23	[0.20, 0.26]	1.26	< .001
Philosophy Major	1.26	[1.08, 1.44]	3.53	< .001
SAT × Philosophy Major	-0.09	[-0.18, 0.01]	0.92	.101

Note. "b" and "95% CI" indicate coefficient estimates and 95% confidence intervals. "OR" indicates the odds ratio. All independent variables apart from Philosophy Major were z-scored for ease of interpretation.

Table A2. Results of Mixed-Effects Regressions

Independent variable	b	95% CI	p
<i>GRE Verbal</i>			
Intercept	539.83	[537.09, 542.57]	< .001
Philosophy Major	33.25	[23.86, 42.64]	< .001
SAT Verbal	61.31	[59.82, 62.8]	< .001
SAT Math	7.14	[5.91, 8.37]	< .001
<i>GRE Quantitative</i>			
Intercept	617.97	[616.12, 619.83]	< .001
Philosophy Major	-3.91	[-18.51, 10.7]	.619
SAT Verbal	3.41	[1.66, 5.17]	.006
SAT Math	81.65	[80.39, 82.92]	< .001
<i>LSAT</i>			
Intercept	155.09	[154.85, 155.33]	< .001
Philosophy Major	2.12	[1.46, 2.78]	< .001
SAT Verbal	2.97	[2.81, 3.14]	< .001
SAT Math	3.06	[2.88, 3.23]	< .001
<i>Senior Year Habits of Mind</i>			
Intercept	0.00	[-0.02, 0.02]	.839
Philosophy Major	0.21	[0.16, 0.25]	< .001
Freshman Year Habits of Mind	0.44	[0.44, 0.45]	< .001

(Continued)

Table A2. *Continued*

Independent variable	<i>b</i>	95% CI	<i>p</i>
<i>Senior Year Pluralistic Orientation</i>			
Intercept	−0.02	[−0.04, −0.01]	.002
Philosophy Major	0.19	[0.15, 0.24]	< .001
Freshman Year Pluralistic Orientation	0.41	[0.41, 0.42]	< .001

Note. “*b*” and “95% CI” indicate coefficient estimates and 95% confidence intervals. All independent variables apart from Philosophy Major were *z*-scored for ease of interpretation.

Propensity Score Analysis

In the main text, we used covariate adjustment to control for baseline differences between philosophy majors and their peers. However, there are other, more sophisticated methods for supporting causal inferences from non-experimental data. In particular, methods using “propensity scores” are increasingly used in fields like economics, epidemiology, educational research, and so on (Austin 2011). To confirm the robustness of the results reported in the main text, we compared them with the results of analyses using propensity scores.

A propensity score is an individual’s probability of receiving a treatment, conditional on their baseline characteristics. In randomized experiments, propensity scores are fixed and known. For example, if treatment condition is determined by a coin flip, then all participants have propensity scores of .5. In non-randomized studies, propensity scores can be *estimated* using observed baseline characteristics. They can then be used when estimating treatment effects. For example, “propensity score matching” involves identifying pairs of treated and untreated participants with the same (or very similar) propensity scores. This produces two groups, one that received the treatment and one that did not, with the same average conditional probability of receiving treatment—just like a randomized experiment. Another approach, which we use here in the Appendix, is called “inverse probability of treatment weighting” and involves creating weights for different observations based on the propensity scores (Leite 2017). The R code used in these analyses is available online (<https://osf.io/6yz8m>).

We used different propensity score models for the standardized tests versus self-reports since our data for these two kinds of measures come from non-overlapping groups of students (1994–2004 cohorts for the tests and 2010–2019 cohorts for the self-reports). Both were logistic regressions in which Philosophy Major was the dependent variable. The independent variables were SAT Verbal, SAT Math, academic self-concept,² sex, race, religion, household income, political ideology, father’s educational attainment, mother’s educational attainment, and the institution that participants attended. We extracted the fitted values from these models.

² This is a construct created by the Higher Education Research Institute, designed to be “a unified measure of students’ beliefs about their abilities and confidence in academic environments” (<https://www.heri.ucla.edu/monographs/TheAmericanFreshman2014.pdf>). The score is composed of self-ratings of “academic ability,” “self-confidence – intellectual,” “mathematical ability,” and “drive to achieve.”

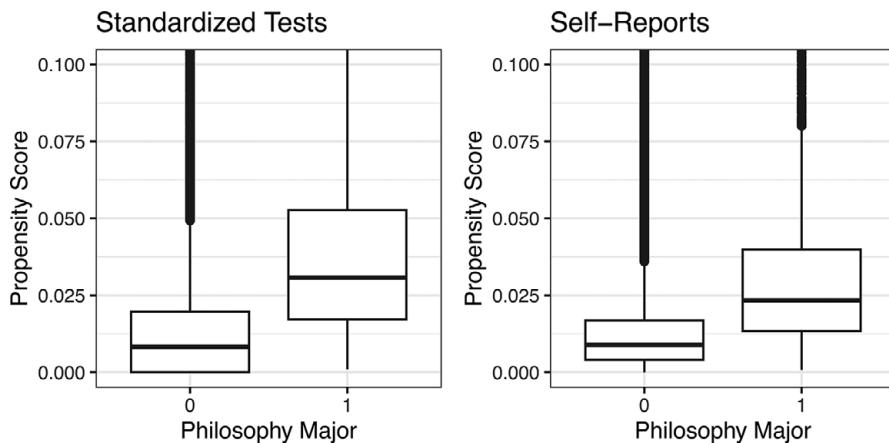


Figure A1. Boxplots of propensity scores. Thick horizontal lines indicate medians, and the boxes around them encompass the interquartile range (i.e., 75% of the observations).

Table A3. Results of IPTW models

Dependent Variable	Estimated effect	95% CI	<i>p</i>
GRE Verbal	30.61	[22.15, 39.06]	.001
GRE Quantitative	-9.99	[-23.87, 3.89]	.228
LSAT	2.30	[1.55, 3.04]	.003
Habits of Mind	0.29	[0.27, 0.30]	< .001
Pluralistic Orientation	0.28	[0.26, 0.29]	< .001

Note. The standardized tests were left in their raw units, whereas Habits of Mind and Pluralistic Orientation scales were z-scored. Hence, for the latter, the estimated effect is reported in standard deviations.

Boxplots of these propensity score estimates (from each model) are presented in Figure A1 below. Although, unsurprisingly, most of the philosophy majors had higher scores than most of the non-philosophy majors, there remained a substantial region of common support.

To estimate the causal impact of majoring in philosophy on each outcome, we computed inverse probability of treatment scores (i.e., the propensity score for non-philosophy majors and the inverse of the propensity score for philosophy majors) and used these as weights in mixed-effects regression models with a random effect of institution. The results, which are presented in Table A3 below, were fully consistent with the results of the models reported in the main text. This analysis suggests that majoring in philosophy rather than another field increases GRE Verbal scores by about 30 points, LSAT scores by about 2 points, and Habits of Mind and Pluralistic Orientation by about .3 standard deviations. We find no significant effect on GRE Quantitative scores.