


## Who Benefits From a Foundational Logic Course? Effects on Undergraduate Course Performance

Rafael Quintana & Christian Schunn

To cite this article: Rafael Quintana & Christian Schunn (2019) Who Benefits From a Foundational Logic Course? Effects on Undergraduate Course Performance, Journal of Research on Educational Effectiveness, 12:2, 191-214, DOI: [10.1080/19345747.2018.1543372](https://doi.org/10.1080/19345747.2018.1543372)

To link to this article: <https://doi.org/10.1080/19345747.2018.1543372>

 [View supplementary material](#) 

 Published online: 10 May 2019.

 [Submit your article to this journal](#) 

 [View Crossmark data](#) 



## Who Benefits From a Foundational Logic Course? Effects on Undergraduate Course Performance

Rafael Quintana<sup>a</sup> and Christian Schunn<sup>a</sup>

### ABSTRACT



Being able to understand and evaluate arguments in different modalities and in different disciplines is thought to be a key component of students' academic success in college. However, many students do not receive explicit instruction in the basic concepts and rules of argumentation. Using a difference-in-differences approach with a multicohort longitudinal data set of almost 15,000 undergraduates beginning in health and science, technology, engineering, and mathematics (STEM)-related fields at a research university, we examined changes in relative performance of students after enrolling in an introductory logic course. We find that students improved their grade point average (GPA) after taking the course, especially if they begin college with low academic achievement (Cohen's  $d=0.18$ ). Our results are consistent with the idea that acquiring foundational skills, in particular general skills in argumentation, prepares STEM students for future learning.

### KEYWORDS


argumentation  
critical thinking  
basic skills  
academic success  
higher education

In order to succeed in college, students need some foundational skills and strategies that allow them to learn material at a deep level and make connections across subjects (Conley, 2010, 2015; McNamara, 2010). In addition, these foundational skills and strategies cannot be reduced to content knowledge and basic skills in reading and mathematics (Conley, 2015; Darling-Hammond & Adamson, 2010; Darling-Hammond, Herman, & Pellegrino, 2013; Pellegrino & Hilton, 2013). In particular, critical thinking skills are commonly regarded as a necessary condition for deep understanding and learning (Conley 2011, 2015; Darling-Hammond, Wilhoit, & Pittenger, 2014; Farrington, 2013).

Though critical thinking has been defined in various ways (see, e.g., Thomas & Lok, 2015), most researchers agree that a key (or even the main) component of critical thinking is argumentation (e.g., Astleitner, 2002; Dick, 1991; Hatcher, 2011; Ikuenobe, 2001; van Gelder, Bissett, & Cumming, 2004). In fact, some researchers have even suggested that argumentation entails critical thinking (e.g., Davies & Barnett, 2015), given that, as Andrews (2015, p. 50) puts it, “to think clearly is to be critical.”

**CONTACT** Rafael Quintana  [raq5@pitt.edu](mailto:raq5@pitt.edu)  Learning Research and Development Center, University of Pittsburgh, LRDC Rm 821, 3939 O'Hara St., Pittsburgh, PA 15260.

<sup>a</sup>Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

 Supplemental data for this article can be accessed on the [publisher's website](#).

© 2019 Taylor & Francis Group, LLC

Argumentative skills can be considered foundational in the sense that they are required for different tasks in different subjects. Thus, successful undergraduate students tend to be those who can understand and learn from argumentative texts and produce arguments in speech or writing, whatever their discipline (Andrews, 2015). Different studies suggest that argument analysis is indeed required in various disciplines, including science (Driver, Newton, & Osborne, 2000), mathematics (Morsanyi, Devine, Nobes, & Szűcs, 2013), nursing (Simpson & Courtney, 2002), economics (Heijltjes, Van Gog, & Paas, 2014), and history, medicine, and law (Jones, 2015). What this suggests is not that argumentation can somehow substitute disciplinary knowledge but rather that argumentative skills (e.g., stating clear claims; proving reasons that support a claim; evaluating reasons and evidence; identifying counterexamples) are relevant for different tasks across disciplinary contexts.

The foundational nature of argumentative skills is reflected in the weight they have in standardized assessments in higher education. For example, the Collegiate Learning Assessment, used by Arum and Roksa (2011) in their highly cited analysis of students' learning in college, has a major focus on critical thinking abilities. The Collegiate Learning Assessment focuses on these general abilities because "they cut across academic majors and they are mentioned in almost every college's mission statement" (Klein et al., 2007, p. 417). In addition, argumentative skills have a significant weight in standardized tests for graduate or professional admissions like GRE, GMAT, LSAT, and MCAT (Deane, Quinlan, Odendahl, Welsh, & Bivens-Tatum, 2008; Hurley, 2014); for example, approximately 50% of the LSAT relates to questions involving argumentation (Hurley, 2014).

## Teaching Argumentative Skills in College

Universities are increasingly concerned with students' academic underpreparedness. Nationally, it is estimated that only 32% of all students leave high school qualified to attend 4-year colleges (Greene & Foster, 2003). This suggests that many students graduate from school without the foundational skills to succeed in college (ACT, 2013; Conley, 2015). A common strategy that colleges and universities have adopted to improve students' academic performance is implementing "Success Courses" or "First-Year Seminars" (hereafter FYS). In fact, it has been estimated that over 90% of colleges and universities offer such programs (Clark & Cundiff, 2011). A 2006 national survey of FYS (Tobolowsky et al., 2008) indicates that the most frequent topics taught in these courses are study skills (41%) and critical thinking (41%), but they also teach about campus resources (38%), academic planning/advising (37%), and time management (29%).

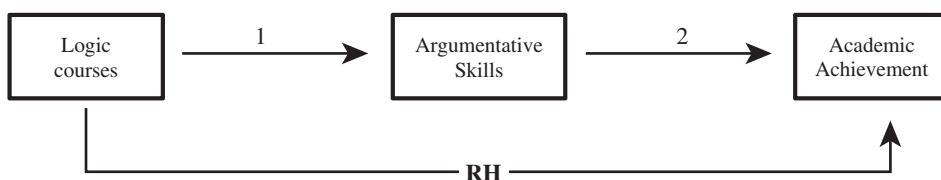
There is scant evidence regarding the effectiveness of FYS (Permazdian & Credé, 2016; Rasmussen, 2013; What Works Clearinghouse, 2016), and some studies have not found any positive effects on student learning outcomes (Clark & Cundiff, 2011; Rasmussen, 2013; Rutschow, Cullinan, & Welbeck, 2012). One reason why FYS represent a very limited solution to students' academic underpreparedness may be that they normally offer a combination of academic and social services in a single course and therefore are not strong enough interventions in academics to affect students' achievement (Rutschow et al., 2012). In particular, it seems unlikely that these courses would have a substantial impact on students' argumentative and critical thinking skills, given

that (a) most faculty have only a vague idea of the concrete skills that these constructs entail (Paul, 2004) and (b) developing argumentative skills—like any other higher-order skill—requires regular, extended, and effortful activity (Ericsson & Charness, 1994; van Gelder, 2005; van Gelder et al., 2004). Consequently, some scholars have argued that critical thinking should be studied and practiced in its own right, because it is composed of particular skills that need to be developed in a structured and deliberate manner (van Gelder, 2005; van Gelder et al., 2004).

Various meta-analyses (Abrami et al., 2008, 2015; Niu, Behar-Horenstein, & Garvan, 2013; Tiruneh, Verburgh, & Elen, 2013) have shown that interventions that teach critical thinking principles explicitly are more effective for improving critical thinking than interventions that do not teach these principles explicitly; for example, interventions that focus on subject-matter instruction and expect to develop critical thinking as a by-product. These findings validate the argument schema theory, which suggests that learning argumentative abilities involves the generation of abstract principles and concepts related to argumentation (Larson, Britt, & Kurby, 2009; Larson, Britt, & Larson, 2004; Reznitskaya & Anderson, 2002; Reznitskaya, Anderson, & Kuo, 2007). More generally, research has shown that the possession of abstract concepts facilitates the transfer of knowledge (Fong & Nisbett, 1991; Goldstone & Son, 2005; Reed, 1993). That is, both general and specialized research suggests that proficiency in argument analysis (at least for nonexpert students) is facilitated by the possession of some general argumentative knowledge, involving understanding general concepts (e.g., *argument*, *claim*, *reason*, and *evidence*) and rules (e.g., the rules of inference). Thus, from a theoretical standpoint, courses focused on argumentation should have broader effects. However, there is relatively little empirical evidence that supports the idea that directly teaching such critical thinking skills has significant effects on other coursework.

### Overview of the Study

The purpose of this study is to evaluate the effect of introductory logic courses on general academic performance in science, technology, engineering, and mathematics (STEM)-related majors. Our theory of action involves two main components (see Figure 1). First, we hypothesize that logic courses have an effect on students' argumentative skills. Argument schema theory and a set of critical thinking interventions suggest that learning and transfer of argumentative abilities involves the generation of abstract principles and concepts related to argumentation. As a result, one can expect that logic courses improve students' argumentation abilities because logic is the



**Figure 1.** A schematic representation of the two components of our theory of action, as well as our main research hypothesis (RH).

discipline dedicated to define, theorize, and teach these abstract concepts. As Hurley (2014) explains, “Logic may be defined as the organized body of knowledge, or science, that evaluates arguments” (p. 1). Though logic courses might vary in their format and tend to focus on formal logic (which is not necessarily related to natural argumentation; see Johnson, 2014), introductory logic courses in college normally begin by teaching the general concepts of argumentation (see, e.g., the classic and commonly used introductory books by Copi, Cohen, & McMahon, 2013; Gensler, 2012; Govier, 2013; Hurley, 2014; Salmon, 2012). Following the argument schema theory, one would then expect that an introductory logic course would improve students’ argumentative abilities.

The second component of our theory of action implies that developing argumentative skills can have a positive effect on students’ general academic achievement. This is supported by (a) various competency models of crosscutting abilities (in particular reading and writing), which consider argumentation a central component (e.g., Deane et al., 2008; Deane, Quinlan, & Kostin, 2011; National Assessment Governing Board, 2015; Organization for Economic Cooperation and Development, 2013); (b) research that highlights the importance of argumentation in different disciplines (e.g., Jones, 2015); (c) the weight that these skills have in standardized tests for graduate and professional admission (e.g., Hurley, 2014); and (d) the widespread belief that critical thinking is a key component for students’ academic success (see, e.g., Ennis, 2015; Lai, 2011; Paul, 2004; ten Dam & Volman, 2004).

Though there is a wide agreement on the importance of argumentation for different learning outcomes, there is no direct evidence demonstrating that learning abstract concepts and principles of argumentation can actually improve students’ academic achievement. Studies that evaluate argumentation (or critical thinking) interventions normally consider outcomes related to argumentation (or critical thinking), rather than general academic performance (see, e.g., Abrami et al., 2008, 2015). Furthermore, many of these studies (e.g., Couzijn & Rijlaarsdam, 2004; Heijltjes, Van Gog, Leppink, & Paas, 2014; Larson et al., 2004, 2009; Marin & Halpern, 2011; Reznitskaya et al., 2007) evaluate short interventions, which commonly do not generate robust learning effects and are more difficult to generalize to real educational settings.

The idea that specific kinds of learning have pervasive effects and enhance general thinking skills (the so-called doctrine of formal discipline) has been a widely contested topic in cognitive psychology (Barnett & Ceci, 2002; Detterman, 1993; Mayer & Wittrock, 1996). An analogous debate is also present in the critical thinking literature, where researchers disagree on whether learning general critical thinking skills can have an impact in learning outcomes in different domains (Abrami et al., 2008; Davies, 2006, 2013). Yet, there is also a paucity of empirical evidence, in particular from experimental or quasi-experimental designs, regarding the effects of learning foundational skills in argumentation and critical thinking in different contexts and domains (van Gelder, 2015). This lack of evidence is especially concerning in view of the steady decline in enrollments and degrees conferred in the humanities (Jaschik, 2016, 2017).

The present study contributes to the literature by examining whether a naturally occurring logic course is associated with an increase in students’ general academic achievement. In addition, we were interested in whether taking the course is more

beneficial for some students than others. Our hypothesis in this regard was that a logic course would be most beneficial for academically weaker students (based on high school performance and entering test scores) who are likely to have poor baseline argumentation abilities and less beneficial for academically stronger students who were already likely to be familiar or proficient with basic argumentation. Finally, we wanted to examine whether there is an advantage in taking the logic course at a particular time; for example, at the beginning (like FYS) or in later years. Our hypothesis was that, given that a logic course provides foundational skills, students would benefit more if they take it early on.

## Method

### *The Studied Introductory Logic Course*

The Philosophy Department at the University of Pittsburgh offers an Introductory Logic Course (PHIL 0500, henceforth “Logic”). Logic is offered every fall and spring semester without any prerequisites, and usually several sections of the course are offered each semester (for example, in spring 2017 there were three section offerings). Different sections in a different semester are typically taught by different instructors, and section sizes can vary from 20 to 120, although the larger sizes are more common. The course is also taught in the summer session. The diversity of semester, section sizes, and instructors improves generalizability of the study.

Features of this particular introductory logic course provide several advantages for testing the effects of acquiring foundational knowledge in logic on academic performance. First, the contents covered in introductory logic courses are highly standardized; in particular, they normally cover the basic concepts of argumentation (e.g., *argument*, *proposition*, *premise*, *conclusion*, *validity*, etc.). Second, at the University of Pittsburgh, it is a course that a large number of students outside the Philosophy Department take to fulfill the Quantitative and Formal Reasoning Requirement. Because it fulfills a general education requirement, diverse students are enrolled and many students may be enrolled for reasons other than interest; consequently, the effects are more likely to generalize beyond only those students highly interested in philosophy or logic. Third, students can take the course whenever they want, and as a general education requirement there is no strong rationale for taking the courses before or after other courses, which generates some natural (albeit not fully random) variation in the semester in which students take the course.

### *Sample*

The University of Pittsburgh is a public, research-focused doctoral university with a student population of approximately 28,000 undergraduates who are predominantly full-time, residential students. The university implements a selection process that is more selective and lower transfer-in, according to the Carnegie Classification of Institutions of Higher Education (n.d.).

The sample consists of 14,941 full-time undergraduate students who were enrolled in the School of Arts and Sciences at the University of Pittsburgh from 2007 to 2017 and

had taken any introductory chemistry course (General Chemistry 1; General Chemistry 2; Organic Chemistry 1; or Organic Chemistry 2) during that time. The students could have taken the chemistry course in different academic terms, and they were followed from their first to their last recorded semester. The first cohort of students were freshman in 2007 and the last in 2017.

This longitudinal multicohort data set comes from a larger project analyzing the academic trajectories of students who had enrolled in any introductory chemistry course, a subset of which is required for most science majors and all students pursuing health-related careers. Thus, the students in this sample are more likely to pursue health or STEM-related careers than careers in social sciences or humanities. More specifically, based on surveys conducted by the larger research project, 37% of the students were interested in pursuing a health-related major and 30% some STEM major; 12% were interested in pursuing a major in social sciences or humanities; 3% were attracted to other fields; and around 18% were undecided.

The advantage of considering this sample (rather than all full-time enrollees) is that they are less likely to enroll in courses in social sciences or humanities, some of which overlap with the contents covered in Logic (e.g., courses with an emphasis in reading or writing argumentative texts). By focusing on students pursuing STEM or health-related majors we can assume, then, that the students who enrolled in Logic had had minimal exposure to this kind of training before and that the students not enrolled in Logic had little exposure to the same content in another course. In other words, we reduce the possibility of possible confounders, which would help us obtain a better estimate of the effect of Logic on student learning outcomes.

Within the sample, 758 students (approximately 5% of the full sample) took Logic, and only 28 students (i.e., less than 4% of course enrollees) took it more than once. We restricted the sample to include only students with a passing grade in Logic (approximately 95% of the course enrollees). Our theory of action implies that Logic has an effect on student learning outcomes by imparting particular skills, and it is likely that the students who failed the course or withdrew from the course in the middle of a semester did not acquire (or acquired in a very limited or inadequate fashion) those skills. In addition, we excluded the 131 students who took Logic in their first semester, because they did not have a prior GPA and therefore the effect of Logic cannot be estimated (see details below). The majority of students (around 70%) took Logic in the first 2 years: 17% in the first semester; 19% in the second; 18% in the third; and 16% in the fourth. After these reductions, the final sample of students who took Logic (i.e., the “treated” sample) was 588.

Table 1 displays descriptive statistics for the students who took Logic and for the students who did not. The two samples have a similar proportion of Hispanic and African American students. However, only 42% of students who took Logic are female, compared to 58% in the entire population. The socioeconomic characteristics of the two samples are very similar. We also perceive an interesting pattern regarding the academic background of these two samples: though students who took Logic have, on average, higher verbal, writing, and mathematics SAT scores, they also have lower high school GPAs.

**Table 1.** Descriptive statistics for students who enrolled in Logic and for students who did not.

Stable characteristics	Took logic ( <i>N</i> = 588)		Did not take logic ( <i>N</i> = 14,183)		Logic – None difference ( <i>SD</i> )
	Mean	<i>SD</i>	Mean	<i>SD</i>	
<b>Demographic</b>					
Female	0.42	0.49	0.58	0.49	–0.33
Age at enrollment	19.42	0.84	19.52	1.27	–0.10
Black (%)	0.09	0.29	0.8	0.27	0.06
Hispanic (%)	0.03	0.16	0.03	0.18	–0.04
<b>Academic</b>					
SAT Mathematics (standardized)	0.03	0.98	–0.01	1.00	0.04
SAT Writing (standardized)	0.11	1.02	–0.01	1.00	0.12
SAT Verbal (standardized)	0.12	0.98	–0.01	1.00	0.14
High school GPA	3.86	0.46	3.94	0.49	–0.16
<b>Socioeconomic</b>					
Average gross income (ln)	11.46	0.85	11.47	0.89	0.00
First generation (%)	0.12	0.33	0.13	0.34	–0.02
Minority in school (%)	12.17	15.19	12.77	16.59	–0.04
Free lunch in school (%)	23.97	17.68	24.10	17.27	–0.01

## Measures

### Demographics

The student records provided by the University of Pittsburgh contain a wide range of demographic and academic information about students. Demographic covariates include gender, ethnicity, date of birth, family’s adjusted gross income, parental education, and high school identification number. We used this high school identification number to retrieve information from the Department of Education’s Common Core of Data (National Center for Education Statistics, 2014) regarding the high school, in particular the percentage of African American and Latino students in the school and the percentage of students eligible for free or reduced-price lunches. These school-level variables serve as proxies for students’ likely access to quality education (e.g., Bohrnstedt, Kitmitto, Ogut, Sherman, & Chan, 2015; Fryer & Levitt, 2004; Kao & Thompson, 2003; Reardon & Galindo, 2009), such as Advanced Placement courses, which include some argumentative writing.

### Prior Academic Achievement

The student records included several academic achievement covariates, such as the students’ scores in the three main components of the SAT (verbal, writing, and mathematics) used by many U.S. universities to make admittance decisions to undergraduate programs, as well as high school GPA. For the Aptitude  $\times$  Treatment interaction analyses, we generated a standardized indicator of students’ baseline academic achievement by averaging the students’ standardized average SAT scores in reading, mathematics, and writing with their standardized high school GPAs. We used this composite measure as a comprehensive indicator of students’ baseline academic performance.

### Treatment

The students’ academic records included which courses were taken, in which semester, and which grades were obtained (as well as the number of credits associated with each class). We used this information to generate our main time-varying predictor, which is coded as 0 for students who never took Logic and changes from 0 to 1 for



students who took Logic in the semester they enrolled in the course. That is, we are interested in examining the effect of logic both during and after the students take the course.

#### *GPA as an Outcome*

We also used the academic record information to calculate the students' GPAs in every semester (the outcome of interest). We estimated the effect of Logic on students' GPAs in each semester, weighted by the number of credits of each class. The GPA is based on a 5-point scale (A = 4, B = 3, C = 2, D = 1, F = 0) and follows a censored normal distribution with a mean of 3.2 and standard deviation of 0.7. The GPA was calculated without the grade that students earned in the logic course.

### **Research Question**

As explained above, argumentative skills can be considered foundational, given that they are required by different disciplines and are a component of key crosscutting skills such as reading and writing. As a consequence, it is reasonable to assume that the Introduction to Logic course has an impact on a variety of courses with at least some emphasis on argument analysis, argumentative reading, and argumentative writing. Considering GPA as an outcome entails two complications: first, it is a combination of different classes, some of which might not be influenced by Logic, and, second, students can choose among courses that vary in difficulty. We dealt with the latter problem by replicating the results using controls for course difficulty.

At the same time, considering GPA as an outcome provides two main advantages: first, by averaging multiple grades, it represents a more valid and reliable indicator of academic performance than the grade of particular courses and, second, given that students have a GPA in every semester, we can identify entire academic trajectories and examine whether Logic has any effect in these trajectories. Furthermore, considering GPA as an outcome is consistent with our main research hypothesis, namely, that Logic teaches foundational skills in argumentation and that, as a consequence, they have an impact on a wide variety of courses. Our main research question is, then, whether taking a Logic course has an effect on STEM-related students' academic achievement, as measured by their GPAs.

### **Identification Strategy**

According to the potential outcome framework formalized by Rubin (1974), in order to estimate treatment effects, one needs to estimate the counterfactual states for particular units. Given that the counterfactual states are, by definition, unobservable, one needs to infer these counterfactuals by imposing particular assumptions (West & Thoemmes, 2010). One of the main assumptions for drawing valid causal inferences is statistical independence in potential outcomes, which implies that by knowing a unit's treatment status one cannot infer anything about its potential outcome. This condition is hardly met in observational studies and nonequivalent control group designs, which are characterized by self-selection into treatment and control conditions and therefore pretreatment differences in observed and unobserved covariates.

The bias resulting from differential selection can be solved by conditioning treatment selection on all confounding covariates. In this case, one can assume conditional independence and strong ignorability (Steiner, Cook, Shadish, & Clark, 2010). Given that in our case the selection mechanism is not well understood, it is difficult to utilize a conditioning strategy to eliminate selection bias. That is, if we want to estimate the treatment effect we would not be able to assess whether strong ignorability holds, even if we control for a rich set of covariates. However, the longitudinal structure of the data allows us to find valid approximations of the true counterfactual under less restrictive assumptions. In this section, we describe two strategies that we used to identify the treatment effect of taking a Logic course. First, we used a difference-in-differences (DiD) approach to compare the changes in GPAs of the students who took Logic (henceforth the treated sample) with the changes in GPAs of the students who did not take Logic (henceforth the control sample). Second, we took advantage of the time-varying treatment structure, and in an individual fixed-effects framework we used students in the treated sample as their own controls. These two strategies are described below.

### Empirical Specifications

Using the entire sample of students, we used a DiD approach to identify the treatment effect of taking a logic course. Following Morgan and Winship (2014), we distinguished between two different treatment indicators: a time-invariant dummy variable ( $D_i^*$ ) indicating whether a student ever enrolled in Logic at any point in the time span under study and a time-varying dummy variable ( $D_{it}$ ) indicating whether individual  $i$  enrolled in Logic in semester  $t$ . Given that we are interested in estimating the effect of treatment exposure,  $D_{it}$ , we need to control for the selection bias related to treatment status,  $D_i^*$ . The basic idea behind the DiD strategy is to examine whether the *changes* in GPAs of the treated sample after taking Logic are significantly different from the *changes* in GPAs of the control sample in the same period. As Morgan and Winship (2014) explain, in this strategy one uses the average observed values in the control group in the posttreatment time periods in order to predict the average counterfactual values in the treatment group in the same time periods. More specifically, we implemented a group-level panel regression analysis of the following form:

$$Y_{it} = \mu + \tau D_i^* + \sum_{t=1}^T (D_i^* \times d_{it}) \delta_{1t} + \alpha D_{it} + X_i \Gamma + \varepsilon_{it} \quad (1)$$

where  $Y_{it}$  represents the GPA earned by student  $i$  in semester  $t$ ,  $\tau$  represents a fixed parameter measuring unobserved and time-invariant differences between the treatment and control groups,  $\delta_{1t}$  represents group-specific time trends,  $\alpha$  represents the average treatment effect,  $X_i$  are baseline covariates, and  $\mu$  is an intercept. Given that in this strategy it is crucial to model effectively the group-specific differences in GPA across semesters, we used a full nonparametric approach to measure time trends, rather than relying on functional form assumptions.

The key identifying assumption in this approach is that the changes in GPA would be the same in both groups in the absence of treatment. This assumption would be incorrect if there are omitted group-specific and time-varying effects correlated with

treatment status. In order to test this assumption, we performed a second analysis using only the treated sample. More specifically, we took advantage of the fact that students in the treated sample enrolled in Logic in different semesters, so we could use the average observed values of  $Y_{it}$  for the “late enrollees” to predict the average counterfactual values of  $Y_{it}$  for the “early enrollees.” For this purpose, we implemented a model including both semester- and person-specific fixed effects, which can be written as

$$Y_{it} = \mu + \delta_t + \tau_i + \alpha D_{it} + \varepsilon_{it} \quad (2)$$

where  $\delta_t$  represents semester-specific time trends,  $\tau$  represents a fixed parameter measuring time-invariant person-specific unobservables,  $\alpha$  represents the average treatment effect, and  $\mu$  is an intercept.

This empirical specification eliminates potential confounding variables related to treatment group membership, unobserved and stable person-specific characteristics, and common time trends. However, this approach does not remove potential bias related to time-varying confounds. In particular, one might worry about potential dynamic processes, where students’ selection into treatment might be related to their previous GPAs. In order to test this possibility, one can include the lag value of  $Y_{it}$ ,  $Y_{i,t-1}$ , in Equation 2. Incorporating the lagged dependent variable can be also justified by the fact that students’ academic achievement in semester  $t$  might be determined by the students’ achievement in semester  $t - 1$ . That is, including  $Y_{i,t-1}$  can be a good way of controlling for temporal dependencies. However, as is well known in the econometric literature, including the lagged dependent variable in a fixed effects model generates biased coefficients, because  $Y_{i,t-1}$  is related to the average value of the error term (see, e.g., Angrist & Pischke, 2008). Consequently, in order to include  $Y_{i,t-1}$  while controlling for individual unobserved and stable characteristics, we took advantage of the multiple waves in the data set and used the Arellano-Bond estimator, which uses lagged values and lagged differences of the dependent variable as instruments (see Arellano & Bond, 1991).

## Results

### *Main Effects of Logic*

Table 2 displays the main results of our two analytic strategies.<sup>1</sup> Columns 1 and 2 present the estimates using the whole sample and a DiD approach. By comparing these columns, one can perceive that the estimated effect of Logic decreases from 0.08 to 0.07 when the stable covariates are included. The results also indicate that, on average, the GPA of the students in the treatment group is around 0.27 points higher than the GPA of the students in the control group, even after controlling for baseline achievement and other invariant characteristics.

The DiD strategy relies on the parallel trend assumption, according to which the outcome variable follows the same time trend in the control group and the counterfactual states of the treatment group. Though this assumption cannot be empirically demonstrated, a common way of testing it is by comparing the time trends of the treatment and control groups before treatment (Angrist & Krueger, 1999). If the assumption holds,

---

<sup>1</sup>Full regression results are included in the Supplemental Material.

**Table 2.** Estimated effect of taking a Logic course on academic achievement.

	Whole sample DiD		Logic takers two-way fixed effects	
	1. No covariates	2. Covariates included	3. No lag effect	4. Lag included
Logic effect	0.080** (0.029)	0.071* (0.028)	0.072* (0.030)	0.072 (0.039)
Treatment group	0.225*** (0.046)	0.268*** (0.045)		
Lag GPA				0.074 (0.151)
Control variables				
Semester dummies	Yes	Yes	Yes	Yes
Stable covariates	No	Yes	No	No
R <sup>2</sup>	0.04	0.17	0.07	0.04
Number of students	14,771	14,771	588	580
Number of semesters	79,852	79,852	4,012	2,767
Average number of semesters	5.4	5.4	6.8	6.8

*Note.* Columns 1 and 2 were estimated using a DiD strategy, and columns 3 and 4 were estimated using an individual two-way fixed effects framework. Given that one cannot estimate the effect of stable characteristics in a within estimator, stable covariates were not included in the subsample analysis. The  $R^2$  in columns 1 and 2 represents the overall  $R^2$ , and in columns 3 and 4 we display the within  $R^2$ . We report both the total number of observed semesters, as well as the average number of semesters per student. The standard errors were constructed to allow for heteroskedasticity and clustering at the student level.

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

then one should not find any pretreatment trend differentials between the treatment and control groups (i.e., the difference between the groups should remain constant over time). In order to assess this assumption, we included a series of pretreatment indicators (or “leads” of the treatment) in Equation 1. The first lead, for example, captures the difference in outcomes between the treatment and control groups one semester before treatment. As expected, we found no evidence of pretreatment trends,  $\chi^2(6) = 5.56$ ,  $p = .475$ . As Table 2A in the Supplemental Material shows, the coefficients of each lead term were far from being statistically significant. This result supports our DiD strategy, because there is no evidence of any trend that is specific to the treated sample and that could explain the estimates displayed in columns 1 and 2.

Columns 3 and 4 in Table 2 present the estimates using only the treated sample in a two-way fixed effects framework. The estimated effect of Logic in these two models is consistent with the results obtained using the whole sample. Furthermore, these coefficients remained stable when we varied the model specifications; for example, when we added individual slopes (the results of this model are included in the Supplemental Material). Column 4 presents the results after including the lag value of the dependent variable using the Arellano-Bond estimator.<sup>2</sup> One can perceive that the coefficient of the lag GPA is not significant, and its inclusion in the model has a negligible effect on the main predictor (even though it pushes its  $p$ -value slightly above the .05 significance level). What is important, however, is that we do not observe any sign of dynamic selection or time-varying confounds that are related to the lag GPA.

<sup>2</sup>The model met the two main assumptions of this procedure (see Arellano & Bond, 1991). First, there was no second-order autocorrelation among the residuals, because the null hypothesis of no autocorrelation was not rejected ( $z = 0.754$ ,  $p = .451$ ). Second, we could not reject the hypothesis according to which the overidentifying restrictions were valid,  $\chi^2(20) = 22.45$ ,  $p = .317$ . That is, the instruments used in the procedure appear to be truly exogenous, because they are not related to the second-stage residuals.

**Table 3.** Estimated effect of taking a Logic course controlling for course difficulty.

	GPA		Standardized GPA	
	1. DiD	2. Fixed effects	3. DiD	4. Fixed effects
Logic effect	0.134*** (0.026)	0.143*** (0.028)	0.109*** (0.023)	0.117*** (0.025)
Treatment group	-0.113** (0.042)		-0.253*** (0.035)	
Average GPA	0.782*** (0.011)	0.776*** (0.046)		
Control variables				
Semester dummies	Yes	Yes	Yes	Yes
Stable covariates	Yes	No	Yes	No
R <sup>2</sup>	0.29	0.19	0.16	0.02
Number of students	14,771	588	14,771	588

Note. The dependent variable in columns 1 and 2 is an unstandardized GPA, and in columns 3 and 4 the dependent variable is a standardized GPA. Columns 1 and 3 were estimated using the whole sample and a DiD strategy, and columns 2 and 4 were estimated using a subsample and an individual two-way fixed effects framework. The standard errors were constructed to allow for heteroskedasticity and clustering at the student level.

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

This set of results provides a robust estimate of the effect of Logic on GPA of around 0.07 points, which is equivalent to 10% of a standard deviation in the overall distribution. Borrowing from Donato and Thomas’s (2017) discussion of effect sizes, we can say that 0.07 points is approximately one third as large as the difference between B and B+ (0.25 grade points).

### Controlling for Course Leniency

One particular threat to validity that we have not addressed directly is related to course choice behavior and the fact that students can selectively choose leniently or harshly graded courses (Donato & Thomas, 2017). This would represent a threat only if course leniency is correlated with treatment status (in the case of our group-level strategy), as well as to treatment exposure. In order to examine this potential threat, we tried two different strategies. First, we included in Equations 1 and 2 a time-varying and person-specific measure of course leniency,  $\gamma_{it}$ , representing the average GPA of the classes that student  $i$  took in semester  $t$ . For example, if  $\gamma_{it}$  equals a mean course grade of 2, then the student took “difficult” classes in that semester, whereas if  $\gamma_{it}$  equals 4, the student enrolled in “easy” classes. Column 1 in Table 3 displays the results of this strategy using Equation 1 (i.e., the group-based approach), and column 2 displays the results using Equation 2 (i.e., the individual-based approach). In both cases, the estimated effects of taking Logic grow to be approximately twice as large (0.134 and 0.143, respectively).

Our second strategy consisted of predicting a standardized GPA, based on the mean and standard deviation of each particular course. Column 3 in Table 3 presents the results of this strategy using Equation 1, and column 4 presents the results using Equation 2. Again, in both cases the estimated effects of taking Logic become slightly larger (0.109 and 0.117, respectively) compared to our previous standardized results ( $SD = 0.10$ ). Given that the estimated effects get larger when one adjusts for course difficulty or teacher grading, one can conclude that the initial coefficients might be biased downward.

**Table 4.** Estimated timing effects of taking a Logic course.

	Semester 2	Semester 3	Semester 4	Semester 5	Semester 6	Semester 7	Semester 8
Logic effect	0.352*** (0.079)	0.302*** (0.084)	0.297*** (0.077)	0.055 (0.091)	0.071 (0.075)	0.195 (0.168)	0.154 (0.143)
Treatment group	0.013 (0.061)	0.009 (0.056)	0.144* (0.057)	0.158* (0.072)	0.288*** (0.075)	0.131 (0.136)	0.255 (0.147)
$R^2$	0.18	0.18	0.18	0.18	0.18	0.18	0.18
Total number of students	14,324	14,311	14,296	14,277	14,248	14,211	14,202
Number of students in the treatment group	141	128	113	94	65	28	19

Note. All models were estimated using the DiD strategy and include semester fixed effects and stable covariates. Each model was estimated with a different treatment group, associated with the semester in which the students took the logic course. The standard errors were constructed to allow for heteroskedasticity and clustering at the student level. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

### Timing Effects of Taking Logic

In addition to identifying the average effect of Logic, we were interested in examining potential “timing effects”; that is, differential effects of taking the course in different semesters. In order to examine this issue, we applied the DiD strategy described by Equation 1 for each subgroup of students separately (i.e., for the students who took Logic in the second semester, in the third semester, and so forth). We defined seven treatment groups, associated with the seven semesters in which students could have enrolled in the course (starting in the second semester). It is worth noting that in each of these subsamples all students took Logic at the same time and, as a consequence, we could not use the specification described by Equation 2, because in that strategy we exploit variations in timing between students in the treatment group.

Table 4 presents the estimates of taking Logic for the seven different treatment groups using the DiD approach. One can perceive that the effect of Logic is positive for all treatment groups, but it is only statistically significant for the students who took the course in the first three semesters. In addition, as Figure 2 illustrates, the effect is higher in the initial semesters and decreases over time. The effect in the first semesters is very high; in particular, the difference in GPA for the students who took Logic in the second semester (0.35 grade points) represents around 50% of a standard deviation, which is more than the difference between B and B+ (0.25 points). The estimated effect for the students who took the course in the third semester is equivalent to 43% of a standard deviation, and the effect for those who took in their fourth semester is equivalent to 43% of a standard deviation.

### Heterogeneous Effects by Baseline Academic Achievement and Subject-Matter GPAs

We were also interested in examining potential heterogeneity in the effects of the course conditional on the students’ baseline academic achievement. In order to examine this issue, we divided the population in quartiles based on the students’ composite standardized academic achievement score and then computed the effect of taking Logic for

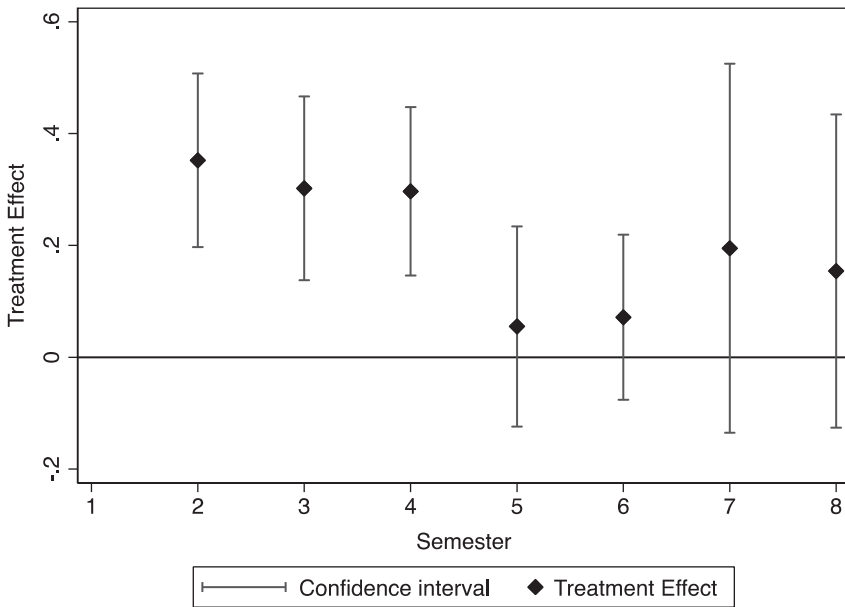


Figure 2. Estimated timing effects of taking Logic.

Table 5. Estimated effect of Logic conditional on students' baseline academic achievement.

Quartile	Number of students		Predicted achievement		Treatment effect		
	Control	Treatment	Mean	SE	Estimate	$\chi^2$	p-Value
1	3,532	160	2.83	0.010	0.129	13.77	.000
2	3,493	141	3.05	0.009	-0.008	0.03	.855
3	3,621	142	3.20	0.009	-0.038	0.80	.371
4	3,537	145	3.47	0.008	0.081	5.07	.024

Note. *p*-Values indicate the probabilities that the treatment and control groups are equal. The predicted achievement and the treatment effects were estimated using the entire sample ( $N=14,771$ ) and the DiD strategy described by Equation 1. All of the baseline covariates were included except for high school GPA and average SAT to avoid collinearity. The standard errors were constructed to allow for heteroskedasticity and clustering at the student level.

each of these groups, using our DiD strategy. In other words, we estimated the effect of treatment exposure for four different treatment groups, using the corresponding control group to predict their respective counterfactual. As in our main specification described by Equation 1, the four achievement groups had a fixed parameter measuring unobserved differences, and the eight subgroups (created by the treatment status and achievement combination) had a specific nonparametric time trend.

Table 5 presents the number of students in each group as well the average predicted GPA in each quartile, holding constant the other covariate values at their analytic sample means. The results indicate that the treatment effect is only significant in both ends of the achievement distribution, with an estimated effect of 0.129 ( $p < .001$ ) for the lowest quartile and 0.081 ( $p < .05$ ) for the highest quartile. Thus, taking a logic course is especially beneficial for students with low academic achievement, with an effect of 18% of a standard deviation, which is equivalent to around half as large as the difference between B and B+.

**Table 6.** Effects of taking Logic on subject-matter GPAs.

Sample	Math GPA		Social sciences GPA		Natural sciences GPA	
	Treatment effect	Average number of GPAs per student	Treatment effect	Average number of GPAs per student	Treatment effect	Average number of GPAs per student
Q1	0.192 (0.134)	2.2	-0.054 (0.079)	3.1	0.062 (0.089)	3.3
Q2	0.126 (0.135)	2.1	-0.116 (0.088)	3.0	0.124 (0.085)	3.6
Q3	-0.125 (0.123)	1.9	-0.06 (0.092)	2.8	-0.092 (0.078)	3.9
Q4	0.303** (0.116)	1.8	0.035 (0.088)	2.7	0.087 (0.073)	4.3
Whole sample	0.133* (0.068)	2.0	-0.034 (0.045)	2.9	0.054 (0.041)	3.8

Note. All models were estimated using the DiD strategy and include semester fixed effects and stable covariates. The standard errors were constructed to allow for heteroskedasticity and clustering at the student level.

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

Though we hypothesized that the students with low baseline academic achievement would benefit from taking a logic course, we were not expecting a significant effect among the students with high baseline academic achievement. In order to understand why this occurs, we examined the extent to which taking Logic affects the students' performance in particular courses. For this purpose, we created three different GPAs, based on broad disciplinary categories: a math GPA, which included all classes in mathematics and statistics<sup>3</sup>; a natural sciences GPA, which included classes in biology, chemistry, neuroscience, and physics; and a social sciences GPA, which included classes in economics, sociology, history, psychology, and political science. Given that the threats related to course leniency increase when one considers the students' performance in specific courses, we standardized these GPAs based on the mean and standard deviation of each particular course.

Using our DiD strategy, we computed the effect of taking Logic on the different GPAs for each of the four achievement groups as well as the whole sample. As Table 6 indicates, the point estimates were not precisely estimated, because in general they have relatively large standard errors. This might be related to a variety of factors, notably a smaller sample size and differences in the distribution of the courses used to compute the different GPAs. The average number of GPAs by student ranges from 1.8 to 4.3 (compared to the average number of 5.4 semesters used in our main analysis). In addition, the observed GPAs of different students might correspond to different semesters.

With these caveats in mind, however, one can perceive that taking Logic is associated with a statistically significant increase in performance in the math GPA for the whole sample (0.133), as well as for the students in the highest quartile of baseline academic achievement (0.303). In addition, the point estimates associated to the natural sciences GPA tend to be higher than those related to the social sciences GPA, but none of these effects are statistically significant. These results suggest that taking Logic helps high-achieving students improve their performance in math courses. At the same time, we do

<sup>3</sup>The subject-matter classifications are included in the transcripts provided by the university.



**Table 7.** Estimated effect of taking a Music course on academic achievement.

	1. Whole sample DiD	2. Music takers fixed effects
Music effect	0.017 (0.015)	0.028 (0.016)
Treatment group	0.357*** (0.025)	
Control variables		
Semester dummies	Yes	Yes
Stable covariates	Yes	No
$R^2$	0.17	0.07
Number of students	14,677	1,525
Number of semesters	79,420	10,607
Average number of semesters	5.4	7.0

Note. Column 1 was estimated using a DiD, and column 2 was estimated using an individual two-way fixed effects framework.

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

not find evidence suggesting that the skills acquired in Logic transfer to natural or social sciences courses.

### Model Testing

In order to increase confidence in the estimated effects and assess our empirical specifications, we conducted a placebo test by estimating the effect of a course that we considered would not have a direct effect on student achievement. For this purpose, we chose the course Introduction to World Music (hereafter “Music”), which can provide important knowledge and skills but is not intended to cover foundational academic skills. In addition, this course provides several methodological advantages (similar to the characteristics of the Logic course): (a) a large number of students in our sample took the course ( $n = 1,525$ ); (b) the course belongs to a non-STEM major (so we can expect that the control group did not acquire the knowledge and skills in other courses); and (c) students enrolled in the course in different semesters (with a maximum of 381 in the second semester and a minimum of 69 in the eighth semester).

As with Logic, we excluded the students who took the Music course in their first semester, as well as the students who had a failing grade in the course. In order to estimate the effect of Music, we followed the same procedure described above for estimating the Logic course effects. Columns 1 and 2 in Table 7 display the estimated effects using our two main empirical specifications. One can perceive that, according to our hypothesis, the effect of taking a music course on academic achievement is not significantly different from zero, even though the treatment group is different from the nontreatment group in overall GPA. Further, standard errors are small enough to establish that any effect of Music would likely be much smaller than the estimated effect of Logic. This placebo test contributes, then, to a validity argument in support of our empirical specifications and the conclusions we derived from them.

### General Discussion

Argumentative abilities are widely considered foundational skills that are key components of academic success across modalities and disciplines. Consequently, an important question that arises is whether argumentative abilities can be explicitly taught and have

an effect on students' general academic success. Multiple meta-analyses on argument and critical thinking interventions, as well as a range of studies in the learning sciences, suggest that teaching abstract concepts and rules in argumentation can help students acquire argumentative abilities and facilitate the application of this knowledge in different contexts. However, there is no evidence suggesting that explicit instruction in argumentation can improve students' academic achievement, because most studies either consider laboratory-based interventions with limited scope and generalizability or do not examine transfer effects.

This study provides evidence that taking a naturally occurring Introductory Logic course is associated with a positive increase in students' academic achievement. In order to estimate treatment effects, we used two different strategies and analytics samples: a DiD approach using the whole sample and an individual fixed effects approach using only the treated sample. The two analyses provided highly consistent results, which strengthens confidence in the estimated effects. We also successfully conducted a placebo test in order to assess our empirical specifications.

The estimated overall effect of the logic courses on students' GPA is 0.07 ( $d=0.10$ ), which can be considered a small effect by traditional standards (Ellis, 2010). In order to explore effect heterogeneity, we divided the sample into four groups according to their composite measure of academic achievement and calculated the effect of taking Logic for each of these groups. For the lowest quartile, taking Logic is associated with an increase of 0.13 points in GPA ( $d=0.18$ ). Taking Logic is also associated with a 0.08 positive difference in GPA ( $d=0.12$ ) for students in the highest quartile. However, taking Logic is not associated with a significant difference in GPA for students in the two middle quartiles of baseline academic achievement. We further explored the heterogeneity in the effects and found that the effect is significant only for the students who took the course in the second, third, and fourth semesters ( $d=0.50$ ,  $d=0.43$ , and  $d=0.42$ , respectively). These effects, corresponding to 0.35, 0.30, and 0.29 grade points, respectively, represent more than the difference between B and B+ (0.25 points).

To put these effect sizes into context, we can compare them with previous findings regarding the effectiveness of FYS, another kind of foundational skills intervention for university students taught through a single course. According to a recent meta-analysis (Permzadian & Credé, 2016), the overall effectiveness of FYS on the first-year GPA is very small ( $d=0.02$ ), and it varies conditional on a series of subgroup moderators; for example, whether they provide academic content ( $d=0.09$ ) or target academically underprepared students ( $d=0.02$ ). Thus, if one conceives Introductory Logic courses as foundational courses that should be taken early, especially by academically underprepared students, our results suggest that these courses are more effective than FYS (given that our effect sizes with similar moderators range from 0.18 to 0.50).

In order to better understand the heterogeneous effects by baseline academic achievement, we examined the extent to which taking Logic affects students' performance in particular courses. For this purpose, we used the DiD strategy to estimate the effects of taking Logic on students' GPAs in mathematics, natural sciences, and social sciences independently. Though these estimates were less precise (due, among other things, to differences between the distribution of the courses used to calculate the different GPAs),

the results suggest that taking Logic is beneficial for the students' performance in mathematics and in particular for students with high baseline academic achievement.

This result is not entirely surprising, given that mathematical abilities are strongly related to logical reasoning skills (Ayalon & Even, 2008; Morsanyi et al., 2013). Furthermore, high-achieving students might benefit more from logic courses, because rigorous logical proof is especially relevant in advanced work in mathematics. At the same time, the absence of significant transfer effects to social sciences and natural sciences courses reinforces the idea that, by focusing on formal languages, common logic courses might not necessarily influence natural argumentation (which is mostly composed by informal arguments; see, e.g., Blair, 2015). The limitations of a purely formal approach to argumentation (and reasoning in general) have been pointed out by some authors (e.g., Johnson-Laird, 2010; Johnson-Laird, Khemlani, & Goodwin, 2015), and several studies suggest that individuals are more likely to apply relevant knowledge in new situations if the information encoded contains both abstract and concrete representations (e.g., Duschl & Osborne, 2002; Goldstone & Son, 2005).

Notwithstanding the limitations of formal instruction in argumentation, our findings suggest that some students benefit from a foundational course in logic. Consistent with our research hypotheses, taking a logic course is especially beneficial for students who take the course early on and for students with low baseline academic achievement. From a practical standpoint, our results suggest that higher education institutions should consider ways of gauging STEM students' basic argumentative and logical skills and, based on this information, help these students make adequate course choices. At the same time, further research is needed to examine more effective ways of teaching basic argumentative skills that can transfer to different modalities and domains.

### **Limitations**

It is worth mentioning that though we estimated the effect of Logic using multiple methods and obtained similar results, we should be cautious about drawing strong causal inferences from observational data. The study used a variety of strategies for improving the strength of the claims: longitudinal rather than cross-sectional data; fixed parameters measuring unobserved and time-invariant differences between treatment and control units at the group or individual level; inclusion of many performance-related covariates; and a context in which a logic course is taken for a variety of reasons and at different time points. However, other causal method designs should also be conducted to verify the current findings and eliminate potential sources of bias that may not have been fully controlled within our analyses.

Compared to other laboratory-based research or curricular interventions, the present study contributes to the literature by examining the effects of naturally occurring courses on students' performance in other courses. However, a limitation of this approach is that we cannot fully test the theory of action behind our research hypothesis. That is, we cannot categorically affirm that, as our theory of action suggests, students (and in particular students with low baseline academic achievement) improve their performance in other subjects by acquiring abstract knowledge related to argumentative reasoning. Though this is a plausible explanation given the extant literature in

argumentation as a generic skill, the specific mechanisms through which logic courses affect academic performance remain vaguely delineated. In order to examine these mechanisms in more detail, one could randomly assign students to specific kinds of instruction in argumentation and measure learning gains using appropriate measures of argument performance. Ideally, these measures would also assess argumentative skills in different modalities and embedded in different disciplinary contexts, so we can obtain more precise estimates of transfer effects.

Separately, given that the majority of students in our sample were pursuing STEM- or health-related majors at a strong research university, replications are needed to examine to what extent these results are generalizable to other kinds of students (e.g., social sciences or humanities students or students with even lower academic preparation). Based on claims in the literature that argumentation is broadly useful, similar benefits should also be expected in students in the social sciences or humanities. However, given that according to our results logic courses have larger effects on student performance in mathematics, it is unclear to what extent students from different disciplines would benefit from these courses.

## Conclusion

The findings of the present study support the idea that educational institutions should provide students with the foundational skills necessary for learning material at a deep level. This contradicts the “learning by consumption” model of education; that is, the “overwhelming assumption in our educational system that the most important thing to *deliver* to students is content” (Mcnamara, 2010, p. 341). Now, though many researchers have stressed the importance of teaching foundational skills that can be transferred across disciplines and help students learn content at a deep level, there is no agreement regarding how to define or identify these general skills (Rychen & Salganik, 2003; Weinert, 2001). The available research evidence regarding the effects of acquiring particular skills on future outcomes is mostly correlational, and many studies investigate competencies that are not clearly defined (Pellegrino & Hilton, 2013). In addition, most studies examine learning outcomes within particular subjects, providing little evidence regarding the extent to which developing generic knowledge and skills can be transferred across contexts and disciplines (Pellegrino & Hilton, 2013).

In this study, we tackled these issues by using a quasi-experimental design to examine how the performance in actual courses is affected by taking a naturally occurring course that teaches a well-defined set of skills. Our findings suggest that learning the basic concepts and rules of argumentation, as taught in an introductory logic course, can improve students’ academic achievement. Our findings also support the hypotheses that this kind of instruction is especially beneficial for students with a low baseline academic achievement. All of these results are consistent with the idea that argumentative skills prepare students for future learning.

At the same time, the present study highlights the need for rigorous experimental or quasi-experimental investigations on the effects of teaching particular skills on students’ academic performance. The majority of studies related to postsecondary success evaluate program participation rather than the acquisition of particular skills, and these studies

assess the impact on grades or retention rather than well-defined learning outcomes. However, as Morgan and Winship (2014) note, a causal effect is properly defined only when all of its nested mechanisms have been correctly identified. In the case of academic success, this might entail opening up the black box related to the foundational skills that enable durable and flexible learning.

## Acknowledgments

The authors gratefully acknowledge very helpful comments from Laura Betancur, Rip Correnti, Steven Finkel and Lindsay Page.

## ARTICLE HISTORY

Received 23 September 2017

Revised 4 July 2018

Accepted 26 October 2018

## References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85*(2), 275–314. doi:10.3102/0034654314551063
- Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102–1134. doi:10.3102/0034654308326084
- ACT. (2013). The reality of college readiness 2013: National. Retrieved from <http://www.act.org/content/dam/act/unsecured/documents/Reality-of-College-Readiness-2013.pdf>
- Andrews, R. (2015). Critical thinking and/or argumentation in higher education. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 49–63). New York, NY: Palgrave MacMillan.
- Angrist, J., & Krueger, A. (1999). Empirical strategies in labor economics. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3A, pp. 1277–1366). Amsterdam: North-Holland.
- Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics*. Princeton, NJ: Princeton University Press.
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies, 58*(2), 277–297. doi:10.2307/2297968
- Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.
- Astleitner, H. (2002). Teaching critical thinking. *Journal of Instructional Psychology, 29*(2), 53–77.
- Ayalon, M., & Even, R. (2008). Deductive reasoning: In the eye of the beholder. *Educational Studies in Mathematics, 69*(3), 235–247. doi:10.1007/s10649-008-9136-2
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. doi:10.1037//0033-2909.128.4.612
- Blair, A. (2015). What is informal logic? In F. van Eemeren & B. Garssen (Eds.), *Reflections on theoretical issues in argumentation* (pp. 27–43). Dordrecht: Springer.
- Bohrstedt, G., Kitmitto, S., Ogut, B., Sherman, D., & Chan, D. (2015). School composition and the Black-White achievement gap. NCES 2015–018. National Center for Education Statistics.

- Clark, M. H., & Cundiff, N. L. (2011). Assessing the effectiveness of a college freshman seminar using propensity score adjustments. *Research in Higher Education*, 52(6), 616–639. doi:10.1007/s11162-010-9208-x
- Conley, D. T. (2010). *College and career ready: Helping all students succeed beyond high school*. San Francisco, CA: Jossey-Bass.
- Conley, D. T. (2011). Crosswalk analysis of deeper learning skills to common core state standards. Educational Policy Improvement Center (NJ1).
- Conley, D. T. (2015). A New Era for Educational Assessment. *Education Policy Analysis Archives*, 23, n8. doi:10.14507/epaa.v23.1983
- Copi, I. M., Cohen, C., & McMahon, K. (2013). *Introduction to logic*. New International Edition ed. London: Pearson Education Limited.
- Couzijn, M., & Rijlaarsdam, G. (2004). Learning to read and write argumentative text by observation of peer learners. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Studies in writing: Vol. 14. Effective learning and teaching of writing. A handbook of writing in education* (2nd ed., pp. 241–258). Boston: Kluwer.
- Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, L., Herman, J., & Pellegrino, J. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). Accountability for college and career readiness: Developing a new paradigm. *Education Policy Analysis Archives*, 22, n86. doi: 10.14507/epaa.v22n86.2014
- Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research & Development*, 32(4), 529–544. doi:10.1080/07294360.2012.697878
- Davies, W. M. (2006). An ‘infusion’ approach to critical thinking: Moore on the critical thinking debate. *Higher Education Research & Development*, 25(2), 179–193. doi:10.1080/07294360600610420
- Davies, M., & Barnett, R. (2015). What is critical thinking in higher education? In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 27–29). New York, NY: Palgrave MacMillan.
- Deane, P., Quinlan, T., & Kostin, I. (2011). *Automated scoring within a developmental, cognitive model of writing proficiency (ETS Research Report No. RR-11-16)*. Princeton, NJ: ETS.
- Deane, P., Quinlan, T., Odendahl, N., Welsh, C., & Bivens-Tatum, J. (2008). *Cognitive models of writing: Writing proficiency as a complex integrated skill. CBAL literature review writing (ETS Research Report No. RR-08-55)*. Princeton, NJ: ETS. doi:10.1002/j.2333-8504.2008.tb02141.x
- Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 1–24). Norwood, NJ: Ablex.
- Dick, R. D. (1991). An empirical taxonomy of critical thinking. *Journal of Instructional Psychology*, 18(2), 79–92.
- Donato, A. D., & Thomas, J. (2017). The effects of Greek affiliation on academic performance. *Economics of Education Review*, 57, 41–51. doi:10.1016/j.econedurev.2017.01.004
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287–312.
- Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse. *Studies in Science Education*, 38(1), 39–72. doi:10.1080/03057260208560187
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, England: Cambridge University Press.
- Ennis, R. (2015). Critical thinking: A streamlined conception. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 49–63). New York, NY: Palgrave MacMillan.
- Ericsson, K. A., & Charness, N. (1994). Expert performance. *American Psychologist*, 49(8), 725–747. doi:10.1037//0003-066X.49.8.725

- Farrington, C. A. (2013). *Academic mindsets as a critical component of deeper learning*. University of Chicago: Consortium on Chicago School Research.
- Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, 120(1), 34–45. doi:10.1037//0096-3445.120.1.34
- Fryer, R. G., Jr., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics*, 86(2), 447–464.
- Gensler, H. J. (2012). *Introduction to logic*. London: Routledge.
- Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, 14(1), 69–110. doi:10.1207/s15327809jls1401\_4
- Govier, T. (2013). *A practical study of argument*. Belmont, CA: Wadsworth.
- Greene, J., & Foster, G. (2003). *Public high school graduation and college readiness rates in the United States (Manhattan Institute, Center for Civic Information, Education Working Paper, No. 3)*. New York: Manhattan Institute.
- Hatcher, D. L. (2011). Which test? Whose scores? Comparing standardized critical thinking tests. *New Directions for Institutional Research*, 2011(149), 29–39. doi:10.1002/ir.378
- Heijltjes, A., Van Gog, T., Leppink, J., & Paas, F. (2014). Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. *Learning and Instruction*, 29, 31–42. doi:10.1016/j.learninstruc.2013.07.003
- Heijltjes, A., Van Gog, T., & Paas, F. (2014). Improving students' critical thinking: Empirical support for explicit instructions combined with practice. *Applied Cognitive Psychology*, 28(4), 518–530. doi:10.1002/acp.3025
- Hurley, P. (2014). *A concise introduction to logic*. Boston: Wadsworth.
- Ikenobe, P. (2001). Teaching and assessing critical thinking abilities as outcomes in an informal logic course. *Teaching in Higher Education*, 6(1), 19–35. doi:10.1080/13562510020029572
- Jaschik, S. (2016). The Shrinking Humanities Major. Inside Higher Ed. Retrieved from: <https://www.insidehighered.com/news/2016/03/14/study-shows-87-decline-humanities-bachelors-degrees-2-years>
- Jaschik, S. (2017). Humanities Majors Drop. Inside Higher Ed. Retrieved from: <https://www.insidehighered.com/news/2017/06/05/analysis-finds-significant-drop-humanities-majors-gains-liberal-arts-degrees>
- Johnson, R. H. (2014). *The rise of informal logic: Essays on argumentation, critical thinking, reasoning and politics* (Vol. 2). Windsor: Windsor Studies in Argumentation.
- Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18243–18250. doi:10.1073/pnas.1012933107.
- Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, 19(4), 201–214. doi:10.1016/j.tics.2015.02.006
- Jones, A. (2015). A disciplined approach to critical thinking. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 169–182). New York, NY: Palgrave MacMillan
- Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual Review of Sociology*, 29, 417–442. doi:10.1146/annurev.soc.29.010202.100019
- Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. *Evaluation Review*, 31(5), 415–439. doi:10.1177/0193841X07303318
- Lai, E. R. (2011). Critical Thinking: A Literature Review. Pearson's Research Reports. Retrieved in 2016, March, from <http://images.pearsonassessments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf>
- Larson, A. A., Britt, M. A., & Kurby, C. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education*, 77(4), 339–366. doi:10.3200/JEXE.77.4.339-366
- Larson, M., Britt, M. A., & Larson, A. (2004). Disfluencies in comprehending argumentative texts. *Reading Psychology*, 25(3), 205–224. doi:10.1080/02702710490489908

- Marin, L. M., & Halpern, D. F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity*, 6(1), 1–13. doi: [10.1016/j.tsc.2010.08.002](https://doi.org/10.1016/j.tsc.2010.08.002)
- Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 45–61). New York: Macmillan.
- McNamara, D. S. (2010). Strategies to read and learn: Overcoming learning by consumption. *Medical Education*, 44(4), 340–346. doi:[10.1111/j.1365-2923.2009.03550.x](https://doi.org/10.1111/j.1365-2923.2009.03550.x)
- Morgan, S. L., & Winship, C. (2014). *Counterfactuals and causal inference*. Cambridge: Cambridge University Press.
- Morsanyi, K., Devine, A., Nobes, A., & Szűcs, D. (2013). The link between logic, mathematics and imagination: Evidence from children with developmental dyscalculia and mathematically gifted children. *Developmental Science*, 16(4), 542–553. doi:[10.1111/desc.12048](https://doi.org/10.1111/desc.12048)
- National Assessment Governing Board (2015). Reading Framework for the 2015 National Assessment of Educational Progress. Available online: <http://www.nagb.org/content/nagb/assets/documents/publications/frameworks/reading/2015-reading-framework.pdf>
- National Center for Education Statistics. (2014). Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2001 to 2014. U.S. Department of Education. Retrieved from <https://nces.ed.gov/ccd/pubschuniv.asp>
- Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational Research Review*, 9, 114–128. doi:[10.1016/j.edurev.2012.12.002](https://doi.org/10.1016/j.edurev.2012.12.002)
- Organization for Economic Cooperation and Development. (2013). PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy. OECD Publishing. <http://dx.doi.org/10.1787/9789264190511-en>
- Paul, R. (2004). The State of Critical Thinking Today: The Need for a Substantive Concept of Critical Thinking. Available from <http://www.criticalthinking.org/pages/the-state-ofcritical-thinking-today/523>.
- Pellegrino, J. W., & Hilton, M. L. (Eds.). (2013). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: National Academies Press.
- Permzadian, V., & Credé, M. (2016). Do first-year seminars improve college grades and retention? A quantitative review of their overall effectiveness and an examination of moderators of effectiveness. *Review of Educational Research*, 86(1), 277–316. doi:[10.3102/0034654315584955](https://doi.org/10.3102/0034654315584955)
- Rasmussen, T. (2013). *A meta-analysis of academic success courses in postsecondary institutions* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3572090)
- Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, 46(3), 853–891. doi:[10.3102/0002831209333184](https://doi.org/10.3102/0002831209333184)
- Reed, S. K. (1993). A schema-based theory of transfer. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on Trial: Intelligence, cognition, and instruction* (pp. 39–67). Norwood, NJ: Ablex.
- Reznitskaya, A., & Anderson, R. C. (2002). The argument schema and learning to reason. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 319–334). New York: Guilford.
- Reznitskaya, A., Anderson, R. C., & Kuo, L. J. (2007). Teaching and learning argumentation. *The Elementary School Journal*, 107(5), 449–472. doi:[10.1086/518623](https://doi.org/10.1086/518623)
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701. doi:[10.1037/h0037350](https://doi.org/10.1037/h0037350)
- Rutschow, E. Z., Cullinan, D., & Welbeck, R. (2012). *Keeping students on course: An impact study of a student success course at Guilford Technical Community College*. New York: MDRC.
- Rychen, D. S., & Salganik, L. H. (Eds.). (2003). *Key competencies for a successful life and well-functioning society*. Goettingen: Hogrefe & Huber Publishers..
- Salmon, M. (2012). *Introduction to logic and critical thinking* (6th ed.). Wadsworth: Cengage Learning.



- Simpson E., & Courtney, M. (2002). Critical thinking in nursing education: Literature review. *International Journal of Nursing Practice*, 8(2), 89–98. doi:10.1046/j.1440-172x.2002.00340.x
- Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, 15(3), 250–267. doi:10.1037/a0018719
- ten Dam, G., & Volman, M. (2004). Critical thinking as a citizenship competence: Teaching strategies. *Learning and Instruction*, 14(4), 359–379. doi:10.1016/j.learninstruc.2004.01.005
- The Carnegie Classification of Institutions of Higher Education. (n.d.). About Carnegie classification. Retrieved from <http://carnegieclassifications.iu.edu/>.
- Thomas, K., & Lok, B. (2015). Teaching critical thinking: An operational framework. In *The Palgrave handbook of critical thinking in higher education* (pp. 93–105). New York, NY: Palgrave MacMillan.
- Tiruneh, D. T., Verburgh, A., & Elen, J. (2013). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies*, 4(1), 1–17. doi:10.5539/hes.v4n1p1
- Tobolowsky, B. F. & Associates. (2008). 2006 National Survey of First-Year Seminars: Continuing innovations in the collegiate curriculum (Monograph No. 51). Columbia, SC: University of South Carolina, National Resource Center for The First-Year Experience and Students in Transition.
- van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching*, 53(1), 41–46. doi:10.3200/CTCH.53.1.41-48
- van Gelder, T. (2015). Using argument mapping to improve critical thinking skills. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 169–182). New York, NY: Palgrave MacMillan.
- van Gelder, T., Bissett, M., & Cumming, G. (2004). Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology*, 58(2), 142–152. doi:10.1037/h0085794
- Weinert, F. (2001). Concept of competence. A conceptual clarification. In D. Rychen & L. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–66). Seattle: Hogrefe and Huber.
- West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, 15(1), 18–37. doi:10.1037/a0015917
- What Works Clearinghouse. (2016). First Year Experience Courses. WWC Intervention Report. Retrieved from <http://ies.ed.gov/ncee/wwc/Intervention/8>