The effect of philosophy on critical reading: Evidence from initial teacher education in Colombia

Alejandro Farieta ¹,*, Marcos Delprato ²

Centre for International Education and Development, University of Sussex, Hove, Brighton, United Kingdom

A R T I C L E  I N F O
Keywords:
Philosophy
Critical reading
Educational policy
Initial teacher education
Standardized tests
Propensity score matching

A B S T R A C T
Teacher quality, its effect on students’ outcomes, and the association of these with economic growth, is the core of recent discussions in Latin America given the region’s weak results in international learning assessments. This paper investigates whether there is an effect of philosophy on the outcomes of critical reading for students in B.Ed. programs in Colombia. Relying on exact matching combined with propensity score matching with regression adjustment, we use national data from Colombia to show that students in B.Ed. in philosophy outperformed students in other B.Ed. in critical reading test (0.401–0.124 SD), and, importantly, with higher effects observed for students with lower prior academic achievement (0.44 SD). This suggests that philosophy can help to narrow educational outcomes of students whose socioeconomic conditions are disadvantageous, contributing to social justice in education.

1. Introduction

Teacher quality has become one of the most important concerns in the educational policy domain because of their link to economic growth (Hanushek and Woessmann, 2015a). According to the human capital theory (Becker, 2009; Hanushek, 2011), improving the quality of teachers will enhance learning outcomes (with literacy being one of the most important), and in turn it will boost social and economic conditions of their students in the long run. The cognitive skills of a population are a crucial factor for a country’s economic growth, as shown in a series of longitudinal studies about the association between PISA and TIMMS international tests and national GDP growth (Hanushek and Woessmann, 2012a, 2015a, 2015b; Hanushek, Piopiunik, and Wiederhold, 2019). A clear example of the effects of educational outcomes on economic growth can be found across the East Asian countries, which have been in the top bracket of performance in PISA results over the last few years and they have also significantly increased their economies compared to two generations ago (Hanushek and Woessmann, 2016).

A vital component behind this economic expansion process is higher teacher quality. Indeed, better teachers in the classroom can boost students’ learning and so help them to get better-paid jobs, thereby boosting economic growth (Chetty, Friedman, and Rockoff, 2014a, 2014b; Hanushek, Piopiunik, and Wiederhold, 2019). Teacher experience, specific education, and cognitive skills are determining factors of school performance (Bardach and Klassen, 2020; Burroughs et al., 2019; Coenen et al., 2018; OECD, 2018a).

Due to the importance of teacher quality, initial teacher education has become central to educational policy in recent years. According to Cochran-Smith, since the end of the 1990 s, educational policies about teacher education have become “the public policy problem”, trying to identify which elements can be manipulated from public policy about initial teacher education to improve the quality of education at a large scale (Cochran-Smith, 2005, 2021; Furlong et al., 2008; Tattoo and Menter, 2019). The underlying argument is that improving undergraduate teacher education will have an impact on future teacher quality which, as a result, will have an impact on the economy and development at a national level (Barber and Mourseshed, 2007; Rockoff, 2004).

Latin America is a region where low teacher quality and cognitive skills have a bearing on the observed stagnation of GDP growth. In this region, despite the increase in enrolment in the last 50 years, the economic effect has been significantly lower than other regions in the world, especially compared to southeast Asia, with similar growth in enrolment but considerably higher economic outcomes (Bruns and Luque, 2015). Consequently, several educational policy decisions...
recommendations for Latin America advise to improve teacher education, recruitment, and remuneration to boost economic growth (Hanushek and Woessmann, 2012b, OREALC/UNESCO, 2013, 2016). This is the case of Colombia, where students in initial teacher education programs—equivalent to a Bachelor in Education, B.Ed.—, have lower academic outcomes compared with students in other undergraduate majors (Barrera-Osorio et al., 2012; Bonilla-Mejia, 2018; Garcia J. et al., 2014). Recently, this has motivated important national reforms of educational policy to improve the educational outcomes of future teachers (Colombia, Ministry of Education [MEN], 2016, Feb. 3; 2017, Sep. 15; Arias G. et al., 2018).

In this paper, we investigate a novel unexplored mechanism that can improve the outcomes of B.Ed. students; that is, what is the impact on teaching quality if future teachers study more philosophy. This is our main research question. Our motivation is to add to the body of empirical evidence about the effectiveness of philosophy undergraduate majors, which is almost null. The only exception is the study of Farieta (2022) showing that, in Colombia, more philosophy credits and subjects in B.Ed. are associated with better outcomes in reading, citizenship education, and writing. Our study further contributes to this for the Colombian case, relying on a robust methodology approach, that is: matching with regression adjustment.

Philosophy has been a valuable knowledge area for western culture and history since its rise in ancient Greece. Its influence and importance are undeniable for developing other knowledge areas, such as mathematics, physics, medicine, politics, economics, psychology, and sociology, among many others. Nevertheless, in the last fifty years, the inclusion of philosophy modules or credits in initial teacher majors has declined, especially in countries like the UK, where most of them are optative (Barrow, 2019). Additionally, many philosophy faculties, departments, and undergraduate programs have been closed or threatened to be closed, or they are experiencing spending cuts. According to www.dailynous.com, a website focused on the diffusion of philosophy, leading universities around the world have experienced some of these menaces in the last five years in countries such as the United States, the United Kingdom, Australia, Serbia, Hungary, and Spain (Weinberg, 2021, Oct. 6). In Colombia, specifically, ten philosophy programs were closed in the last ten years (Farieta, 2018; Farieta et al., 2019); the number of students in these programs has been significantly reduced (Farieta et al., 2024), and the hours spent in high school teaching philosophy have also been reduced (Herrera, 2022; Prada-Dussan, 2023).

This article empirically shows that philosophy has a national economic potential because of its impact on literacy, which is seen as one of the most valuable skills in economic development and democratic stability (OECD, 2019). Literacy is deeply linked to democratic values and the health of political institutions because it is necessary to dialogue and to solve differences between humans in a non-violent way (Morais, 2018). The importance of philosophy for democracy, along with the arts and humanities, has been defended vehemently by authors like Nussbaum (2012) and others (Belfiore, 2015; Machura, 2018; O’Brien, 2015; Olmos-Penuela et al., 2015). Plainly, the leading thread of the paper is to show that philosophy, in a global south country context, is not only important for maintaining democratic stability, but also it could have a significant economic potential through boosting educational quality. We attempt to answer the following three research questions (RQ):

- **RQ1:** Do students in B.Ed. in philosophy obtain higher scores in critical reading test than students attending a different B.Ed.?
- **RQ2:** Do students who have attended B.Ed. in philosophy obtained better scores than students in B.Ed. more related to literacy, such as literature, Spanish language, linguistics, etc.?
- **RQ3:** Is the effect of philosophy homogeneous or, rather, are there gaps in critical reading outcomes depending on students’ socioeconomic and contextual characteristics or prior academic achievement?

Beyond the generic arguments on teaching quality leverage impacts on economic growth and democratic values mentioned above, the Colombian context is particularly suitable for answering these questions. It has a long tradition in B.Ed. programs that instruct future teachers for elementary and high schools in specific subjects like mathematics, natural sciences, social sciences, child education, language, philosophy, special education, etc. (Colombia, MEN, 2016, Feb. 3; 2017, Sep. 15). Moreover, Colombia has mandatory standardised tests at the end of high school (Colombia, Congress of the Republic [CRC], 2009, Jul. 13) and at the end of undergraduate programs (Colombia, Presidency of the Republic [PRC], 2009, Oct. 14). So, it is possible to estimate the added value of the undergraduate programs controlling from prior academic achievement—a leading drive of educational performance—and net of an array of observable characteristics, making the empirical contribution of the paper more robust. Additionally, it is possible to assess of whether the impact of philosophy is heterogeneous and varies by students’ socioeconomic conditions, thereby highlighting possible mechanisms behind how philosophy can translate into boosted educational outcomes among the most disadvantaged students.

The structure of the paper is as follows. In Section 2 we present a review of the effects and consequences of a better teacher education for students, with a greater focus on a review of the effects of philosophy and related matters on student outcomes. In Section 3, we show the characteristics of the Colombian B.Ed. programs and how they are appropriate to answer the paper’s research questions. In Section 4, we introduce the methodological design, i.e., the data analysis and the matching strategy. Section 6 includes the results and Section 7 offers conclusions and policy recommendations from the analysis.

## 2. Conceptual framework

### 2.1. Initial teacher education and student outcomes

Several studies show how the performance of students depends on the training and quality of teachers, which, beyond socioeconomic, familiar, and contextual conditions, are the main determinants of student achievement (Hanushek and Woessmann, 2015a; Burroughs et al., 2019; Coenen et al., 2018). Teachers with better training—a significant malleable entry point for an educational policy design—and with more experience have equally positive effects on students’ education trajectories—such as more chances to enter higher education—and their social outcomes—e.g., a lower risk of teenage pregnancy, better-paid jobs, etc.—and, in the long term, on living in areas where people have higher incomes and better retirement plans (Chetty et al., 2014b).

According to a meta-analysis done by Hanushek and Rivkin (2010), the added value of a teacher one SD above the mean of the teacher effectiveness distribution could improve student outcomes in reading tests by 0.10–0.26 SD. The long-term economic effect of student outcomes’ increase is that one standard deviation in scores translates into 10–20% higher annual earnings throughout the lifetime of students (Hanushek, 2011). Likewise, in a study across 31 countries, Hanushek et al. (2019) found that teacher cognitive skills have effect on student performance of 0.145–117 SD in mathematics and 0.092–148 SD in reading scores, controlling by parents’ cognitive skills, student and school characteristics, and country variations.

Professional development and experience of in-service teachers is often considered one of the most central strategies to improve students’ outcomes (Didion et al., 2020), but there is also evidence of the effects of the quality of pre-service teacher training programs on student outcomes (Goldhaber et al., 2013; Koedel et al., 2015; Mihaly et al., 2013). Additionally, academic achievement of future teachers predicts teacher performance and, therefore, it has a positive effect on future students’ cognitive outcomes (Corcoran and O’Flaherty, 2018).
2.2. Philosophy and effectiveness

Research about the effectiveness of philosophy undergraduate programs is scarce. A notable exception is the Colombian study of Farieta (2022), which through multilevel regression analysis, found that there is an association between more credits in philosophical modules and higher outcomes in critical reading, writing, and citizenship competencies (but lower outcomes in quantitative reasoning).

Nevertheless, critical thinking and argument processing are more widely studied, usually with positive effects in different educational levels (Abrami et al., 2015). Logic, argumentation, and critical thinking have been rooted in philosophy since the ancient Greek philosophers (Castagnoli and Fait, 2022), and nowadays they are mandatory in any philosophy undergraduate program. Logic and argumentation are considered central in teaching critical thinking (Haussman et al., 2021; Salmon, 2012). A study carried out at a university in the United States by Quintana and Schunn (2019) concluded that logic modules in first-year students of undergraduate STEM-related majors improve their academic achievement, especially for students with lower academic background. According to Quintana and Schunn (2019), the mediator for obtaining critical thinking skills is the argumentation linked to the logic modules. A recent study shows that argument processing can be improved with online programs, compensating student’s lack of previous formal instruction in scientific literacy skills (Münchow et al., 2023). In a different context, a study by Sultan et al. (2017) with pre-service language teachers in Indonesia, showed that a critical literacy training approach had a significant effect on reading skills.

Even though philosophy has not been widely researched in terms of its impact on efficiency in higher education, there is evidence on its positive effects for different educational levels. For instance, in programs and projects as “Philosophy for Children” [P4C] (Bynum and Lipman, 1976; Lipman and Sharp, 1978), and “Philosophy with Children” [PwC] (Cassidy and Christie, 2013; Kennedy and Kennedy, 2011; Vansiegleghem and Kennedy, 2011). PwC has shown to be very effective when applied in schools in terms of improving critical and creative thinking skills. The meta-analysis of García-Moriyón et al. (2005) reported positive effects on students’ cognitive skills, with an average difference between treated and untreated groups of about half a standard deviation (d = 0.58). Likewise, important effects on basic cognitive skills, personality traits and academic achievement of students attending a PwC programme in Madrid (Spain) were found in a long longitudinal study ten years after the intervention (Colom et al., 2014). A study in the UK also found that the results obtained by PwC are better for students from disadvantaged backgrounds (Ventista, 2019).

Some of the evaluations of PwC programmes have been questioned for not being rigorous enough or for having some conflict of interest (García-Moriyón et al., 2005; Colom et al., 2014). However, more recent studies show better and more reliable evidence not only in critical thinking and cognitive skills (İşiklar and Öztürk, 2022), but also in noncognitive, social, and ethical skills (Karadag and Demirras, 2018; Ventista, 2019) like self-regulation, engaged participation in the school, and collaborative dialogue with their peers in children with social, emotional, and behavioural needs (Cassidy et al., 2018).

3. The Colombian Context

3.1. Factors associated with student academic achievement

One of the main drivers of students’ performance in Colombia is their socioeconomic status, whose explanatory power is very high in Latin American countries, accounting for almost 30% of the variation of the scores (Avendano et al., 2016). Latin American education systems are highly segregated, and of the total variation of students’ achievement, around 40% is attributable to school wealth composition and a further 10% to additional individual and school factors (Delprato et al., 2015). This shows the high levels of inequality in the region’s educational systems, especially in Colombia (OECD, 2016, 2018b; García V. et al., 2013; García J. 2015; Rodríguez et al., 2014; Timarán-Pereira et al., 2016). Socioeconomic status is related to parental education, another important factor explaining student performance in the Colombian context (García-Gonzalez and Skrita, 2019; Rangel and Lleras, 2010). Students with more educated parents have more support when it comes to assignments and other academic tasks and they have higher income, better house infrastructure and more books and other academic resources at home (Camacho et al., 2016; Hernández G and Padilla G, 2019). Generally, students of B.Ed. came from families with lower educational levels than other undergraduate programs (García J. et al., 2014).

An additional essential contextual characteristic is where a student lives (region and city, rural/urban location) and the neighbourhood income level where they come from. Colombia is a highly unequal and segregated country, and students’ educational outcomes are a result of this embedded and cumulative geographical inequalities (Arias-Velandia et al., 2021; García J. et al., 2015). Teachers with lower standardized scores—a proxy for teacher quality—are mostly located in the poorest regions of the country, which makes inequalities due to education more difficult to tackle (Bonilla-Mejía et al., 2018). Students in richer regions have better educational outcomes, since these regions have universities with better educational infrastructure, higher-paid staff, and a better reputation (Cvečec et al., 2019; Gibbons y Vignoles, 2012; Helbig et al., 2017). Education segregation even happens within big cities, where students’ achievement is associated with socioeconomic inequalities, and students with higher economic resources are the ones who can attend universities with improved resources and better staff (Rojas, 2019).

Gender is also associated with student performance, with gender inequality persisting due to culturally stereotyped behaviour (Carcamo et al., 2020; Morris, 2012; OECD, 2018b). According to the literature, gender performance in high school is positive for women in literacy, but negative in mathematics (Abadía and Bernal, 2016; Cárcomo, and Mola, 2012; Correa, 2016; Woesssmann, 2010). In higher education, the gap increases in mathematics especially in STEM programs (Gómez et al., 2020). Teacher education in Colombia has prominently more female students (63%), more than other undergraduate courses, whose female student percentage is 54% (Arias G. et al., 2018). However, and contrary to the rest of the B.Ed., in philosophy courses the number of female students and their outcomes in reading are lower than those of male students (Farieta, 2022). Further, age is linked to educational performance (Castro et al., 2018; Rodríguez et al., 2014) because of the opportunity costs of studying for worse-off students’ groups, and this means that students in B.Ed. are usually 1–2 years older in than students in other undergraduate programs (Rodríguez et al., 2014).

At the undergraduate level, one of the leading factors explaining educational outcomes is student’s prior achievement (Schneider and Preckel, 2017). In Colombia it can be obtained from the student score on the Saber 11 test, a mandatory test to obtain a high school diploma and the main benchmark to measure the added value of undergraduate majors (Rodríguez-Revilla and Vallejo-Molina, 2022; Sarmiento et al., 2019). Students in B.Ed. have lower scores on the Saber 11 test than students in other undergraduate programs (García J. et al., 2014; Rodríguez et al., 2014).

In addition, modality of education (on/campus/distance) is found to be an important determinant of undergraduate student performance, with on-campus students having lower performance than on-campus students (Aguilera-Prado, 2017; Arias-Velandia et al., 2018; Rodríguez, Gómez, and Ariza, 2014). Students in distant programs mostly come from rural areas, smaller and poorer towns of lower socioeconomic status and, compared with students in similar conditions in on-campus programs, they attain lower scores (Arias-Velandia et al., 2021; Pineda and Cellis, 2018; Timarán-Pereira et al., 2016).

Other quality indicators associated with student outcomes is the high-quality accreditation. Colombian higher education system has two
types of accreditations: institutional and program accreditation (Colombia, MEN/Comisión Nacional de Acreditación [CNA], 2013). Usually, the institutional accreditation is associated with better student outcomes (Bayona et al., 2018; Cayón et al., 2020), nevertheless, the evidence about the program accreditation is not conclusive, and some studies evidence positive outcomes (Camacho et al., 2016) but others don’t (Sarmiento et al., 2015). In philosophy, the association hasn’t been found (Farieta, 2020). In 2015, the Colombian government compelled all the B.Ed. programs to obtain high-quality accreditation (Colombia, CRC, 2015, Jun. 9) to improve the student achievement. Other institutional conditions related to student scores in undergraduate programs are teachers’ characteristics (educational level, teaching experience, evaluation, etc.) (Ordoñez et al., 2019; Sáenz-Castro et al., 2021).

3.2. Philosophy programs in Colombia

There are two different kinds of philosophy majors in Colombia: the traditional B.A. in philosophy, but also a different type of program called “Licenciatura”, equivalent to a Bachelor of Education (B.Ed.) in the anglophone educational systems, which are aimed at teaching and training schoolteachers. Some universities offer both programs. Usually, the first type is shorter (4 years) and the second is longer (5 years) (Farieta et al., 2015) because it includes school practices or modules related to learning, teaching, and pedagogy. In this study, we will focus only on the B.Ed. in philosophy to compare them with different B.Ed. programs. Some of the B.Ed. in philosophy include other subjects like language, religious studies, history, political science, or other humanities (Farieta et al., 2015; Farieta, 2018). There are slight differences in curricular terms, and there are big discussions about the relevance of their strong European tradition in the Latin American context and other curricular issues (Bernal-Ríos, 2022). Additionally, these programs are strongly masculinised, with only an average 24.8% of professors and lecturers of philosophy programs being female (Acevedo-Zapata and Rivera-Sanín, 2023).

4. Methodological design

4.1. Data and Sample

The educational system in Colombia is a fitting case to assess the hypothesis that students who attend a philosophy program have better outcomes. First, Colombia has a mandatory test called Saber Pro for all students in the final year of their undergraduate studies (Colombia, PRC, 2009, Oct. 14), and one of the modules in the test is “Critical Reading.” Second, since the 1960s Colombia has had a very broad range of initial teacher education programs in all disciplines and knowledge areas which are mandatory in basic and middle education: mathematics, natural sciences, social sciences, philosophy, arts, early childhood, special education, etc. (Colombia, MEN, 2016, Feb. 3; 2017, Sep. 15). There are currently 558 B.Ed. in the country, 29 of which are in philosophy or philosophy with another discipline—such as theology, religious sciences, political theory, humanities—(Farieta et al., 2015; Farieta, 2018). Some of these are inactive or in process of being closed, due to new regulations for Colombian B.Ed. (Arias G. et al., 2018) and other issues related to financing through tuition fees in private universities combined with the lack of getting enough students to open cohorts (Farieta et al., 2024; Herrera, 2022; Prada-Dussán, 2023; Valderrama-Leongómez et al., 2019). This allows a comparison of the outcomes of the students in a B.Ed. in philosophy with those of students in other B.Ed. programs. The main data used came from the ICFES, a governmental organisation in charge of the design and assessment of all the national educational tests. All the data is public and is available at www.icfes.gov.co.

Table 1 describes total population that presented the Saber Pro test during the 2012–2021 period by year and program type. In total, 225,984 B.Ed. students presented the Saber Pro test over this period. Out of this total population, 5240 were students in B.Ed. in philosophy [ph.] or with the noun “philosophy” in the program name: ph. and theology; ph. and religious studies; ph. and ethics; ph. and humanities, etc. (Farieta et al., 2015; Farieta, 2018). Table 1 shows that the number of B.Ed. in Philosophy students has reduced throughout the years, reaching a maximum of 728 in 2014, and a minimum of 302 in 2021, with a reduction in its proportion compared to the rest of B.Ed. students (from 3.16% in 2012 to 1.41% in 2021). Similarly, there has been a decreasing uptake for students in B.Ed. in literature (including linguistics, literature, Spanish, and other related areas presented the test), dropping from 15.4% in 2012 to 10.53% in 2021.

The Saber Pro test changed the scoring system in 2016 (ICFES, 2017). The previous version (2012–2015) was designed to have a standard mean of 10 points and a standard deviation of 1, with a range between 5.2 and 15.8. In the more recent version (2016–2021), scores’ interval is 0 to 300 (ICFES, 2018). Since both scores are normally distributed, we standardise the results for comparison so that to avoid any bias due to changes in the test. Methodologically, we use exact matching for the year of the presentation of the exam (see subsection 4.2).

Most of the control variables we employ in our analytical approach (i.e., matching) are drawn from earlier studies (Arias-Velandia et al., 2021; Avendaño et al., 2016; García J. et al., 2014, 2015; Timar-Pereira et al., 2016). The control variables can be classified in two groups: (1) student and context, and (2) institution and program. The most crucial control variable we use is students’ prior academic achievement, which comes from the Saber 11 test in the reading module. As mentioned above, all students in the last year of high school (11th grade, according to the Colombian educational system) must present the Saber 11 test (Colombia, CRC, 2009, Jul. 13). The test is run twice a year. We merged the databases of Saber Pro and Saber 11 to include students’ prior academic achievement in the analysis.

Table 1

<table>
<thead>
<tr>
<th>Year</th>
<th>B.Ed. in Philosophy</th>
<th>Other B.Ed.</th>
<th>B.Ed. in Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>2012</td>
<td>487</td>
<td>0.035</td>
<td>1.077</td>
</tr>
<tr>
<td>2013</td>
<td>702</td>
<td>0.053</td>
<td>1.104</td>
</tr>
<tr>
<td>2014</td>
<td>728</td>
<td>0.070</td>
<td>1.120</td>
</tr>
<tr>
<td>2015</td>
<td>563</td>
<td>0.269</td>
<td>1.095</td>
</tr>
<tr>
<td>2016</td>
<td>503</td>
<td>0.372</td>
<td>0.946</td>
</tr>
<tr>
<td>2017</td>
<td>566</td>
<td>0.454</td>
<td>1.057</td>
</tr>
<tr>
<td>2018</td>
<td>519</td>
<td>0.427</td>
<td>0.980</td>
</tr>
<tr>
<td>2019</td>
<td>460</td>
<td>0.535</td>
<td>0.940</td>
</tr>
<tr>
<td>2020</td>
<td>410</td>
<td>0.465</td>
<td>1.013</td>
</tr>
<tr>
<td>2021</td>
<td>300</td>
<td>0.577</td>
<td>0.905</td>
</tr>
<tr>
<td>Total</td>
<td>5238</td>
<td>220,629</td>
<td></td>
</tr>
</tbody>
</table>

Note: * p < .05, * * p < .01, * * * p < .001
When students register to present the Saber Pro test, they must complete a socioeconomic survey and the ICFES includes this information with students’ scores. From the ICFES survey, we select the following control variables: gender, age, Neighbourhood Socio-Economic Level (NSEL) —also known as socioeconomic strata, (estrato socioeconómico in Spanish)—. We also include socioeconomic indicators by national departments (Colombian administrative regions) and municipalities of students to account for socioeconomic disparity within the country at a more aggregated level. Specifically, we use the Index of Multidimensional Poverty (IMP) from the data from the 2018 national census, calculated by the National Department of Statistics [Departamento Nacional de Estadística – DANE] (2018). The IMP considers the conditions of the population regarding health, housing, access to basic services (water, electricity, telephone, internet, and domiciliary gas), educational conditions, and conditions of children and youth. The IMP is calculated for the 33 administrative departments of Colombia. Additionally, we use the Index of Unsatisfied Basic Needs (IUBN) calculated by the DANE from the national census of 2018 for the 1103 Colombian municipalities (DANE, 2018). The IUBN measures the living conditions in aspects like inadequate or overcrowded housing, children and youth not attending school, economic dependency, and access to basic public services. We do not use the “rural/urban living place” dichotomous variable because the population of students living in rural areas is quite low, hindering balancing the matching groups. Instead, we split the sample into students living in big cities, capitals of departments (or metropolitan areas) and students living in small towns or rural areas, which with social inequality measured at different levels is enough to capture socio-economic factors behind B.Ed. choices and supply of courses.

Besides, we include variables of the institutions and programs, such as the accreditation status of the institution and the program, since they have been reported by the literature having better added value for student outcomes (Bayona et al., 2018; García J. et al., 2014). We also use the program modality (on-campus or distance) as students in distance courses have shown lower scores in Saber Pro (Aguilera-Prado, 2017; Farieta, 2020). Descriptive statistics are presented in Table 3.

Summary statistics for control variables by the treatment and control groups are shown in the Table 3. Statistics show that students in B.Ed. in philosophy come from less disadvantaged background in terms of living area, education of parents, NSEL, IMP, IUBN, and prior academic achievement (Saber 11 scores) in comparison with students of other B.Ed., as well as students in language B.Ed. Also, they are more likely to be male, (63%) in comparison with students in other B.Ed. (32%) or language B.Ed. (24%). Additionally, more of them come from programs and institutions with accreditation and on-campus programs. These differences support the necessity to implement a matching technique to narrow the characteristics’ differences across the two groups, offering a better control of differences on observables. Since the Saber 11 and the Saber Pro tests have had several changes across the years, we use exact matching for the year of presentation of both tests. This, too, offers a better control by weakening the impact of changing cohorts across time.

**Table 2** shows the working sample after merging, by period of presentation of the Saber 11 test and the type of program. This test had significant changes in 2014 to align all the national tests, and the previous scores from 2012 to 2014 were recalculated with the new structure and scoring system (ICFES, 2013). To avoid bias due to changes in the test, we apply exact matching by period of presentation, as we explain in the next subsection. Also, to improve comparability, we only keep in the sample students whose test time gap were more than three years and less than ten. Using this data, we can determine if students with similar Saber 11 test scores (and other similar factors associated with student performance) have better outcomes in the Saber Pro reading module when they attend a B.Ed. in philosophy, in comparison with those who went to a different B.Ed.

When students register to present the Saber Pro test, they must complete a socioeconomic survey and the ICFES includes this information with students’ scores. From the ICFES survey, we select the following control variables: gender, age, Neighbourhood Socio-Economic Level (NSEL) —also known as socioeconomic strata, (estrato socioeconómico in Spanish)—. We also include socioeconomic indicators by national departments (Colombian administrative regions) and municipalities of students to account for socioeconomic disparity within the country at a more aggregated level. Specifically, we use the Index of Multidimensional Poverty (IMP) from the data from the 2018 national census, calculated by the National Department of Statistics [Departamento Nacional de Estadística – DANE] (2018). The IMP considers the conditions of the population regarding health, housing, access to basic services (water, electricity, telephone, internet, and domiciliary gas), educational conditions, and conditions of children and youth. The IMP is calculated for the 33 administrative departments of Colombia. Additionally, we use the Index of Unsatisfied Basic Needs (IUBN) calculated by the DANE from the national census of 2018 for the 1103 Colombian municipalities (DANE, 2018). The IUBN measures the living conditions in aspects like inadequate or overcrowded housing, children and youth not attending school, economic dependency, and access to basic public services. We do not use the “rural/urban living place” dichotomous variable because the population of students living in rural areas is quite low, hindering balancing the matching groups. Instead, we split the sample into students living in big cities, capitals of departments (or metropolitan areas) and students living in small towns or rural areas, which with social inequality measured at different levels is enough to capture socio-economic factors behind B.Ed. choices and supply of courses.

Besides, we include variables of the institutions and programs, such as the accreditation status of the institution and the program, since they have been reported by the literature having better added value for student outcomes (Bayona et al., 2018; García J. et al., 2014). We also use the program modality (on-campus or distance) as students in distance courses have shown lower scores in Saber Pro (Aguilera-Prado, 2017; Farieta, 2020). Descriptive statistics are presented in Table 3.

Summary statistics for control variables by the treatment and control groups are shown in the Table 3. Statistics show that students in B.Ed. in philosophy come from less disadvantaged background in terms of living area, education of parents, NSEL, IMP, IUBN, and prior academic achievement (Saber 11 scores) in comparison with students of other B.Ed., as well as students in language B.Ed. Also, they are more likely to be male, (63%) in comparison with students in other B.Ed. (32%) or language B.Ed. (24%). Additionally, more of them come from programs and institutions with accreditation and on-campus programs. These differences support the necessity to implement a matching technique to narrow the characteristics’ differences across the two groups, offering a better control of differences on observables. Since the Saber 11 and the Saber Pro tests have had several changes across the years, we use exact matching for the year of presentation of both tests. This, too, offers a better control by weakening the impact of changing cohorts across time.
We use Propensity Score Kernel Matching with Regression Adjustment, combined with exact matching for the categorical variables to improve the quality of the matching (Jann, 2017a; Rubin and Thomas, 2000). The technique allows us to estimate the Average Treatment of the Treated (ATT) after pairing students in the B.Ed. in philosophy (treatment group) with students with same or extremely similar characteristics in different B.Ed. (control group). The regression adjustment adds double robustness property. The matching double robust estimator (Nguyen et al., 2017) uses the same specification for estimating the propensity score and the outcome model. This estimator has the theoretical advantage that it yields unbiased estimates if either or both the propensity and outcomes are correctly specified—the double robust property which offers more protection against misspecification (Jann, 2017a; Rubin and Thomas, 2000). To avoid bias due to the different years of the test, we use exact matching for the year of the SABER test, but also for the SABER II test, because, as mentioned above, these tests have had significant changes through the time. The exact matching by year also avoids any bias caused by the time changes or contextual issues such as the pandemics occurred during 2020 and 2021 which caused the tests to be presented virtually in 2020. We also use exact matching for all the dichotomous variables: gender, living area (capital city or metropolitan area/rural or small city), institutional accreditation, program accreditation, and program modality (on-campus/distance).

The aim of the matching process is to obtain a counterfactual of what the average score would have been if students had attended a program different to the B.Ed. in philosophy. We define the average treatment effect of attending the program as:

\[
E(Y_i | D = 1) - E(Y_i | D = 0) = \\
E(Y_u | D = 1) - E(Y_u | D = 0) + E(Y_n | D = 1) - E(Y_n | D = 0)
\]

(1)

Where \(Y_i\) denotes student scores, and \(D\) is a dummy for program selection, which equals 1 when a student is enrolled in a B.Ed. in philosophy and 0 otherwise. The outcome variable \(Y_u\) is the outcome of a student in a different program, \(Y_n\) is the outcome of a student in a different program. \(E(Y_i | D = 1) - E(Y_i | D = 0)\) is the average treatment effect (ATE), and \(E(Y_u | D = 1)\) is the unobserved counterfactual that has to be constructed using matching estimators. The estimation of the ATE requires two assumptions: the conditional independence assumption (CIA) and common support (Rosenbaum and Rubin, 1983). CIA assumption implies that potential outcome variables \((Y_u, Y_i)\) are independent of the treatment (here, the decision of being enrolled in a B.Ed. in philosophy) when conditioning on a set of observable covariates \(X\); in other words: CIA: \((Y_i, Y_u) \perp D \mid X\). The second assumption entails common support (or overlap), which means that, for each student in a B.Ed. in philosophy, there is a positive probability of a match within a group of non-philosophy students with a similar set of covariates \(X\); that is: \(0 < \Pr(D = 1 \mid X) < 1\).

If both assumptions hold, meaning that \(F_i(Y | X, D = 1) = F_u(Y | X, D = 0)\), the counterfactual \(E(Y_u | D = 1)\) can be built independent of the decision of the program attended after conditioning for the propensity score. The propensity score \(\Pr(D = 1 | X = \alpha)\) is the conditional probability of choosing a B.Ed. in philosophy, given the pre-treatment variables \(X\) (Rosenbaum and Rubin, 1983):

\[
\rho(X) = \Pr(T = 1 | X) = E(T | X) = \int 1_{|X|} dF(X)
\]

(2)

The next step is to match students in B.Ed. in philosophy with peers in a different B.Ed. based on the propensity score of receiving the treatment calculated from the covariates. If CIA holds, the ATT is defined as

\[
ATT = E(Y_u - Y_i) \mid D = 1 = E(Y_u - Y_i)
\]

(3)

To determine the matching weight, we use propensity score kernel matching, where the matching weight is defined as:
Where $K$ is the Epanechnikov kernel density function: $K(x) = \frac{3}{4}(1 - x^2)$, for $|x| < 1$, $a$ is the bandwidth parameter, $P_1$ is the propensity score of the case $j$ in the control group, $P_0$ denotes the propensity score of case $i$ in the treated group, and $P_i - P_j$ represents the distance between propensity scores. The bandwidth is calculated by weighted cross-validation with respect to $Y$ (Fröhlich, 2004, 2005; Galdo et al., 2008). The weights are normalized so, for each student that was enrolled in a B.Ed. in philosophy, $\sum_{j \in D_0} W_{ij} = 1$. The counterfactual estimate is the weighted average of the observed outcome of the students in B.Ed. different to philosophy:

$$f_j = \frac{K(P_i - P_j)/a}{\sum_{i \in D_0} G(P_i - P_j)/a}$$

(4)

We estimate the ATT with the weight from Eq. (3) as:

$$\bar{\text{ATT}} = \frac{1}{N_{\text{pair}}} \sum_{i, j \in D_0} \left( Y_i - \frac{\sum_{j \in D_0} W_{ij} Y_j}{\sum_{j \in D_0} W_{ij}} - D = 0 \right)$$

(5)

The ATT equals the weighted average of the differences between observed and potential outcomes, which represents the difference in scores of a student in a B.Ed. in philosophy compared with the counterfactual that attended a different B.Ed.

Entropy balancing is applied to achieve the covariate balance of the matched data improve significantly, reducing covariate bias, since the matched selection imbalance was very close to zero. In the third and fourth step, we check that there are no significant differences between the original treated group and the matched to verify there is not selection bias after the matching in terms of observables. We check that the SMD between treated and matched groups across the covariates is below the threshold of $<0.2$ SMD (Rubin and Thomas, 2000) (Fig. 3) and, hence, there is not significant changes in terms of observables after the matching.

Lastly, we compare the density of the treated group. Densities of the entire treated group and the matched group are similar, ensuring there is no significant changes or selection bias after matching (Fig. 4). In summary, we argue that empirical checks carried out entail robustness for the matching approach we implement.

5. Results

5.1. RQ 1 – Students in B.Ed. in philosophy compared to all other B.Ed

We verify the common support condition to test the validity of the estimation in four different steps. First, we conduct balancing tests of the covariates between the treated and untreated groups before and after the matching. We evaluate the balance of each covariate across the treatment groups in each matched sample, calculating the standardized mean difference (SMD):
there are no significant differences between the untreated and the treated after matching (Fig. 1a). Also, there is enough common support, since the untreated group has a similar density that the treated group after matching is like the original one, since there are no differences in propensity score density before and after matching, and the differences in propensity score density but also according to the covariates, which reduces substantially residual confounding. Likewise, the treatment group after matching as the SMD between covariates is below the threshold (Fig. 1a), and densities are practically the same for the raw treatment group and the matched group (Fig. 3a), as well as in terms of the propensity score kernel density (compared with focusing only on learning language from a linguistic approach).

As far as the heterogeneity analysis is concerned, we find significant differences between the philosophy students according to the number of philosophical credits (Table 5, columns 4–9), prior academic achievement, and gender. The group balancing (treated/untreated) fits well after matching for all the comparison groups in terms of covariates SDM (Fig. 3c–d), as well as in terms of the propensity score kernel density (Fig. 2b–c). Also, the selection bias is minimum for the treated groups matching, since the standard mean differences by covariates are no higher than 0.1 (Fig. 3b), and there is also no difference in terms of propensity score distribution (Fig. 4b), and all of which ensures that there is no selection bias driven by observables after matching.

5.3. RQ 3 – Effect of philosophy by subpopulations

As far as the heterogeneity analysis is concerned, we find significant differences between the philosophy students according to the number of philosophical credits (Table 5, columns 4–9), prior academic achievement, and gender. The group balancing (treated/untreated) fits well after matching for all the comparison groups in terms of covariates (Fig. 1c–d), as well as in terms of the propensity score kernel density (Fig. 2b–c). Also, the selection bias is minimum for the treated groups after matching, since these are extremely akin to the original treatment group in covariates SDM (Fig. 3c–d), but also in propensity score density (Fig. 4c–d). Conversely, we do not find statistically significant differences in estimates for the ATT in terms of education of the parents, living region, NSEL, IMP, IUBN, program modality, or institutional or program level.

5.2. RQ 2 – Philosophy compared with Literature and language B.Ed

The estimated ATT for philosophy students compared to language students is 0.124 SD and statistically significant (SE 0.026; p < 0.001) (Table 5, columns 4–9). From the treatment group, 1836 out of 2119 students were matched, whereas for the control group 11,132 students out of 19,176 in language B.Ed. were matched (Table 5, column 3). Even though estimates are comparatively lower than the one we obtained for all B.Ed., a significant difference still holds which suggests a positive impact of philosophy for literacy and, especially, for critical reading (compare with focusing only on learning language from a linguistic approach).
5.3.1. Effect by philosophical credits

Students in B.Ed. in philosophy with more philosophical credits surpass students in lower credit programs by 0.1 SD (Tables 5, columns 4 and 5; Fig. 5). This means the effect of philosophy relies not only in attending the major but also in the intensity of the treatment. The effect of philosophy on critical reading is not linear, contrary to what previous studies show (i.e., Farieta, 2022) as the difference of the more exposed group is not twofold when compared to the less exposed group (0.444 SD vs. 0.344 SD), perhaps indicating a ceiling effect on terms of credits for critical reading scores operating within the higher credits' interval sub-group (81–126). Nevertheless, even a lower exposition to philosophy has a large impact on student achievement, which can be worthwhile for a student interested in a multidisciplinary higher education and with a strong philosophical core.

5.3.2. Effect by prior academic achievement

We find contrasting effects according to students’ prior academic achievement. Weaker students (i.e., with lower outcomes in the Saber 11 test) outperform stronger students in 0.11 SD (Table 5, columns 6 and 7). Saber 11 test scores are the main predictor of higher education achievement.

Table 5

<table>
<thead>
<tr>
<th>Critical reading std. scores</th>
<th>Comparison group</th>
<th>Philosophical credits</th>
<th>Saber 11 score</th>
<th>Gender</th>
<th>Program Accreditation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All B.Ed.</td>
<td>Language B.Ed.</td>
<td>Low (40-80)</td>
<td>High (80-120)</td>
<td>Low (≤ 5 SD)</td>
</tr>
<tr>
<td>Wald test</td>
<td>ATT</td>
<td>NATE</td>
<td>Total</td>
<td>Matched</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>0.401 **</td>
<td>0.124 **</td>
<td>0.019</td>
<td>0.211</td>
<td>121,006</td>
</tr>
<tr>
<td>F2</td>
<td>0.344 **</td>
<td>0.444 **</td>
<td>0.024</td>
<td>0.211</td>
<td>71,444</td>
</tr>
<tr>
<td>F3</td>
<td>0.440 **</td>
<td>0.334 **</td>
<td>0.027</td>
<td>0.030</td>
<td>49,562</td>
</tr>
<tr>
<td>F4</td>
<td>0.417 **</td>
<td>0.337 **</td>
<td>0.026</td>
<td>0.026</td>
<td>0.019</td>
</tr>
<tr>
<td>F5</td>
<td>0.664 **</td>
<td>0.444 **</td>
<td>0.028</td>
<td>0.031</td>
<td>0.019</td>
</tr>
<tr>
<td>F6</td>
<td>0.497 **</td>
<td>0.937 **</td>
<td>0.031</td>
<td>0.032</td>
<td>0.019</td>
</tr>
<tr>
<td>F7</td>
<td>0.646 **</td>
<td>0.444 **</td>
<td>0.030</td>
<td>0.032</td>
<td>0.019</td>
</tr>
<tr>
<td>F8</td>
<td>0.704 **</td>
<td>0.724 **</td>
<td>0.032</td>
<td>0.035</td>
<td>0.019</td>
</tr>
<tr>
<td>F9</td>
<td>0.720 **</td>
<td>0.564 **</td>
<td>0.030</td>
<td>0.029</td>
<td>0.019</td>
</tr>
<tr>
<td>F10</td>
<td>0.720 **</td>
<td>0.564 **</td>
<td>0.030</td>
<td>0.029</td>
<td>0.019</td>
</tr>
<tr>
<td>F11</td>
<td>0.720 **</td>
<td>0.564 **</td>
<td>0.030</td>
<td>0.029</td>
<td>0.019</td>
</tr>
</tbody>
</table>

Note: ** p < 0.001, * p < 0.01, * p < 0.05.

accreditation (for details, see the Appendix).

 caveat for critical reading scores operating within the higher credits' interval sub-group (81–126). Nevertheless, even a lower exposition to philosophy has a large impact on student achievement, which can be worthwhile for a student interested in a multidisciplinary higher education and with a strong philosophical core.

5.3.2. Effect by prior academic achievement

We find contrasting effects according to students’ prior academic achievement. Weaker students (i.e., with lower outcomes in the Saber 11 test) outperform stronger students in 0.11 SD (Table 5, columns 6 and 7). Saber 11 test scores are the main predictor of higher education achievement.
outcomes (Sarmiento et al., 2019). This means that for students with learning deficits at the end of high school, philosophy has the potential to allow them to catch up with students that had higher scores at that point and, therefore, to remedy some of the gaps accumulated by a deficient basic education or other contextual disadvantages (e.g., lower parental education or adverse socioeconomic conditions). Philosophy therefore has a vital role on reducing existing learning-driven ability gaps.

Fig. 2. Kernel density plot before and after matching.

Fig. 3. Standardized Mean Difference (SMD) between the entire treated group and the matched by covariates.

*Note: S11: Saber 11 reading std. score; PaEd: Parents education; NSEL: Neighborhood socioeconomic level; MUN: Type of Municipality; IUBN: Index of Unsatisfied Basic Needs by Municipality; IMP: Index of Multidimensional Poverty by Region; IAC: Institutional Accreditation; PAC: Program Accreditation; CM: Course Modality*
5.3.3. Effect by gender

We also find differences on the estimated ATT by gender. Our estimates indicate that male students achieve 0.08 SD higher than female students (Table 5, columns 8 and 9). This result is concerning since it shows a gender gap inside philosophy programs. In Colombia, women tend to outperform men in reading and literacy at the high school level (Abadía and Bernal, 2016; Cárcamo et al., 2020; Correa, 2016), but the results show that after majoring in philosophy, the situation changes drastically against women. It is important to consider the low percentage of women in B.Ed. in philosophy (32.03%) compared to the rest of B.Ed. (68.41%) (Table 3). In Colombia, 75% of the professors and staff in philosophy departments are male (Acevedo-Zapata and Rivera-Sanín, 2023), and this low female representation is likely to cause women’s underperformance, despite other reasons related to a highly masculinized environment (Bernal-Ríos, 2022).

5.3.4. Effect by program accreditation

Students in non-accredited programs outperformed students in accredited ones by 0.079 SD (Table 5, columns 10 and 11). This result is at odds with the accreditation criteria, according to which the added value of the programs is a criterion to obtain the high-quality accreditation status (Colombia, MEN/CNA, 2013). A possible explanation for
this is the fact that students with lower prior academic achievement are more likely to attend programs without accreditation \((n = 620)\), in comparison with students with higher prior academic achievement attending the same type of courses \((n = 383)\). This is aligned with the earlier finding about the relatively higher impact on learning of studying philosophy for the lower prior academic achievement group.

6. Discussion

We found conclusive evidence about the advantages of studying philosophy in terms of critical reading for initial teacher education. Our results are in line with previous studies that applied multilevel fixed-effects linear regression (Farieta, 2022), and with studies about the effect of programmes like Philosophy with Children (García-Moríyón, Rebollo, and Colom, 2005), or the effect of logic and argumentation modules in undergraduate programs (Quintana and Schunn, 2019), or the research related to the effects of critical thinking in student achievement. The most remarkable point is that the effect size found is considerably large \((0.401 \text{ SD})\) since it’s rare to find in the literature a higher effect of interventions of critical thinking in reading or literacy, as the meta-analysis of Abrami et al. (2015) shows. The large sample we employed, and the fact that the analysis was done with the entire population of students who graduated from a B.Ed. in the country during the last 10 years, give additional validity to our estimates, allowing us to answer affirmatively RQ1: students in philosophy B.Ed. are in fact having better scores than students in a different B.Ed.

Regarding RQ2 — if students in B.Ed. in philosophy are obtaining higher scores than students in other B.Ed. programs focused specifically on language —, the answer is equally positive. It is worthy to consider that some of the programs with lower levels of philosophical credits combine philosophy and language, most of them with literature (Farieta, 2018; Farieta et al., 2015). This shows that in curricular terms, it would be a good idea to introduce more credits or subjects of philosophy in the B.Ed. in language, especially if these are related to philosophy or language, semiotics, argumentation, or other similar topics. It is likely that some of these programs already have these modules in their study plans, but a significant increment of these would have an impact on improving student outcomes in critical reading.

There are some important findings from RQ3 — i.e., if there are differences in outcomes according to student and program conditions. The main one is that students with lower scores on the Saber 11 test have a higher improvement in their outcomes at the end of the undergraduate program. The findings are consistent with the literature, supporting the idea that those with lesser prior academic achievement can benefit from philosophy (Quintana and Schunn, 2019; Ventista, 2019). It is also very rare to find an intervention that faces lower prior academic achievement with the impact that philosophy have in our study (Schneider and Preckel, 2017). This a key finding of the paper from a policy perspective, namely, Philosophy has a remarkable potential to reduce academic gaps driven by socioeconomic conditions previously to being enrolled in an undergraduate program, and in consequence, to contribute to social justice.

There are some warnings from our analysis about how philosophy impacts on sub-populations, however. The gender gap is a critical concern, despite being a common issue inside the philosophy courses around the world (Antony, 2012; Beebee, 2023; Dougherty et al., 2015; Hutchinson and Jenkins, 2013). Issues like stereotyped threats or implicit bias reported in the literature (Saül, 2015), seem to be affecting women’s scores. It is worth noting that the gender gap is like the one found in STEM programs in Colombia for mathematics (Gómez S. et al., 2020), with similar issues like the lack of female role models and a highly masculinized environment (Acevedo-Zapata and Rivera-Sanín, 2023; Bernal-Ríos, 2022). This means that universities should take the staff gender gap seriously in these programs. All in all, this suggests that the gender gap in outcomes deserves more detailed research to gauge more precisely how wide the gap is, what its causes are, and therefore what are possible ways to address it.

The second important warning has to do with the high-quality accreditation. According to the literature, institutional accreditation is usually associated with higher student scores (Bayona et al., 2018; Cayón et al., 2020). Yet, the evidence is not conclusive about program accreditation, and some studies show positive effects on student outcomes (Camacho et al., 2016), but others show no association (Sarmiento et al., 2015), especially in philosophy (Farieta, 2020). As we said before, the possible reason of this negative effect has to do with the nature of self-selection into the program, that is: the fact that more students with low prior academic achievement attend non-accredited programs. Other explanation is that, since the accreditation is a peer review very demanding process in terms of resources and time (Arias G. et al., 2018; Rodríguez-Avila, 2021), the programs could be investing more time in the bureaucracy involved and management tasks instead of teaching and research (Hanushek and Woessmann, 2015e). The main concern is that if this is true, the accreditation is favouring the ability of the programs to select students with higher student performances, but at the expense of taking into account their added value. This is especially concerning since the Colombian government imposed mandatory accreditation to the B.Ed. programs (Colombia, CRC, 2015, Jun. 9) to improve student achievement, but as it is shown, this has not been the case. However, since program accreditation was not the focus of this study, the conclusions should be considered with caveats as deeper research about the association between program accreditation and student outcomes is needed.

This research was limited to a specific context in a low-middle income country. Further research is needed to determine if in a different context or country the philosophy has the same effect in students’ outcomes. Other of the limitations of the study are different factors reported by the literature that can improve the outcomes of philosophy students, like the class size or the teacher experience, educational level, or other characteristics of the programs (Ordóñez R et al., 2019; Sáenz-Castro et al., 2021).

7. Policy Recommendations

Recent changes in policy regulations for B.Ed. courses in Colombia (Colombia, CRC, 2015, Jun. 9) compelled these to obtain high-quality accreditation, but the results show a negative effect for courses with this award, not necessarily improving student outcomes. A revision of the accreditation criteria is needed to revise this issue, along with more research about its association with student outcomes. Additionally, these policy regulations asked the programs to increase the school practices (Colombia, MEN, 2016, Feb. 3; 2017, Sep. 15) which made them reduce the disciplinary credits (Arias G. et al., 2018; Valderrama-Leongómez et al., 2019; Farieta et al., 2024). This is concerning since the reduction can negatively impact student outcomes, along with other negative effects such as the reduction of institutional autonomy and a higher risk of closing the programs in regions more in need (Rodríguez-Avila et al., 2021). The policy regulations were cancelled (Colombia, CRC, 2019, May 25), but our findings are proof that these policies could be more harmful than beneficial, and a wider discussion with more evidence about teacher education is needed.

On the other side, many philosophy programs have either been closed or threatened to be closed in the last ten years around the world. Sometimes these threats occur on a national level, like in Brazil, where a former President announced the defunding of philosophy programs in all public universities (Bolsonaro, 2019, Apr. 26), asserting that these do not have impact in the national economy. There were reactions against this announcement from academics in Brazil, Latin America, and the entire world (Weinberg, 2019, Apr. 30), and finally the programs weren’t closed. Our results show that these reactions were not only conceptual correct but empirically sounded, and the widespread opinion that philosophy do not have any potential impact on economic growth is false.
The risks of eliminating the arts and humanities, and especially philosophy, are imminent for democracies. Some authors claim that due to a lack of critical reading skills, some political situations affecting democracy worldwide have occurred due to the radicalization of social or cultural differences and the rise of intolerant or politically polarized environments, which happened, for example, during Brexit or the last elections in the US (Barton, 2019; Oelkers, 2017). We agree with the philosophical tradition that reading, critical thinking, and philosophy, have value per se and are fundamental for human flourishing, and are additionally important values for a democratic, diverse, and pluralistic society (Nussbaum, 2012). But besides that, it is important to note how many of the main critics of these are also misled and biased.

This study has proved how philosophy can be a vital mechanism to improve student performance in critical reading. If the premises coming from the economics of education are accepted and better teachers have positive effects on the economic future of the students (Chetty et al., 2019), there is no doubt that philosophy has an important role to improve national economies in the long term. This is particularly relevant in the Colombian context, where teacher quality and academic achievement have strong effects on future student outcomes (Bonilla-Meja et al., 2018). The number of students in B.Ed. in philosophy has been reduced by year (Farieta et al., 2024; cf. Table 1) which should be a big concern because, for the sake of educational quality, they should be increasing and the enrolling on these programs should be promoted. In terms of educational public policy, this is a central argument to support the programs at risk of being closed, but also to promote the opening of more philosophy programs in regions where they do not exist, not only in Colombia, but also in other countries with emerging and poor economies. The recommendation is also to increase the content of philosophy, critical thinking, and argumentation in other B.Ed. and other programs whose students have lower prior academic achievement.

### Appendix. Estimates by subgroups without significant differences

<table>
<thead>
<tr>
<th>Program Modality</th>
<th>Institutional Accreditation</th>
<th>Critical reading std. scores</th>
<th>Distance</th>
<th>On-Campus</th>
<th>Wald test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
</tr>
<tr>
<td>Lower</td>
<td>Lower</td>
<td>ATT</td>
<td>0.408 **</td>
<td>0.365 **</td>
<td>0.399 **</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NATE</td>
<td>0.692 **</td>
<td>0.639 **</td>
<td>0.647 **</td>
</tr>
<tr>
<td>Higher</td>
<td>Lower</td>
<td>ATT</td>
<td>0.296 **</td>
<td>0.336 **</td>
<td>0.393 **</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NATE</td>
<td>0.502 **</td>
<td>0.546 **</td>
<td>0.647 **</td>
</tr>
<tr>
<td>Matched</td>
<td>Yes</td>
<td>ATT</td>
<td>0.449 **</td>
<td>0.402 **</td>
<td>0.423 **</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NATE</td>
<td>0.306 **</td>
<td>0.336 **</td>
<td>0.401 **</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>ATT</td>
<td>0.496 **</td>
<td>0.429 **</td>
<td>0.437 **</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NATE</td>
<td>0.649 **</td>
<td>0.621 **</td>
<td>0.665 **</td>
</tr>
</tbody>
</table>

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. However, the authors are thankful to the Commission on Education of the Colombian Society of Philosophy, whose members read a draft of this paper and made valuable comments to improve the final version, the discussion, and the policy recommendations. The authors are also grateful for the comments provided by two anonymous referees. Also, thanks to the philosophical community in Colombia for their valuable comments and support, but also for keeping on doing, teaching, and spreading philosophy, just for the sake of wisdom and a better life for Colombians, humans, living things, and the planet, even in the face of unfavorable circumstances.

### CRediT authorship contribution statement

**Farieta Alejandro**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.  **Delprato Marcos**: Formal analysis, Investigation, Methodology, Supervision, Writing – review & editing.

### References


References

tarienagov.co/senado/basedic/ley1515v2015.html.
tarienagov.co/senado/basedic/ley155v2019.html.
Faruji, A., 2022. La incidencia de la filosofía en la lectura crítica y otras competencias psicológicas en los programas de formación de maestros en Colombia. W. Herrera.
Rodríguez-Revilla, R., Vallejo-Molina, R.-D., 2022. Valor agregado y las competencias
Rubin, D., Thomas, N., 2000. Combining propensity score matching with additional
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in
Rangel, C., Lleras, C., 2010. Educational inequality in Colombia: family background,
A. Farieta and M. Delprato
Saul, J., 2013. Implicit Bias, Stereotype Threat, and Women in Philosophy. In:
Hutchison, K., Jenkins, F. (Eds.), Women in Philosophy. What needs to change?
Schneider, M., Preckel, F., 2017. Variables associated with achievement in higher
https://doi.org/10.1037/bul0000098.
literacy approach on pre-service language teachers’ critical reading skills. EurAsia J.
Tatto, M., Menter, I., 2019. The significance of teacher education. In: Tatto, M.,
Menter, I. (Eds.), Knowledge, policy, and practice in teacher education: A cross-
Timarán-Pereira, S.R., Hernández-Arteaga, I., Caicedo-Zambrano, S.J., Hidalgo-
Troya, A., Alvarado-Pérez, J.C., 2016. Descubrimiento de patrones de desempeño
decano con Arboles de decisión en las competencias genéricas de la formación
profesional. Universidad Cooperativa de Colombia. https://doi.org/10.16925/
9789587600490.
Licenciaturas en filosofía: disputas y articulaciones entre lo disciplinar y lo
doi:10.28970/9789585498273.02.
Vansieglehm, N., Kennedy, D., 2011. What is philosophy for children, what is
Ventista, O.M. (2019). An evaluation of the “Philosophy for Children” programme: the
impact on cognitive and non-cognitive skills. Doctoral Dissertation, Durham University.
http://etheses.dur.ac.uk/13121/1/Ventista_Thesis.pdf?DDD29
Weinberg, J. (2019, Apr. 30). Open Letter Regarding President Bolsonaro’s Recent
Pronouncements on Defunding Philosophy and Sociology (updated).
http://etheses.dur.ac.uk/13121/1/Ventista_Thesis.pdf?DDD29
Weinberg, J. (2019, Oct. 6). Philosophy Departments Under Threat: Information, Pro-
argentina and colombia in an international perspective. Appl. Econ. 42 (21),