# Is Studying Latin Associated with (Non-)Linguistic Cognitive Transfer? A Large-Scale Cross-Sectional Study

C. Hauspie[a], W. Duyck[a, b, c], S. Schelfhout[a, b, d], A. Vereeck[a, e, f], M. Janse[g] & A. Szmalec[a, h, i]

[a] Department of Experimental Psychology, Ghent University

[b] Interdepartmental Research Group Vocational and Personnel Psychology, Ghent University

[c] NVAO, The Accreditation Organisation of the Netherlands and Flanders

[d] Department of Work, Organization and Society, Ghent University

[e] Department of Linguistics, Ghent University

[f] Department of Translation, Interpreting and Communication, Ghent University

[g] Faculty of Modern and Medieval Languages and Linguistics, University of Cambridge

[h] Psychological Sciences Research Institute, Université catholique de Louvain

[i] Institute of Neuroscience, Université catholique de Louvain

**Abstract**

Despite ongoing discussions regarding the relevance of Latin in modern education, this language still holds a prominent role in European secondary school curricula. While studying Latin is commonly believed to yield cognitive and linguistic benefits, this argument primarily relies on dated research that often uses methodologies that do not allow to make strong claims justifying the widespread use in education. It also remains unclear to which extent the benefits associated with Latin studies are due to Latin students' superior pre-existing abilities (preselectivity), or to cognitive transfer effects elicited by studying the language. To delve further into the presence and nature of a potential cognitive advantage of Latin, we gathered data from $N$ = 1,731 secondary school students across three grades. We explored whether a 'Latin advantage' exists, and if so, for which subjects, when this advantage arises and how it evolves throughout secondary education. We found that first-year Latin students exhibited higher intelligence scores, superior native language competencies and higher meta-linguistic awareness compared to non-Latin peers, which is in line with the preselectivity account. This performance difference was larger in the second year, but smaller in the last year of secondary education, thereby challenging the notion of cognitive transfer effects attributed to Latin studies. Only one variable, vocabulary, demonstrated a trend in line with cognitive transfer benefits. Longitudinal work is needed to further investigate whether Latin studies result in persisting benefits or whether the 'Latin advantage' is merely a reflection of preselection biases.

*Keywords:* Latin, secondary education, preselection, cognitive transfer, cognitive advantage

Although Latin is often called a dead language (Jessner et al., 2018), it still is very much alive as a language in European secondary education. In fact, Latin is studied by over three million children (EULALIA, n.d.), which translates to no less than 8% of all European lower and upper secondary education students (European Commission - European Education and Culture Executive Agency, 2023). Furthermore, 76% of all countries in the European Union and the European Free Trade Association have established course regulations for this ancient language. Remarkably, in 14% of these countries, Latin is a compulsory course for all secondary education students (European Commission - European Education and Culture Executive Agency, 2023). Moreover, Latin education is still on the rise in some countries, like in the French-speaking part of Belgium where from 2027 on, a minimum of 2 hours per week of Latin will be compulsory for any pupil in the second or third year of secondary education. The underlying idea is that Latin has a unique contribution to intellectual and linguistic development, an advantage that every child should be able to enjoy (Federation Wallonia-Brussels - Ministry of Education, 2022). Still, the continued presence of the language in modern education has repeatedly been called into question, due to longstanding public debates about the subject's educational value (Vereeck et al., 2024). In a society where modern foreign language education is flourishing by focusing on practical applications of the language at hand, the instruction of a dead language is sometimes dismissed as outdated and useless (Bracke & Bradshaw, 2020; Brunello et al., 2023).

In response to criticisms questioning the perceived lack of applicability of Latin, classicists have put forward arguments stating that the study of this classical language fosters transferable skills across various domains (Vereeck et al., 2023). Such claims draw upon *cognitive transfer theory*, which proposes that training skills in one domain can also enhance skills in other domains, be it closely (near transfer) or distantly related (far transfer) (Barnett & Ceci, 2002; Harrison et al., 2013). And indeed, some studies suggest that Latin studies elicit transfer effects to both related and unrelated cognitive functions. More specifically, studies have shown that Latin studies increase students' native language (L1) aptitude (Kennedy, 2006), as well as their modern foreign language abilities (De Vane, 1997; Sparks et al., 1995). Furthermore, other research claimed that studying Latin also contributes to enhanced critical thinking skills and mathematical achievement (De Vane, 1997; Sheridan, 1976). However, the majority of these studies is over 20 years old and used small samples. Furthermore, the methodology of most studies does not allow strong claims about isolated or causal effects of Latin studies (Vereeck et al., in press), thus leaving this research field in dire need of new empirical evidence. Hence, the objective of the present study is to gain renewed, state-of-the-art insights into the linguistic and broader cognitive effects of studying Latin in secondary education. As such, we employ a methodologically rigorous cross-sectional design that leverages a uniquely large sample (N = 1,731) of secondary school students. Our findings can inform debates or inspire research on the cognitive benefits associated with other dead languages taught across the globe, such as Ancient Greek, Old Norse or Sanskrit. While such languages might mainly be taught in the context of cultural considerations, it is likely that cognitive benefits are expected to ensue as well. Hence, our study is a necessary effort to empirically guide the discussion about the cognitive value of dead languages in modern study curricula.

**Cognitive Transfer or Preselectivity?**

Although various studies have reported that Latin students outperform their peers in several educational outcomes (Brunello et al., 2023; Hauspie et al., 2024; Kennedy, 2006), the causal mechanism behind this advantage is less easily agreed upon. A first explanation would be that of a preselection mechanism. This perspective, as brought up by Thorndike already in 1924, entails that students with higher cognitive capacities are more inclined to study Latin, as Latin is traditionally regarded as a challenging study program (Bennett, 2021).

As such, the student population that chooses Latin studies would not be random, but determined by certain student characteristics (e.g., cognitive abilities). Therefore, the benefits Latin students show over other students would be due to a priori differences in student characteristics, rather than a consequence of studying the language (Vereeck et al., 2023). Besides an actual cognitive advantage, pupils who choose Latin may also present different demographic profiles than their non-Latin peers (Bracke & Bradshaw, 2020), which may influence academic achievement. For instance, research in Flanders has already shown that Latin students belong to families with higher socio-economic statuses (SES) (Flemish Department of Education, 2011). Given that SES and academic performance are related (Karadag, 2017), the lack of research that takes this association into account is remarkable. Finally, the students that enter Latin studies may also present different attitudinal profiles that make them more likely to achieve academic success. Research has shown that attitudinal variables, such as motivation or conscientiousness, characterize one's approach to learning, and thus influence scholastic performance (Schneider & Preckel, 2017). In other words, a preselection in terms of attitudinal characteristics would make Latin students more likely to demonstrate behaviors that align well with improved academic performance. In sum, preselection could occur in students' cognitive, socio-economic, and attitudinal profiles, which all have been linked to academic achievement. Hence, with our first research question, we aim to test this preselectivity notion by exploring whether Latin students already demonstrate higher scores on intelligence, linguistic, attitudinal and SES measures at the onset of Latin instruction.

The aforementioned preselectivity-account is at odds (but could co-occur) with the prevailing belief at the start of the twentieth century, namely that specific types of learning may enhance generalized thinking and memory skills (Thorndike, 1914). This concept was later formalized as the *cognitive transfer theory* (Harrison et al., 2013). According to this theory, near transfer occurs when the skills learned in one domain positively impact learning and performance in a closely related domain (e.g., learning Latin and subsequently enhanced L2 skills). When the two domains are very dissimilar or only distantly related, however, such as learning Latin and improved mathematics skills, this phenomenon is termed far transfer (Sala et al., 2019). While the overall notion of near transfer has gathered much support in cognitive psychology (Gobet & Sala, 2023), unrelated to the question of Latin advantages, far transfer remains a contested phenomenon. Although some studies have reported far transfer effects, the available meta-analytical evidence agrees that the true effect size of far transfer effects is zero (Gobet & Sala, 2023; Sala et al., 2019). So why would Latin be an exception?

Advocates of Latin assume that learning this ancient language has inherent cognitive properties that facilitate both near and far transfer (Vereeck et al., 2023). Such transfer would not be found after studying modern foreign languages, as the instruction of Latin is supposed to have a different, strongly analytical approach. Indeed, the emphasis when learning Latin is on cultivating a receptive understanding of its complex vocabulary and grammar, rather than active language production. In this view, the study of the language's intricate grammatical structures would facilitate transfer effects that extend beyond the linguistic domain, more than any other foreign language would (Vereeck et al., 2023). Our study presents the ideal setting to test such claims, as Flemish education has a strong emphasis on (foreign) language learning. Starting from the age of 10, all Flemish students (mostly native Dutch speakers) study French, followed by English from 12-13 years on. Depending on their study program, pupils may also study a third foreign language, such as German or Spanish (European Commission - European Education and Culture Executive Agency, 2023). So, Flanders offers a conservative setting to test unique effects of studying Latin, given that all students also study other foreign languages. If, from a cognitive transfer perspective, Latin distinguishes itself from modern foreign languages through its complex grammar, we should still be able to observe near or far transfer effects. Hence, with our second research question, we aim to explore the evolution of performance differences between Latin and non-Latin students on various constructs (L1 proficiency, meta-linguistic awareness and intelligence) throughout secondary education.

**Benefits Associated with Studying Latin**

A large body of literature has reported that the study of Latin leads to near transfer effects, such as improved L1 competencies (for a review, see Vereeck et al., in press). For instance, Carlisle and Liberman (1989) demonstrated that students who had studied Latin for two years performed better on L1 spelling tests, compared to students who studied a modern foreign language for a similar amount of time. Importantly, the two groups did not differ significantly in terms of verbal aptitude prior to studying Latin or in academic achievement at the time of testing. Similarly, a longitudinal study by Haag and Stern (2000) matched two groups of German students in terms of (non-)verbal intelligence and academic achievement, before they started studying their first foreign language (L2). Four years later, the participants who chose Latin as their foreign language performed significantly better on L1 grammar tests, compared to those who chose French. Notably, the studies by Carlisle and Liberman (1989) and Haag and Stern (2000) resolved pre-existing differences between the control and experimental group by either matching the groups in terms of intelligence, or by incorporating intelligence as a control variable in the analyses. Unfortunately, these two studies did present small samples ($N = 30$ and $N = 50$, respectively). Many other studies also display such small samples, on top of a lack in methodological rigor (in stark contrast to the present study) and are thus ill-equipped to capture the potential effects of Latin education. For instance, the studies by Fromchuck (1984) and Gilliland (1922) report what appear to be causal claims about how Latin education leads to improved vocabulary and increased reading comprehension in L1. However, these studies failed to assess pre-existing differences amongst pupils, such as higher intelligence or linguistic abilities, even though such differences may explain the ‘benefits’ associated with studying Latin.

Besides near transfer effects to L1 skills, research suggests that Latin instruction also positively impacts foreign language acquisition skills and meta-linguistic awareness (Jessner et al., 2018; Sparks et al., 1995). Regrettably, most of these studies did not consider pre-existing or baseline differences between Latin and non-Latin students either. One of the only studies that did take such pupil characteristics into account was the study by Haag and Stern (2003), who made sure to match both groups of students in terms of verbal intelligence and L2 achievement. Interestingly, no near transfer effects were found, as the findings revealed that German students who opted to study Latin (as opposed to French) actually performed worse when they started to study a new foreign language, namely Spanish. More specifically, Latin students made more vocabulary and grammar errors on a Spanish test, compared to French students.

Finally, we observe that the body of research exploring how studying Latin leads to far transfer effects is limited at best. A dated study by Sheridan (1976) with 432 participants demonstrated that students who engaged in Latin studies for a mere five-month period exhibited progress equivalent to approximately six months in formal areas, such as math computation, when compared to a control group that was “quite reasonably well matched” (p. 18) in terms of socioeconomic background and academic profile (i.e. word knowledge, reading, spelling, language, mathematics skills, science and social studies). Contrasting evidence is reported by Haag and Stern (2003), who did not report any discernible far transfer effects of studying Latin on IQ, mathematics achievement, or inductive and deductive reasoning. However, recent work by Hauspie et al. (2024) reported that pupils who studied 6 years of Latin in secondary education exhibited increased academic achievement in higher education, compared to pupils who studied Latin for a shorter period of time or not at all.

In sum, existing research suggests that the study of Latin has the potential to yield both linguistic and non-linguistic cognitive advantages. In Flanders, where the current study was conducted, Latin's perceived benefits are still promoted on governmental education sites (Catholic Education Flanders, 2023; Onderwijskiezer, 2023b), as is the case in many other regions. However, the (limited) evidence for such assertions is mostly outdated and limited to English-speaking countries. Furthermore, various studies show important methodological shortcomings, such as small samples and a lack of matched control groups. As such, the objective of the present study is to examine whether the reported benefits of studying Latin are still found in a methodologically sound, well-controlled and large-sampled study.

**The Present Study**

Our first research question investigates whether Latin students already outperform their non-Latin peers on intelligence, linguistic (i.e., L1 abilities and meta-linguistic awareness), SES and attitudinal measures at the onset of Latin instruction. If the preselectivity hypothesis is correct, we predict that Latin students will demonstrate higher scores on these measures, compared to their non-Latin peers, already at the start of secondary education. Our second research question aims to explore how the expected discrepancy between Latin and non-Latin students evolves over the course of secondary education. If the cognitive transfer hypothesis is justified, we expect that any performance difference between Latin and non-Latin students that already exists at onset, will progressively increase throughout secondary education, for both intelligence (far transfer) and linguistic (near transfer) measures.

Our study adopts a well-controlled cross-sectional framework to explore and compare educational outcomes across large samples in three different grades. Little research is available that has directly examined the impact of Latin education, while also controlling for other influential variables, such as SES or motivation. Nevertheless, the link between various demographic and attitudinal variables and improved academic performance is well established (Schneider & Preckel, 2017). Hence, we will consider these potentially confounding factors when interpreting the impact Latin instruction has on intelligence and linguistic measures. Importantly, we examine the effects of Latin education in a uniquely large sample ($N$ = 1,731), which sharply contrasts with the former literature.

## Method and Materials

### Data

Over 50 schools were asked to participate through various channels (e-mails, newsletters, Facebook, personal contacts). Inclusion criteria were that the schools (1) were part of one of the two biggest school networks in Flanders (i.e., catholic or governmental) and (2) had a minimum of 20 Latin students in the second year of secondary school. Eventually, data from $N$ = 1,731 Flemish secondary school students were collected in 13 Flemish general secondary[1] schools across Flanders, which are schools that prepare students for higher education. The sample was distributed across the first, second and sixth (last) year of secondary education (the US equivalents are the seventh, eighth, and twelfth grade, respectively; see Table 1)[2]. The students were tested between September 2022 and June 2023, and they were not aware that they were taking part in a study that specifically assesses the effects of Latin education. Instead, the participants were told that they were participating in a study that explored cognitive development across various study programs. This was supported by the fact that data collection took place across different study programs in the same schools. This study was approved by the Ethical Committee of the Faculty of Psychology and Pedagogical Sciences at Ghent University, with reference 2022-019-Cathy

[1] In Flanders, students that enter secondary education choose between four tracks: general, technical, vocational, or artistic secondary education. Only students in general secondary education can choose to study Latin.
[2] A decline is observed in the number of Latin students throughout the duration of secondary education in Flanders (Classica Flanders, 2022). Hence, to ensure similar participant numbers between the Latin and the non-Latin group, the samples are smaller in the sixth year, compared to the first or second year.

Hauspie. The school board, the parents of the students, and the students themselves all gave their informed consent prior to starting the study. The responsible researcher or associates were always present during the testing moments.

**Table 1**

*Sample Details*

| Year | Number of students | | | Mean age | Standard deviation age | % male students |
|---|---|---|---|---|---|---|
| | Total | Latin | Non-Latin | | | |
| First | 801 | 362 | 439 | 11.96 | 0.45 | 48.19 |
| Second | 684 | 304 | 380 | 12.99 | 0.54 | 47.37 |
| Sixth | 246 | 126 | 120 | 16.85 | 0.45 | 36.59 |

In Flanders, all students in the first or second year of general secondary education have a uniform curriculum, with 5 designated lesson periods per week for differentiation. For all Latin students, these 5 periods are devoted to Latin, while students from the control group can take various courses specific to, for example, economics. By the time students enter their final, sixth year of secondary education, the available study programs are more specific and tailored to their interests. A core curriculum still exists, but students can differentiate by choosing between several main tracks (e.g., Latin, sciences, economy and humanities), in which various options are available (e.g., focus on languages, mathematics) (Onderwijskiezer, 2023a). Latin is a widespread study program in Flanders, with 8.82% of all secondary education students (excluding first year pupils) studying the language (Classica Flanders, 2022).

As we are interested in the effects of Latin studies, we compare two major groups: students from the Latin track versus an aggregated group of students from the other, non-Latin tracks within general secondary education. In other words, for the purpose of this study, we consider all general secondary education students that do not study Latin as one control group. All Latin students had started learning Latin from the start of secondary education on, while students in the control group (i.e., the non-Latin group) had not studied any Latin prior to testing[3].

The testing moment was organized in groups and took a maximum of 3 lesson periods (150 minutes). A break was included. All tests were administered online through Qualtrics and the CoVaT-3 testing platform[4].

**Measures**

To decide which measures had to be included in the test battery, we focused on tests that would adequately address the claims made in the existing literature about the benefits of studying Latin on L1 proficiency, meta-linguistic awareness, and intelligence. Besides those, we included several attitudinal and background variables. All measures are listed in Table 2.

Due to occasional technical problems with the testing platforms, some participants were not able to complete the full test battery. Hence, varying sample sizes are reported for the different dependent variables. To depict internal consistency reliability, we include Cronbach's alpha measures. A value of $\alpha > .70$ is considered acceptable reliability, while $\alpha > .80$ indicates good reliability (George & Mallery, 2003). Note that Cronbach's alpha is heavily influenced by the length of the scale or subscale (Piedmont, 2014; Tavakol & Dennick, 2011). Given that this study involves some scales with a relatively small number of items, we also included the average inter-item correlation (AIIC) as an additional measure of internal consistency reliability, when relevant. The AIIC is not influenced by scale length and should be within the range of .15 to .50 (Clark & Watson, 2019).

[3] To ensure a conservative approach, we excluded sixth year students who had studied Latin for less than six years (i.e., students that switched to a non-Latin program after 1 to 5 years of studying Latin). Second year students that switched to a non-Latin program after 1 year of studying Latin were also excluded.
[4] The CoVaT-3 testing platform is the platform where the intelligence measures were administered. For more information, we refer to the measures-section of this paper.

**Table 2**

*Pool of measures*

| Category | Variables |
|---|---|
| Intelligence [a] | General intelligence (G)<br>Fluid intelligence (Gf)<br>Crystallized intelligence (Gc)<br>Visual information processing (Gv)<br>Verbal short-term memory (Gsm) |
| Native language | Spelling<br>Vocabulary<br>Comprehensive reading |
| Meta-linguistic awareness | Phonological awareness<br>Syntactical awareness |
| Attitudinal variables | Autonomous motivation<br>Academic self-efficacy<br>Conscientiousness |
| Background variables | Gender<br>Socio-economic status<br>Native language<br>Learning disorder<br>Study program<br>Year<br>Testing moment |

Note. [a] All intelligence measures need to be interpreted with caution, as no full intelligence test battery was administered. The included tests provide an estimation for each broad cognitive ability (Gf, Gc, Gv and Gsm), but are no comprehensive reliable measurements of these constructs. For more information, we refer to the Measures section of this paper.

### ***Linguistic Measures***

We included data from three L1 (i.e., Dutch) tests. First, a test from the Gletschr test battery (Depessemier & Andries, 2009) was used to assess spelling. The participants heard 30 words and had to write them down correctly (maximum score = 30, *M* = 17.88, *SD* = 5.25, Cronbach's $\alpha$ = .84, *AIIC* = .39). Second, a vocabulary test was developed based on the work of Vander Beken et al. (2018). 30 words were presented, along with four multiple-choice (MC) options. Participants had to indicate the correct meaning of the word. Some items were removed to increase reliability, thus resulting in a total of 26 items (maximum score = 26, *M* = 12.18, *SD* = 4.77, Cronbach's $\alpha$ = .78, *AIIC* = .35). Finally, to assess reading comprehension, students were presented with one of three texts[5], each with ten corresponding multiple-choice questions. These tests were retrieved from the study of van Gelderen et al. (2007) and slightly altered to fit modern-day contexts. Some questions were removed to increase reliability. As such, text 2 (*n* = 1284, Cronbach's $\alpha$ = .70, *AIIC* = .43) had 10 questions, while text 1 (*n* = 149, Cronbach's $\alpha$ = .58, *AIIC* = .36) and text 3 had nine questions (*n* = 300, Cronbach's $\alpha$ = .66, *AIIC* = .42). As the AIIC was satisfactory for all three texts, one measure for reading comprehension was made, depicting the student's performance on one of the three texts (scaled to 10, *M* = 6.20, *SD* = 2.39).

Meta-linguistic awareness was assessed through two tests. First, we included a test on phonological awareness (*M* = 0.14, *SD* = 0.04) from the Gletschr test battery (Depessemier & Andries, 2009). In this timed test, the participants heard 20 non-word pairs (e.g., "pral" – "larp") and had to indicate whether the second word was the correct inversion of the first one. The total score was calculated by dividing the amount of correctly guessed inversions by the total test time (in seconds). Because of this division by total test time, which results in one final score per participant, no normal reliability analyses could be done on this data (as such, no maximum score is available for this test either). Hence, we calculated reliability without taking reaction times into account (Cronbach's $\alpha$ = .80, *AIIC* = .41). Second, a syntactical awareness test (maximum score = 8, *M* = 4.07, *SD* = 1.75, Cronbach's $\alpha$ = .58, *AIIC* = .38) was administered. This test was developed based on the 'morphology and syntaxis' test from the Gletschr test battery (Depessemier & Andries, 2009). The participants were presented with eight Dutch sentences and had to assess whether these sentences were grammatically

[5] All students from the first and sixth year were presented with text 2. Students from the second year are included in a follow-up study and were thus tested with one of three texts. Hence, the sample sizes for the three texts differ largely.

correct. If this was not the case, the participants had to provide the corrected sentence. The sentences included common mistakes against Dutch grammar, such as double negation (e.g., *"Hij zegt dat hij nooit geen dorst heeft na het sporten"*, which corresponds to *"He says that he is never not thirsty after sports")*.

### *Intelligence Measures*

Intelligence was conceptualized through subtests of the COVAT-3 test (Magez et al., 2023). This Dutch digital cognitive test battery is based on the Cattell-Horn-Carroll (CHC) model of intelligence (Schneider & McGrew, 2012). In the CHC model, general intelligence (*G*) is composed of various broad cognitive abilities. We used four subtests (dot sequences, opposites, turned shapes, and verbal memory) to get an estimation of four broad cognitive abilities (fluid intelligence, crystallized intelligence, visual information processing and verbal short-term memory). All tests were scaled to 100. While we refer to these tests by using the broad cognitive abilities to which they correspond, this is only done for ease of interpretation. In order to obtain fully reliable scores on these broad cognitive abilities, the full COVAT-3 test battery would need to be administered, which was impossible due to time limits, as we prioritized a very large sample size. Hence, these tests cannot be interpreted in the same way as traditional intelligence measures. Nevertheless, we did achieve a reliable assessment of students' performances that allows for group comparisons, which was more important for the present study than the calibration of default IQ scales ($M = 100$, $SD = 15$).

The first task, dot sequences ($M = 61.32$, $SD = 10.78$, Cronbach's $\alpha = .77$), corresponds to the broad cognitive ability fluid intelligence (*Gf*). In this task, participants had to continue various dot patterns based on an underlying rule that had to be discovered. The second task, opposites ($M = 57.70$, $SD = 11.44$, Cronbach's $\alpha = .75$), corresponds to the broad cognitive ability crystallized intelligence (*Gc*). Participants had to choose which word from a list of four alternatives was the most opposite of a given word. The third task, turned shapes (*M* = 41.20, *SD* = 16.31, Cronbach's $\alpha$ = .91), belongs to the broad cognitive ability visual information processing (*Gv*). During this task, participants had to check which of the various two-dimensional figures were rotations of a given figure. The final task, verbal memory (*M* = 36.09, *SD* = 20.57, Cronbach's $\alpha$ = .90), corresponds to the broad cognitive ability short-term memory (*Gsm*). Participants were given a list of words and instructed to learn them by heart in a set amount of time. Afterwards, they had to reproduce as many of these words as possible in a pre-determined structured way. G (*M* = 317.23, *SD* = 56.06, Cronbach's $\alpha$ = .93), was computed according to the rules set out by the authors of the COVAT-3, namely $G = 2Gf + 2Gc + Gv + Gsm$.

### *Attitudinal Variables*

We considered three attitudinal variables that have repeatedly been positively associated with academic outcomes (Schneider & Preckel, 2017). First, autonomous motivation (*M* = 25.96, *SD* = 4.97, Cronbach's $\alpha$ = .86, *AIIC* = .69), which entails the extent to which a student's behavior is driven by internal factors (Kriegbaum et al., 2018). This construct was assessed with the motivational subscales from the Children's Perceived use of Self-Regulated Learning Inventory (CP-SRLI) (Vandevelde et al., 2013). Seven items were administered (e.g., *"I work hard for school because I want to learn new things"*), which the students had to rate on a 1 to 5 scale (completely disagree – completely agree). Second, we included academic self-efficacy (*M* = 45.13, *SD* = 7.50, Cronbach's $\alpha$ = .81, *AIIC* = .46), which can be defined as an individual's confidence in their ability to achieve desired academic goals (Schneider & Preckel, 2017). For this, we used the self-efficacy subscales from the CP-SRLI (Vandevelde et al., 2013). This test consisted of 13 items, such as *"I am good at doing my homework, even though I find it boring or hard"*, which participants had to rate on a scale ranging from 1 to 5 (completely disagree - completely agree). Finally, conscientiousness (*M* = 42.77, *SD* = 5.91, Cronbach's $\alpha$ = .77, *AIIC* = .47), a personality trait characterized by traits such as orderliness and dependability (Poropat, 2009), was assessed using the 12 corresponding items from the Dutch translation of the International Personality Item Pool-NEO-120 (Blanken et al., 2018; Johnson, 2014). Students indicated the extent to which they found statements like "*I am always prepared*" characteristic of themselves, using a 5-point Likert scale (completely disagree – completely agree).

***Background and Study Variables***

We included various background variables that have previously been associated with study achievement. The first variable is gender, as female students have been reported to outperform their male counterparts in various stages of the educational trajectory (Voyer & Voyer, 2014). Participants could indicate whether they identified as male, female or x. Given that only eight participants identified as x, we excluded them from further analyses to ensure sufficient sample sizes for meaningful statistical analyses, which involve the gender variable. Hence, the variable was coded as 1 = female and -1 = male. Second, we included SES, as this variable is associated with study achievement in secondary education (Çiftçi & Cin, 2017), but also with intelligence, which is one of our dependent variables (Levine, 2011). For this study, each parent was assigned a numerical score ranging from 0 (no degree) to 7 (PhD). To quantify SES, we employed the mean value of the highest educational degree attained by both parents. Besides this, we considered other background variables, such as whether a student is a native speaker of Dutch or not (1 = no, -1 = yes), and whether or not the student has been diagnosed with a learning disorder (e.g., dyslexia or dyscalculia) (1 = no, -1 = yes). Furthermore, we include information regarding the participants' study year (1, 2 or 6) and whether they were enrolled in a Latin study program (1 = yes, -1 = no). Finally, because the data collection spanned an entire school year, we also included the testing moment as a control variable. As such, we could ensure that any effects observed within each year were not solely attributable to the timing of the assessments. Hence, we denoted how many months had elapsed since the beginning of the school year when the participants were tested, with a maximum of 9 months.

**Procedure and Analyses**

All continuous variables were standardized[6] and effect coding was applied to specify contrasts for all categorical variables (i.e., everything is interpretable to the mean). All analyses were run in R (R Core Team, 2023). First, a Tukey HSD (Honestly Significant Difference) procedure was employed to explore the significance of the year-specific (1, 2 or 6) performance differences for each dependent variable between Latin and non-Latin students. We calculated effect sizes[7] to express the magnitude of these differences for each year separately. The effect sizes allowed us to examine whether or not the difference between the Latin and non-Latin students was larger over the course of secondary education. Second, we performed 2-way analysis of variance (ANOVA) tests on each dependent variable, incorporating Year and Study Program (Latin or non-Latin) as independent variables. Finally, in order to examine whether the effect of Latin on the intelligence and linguistic measures was still significant when we controlled for an array of attitudinal and demographic variables, we used the lme4 package in R (Bates et al., 2015) to perform linear mixed model (LMM) analyses. Separate models were made for all intelligence and linguistic measures. We added a random intercept for Class and School, and Study Program was included as a fixed effect. All background and attitudinal variables listed in Table 2 were included as control variables. For the models with linguistic measures as the dependent variable, the intelligence measure G was included as a control variable as well. For the LMM of Gc and Gsm, we also included the scores on the vocabulary tests as a control variable, because of the high correlations between these constructs (reported in Appendix A). ANOVA tests were used to compare the

[6] Except Testing Moment, as this makes the interpretation of the linear mixed models easier.
[7] A Cohen's $d_s$ was used to report these effect sizes, with 0.01 = very small effect, 0.20 = small effect, 0.50 = medium effect, 0.80 = large effect, 1.20 = very large effect and 2.00 = huge effect (Sawilowsky, 2009).

models with and without Study Program included as a predictor. Interactions between Study Program and three variables (Gender, Testing Moment, and Year) were explored and are reported when significant. Variance Inflation Factor values of the predictors were all below 10, indicating that the multicollinearity levels between the predictors lied within the acceptable range (Stevens, 2012). We employed a significance level of $\alpha = .05$, and applied Bonferroni-Holm correction to all *p*-values (Holm, 1979).

## Results

Appendix A contains the correlation matrix of all variables. Study Program (i.e., Latin or no Latin) displayed weak to moderate correlations with the other variables (ranging from $r = -.09$ with Year to $r = .25$ with Gc). With our first research question, we examined whether Latin students already demonstrate higher scores on intelligence, linguistic, attitudinal and SES measures at the onset of Latin instruction. As such, we performed Tukey HSD tests for each dependent variable, which are reported in Table 3. The effect sizes, illustrating the size of the difference between the first-year Latin and non-Latin students are also reported in Table 3. These effect sizes indicated an advantage for the Latin students for all variables. These effect sizes ranged from small to medium for the intelligence measures ($d_s$ = 0.26 – 0.72). However, the Tukey HSD tests revealed that only the three medium effects (G, Gc and Gsm) were significant, while the two small effects (Gf and Gv) were not. The linguistic variables consistently demonstrated significant differences between the two student groups, which ranged from small to medium ($d_s$ = 0.33 – 0.62). The attitudinal variables were characterized by small effect sizes, which were only significant for Autonomous Motivation ($d_s$ = 0.32). SES differences were significant and characterized by a medium effect size ($d_s$ = .58). In sum, these results confirm the existence of pre-selectivity for all tested variables, but this effect was much smaller for the attitudinal measures.

**Table 3**

*Results from Tukey HSD tests and 2-way ANOVA tests*

| Dependent variable | Latin students vs. non-Latin students: Cohen's $d_s$ | | | Main effect: Year | Main effect: Study Program | Interaction effect: Year x Study Program |
|---|---|---|---|---|---|---|
| | First year | Second year | Sixth year | | | |
| G | 0.69*** | 0.89*** | 0.54 | F(2, 1616) = 234.93*** | F(1, 1616) = 226.02*** | F(2, 1616) = 4.64 |
| Gf | 0.28 | 0.47*** | 0.11 | F(2, 1681) = 21.20*** | F(1,1681) = 46.19*** | F(2, 1681) = 3.4 |
| Gc | 0.72*** | 1.01*** | 0.73** | F(2, 1690) = 360.17*** | F(1, 1690) = 294.90*** | F(2, 1690) = 4.86 |
| Gv | 0.26 | 0.32 | -0.07 | F(2,1661) = 3.96 | F(1, 1661) = 23.86*** | F(2, 1661) = 3.21 |
| Gsm | 0.54*** | 0.64*** | 0.51 | F(2, 1667) = 202.95*** | F(1, 1667) = 138.06*** | F(2, 1667) = 1.16 |
| Spelling | 0.62*** | 0.81*** | 0.77*** | F(2, 1708) = 273.11*** | F(1, 1708) = 217.98*** | F(2, 1708) = 1.63 |
| Vocabulary | 0.55*** | 1.09*** | 1.14*** | F(2, 1720) = 885.83*** | F(1, 1720) = 281.51*** | F(2, 1720) = 18.81*** |
| Reading comprehension | 0.56*** | 0.74*** | 0.34 | F(2, 1607) = 154.53*** | F(1, 1607) = 143.84*** | F(2, 1607) = 4.89 |
| Phonological awareness | 0.33* | 0.42*** | 0.24 | F(2, 1715) = 96.53*** | F(1, 1715) = 52.65*** | F(2, 1715) = 0.86 |
| Syntactical awareness | 0.34** | 0.62*** | 0.39 | F(2, 1707) = 185.43*** | F(1, 1707) = 88.64*** | F(2, 1707) = 3.74 |
| Autonomous motivation | 0.32* | 0.38** | 0.10 | F(2,1718) = 16.75*** | F((1, 1718) = 42.99*** | F(2, 1718) = 2.07 |
| Academic self-efficacy | 0.15 | 0.24 | 0.24 | F(2,1712) = 20.63*** | F(1,1712) = 17.74** | F(2,1712) = 1.18 |
| Conscientiousness | 0.19 | 0.37* | -0.07 | F(2, 1711) = 8.36 | F(1, 1711) = 20.57** | F(2 , 1711) = 4.14 |
| SES | 0.58*** | 0.67*** | 0.41 | F(1, 1579) = 1.92 | F(1,1579) = 135.94*** | F(2, 1579) = 1.30 |

*Note*. Due to occasional technical problems, not all participants were able to complete all tests. Hence, the sample size for the analyses differs for each dependent variable. The effect sizes (Cohen's $d_s$) express the advantage of Latin students over their non-Latin peers, the significance levels in this column indicate the p-values from the Tukey HSD tests.

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

Second, we explored the evolution of the discrepancy between Latin and non-Latin students over the course of secondary education. For this, we started by examining the performance difference between Latin and non-Latin students in the second and sixth year separately, in the same way as we did for the first-year students (the results are reported in Table 3 as well). In the second year, the advantage of Latin students was significant for all variables except for Gv and Academic Self-efficacy. The range of the effect sizes was similar to that in the first year, but shifted upwards. For instance, the range of the effect sizes expressing the difference in the intelligence measures shifted from $d_s$ = 0.26 – 0.72 to $d_s$ = 0.32 – 1.01. In other words, Latin students performed better on all tested variables than the non-Latin students, and this difference was even bigger in the second than in the first year. The difference between the two groups of students was again the smallest for the attitudinal variables.

In the sixth year however, the size and range of the effect sizes changed remarkably. For the intelligence measures, the effect sizes ranged from very small to large ($d_s$ = - 0.07 – 0.73) and the difference was only significant for Gc ($d_s$ = 0.73). For Gv, the effect was even reversed, with a very small, but non-significant, advantage for the non-Latin students ($d_s$ = -0.07). For the linguistic variables, the effects ranged from small to very large ($d_s$ = 0.24 – 1.14). However, this difference was only significant for the two variables that exhibited a large difference, namely Spelling and Vocabulary ($d_s$ = 0.77 and 1.14, respectively). The attitudinal variables displayed small differences ($d_s$ = - 0.07 – 0.24), but none of them were significant. The difference in SES was somewhat larger but also not significant ($d_s$ = 0.41).

We noticed that the same three subtests displayed the largest advantage for Latin students for all three years, namely Gc, Vocabulary and Spelling. Unsurprisingly, these tests were all closely correlated (Gc-Vocabulary: $r$ = .69, Gc-Spelling: $r$ = .58, Vocabulary-Spelling: $r$ = .60). Similarly, Gf and Gv ($r$ = .32) consistently displayed no or the smallest advantage for Latin students over their non-Latin peers.

In order to obtain a better idea of the general direction of the performance differences across the three years, we computed the average of these effect sizes as a function of year and testing category (intelligence, linguistic competencies, attitudes; see Table 4[8]). On average, for the intelligence and linguistic measures, Latin students showed great advantages over their non-Latin peers in the first year, which were even larger in the second year (an 43.75% increase in the average effect size of all the intelligence and linguistic measures). In the sixth year however, the advantage of Latin students on these measures was on average 4.17% smaller than what was observed in the first year. Similar trends were observed for the attitudinal variables, albeit with much smaller effect sizes.

**Table 4**

*Average effect sizes expressing the difference between Latin and non-Latin students.*

| Year | Intelligence measures [a] | | Linguistic measures [b] | | Attitudinal measures [c] | |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| First year | 0.45 | 0.22 | 0.48 | 0.14 | 0.22 | 0.09 |
| Second year | 0.61 | 0.30 | 0.74 | 0.25 | 0.33 | 0.08 |
| Sixth year | 0.32 | 0.37 | 0.58 | 0.37 | 0.09 | 0.16 |

*Note*. [a] The intelligence measures include Gf, Gc, Gv and Gsm. G was not included here, as G was calculated based on the four previously mentioned intelligence measures. [b] The linguistic measures include spelling, vocabulary, reading comprehension, phonological and syntactic

[8] An average was not computed for the background variables, as only differences in SES were examined.

awareness. [c] The attitudinal measures include autonomous motivation, academic self-efficacy, and conscientiousness.

We note that the average effect sizes were larger for the linguistic than for the intelligence variables (average $d_s$ across all years = 0.60 and 0.46, respectively), which is in line with the near transfer hypothesis. This difference was consistent across all three years, but it was more prominent in the sixth-year students. Furthermore, when making a distinction between the variables in the linguistic category, we observed that the advantage Latin students had over their non-Latin peers was much higher for L1 competencies than for meta-linguistic awareness (average $d_s$ across all years = 0.74 and 0.39, respectively).

To explore our second research question, regarding the temporal evolution of the performance differences between Latin and non-Latin students, we wanted to examine the presence of interaction effects between Year and Study Program. First, we explored this without considering control variables. For this, we ran two-way ANOVA tests for each dependent variable, with Study Program and Year as the independent variables. The results are summarized in Table 3. A main effect of Study Program was found for all variables. Similarly, a main effect of Year was found for all variables, except Gv, Conscientiousness and SES. Hence, the grade a student was in and the choice of study program significantly influenced nearly all the variables that were tested. A significant interaction between Year and Study Program was found only for Vocabulary. This entails that the effect of Study Program (i.e., Latin instruction) on a student's vocabulary scores was larger in the second than in the first year, and larger in the sixth than in the second year. Second, we wanted to explore the impact of Study Program, and its potential interactions with Year, while simultaneously controlling for various attitudinal and background characteristics (listed in Table 2). For this, we conducted LMM analyses (full models are reported in Appendix B).

We investigated whether Study Program (and its interaction with Year) was a significant addition to the LMM through model comparisons with ANOVA tests. For G, Gc and Gsm, the variable Study Program had incremental predictive value above and beyond 10 attitudinal and background variables, such as Autonomous Motivation and SES. For the L1 competencies (Spelling, Vocabulary and Reading Comprehension), Study Program was also a significant predictor in models that included attitudinal and background measures, as well as a general intelligence factor (i.e., G). We found that including Latin as a predictor was not significant for Gf, Gv, Phonological Awareness and Syntactic Awareness. Vocabulary was again the only variable where an interaction between Study Program and Year was found, meaning that the difference in Vocabulary observed in Year 1 was bigger in Year 2, and even bigger in Year 6, a finding that hints at near transfer. Furthermore, the explained variance of the LMM of Vocabulary was much larger than for the other variables ($R^2$ =.65). The only other variable that showed an $R^2$ of similar magnitude was Gc ($R^2$ =.52), a variable strongly correlated with vocabulary ($r$ = .69).

## Discussion

Latin is a prominent study option in European secondary education (European Commission - European Education and Culture Executive Agency, 2023) and while it is still on the rise in some countries, like in the French-speaking part of Belgium, the subject is oftentimes dismissed as irrelevant and outdated for modern curricula (Bracke & Bradshaw, 2020). Nevertheless, the extent to which studying Latin produces cognitive and/or linguistic benefits remains a topical point of discussion amongst parents, educators, policy makers and scientists (Brunello et al., 2023). Two main theories about the alleged benefits of Latin studies exist. First, according to the preselectivity account, Latin students already demonstrate superior performances before the onset of Latin education, which is the sole reason for their advantageous performances (Thorndike, 1924). Second, according to the cognitive transfer account, the study of Latin has a cognitive value, which prompts near and far transfer effects (Vereeck et al., 2023). These two theories are not mutually exclusive, but within the latter view, any pre-existing performance difference between Latin and non-Latin students should become larger over the course of secondary education. While several studies have attempted to demonstrate the existence of such cognitive transfer effects, the majority suffers from important methodological shortcomings, such as small sample sizes and a lack of consideration for a priori pupil characteristics. Hence, most studies cannot convincingly rule out potentially confounding preselectivity effects (Vereeck et al., 2023). Furthermore, most studies are quite old. Given that issues like preselectivity change over time and interact with educational context or orientation practices, previous work is hard to translate to current educational contexts. For instance, as equal opportunity policies strengthen, preselectivity may now be less pronounced than it used to be. This may especially be the case in open access systems like Flanders, with constitutionally guaranteed free school choice and free education with very low private contributions (European Commission - Eurydice, 2024). As such, the present study aims to actualize this field of research by examining whether a performance difference between Latin and non-Latin students exists, identifying the specific areas for which such a difference is present, the time at which it emerges and finally, how it evolves over time. Furthermore, we present the first large-scale empirical study to examine the predictive effect of Latin education on intelligence and linguistic measures, while simultaneously controlling for various attitudinal and background variables. These analyses were conducted on an extensive dataset ($N$ = 1,731) from students from the first, second and sixth year of secondary education.

In line with the preselectivity account, we found that the students who chose to study Latin were on average more academically adept, already before the study of Latin could exert any effects. Indeed, first-year Latin students performed better on all intelligence and linguistic measures than their non-Latin peers. The advantage first-year Latin students showed over their non-Latin peers was of similar magnitude for the intelligence and linguistic variables (average effect size of $d_s = 0.45$ and $d_s = 0.48$, respectively). First-year Latin students also generally belonged to families with higher SES ($d_s = 0.58$), a finding in line with an older report by the Flemish government (Flemish Department of Education, 2011). Finally, first-year Latin students demonstrated higher average scores on attitudinal characteristics as well, albeit that this difference was much smaller (average effect size of $d_s = 0.22$). In sum, it appears that the entry selection takes place mainly in terms of cognitive, linguistic and SES characteristics, more than in attitudinal traits.

When inspecting the advantage first-year Latin students displayed on the intelligence variables, a discrepancy can be observed between Fluid Intelligence (Gf) and Visual Information Processing (Gv) on the one hand, and Crystallized Intelligence (Gc) and Short-term Memory (Gsm) on the other hand. Gf and Gv are focused on mathematical reasoning and visuo-spatial information processing respectively, two abilities that are only distantly related to the study of Latin. The Latin advantage in Gf and Gv was small. The opposite is true for Gc and Gsm however, as the advantage first-year Latin students displayed when compared to their non-Latin peers was quite large for these two strongly correlated constructs. Unsurprisingly, as Gc and Gsm also correlated strongly with Vocabulary, the advantage first-year Latin students displayed over their peers was of similar magnitude for Vocabulary. While advances in vocabulary knowledge is an effect often attributed to the study of Latin (Bracke & Bradshaw, 2020; Gilliland, 1922), we thus find that Latin students already performed much better on this construct at the start of secondary education. Put differently, we see that the students who chose to study Latin were already better equipped in the domain that Latin is most expected to cultivate. Taken together, these findings again provide robust evidence for the preselectivity account, as first-year Latin students demonstrated greater performances for all intelligence and linguistic variables compared to their non-Latin peers.

Next, with our second research question, we wanted to explore the evolution of these performance differences over the course of secondary education: do they decrease, remain stable, or does studying Latin result in added (transfer) effects, which would then yield even bigger performance differences? We found that no variables showed signs of far transfer, and only one variable demonstrated a trend in line with the near transfer hypothesis, namely Vocabulary. Indeed, Vocabulary was the only variable for which the advantage of Latin over non-Latin students increased towards the end of secondary education. This finding is in line with the literature on near transfer effects of studying Latin on vocabulary development (e.g., Gilliland, 1922). For all variables but Vocabulary, we observed that the difference between the Latin and non-Latin students was the largest in the second year, and the smallest in the sixth year of secondary education[9]. Furthermore, we found that the difference between the two student groups was smaller after six years of studying Latin than at the onset of secondary education, which clearly contrasts any Latin-related transfer hypothesis.

So, if the comparison between the second and sixth year of secondary education points toward the absence of transfer, the question arises as to why the comparison between the first and second year points in an opposite direction? Concretely, we found that the difference between Latin and non-Latin students increased from the first to the second year, while it decreased from the second to the sixth year, remarkably even to a difference level smaller than in the first year. Various explanations for these contradictory findings can be proposed. First, Latin curricula in Flanders are structured in such a way that the majority of basic vocabulary acquisition and grammar education takes place in the initial three years of secondary education. Conversely, the latter three years predominantly focus on translating

[9] For Gv, we even found that the non-Latin group performed better than the Latin group in the sixth year.

works from historical writers, such as Ovid and Vergil. Given the focus on vocabulary and grammar in the initial phase of Latin education, it is possible that all potential transfer takes place during this early period, thus explaining the rather transitory performance difference between Latin and non-Latin students. Second, secondary education in Flanders is characterized by a so-called ‘waterfall-system’. This phenomenon entails that students often start their studies in what is considered the most challenging study program (i.e., Latin). If the students encounter difficulties meeting graduation standards, they reorient ‘downwards’ (cfr. waterfall) to a less cognitively demanding study program (Stevens & Vermeersch, 2010). This phenomenon is clearly reflected in the Flemish enrollment numbers in Latin, as these decrease from the first to the second year (Classica Flanders, 2022). The observed surge in the performance difference between Latin and non-Latin students in the second year may thus be attributed to the fact that the remaining Latin students are the most academically adept, thus amplifying the existing disparity. Hence, the sample in the second year is likely a purified representation of the pre-selectivity effect, rather than an initial indication of cognitive transfer. Finally, a third possibility is that students who choose to study Latin may display accelerated cognitive growth compared to their non-Latin peers, rather than larger cognitive development in the end. If so, Latin students would be further ahead in their cognitive growth than non-Latin students, thus resulting in the first- and second-year performance difference.[10] As a consequence of this accelerated growth, Latin students would be significantly closer to reaching their full potential after the second year, thus leaving less room for improvement. The control group on the other hand, might display a less steep cognitive growth curve than that of non-Latin students in the first few years of secondary education, but those students then still have plenty of potential to exploit throughout the remainder of secondary education. As such, it appears that the non-Latin group ‘catches up’

[10] Such accelerated cognitive growth could also be further facilitated by the study of Latin.

with the Latin students by the time they reach the sixth year of secondary education, which then results in a performance difference potentially smaller than what can be observed in the first year.

In light of the presented evidence, our results suggest that preselection is the main, if not only, driving force behind the ‘Latin advantage’ we observed, rather than actual cognitive transfer. By exploring the cognitive implications of learning an ancient language, the present study contributes to the debate regarding the existence of cognitive transfer effects. We find very limited evidence for the existence of near transfer effects (studying Latin appears to result in increased vocabulary in L1), but we do not find support for the existence of far transfer effects, which is in line with the broader cognitive transfer literature (Gobet & Sala, 2023). As such, it appears that Latin students barely experience more cognitive transfer than their non-Latin counterparts who are also enrolled in general secondary education. Of course, our design does not allow to draw firm conclusions about the actual working mechanism behind the evolution of the observed trends. Only a longitudinal design that explores the trajectory of Latin and non-Latin students, matched in a priori pupil characteristics, would be suited to draw such conclusions. It may be interesting to note that we are in fact also conducting such longitudinal research, as we follow up the second-year students for three consecutive years. This will allow us to thoroughly investigate the existence of cognitive transfer effects due to the study of Latin.

Notwithstanding this need for longitudinal research, our well-controlled cross-sectional study leveraged a uniquely large sample that already allows us to shed light on the question ‘does studying Latin produce linguistic and/or cognitive benefits?’. Besides the theoretical importance of our findings, the present study addresses the need for research to feed the debate on the continued inclusion of Latin within the curricula in European secondary schools. The reasons for studying Latin can be quite diverse: one might opt to study Latin just because it is a challenging subject, for cultural or linguistic enrichment purposes, because of the academic advantage of Latin students in later higher edcuation, etc. (Hauspie et al., 2024). Notably, one of the reasons that is also still largely advertised towards potential stakeholders (i.e. parents and prospective students) is potential cognitive transfer effects (Vereeck et al., 2023) which refers to the belief that studying Latin would make students more cognitively fit. However, the present data suggest that studying Latin does in fact not cause larger cognitive transfer effects than the other educational tracks that were tested (i.e., non-Latin tracks). Hence, based on our findings and to the extent that future research yields similar results, we recommend that the expectations regarding cognitive transfer effects originating from studying Latin should be nuanced. Overall, we advocate for a more refined and realistic understanding of (the origins of) the benefits of studying Latin, and we emphasize the importance of empirical evidence when informing stakeholders about educational programs.

**Limitations and Future Work**

Our study presents several limitations. First, we did not account for Latin students' mastery of the ancient language. Differentiating between these students, based on their Latin proficiency levels, might yield valuable insights. Second, because we made a binary distinction between Latin and non-Latin students in the sixth year, we were not capable of assessing the differences between the specific study programs within these groups. On the one hand, Latin students have diversified their study programs to focus on Latin-sciences, Latin-languages, or Latin-mathematics. On the other hand, students from the control group stem from a variety of different study programs, such as science-mathematics, economy-languages, and science-languages, to name a few. Making a distinction between these tracks could potentially nuance the conclusions that we draw here. Hence, future research should address this by recruiting a larger number of sixth year students, which would allow for meaningful analyses on the specific subsamples. Third, our sample consists of native Dutch speakers. While a significant part of the Dutch lexicon has Latin origins (Van Keymeulen, 2008), exploring whether similar effects would be found for Romance languages (e.g., French) would be very interesting, as these languages have even stronger and clearer Latin origins. Finally, the Latin students in our sample generally studied Latin as their third or fourth language. Perhaps, learning Latin as a third or fourth language does not lead to as significant an effect as had it been the student's second language. Consequently, future work should compare pupils who study Latin as their second language with a comparable cohort of pupils studying a different, modern foreign language as a second language. Such research would allow us to explore whether the beneficial effects of Latin would be more pronounced for students who are exposed to a smaller range of foreign languages than those in our sample.

## Conclusion

The present study shows that Latin students demonstrate advantages in intelligence measures, enhanced native language proficiency and increased meta-linguistic awareness, compared to their non-Latin peers. Furthermore, Latin students demonstrate a small advantage in attitudinal traits linked to academic achievement and they generally stem from families with a higher SES. Importantly, this advantage is already present at the start of secondary education, thus confirming the existence of a cognitive preselection for the Latin study program. The performance difference between Latin and non-Latin students does not seem to widen throughout secondary education, except for vocabulary. Nevertheless, further longitudinal research is warranted to more closely address the existence of near or far transfer effects that may arise as a result of Latin education.

**References**

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. https://doi.org/10.1037/0033-2909.128.4.612

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bennett, D. (2021). A case study into pupil perceptions of Latin, conducted with a mixed-ability Year 9 Latin class at a comprehensive faith school. *Journal of Classics Teaching*, *22*(43), 38–46. https://doi.org/10.1017/S2058631021000052

Blanken, T., Dekker, K., & Van Someren, E. (2018). *How personality profile similarity can improve comparability between assessment formats: An example of the Mini-IPIP and IPIP-NEO-120 in a Dutch community sample*. https://doi.org/10.31234/osf.io/pjtgv

Bracke, E., & Bradshaw, C. (2020). The impact of learning Latin on school pupils: A review of existing data. *The Language Learning Journal*, *48*(2), 226–236. https://doi.org/10.1080/09571736.2017.1400578

Brunello, G., Esposito, P., Rocco, L., & Scicchitano, S. (2023). Do Classical Studies Open Your Mind? *SSRN Electronic Journal*, *15985*. https://doi.org/10.2139/ssrn.4381866

Carlisle, J. F., & Liberman, I. Y. (1989). Does the study of Latin affect spelling proficiency? *Reading and Writing*, *1*(2), 179–191. https://doi.org/10.1007/BF00377470

Catholic Education Flanders. (2023). *Basisoptie klassieke talen [Option Classical languages]*. https://pro.katholiekonderwijs.vlaanderen/I-KlTa-a

Çiftçi, Ş. K., & Cin, F. M. (2017). The Effect of Socioeconomic Status on Students' Achievement. In *The Factors Effecting Student Achievement* (pp. 171–181). Springer International Publishing. https://doi.org/10.1007/978-3-319-56083-0_10

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Classica Flanders. (2022). *Statistieken Vlaams Onderwijs [Statistics Flemish Education]*. http://www.classicavlaanderen.be/informatie/cijfermateriaal/index.html

De Vane, A. K. (1997). Efficacy of Latin Studies in the Information Age. *Educational Psychology*, 1–12.

Depessemier, P., & Andries, C. (2009). *Gletschr. Dyslexie bij +16-jarigen. Test voor Gevorderd Lezen en Schrijven [Gletschr. Dyslexia in +16 year olds. Test for Advanced Reading and Writing]*. Garant, Leuvel/Apeldoorn.

EULALIA. (n.d.). *Project*. Retrieved May 24, 2023, from https://site.unibo.it/eulalia/en/project

European Commission - European Education and Culture Executive Agency. (2023). *Key data on teaching languages at school in Europe*. (Eurydice report). Publications Office of the European Union. https://doi.org/10.2797/737474

European Commission - Eurydice. (2024). *Key features of the Education System in Belgium - Flemish Community*. https://eurydice.eacea.ec.europa.eu/national-education-systems/belgium-flemish-community/overview

Federation Wallonia-Brussels - Ministry of Education. (2022). *Référentiel de français et langues anciennes [Reference of French and ancient languages]*. http://enseignement.be/index.php?page=28597#documents

Flemish Department of Education. (2011). *Sociale ongelijkheid bij de overgang van basis- naar secundair onderwijs: een onderzoek naar de oriënteringspraktijk [Social inequality in the transition from primary to secondary education: a study of orientation practices]*.

https://www.vlaanderen.be/publicaties/sociale-ongelijkheid-bij-de-overgang-van-basis-naar-secundair-onderwijs-een-onderzoek-naar-de-orienteringspraktijk

Fromchuck, A. R. (1984). The Measurable Benefits of Teaching English through Latin in Elementary School. *The Classical World*, *78*(1), 25–29. https://doi.org/10.2307/4349660

Gilliland, A. R. (1922). The Effect of the Study of Latin on Ability to Define Words. *Journal of Educational Psychology*, *13*(8), 501–505. https://doi.org/10.1037/h0072236

Gobet, F., & Sala, G. (2023). Cognitive Training: A Field in Search of a Phenomenon. *Perspectives on Psychological Science*, *18*(1), 125–141. https://doi.org/10.1177/17456916221091830

Haag, L., & Stern, E. (2000). Non scholae sed vitae discimus? Auf der Suche nach globalen und spezifischen Transfereffekten des Lateinunterrichts. [Non scholae sed vitae discimus? In search of global and specific transfer effects of Latin teaching.] *Zeitschrift Für Pädagogische Psychologie*, *14*(2–3), 146–157. https://doi.org/10.1024/1010-0652.14.23.146

Haag, L., & Stern, E. (2003). In Search of the Benefits of Learning Latin. *Journal of Educational Psychology*, *95*(1), 174–178. https://doi.org/10.1037/0022-0663.95.1.174

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working Memory Training May Increase Working Memory Capacity but Not Fluid Intelligence. *Psychological Science*, *24*(12), 2409–2419. https://doi.org/10.1177/0956797613492984

Hauspie, C., Schelfhout, S., Dirix, N., Fonteyne, L., Janse, M., Szmalec, A., Vereeck, A., & Duyck, W. (2024). Does Studying Latin in Secondary Education Predict Study Achievement in Academic Higher Education? *Language Learning*. https://doi.org/10.1111/lang.12639

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Jessner, U., Török, V., & Pellegrini, C. (2018). The role of Latin in multilingual learners' strategies. *Language Education and Multilingualism*, *1*, 85–102. https://doi.org/10.18452/19035.2

Karadag, E. (2017). *The Factors Effecting Student Achievement*. Springer International Publishing. https://doi.org/10.1007/978-3-319-56083-0

Kennedy, C. J. (2006). *The study of Latin and its correlation with improved English abilities* [Doctoral dissertation]. Touro University International, Faculty of the College of Education.

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review*, *25*, 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Levine, S. Z. (2011). Elaboration on the association between IQ and parental SES with subsequent crime. *Personality and Individual Differences*, *50*(8), 1233–1237. https://doi.org/10.1016/j.paid.2011.02.016

Magez, W., Van Parijs, K., Joris, S., & Tierens, M. (2023). *COVAT-3*. Psychodiagnostisch centrum, Toegepaste Psychologie, Thomas More Antwerpen.

Onderwijskiezer. (2023a). *Gewoon secundair onderwijs [Regular secondary education]*. https://www.onderwijskiezer.be/v2/secundair/sec_gewoon_so.php

Onderwijskiezer. (2023b). *Latijn [Latin]*. https://www.onderwijskiezer.be/v2/secundair/sec_detail_2.php?detail=10831&domein=2

Piedmont, R. L. (2014). Inter-item Correlations. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3303–3304). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1493

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–328. https://doi.org/10.1037/a0014996

R Core Team. (2023). *R: A Language and Environment for Statistical Computing.* https://www.R-project.org/

Sala, G., Deniz Aksayli, N., Semir Tatlidil, K., Tatsumi, T., Gondo, Y., & Gobet, F. (2019). Near and far transfer in cognitive training: A second-order meta-analysis. *Collabra: Psychology*, *5*(1). https://doi.org/10.1525/collabra.203

Sawilowsky, S. S. (2009). Very large and huge effect sizes. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599. https://doi.org/10.22237/jmasm/1257035100

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99–144). The Guilford Press.

Sheridan, R. (1976). *Augmenting Reading Skills through Language Learning Transfer. FLES Latin Program Evaluation Reports, 1973-74, 1974-75, 1975-7*. Indianapolis: Indianapolis Public Schools.

Sparks, R. L., Ganschow, L., Fluharty, K., & Little, S. (1995). An Exploratory Study on the Effects of Latin on the Native Language Skills and Foreign Language Aptitude of Students with and without Learning Disabilities. *The Classical Journal*, *91*(2), 165–184.

Stevens, J. (2012). *Applied multivariate statistics for the social sciences.* (5th ed.). Routledge.

Stevens, P. A. J., & Vermeersch, H. (2010). Streaming in Flemish secondary schools: Exploring teachers' perceptions of and adaptations to students in different streams. *Oxford Review of Education*, *36*(3), 267–284. https://doi.org/10.1080/03054981003629862

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Thorndike, E. L. (1914). *The psychology of learning*. Teachers' College, Columbia University, New York.

Thorndike, E. L. (1924). Mental Discipline in High School Studies. *Journal of Educational Psychology*, *15*(1), 1–22. https://doi.org/10.1037/h0075386

van Gelderen, A., Schoonen, R., Stoel, R. D., de Glopper, K., & Hulstijn, J. (2007). Development of Adolescent Reading Comprehension in Language 1 and Language 2: A Longitudinal Analysis of Constituent Components. *Journal of Educational Psychology*, *99*(3), 477–491. https://doi.org/10.1037/0022-0663.99.3.477

Van Keymeulen, J. (2008). Latijnse leenwoorden in het Nederlands en de Nederlandse dialecten. [Latin loanwords in Dutch and Dutch dialects.]. *Van Mensen En Dingen: Tijdschrift Voor Volkscultuur in Vlaanderen*, *6*(1/2), 69–85.

Vander Beken, H., Woumans, E., & Brysbaert, M. (2018). Studying texts in a second language: No disadvantage in long-term recognition memory. *Bilingualism: Language and Cognition*, *21*(4), 826–838. https://doi.org/10.1017/S1366728917000360

Vandevelde, S., Van Keer, H., & Rosseel, Y. (2013). Measuring the complexity of upper primary school children's self-regulated learning: A multi-component approach. *Contemporary Educational Psychology*, *38*(4), 407–425. https://doi.org/10.1016/j.cedpsych.2013.09.002

Vereeck, A., Bracke, E., De Herdt, K., & Janse, M. (2024). Revered and reviled. An outline of the public debate regarding classical language education. *Journal of Classics Teaching*. https://doi.org/10.1017/S2058631024000424

Vereeck, A., Janse, M., De Herdt, K., Hauspie, C., Szmalec, A., & Woumans, E. (in press). Does studying Latin make pupils smarter? Presenting the field of classical language impact studies. *The Classical Journal*.

Vereeck, A., Janse, M., De Herdt, K., Szmalec, A., Hauspie, C., & Duyck, W. (2023). Why Plato needs psychology: Proposal for a theoretical framework underpinning research on the cognitive transfer effects of studying classical languages. *Psihološka Obzorja / Horizons of Psychology*, *32*, 121–130. https://doi.org/10.20419/2023.32.580

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174–1204. https://doi.org/10.1037/a0036620

## Appendix A

**Table A.1**

*Correlation Matrix*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1. Test moment** | - | | | | | | | | | |
| **2. Gender** | .02 | - | | | | | | | | |
| **3. Degree parents** | -.08 | -.09 | - | | | | | | | |
| **4. Dutch** | -.01 | -.02 | 0.12*** | - | | | | | | |
| **5. Learning disorder** | .00 | .00 | 0.05 | 0.03 | - | | | | | |
| **6. Year** | -.18*** | .07 | 0.08 | 0.07 | 0.08 | - | | | | |
| **7. Study program** | -.03 | .03 | 0.24*** | 0.06 | -0.13*** | -0.09 | - | | | |
| **8. Spelling** | .03 | .04 | 0.14*** | 0.13*** | -0.14*** | 0.51*** | 0.22*** | - | | |
| **9. Reading comprehension** | -0.04 | .09 | 0.17*** | 0.08 | 0.01 | 0.43*** | 0.20*** | 0.45*** | - | |
| **10. Vocabulary** | -.04 | .01 | 0.18*** | 0.07 | -0.01 | 0.74*** | 0.18*** | 0.60*** | 0.50*** | - |
| **11. Phonological awareness** | -.05 | .03 | 0.0 | 0.06 | -0.05 | 0.33*** | 0.14*** | 0.42*** | 0.33*** | 0.36*** |
| **12. Syntactic awareness** | -.09 | .12*** | 0.15*** | 0.08 | -0.05 | 0.47*** | 0.12*** | 0.45*** | 0.39*** | 0.49*** |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **13. G** | -.07 | .07 | 0.18*** | 0.09 | -0.05 | 0.50*** | 0.24*** | 0.56*** | 0.52*** | 0.59*** |
| **14. Gf** | -.05 | .09 | 0.07 | 0.03 | -0.01 | 0.18*** | 0.11** | 0.25*** | 0.24*** | 0.20*** |
| **15. Gc** | -.09 | -.01 | 0.23*** | 0.11** | -0.03 | 0.58*** | 0.25*** | 0.58*** | 0.54*** | 0.69*** |
| **16. Gsm** | -.02 | .15*** | 0.07 | 0.05 | -0.04 | 0.48*** | 0.18*** | 0.48*** | 0.38*** | 0.50*** |
| **17. Gv** | -.01 | -.04 | 0.10* | 0.06 | -0.05 | 0.07 | 0.09 | 0.19*** | 0.24*** | 0.17*** |
| **18. Autonomous motivation** | -.08 | .10* | 0.04 | 0.05 | 0.00 | -0.02 | 0.15*** | 0.02 | 0.06 | 0.05 |
| **19. Academic self-efficacy** | -.01 | .08 | -0.01 | 0.02 | -0.02 | -0.18*** | 0.10* | -0.08 | -0.08 | -0.11** |
| **20. Conscientiousness** | -.04 | .08 | -0.02 | 0.07 | -0.04 | -0.06 | 0.10 | 0.01 | 0.02 | -0.01 |

| | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** |
|---|---|---|---|---|---|---|---|---|---|---|
| **11. Phonological awareness** | - | | | | | | | | | |
| **12. Syntactic awareness** | 0.34*** | - | | | | | | | | |
| **13. G** | 0.44*** | 0.45*** | - | | | | | | | |
| **14. Gf** | 0.24*** | 0.19*** | 0.67*** | - | | | | | | |
| **15. Gc** | 0.37*** | 0.47*** | 0.76*** | 0.27*** | - | | | | | |
| **16. Gsm** | 0.34*** | 0.38*** | 0.72*** | 0.24*** | 0.48*** | - | | | | |
| **17. Gv** | 0.23*** | 0.15*** | 0.59*** | 0.32*** | 0.24*** | 0.23*** | - | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **18. Autonomous motivation** | 0.02 | 0.06 | 0.06 | 0.05 | 0.04 | 0.06 | 0.02 | - | | |
| **19. Academic self-efficacy** | -0.05 | -0.05 | -0.10* | -0.04 | -0.11** | -0.06 | -0.05 | 0.53*** | - | |
| **20. Conscientiousness** | -0.01 | 0.05 | 0.02 | 0.02 | 0.00 | 0.00 | 0.04 | 0.49*** | 0.62*** | - |

*Note. N* = 1,547. Effect coding (which was used in the analyses reported throughout the paper) was not applied to the data to develop this correlation matrix: Gender: 1 = female, 0 = male; learning disorder: 0 = none, 1 = one; Dutch: 0 = non-native, 1 = native; study program: 1 = Latin, -1 = non-Latin; Year: 1 = first year, 2 = second year, 6 = sixth year.

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

## Appendix B

**Table B.1**

*Linear mixed models per dependent variable*

| Variable | Adding study program to model | $R^{2\,a}$ | Model |
|---|---|---|---|
| G | $\chi^2$ (1, 1491) = 65.46, p < .001 | .38 | -0.11 – 0.58 x Year 1 – 0.18 x Year 2 – 0.04 x Gender + 0.06 x Degree parents + 0.15 x Learning disorder – 0.13 x Dutch + 0.03 x Testing moment + 0.06 x Conscientiousness + 0.06 x Autonomous motivation – 0.10 x Self-efficacy + 0.24 x Study program |
| Gf | $\chi^2$ (1, 1538) = 9.99, p = .37 | .12 | -0.05 – 0.120x Year 1 – 0.07 x Year 2 – 0.08 x Gender + 0.05 x Degree parents + 0.07 x Learning disorder – 0.06 x Dutch + 0.03 x Testing moment + 0.03 x Conscientiousness + 0.06 x Autonomous motivation – 0.08 x Self-efficacy |
| Gc | $\chi^2$ (1, 1544) = 49.89, p < .001 | .52 | -0.12 – 0.24 x Year 1 – 0.13 x Year 2 + 0.03 x Gender + 0.08 x Degree parents + 0.08 x Learning disorder – 0.19 x Dutch + 0.00 x Testing moment + 0.04 x Conscientiousness + 0.01 x Autonomous motivation – 0.06 x Self-efficacy + 0.45 x Score vocabulary test + 0.16 x Study program + 0.07 x Latin:Female |
| Gv | $\chi^2$ (1, 1531) = 7.22, p = 0.99 | .09 | -0.25 – 0.10 x Year 1 + 0.02 x Year 2 + 0.03 x Gender + 0.07 x Degree parents + 0.16 x Learning disorder – 0.11 x Dutch + 0.01 x Testing moment + 0.09 x Conscientiousness + 0.05 x Autonomous motivation – 0.12 x Self-efficacy |
| Gsm | $\chi^2$ (1, 1527) = 30.44, p < .001 | .33 | -0.06 – 0.37 x Year 1 – 0.08 x Year 2 - 0.11 x Gender – 0.03 x Degree parents + 0.09 x Learning disorder – 0.04 x Dutch + 0.03 x Testing moment + 0.01 x Conscientiousness + 0.03 x Autonomous motivation – 0.02 x Self-efficacy + 0.24 x Score vocabulary test + 0.15 x Study program |
| Spelling | $\chi^2$ (1, 1484) = 39.78, p < .001 | .45 | -0.52 – 0.46 x Year 1 – 0.15 x Year 2 + 0.02 x Gender + 0.03 x Degree parents + 0.40 x Learning disorder – 0.18 x Dutch + 0.31 x G + 0.06 x Testing moment + 0.02 x Conscientiousness – 0.01 x Autonomous motivation – 0.00 x Self-efficacy + 0.15 x Study program |

| | | | |
|---|---|---|---|
| Vocabulary | $\chi^2$ (1, 1491) = 79.25, p < .001 | .65 | 0.10 – 0.72 x Year 1 – 0.26 x Year 2 + 0.05 x Gender + 0.05 x Degree parents + 0.10 x Learning disorder + 0.04 x Dutch + 0.23 x G + 0.03 x Testing moment + 0.00 x Conscientiousness + 0.04 x Autonomous motivation + 0.01 x Self-efficacy + 0.21 x Study program - 0.10 x Year 1:Latin + 0.06 x Year 2:Latin |
| Reading comprehension | $\chi^2$ (1, 1459) = 23.13, p < .001 | .34 | 0.02 - 0.27 x Year 1 – 0.08 x Year 2 - 0.06 x Gender + 0.05 x Degree parents – 0.05 x Learning disorder – 0.07 x Dutch + 0.34 x G + 0.03 x Testing moment + 0.03 x Conscientiousness + 0.03 x Autonomous motivation – 0.06 x Self-efficacy + 0.13 x Study program |
| Phonological awareness | $\chi^2$ (1, 1489) = 7.09, p = .99 | .22 | -0.13 – 0.29 x Year 1 + 0.04 x Year 2 + 0.01 x Gender + 0.00 x Degree parents + 0.17 x Learning disorder – 0.05 x Dutch + 0.29 x G + 0.01 x Testing moment + 0.00 x Conscientiousness + 0.01 x Autonomous motivation – 0.02 x Self-efficacy |
| Syntactic awareness | $\chi^2$ (1, 1485) = 7.55, p = .99 | .31 | -0.14 – 0.39 x Year 1 – 0.09 x Year 2 – 0.09 x Gender + 0.08 x Degree parents + 0.19 x Learning disorder – 0.10 x Dutch + 0.24 x G + 0.00 x Testing moment + 0.08 x Conscientiousness + 0.03 x Autonomous motivation – 0.04 x Self-efficacy |

*Note:* Gender: 1 = female, -1 = male; learning disorder: 1 = none, -1 = one; Dutch: 1 = non-native, -1 = native; study program: 1 = Latin, -1 = non-Latin.

[a] Nagelkerke's $R^2$.