

ὁ δὲ ἀνεξέταστος βίος οὐ βιωτὸς ἀνθρώπῳ

the unexamined life is not worth living

~ Plato, Apology 38a

Non scholae sed vitae?

The Cognitive Transfer Effects of Studying Classical
Languages in Flemish Secondary Education

Alexandra Vereeck

Supervisors:

prof. dr. Evy Woumans (Department of Translation, Interpreting and Communication)

prof. dr. Katja De Herdt (Department of Literary Studies)

prof. dr. Wouter Duyck (Department of Experimental Psychology)

prof. dr. Arnaud Szmalec (Department of Experimental Psychology)

Dissertation submitted to Ghent University in partial fulfilment of the
requirements for the degree of Doctor of Linguistics and of Psychology

2025

Preface: On Being and Becoming a Classicist

Dear reader of this dissertation,

Before I adopt the detached and neutral role of a researcher, striving for an objective account *sine ira et studio* and free from other kinds of biases, I will seize the opportunity a preface provides and consider the personal significance of this doctoral research on the cognitive impact of classical language learning. More particularly, before I approach said topic in a scientific and scholarly manner, I want to take the liberty of articulating my own experiences of becoming, being, becoming anew, and finally being a classicist. In doing so, I will make a genuine attempt to address the question of what it means to be a classicist, beyond holding a degree in classics.

Alternatively, this preface can be seen as a descriptive case study of sorts of one particular classical language learner, as well as an individual-level illustration of the controversiality of classical language education at a societal level, which will be discussed in Chapter 1. Given the autobiographical and introspective nature of the following essay, I have chosen to compose it in Dutch. It was not written as an afterthought to the research conducted, as prefaces often are; rather, it is the product of reflections unfolding throughout my PhD trajectory, in the course of which I have felt the least and the most like a classicist I ever have.

Yours truly,

Ἀλεξάνδρα

Het onderzochte leven

Het is op zich niet ongewoon, zo schijnt me toe, dat wetenschappers en geleerden een persoonlijke connectie ervaren met hun onderzoek. Die betrokkenheid en gevoelens van identificatie gaan in mijn geval evenwel nog een stap verder: ik zou zelfs durven stellen dat ik in zekere zin mijn onderzoeksonderwerp ben. Ik heb namelijk de uitzonderlijke kans gekregen om een doctoraat te puren uit een existentiële kwestie die mij als persoon al veel langer bezighoudt, en op rigoureuze wijze onderzoek te doen naar mijn eigen levenskeuzes. Het is vast geen toeval dat toen mij ooit (jaren voordat van dit doctoraal mandaat zelfs maar sprake was) gevraagd werd mezelf voor te stellen aan de hand van een citaat, ik zonder aarzelen koos voor “een leven dat niet is onderzocht, is het niet waard te worden geleefd”. Deze socratische wijsheid, waarmee mijn oud-leraar Latijn ons indertijd het belang van de filosofie bijbracht, heb ik steeds op alle vlakken ter harte genomen. Een dergelijke houding lijkt me in het bijzonder gepast in mijn hoedanigheid van aspirant-doctor *Philosophiae*.

Deze dissertatie is gewijd aan de bestudering van de studie van klassieke talen. Het vraagstuk van de zin en waarde van klassieketalenonderwijs heeft mij als een schaduw achtervolgd sinds ik, meer dan een decennium geleden, de wens kenbaar maakte om aan de universiteit Latijn verder te studeren. Dat ik Latijn samen met de middelbare school achter mij zou laten, kon ik me simpelweg niet voorstellen. Het was misschien een ietwat aparte droom, maar voor mij lagen in het woord “classicus, -a” een zekere magie en romantiek besloten. Classica, ja, dat was wat ik worden wilde – en liefst geen halve maar een hele, dus koos ik er uiteindelijk voor om Latijn te combineren met een tweede, mij toen nog onvertrouwde, zogenaamd niet-levende taal: het Oudgrieks.

Meteen na mijn enthousiaste liefdesverklaring aan de klassieke talen kwam ik erachter dat dit levenspad, dat ik op het punt stond met volle overtuiging in te slaan, gepaard ging met – *sit venia verbo* – een bepaalde prijs. Tot grote ontzetting van mijn jongere zelf bleek mijn studiekeuze haast als een daad van rebellie te worden beschouwd, tenminste in de context van een samenleving die misschien wel tuk is op de Oudheid in musea of cinema, maar minder opheeft met mensen die hun leven wijden aan de studie van oude talen, terwijl dezen ook dokters, ingenieurs of advocaten hadden kunnen worden. Hoewel ik me stiekem ook een beetje verkneukelde in dat recalcitrante aspect van het classicus-worden, heeft de weerstand die mijn studiekeuze opriep een diepe indruk op mij nagelaten. Vandaar ben ik het altijd als mijn taak

blijven aanvoelen om op een dag met een repliek te komen, en antwoorden te bieden op het “waarom” en “waartoe” van de klassieke talen.

Gedurende enige tijd kon ik mezelf wijsmaken dat deze heugelijke dag van repliek vanzelf en volautomatisch zou volgen op mijn afstuderen. Wanneer ik na vier jaren hard studeren een volleerde classica zou zijn geworden, dan zou ik toch zeker wel tot inzicht zijn gekomen en alle antwoorden in handen hebben? En ik studeerde af, en de openbaring bleef uit. Wel wist ik dat ik nooit spijt had gehad van mijn studiekeuze, zelfs voor geen ogenblik, maar wat nu dé meerwaarde was van klassieke talen in het dagelijkse leven, dat wist ik nog steeds niet, of toch niet zo scherp omlijnd en helder geformuleerd als ik had gewild. Mettertijd werd me duidelijk dat deze zaak veel omvattender en complexer was dan ik aanvankelijk had gedacht, en dat ik haar onmogelijk in mijn eentje zou kunnen beslechten.

Vrouwe Fortuna scheen me dan ook een knipoog te geven toen ik op een goede herfstdag in 2018 kwam aanwaaien aan de UGent, net op een geschikt moment om te horen over sudderende ideeën voor een onderzoeksproject omtrent klassieke talen en cognitieve transfer. Twee jaar en één FWO-aanvraag later kon ik vervolgens – zoals ik het in mijn *motivation statement* omschreef – “beginnen werken aan een wetenschappelijk onderbouwd antwoord op een vraag die niet alleen mijzelf in haar greep hield, maar eveneens van groot belang was voor het Vlaamse en bij uitbreiding Europese onderwijsstelsel”, namelijk: waarom Latijn leren? Of iets preciezer: zijn er bepaalde elementen eigen aan klassieketalenstudie die kunnen transfereren van gespecialiseerde schoolse kennis naar tastbare cognitieve vaardigheden die een breder nut hebben in het leven? De senecaanse tegenstelling tussen *vita* en *schola* is in het onderwijsdebat nimmer veraf.

De zoektocht naar transfer en naar de essentie van classicusschap

Een dergelijke aangelegenheid vraagt in feite niet om de filologische aanpak van een classicus, maar wel om empirisch onderzoek door een psycholoog. Bijgevolg vulde ik beetje bij beetje mijn doorleefde begrip van het onderwerp aan met de vereiste theoretische en methodologische expertise. Ik dook de wereld van de statistiek in, overigens met veel genoegen

en bewondering. Zoals zinsontleding voor mij als leerling de methode bij uitstek was om de woordenwirwar van complexe Latijnse zinnen systematisch te doorgronden, zo bleek statistiek voor ingewijden een krachtige kunst om voor het blote oog volstrekt oninzichtelijke datasets hun onderliggende betekenis te laten prijsgeven. Daarnaast verdiepte ik me in het intrigerende vakgebied van de cognitieve psychologie, van werkgeheugen tot Wechsler schalen, van transferliteratuur tot tweedetaalverwerving, van replicerbaarheidscrisissen tot refereren-volgens-APA, en zoveel meer. Zodoende werd ik stapje voor stapje, met enige hulp maar vooral veel zelfstudie, psychologe.

Terwijl de maanden zich al gauw vermenigvuldigden tot jaren en ik me gretig toelegde op de psychologische kant van mijn doctoraat, kwam ik ondertussen alsmaar meer los van mijn klassieke cocon. Langzaam en ongemerkt verloor ik de voeling met die passie voor Grieks en Latijn, die ik eens als een zo wezenlijk en kenmerkend onderdeel van mezelf had beschouwd. Schijnbaar plotseling was ik op een punt aanbeland dat ik nauwelijks meer wist wat het voor mij inhield om een classica te zijn. Was dan toch de tijd aangebroken, door sommigen lang verhoopt en door mij onmogelijk geacht, dat ik de letteren voorgoed vaarwel zou wuiven? Ik ging de vertwijfeling te lijf zoals iedere andere crisis: met nadenken en schrijven. In de loop van 2023 had ik een handvol inspirerende ontmoetingen met psychologen en classici, zowel oude bekenden als nieuwe gezichten, in binnen- en buitenland, die grondige introspectie onontkoombaar maakten en dit voorwoord deden ontkiemen.

Ik zag me genoopt terug te keren naar een andere vraag waar ik reeds als student mee dolde: wat betekent het eigenlijk classicus te zijn? Deze kwestie zou me, naast de kwestie van transferwaarde, doorheen de tweede helft van mijn doctoraat regelmatig bezighouden. In eerste instantie kwam ik op de proppen met een resem eigenschappen en ervaringen die als karakteristiek kunnen worden beschouwd voor classici. Ik verwoordde ze, met inbegrip van de voetnoten, als volgt.

Classicus-zijn is spontaan in twee- en driedigdheden denken, overal antithesen of parallellismen zien, lettergrepen tellen in zowel je poëzie als proza. Het is de eeuwen in je handpalm houden en keer op keer merken dat er weinig nieuws onder de zon is, omdat mensen altijd mensen zijn geweest en zullen zijn. Het is doordrongen zijn van de gedachte dat het leven meer behelst dan het meet- en verklaarbare, dat wetenschap

datgene mag zijn waarmee we de wereld stelselmatig simplificeren,¹ maar cultuur datgene is wat ons tot mens maakt en ons helpt te leven, al helpt het niet.² Classicus-zijn is ook genoeg scheppen in de veelzijdige schoonheid van je studieobject, in het herkenbare, het exotische, en soms obscure ervan. Het is trouw blijven aan het plezier en de verwondering die de klassieke talen je hebben geschonken, en durven keuzes maken die tegen dominante tendenzen in de maatschappij ingaan. Het is een buitenbeentje zijn, het is steeds opnieuw je bestaansrecht in vraag stellen, het is in permanente identiteitscrisis verkeren.

Kortom, ik begon in te zien dat ik me als classica op een heel andere manier tot mijn vakgebied verhoud dan pakweg het geval lijkt te zijn voor de meeste psychologen. Alles samengenomen voelde ik classicus-zijn aan als een enorm voorrecht, maar ook als een last, die in de loop van mijn doctoraatstraject steeds zwaarder op mij begon te wegen. De last wordt weliswaar enigszins verlicht door de sterke samenhang die classici zo typeert. Behalve de liefde voor klassieke talen zijn de ogenschijnlijk eindeloze verantwoordingsplicht en de voortdurende noodzaak om het vakgebied te verdedigen volgens mij net hetgeen wat classici verbindt en tot een hechte gemeenschap maakt. In andere contexten³ heb ik al gekscherend verwezen naar de “telepatie der classici”. Een collega-classicus leren kennen blijft voor mij dan ook een aparte belevenis. Het is alsof er bij de vaststelling van het gemeenschappelijk classicus-zijn prompt een zeker wederzijds begrip tot stand komt, een soort samenzweerderigheid haast, doordat beide partijen de vreugde en de pijn kennen die het beminnen van de klassieke talen met zich meebrengt.

Gedurende een al bij al korte periode was ik het gevoel van classicus-zijn verloren, maar via reflectie had ik het weer teruggewonnen. Door een ander vakgebied te omarmen en leraren en academici uit verschillende vakgebieden te ontmoeten kwam ik tot het inzicht dat classicus-zijn doorsijpelt in een heleboel kleine en grote dingen: mijn omgang met taal, met medemensen,

¹ Naar Karl Popper: “Science may be described as the art of systematic oversimplification—the art of discerning what we may with advantage omit.”

² Naar Herman De Coninck: “Zoals je tegen een ziek dochttertje zegt: / mijn miniatuurmensje, mijn zelfgemaakt / verdrietje, en het helpt niet; / zoals je een hand op haar hete voorhoofdje / legt, zo dun als sneeuw gaat liggen, / en het helpt niet: / zo helpt poëzie.”

³ Deze alinea is gebaseerd op een stukje van de openingsrede die ik heb uitgesproken op het eerste Interuniversitair Studentensymposium voor klassieke talen en cultuur, dat op 10 november 2023 plaatsvond te Gent.

met de wereld. Als classica heb ik oog voor de stilistische kwaliteiten van een tekst, ben ik initieel sceptisch wanneer ideeën als gloednieuw worden voorgesteld, frons ik mijn wenkbrauwen bij het zien van literatuuroverzichten die niet verder terug gaan dan een paar decennia, en dies meer. De ene dag vraag ik me af of ik niet gewoon een sofist ben geworden, een woordkunstenaar die iedere zwakte kan omzetten in een sterkte, en die hetzelfde onderzoek steeds opnieuw in overtuigende abstracts kan gieten die naadloos aansluiten bij de thema's van de meest uiteenlopende congressen. De andere dag waan ik me te streven naar Platoonse Ideeën van het Ware en het Goede.

Met een zeker pathos, mij niet vreemd, concludeerde ik in de eerste versie van dit voorwoord dat classicus-zijn niets minder is, of alleszins kan zijn, dan een levenshouding. Is dat niet fraai? Als idealistische abituriënt zou ik de essentialistische benadering van de voorgaande alinea's geweldig hebben gevonden, maar ik ben ouder nu, en een aantal illusies armer. Hoe fraai ik dit alles ook vind, en ondanks het feit dat ik het geloof erin niet heb losgelaten, toch becroop me een zekere ontevredenheid. Ik had het onbehaaglijke gevoel dat ik niet verder was geraakt dan een aantal oppervlakkige vakclichés, die soms zelfs dreigden over te hellen naar vakidiotie, zoals het denken in stijlfiguren. Ik kon me niet van de indruk ontdoen dat ik nog niet was doorgedrongen tot de ware kern van de zaak.

Op een gegeven moment had ik een inval. Is dit in wezen niet waarom we klassieke talen studeren, dacht ik plots: om de mens te leren kennen, en onszelf en anderen beter te begrijpen? Zelf geloof ik graag dat ik er een beter mens van ben geworden. Ruimdenkender, empathischer. Daartegenover staat dat, als het al zo is dat ik me dankzij het lezen van antieke teksten beter in anderen kan verplaatsen, dat dan toch niet lijkt te hebben gewerkt voor alle classici die ik ken. Ook classici zijn er van alle slag, zoveel weet ik allang. Ik ken fantastische exemplaren aan wie ik zonder aarzelen mijn leven zou toevertrouwen, en anderen die me zo hebben teleurgesteld dat ik niet anders kan dan me afvragen of ze helemaal niets hebben opgestoken van hun studie. Wie weet zou ook ik altijd al de persoon zijn geworden die ik nu ben, en is "classica" niet meer dan een naam die ik er toevallig aan geef. Misschien is er niet daadwerkelijk zoiets als "de classicus". Misschien bestaan "classici" net zomin als "de oude Grieken" ooit bestaan hebben buiten de hoofden van romantisch aangelegde geleerden.

En toch. Wat zijn de klassieke studies dan een mooie en beproefde weg om jezelf te worden: een weg van taal, kunst en filosofie. Latijn en Grieks zijn bij uitstek vakken waarin je heel de *condition humaine* verkent, in mythologie, geschiedschrijving, retoriek... en dat via het mateloos fascinerende medium van de taal – antieke talen, tot de hoogste kunstvormen gebracht. Als

Latijn een set grafietpotloden is, netjes geordend van 8B tot 8H, dan is Oudgrieks een doos aquarel: grilliger, onvoorspelbaarder, het vereist meer oefening om ermee om te gaan, maar de mogelijkheden zijn eindeloos. Mijn taalkundig-psychologische doctoraat mag dan bijzonder ver afstaan van klassieke filologie, ik houd nog steeds van Latijn en Grieks en alles wat ermee te maken heeft. Een paar jaar geleden heb ik enkele goede vrienden eens gevraagd om mij te beschrijven – een experiment dat ik iedereen alleen maar kan aanraden. Een van hen begon meteen met “je enthousiasme voor *all things classics*”. Het is tot op de dag van vandaag een van mijn favoriete complimenten die ik ooit heb gekregen.

Een onapologetische apologie

Laat ik even op mijn stappen terugkeren. Een van de weliswaar vele redenen waarom ik dit doctoraat had aangevat was om met uitstel mijn studiekeuze te verantwoorden. Ik wenste het proefschrift te schrijven dat ik in het zesde middelbaar had willen kunnen overhandigen aan degenen die me toen het vuur aan de schenen legden. Tijdens mijn beschouwingen over classicusschap begon het me evenwel te dagen dat ik niet alleen de tegenpartij moest overtuigen. De tegenkanting die ik had ondervonden bij mijn keuze om klassieke talen te gaan studeren had ook intern conflict en onzekerheid veroorzaakt. Ergens was ik altijd de sombere vrees blijven koesteren dat de sceptici het bij het rechte eind hadden, en dat de klassieke studies niets meer waren dan een boeiende tijdverspilling. Het kwam er dus op aan mezelf te overtuigen, en aan mezelf verantwoording af te leggen voor mijn levenskeuzes.

Zo kwam ik tot het besef dat dit doctoraatsproject wel eens een verregaande reflectieoefening kon zijn om zélf in het reine te komen met mijn eigen classicus-zijn, een manier om de wonden te helen die ik had opgelopen tijdens mijn studiekeuzeproces voor het hoger onderwijs. De eigenlijke dissertatie zou zich buigen over cognitieve transfer, ingebed in ruimere beschouwingen over de waarde van klassieketalenonderwijs en de discussie daaromtrent, en het voorwoord zou culminerend in een weldoordachte lofrede op de eigenheid van het classicusschap. Zoals hoger duidelijk werd, liep ik echter vast in mijn reflecties. Mijn gedachten over het concept “classici” bevonden zich een tijdlang in een limbo. Wellicht is de vraag naar wat het in algemeen-geldende termen betekent classicus te zijn op zichzelf genomen

essentialistisch, en gedoemd te eindigen in aporie. Hieronder volgt een nieuwe poging om te verwoorden wat het voor mij persoonlijk betekent.

Ik ben Latijn gaan en blijven studeren, en later ook Grieks, niet vanwege enig argument over het nut of de waarde ervan, maar vanuit een oprechte interesse. In het bijzonder in het laatste jaar van de middelbare school brachten het vak Latijn en degene die het mij onderwees een zin in mijn jonge leven waar ik die voorheen miste. Ook van mijn universitaire opleiding heb ik intens genoten (op het blokwerk na, kan ik niet nalaten met een kwinkslag toe te voegen). Nog steeds identificeer ik me graag met mijn diploma klassieke talen. Ik geloof dat er zaken zijn die typisch zijn voor classici, dat de gemiddelde classicus en psycholoog met een andere blik naar de dingen kijken en een andere relatie hebben tot hun vakgebied. Daarom hoeven die verschillen niet per se zo verregaand te zijn dat er sprake is van een andere levenshouding. Ik geloof dat mijn studies me meer hebben bijgebracht dan kennis van Latijn, Grieks of de Oudheid, en dat ze mijn taal en mijn denken fundamenteel hebben vormgegeven.

Bovenal heb ik het verlangen om van betekenis te zijn voor de gemeenschap van classici en met mijn werk andere classici te helpen, al hebben zij maar een fractie van de existentiële twijfel ervaren die ik zelf heb doorgemaakt. Ik wil het klassieketalenonderwijs ten dienste zijn zoals het mij ten dienste is geweest. Op de moeilijke momenten, die onvermijdelijk deel uitmaken van ieder doctoraatstraject, was het met voorsprong dit hogere doel dat me op de been hield en me motiveerde om te blijven doorzetten. Ik ben nooit vergeten waarvoor ik het in de allereerste plaats deed. Ook dat betekent het voor mij om een classica te zijn: de verantwoordelijkheid mijn vak te beschermen.

Aldus bleek doctoreren een manier om opnieuw classica te worden, niet meer met een onderhuids schuldgevoel, maar met dezelfde fierheid als van een tiener die voor het eerst de wijde wereld in wordt gestuurd en niets mooiers kan bedenken dan zich toe te leggen op oude talen en teksten. Om heel eerlijk te zijn, waren het uiteindelijk noch wetenschap noch filosofie die mij heil brachten. Het waren de mensen in mijn omgeving die onbewust een radicale mentaliteitsverschuiving in mij teweegbrachten. Ik zoek nu niet langer goedkeuring om mezelf te zijn, en weiger me nog te verontschuldigen voor wie ik ben. Wat ook de uitkomst van mijn empirische studie moge zijn – op het moment dat ik deze woorden schrijf, is dat nog niet geheel duidelijk – hoop ik eindelijk een begin te maken met voorgoed en vol trots te zijn wat ik worden wilde, misschien soms meer of minder uitgesproken, maar steeds ontegensprekelijk: classica.

Acknowledgements

One of my teachers, who taught me the majority of what I know about the Ancient Greek language, once wrote to me in the course of a mundane e-mail conversation: “Problems exist to be solved”. Looking back, I sometimes think this single sentence changed my life. At its core, pursuing a PhD means becoming a professional problem-solver. The following words are dedicated to those who put problems on my plate and those who helped me solve them. Since they are all Dutch-speaking, as I am myself, I will acknowledge them in Dutch.

De Ander: je kan niet zonder en je kan niet met. De vele mensen met wie ik de afgelopen jaren te maken heb gehad waren aan dit doctoraat het allermooiste, maar ook het allermoeilijkste – de ingewikkeldste statistische analyse of het lastigste verwoordingsvraagstuk verbleekt er volkomen bij. Niettemin kan ik in volle eerlijkheid stellen dat ik doordrongen ben van dankbaarheid, zowel voor de kansen die ik heb gekregen alsook voor de mensen om me heen.

Wie mij een beetje kent, weet hoe belangrijk ik het altijd heb gevonden mijn dankbaarheid te uiten, en ik overdrijf geenszins wanneer ik zeg dat ik vanaf het prille begin van mijn aspirantschap in gedachten bezig ben geweest met mijn dankwoord. Wanneer ik aan mijn dankwoord dacht, stelde ik me stevast de gesproken dankrede voor die ik aan het eind van mijn publieke verdediging zou houden, in de aanblik van zoveel mogelijk dierbare gezichten. Pas in een veel later stadium kwam een spreekwoordelijke euro te vallen: het decorum wil natuurlijk dat ik ook de schriftelijke neerslag van mijn doctoraat met enkele woorden van dank tooi. Aangezien ik toch een en ander wil bewaren voor die langverwachte dankrede, zal ik het hier voor mijn doen onkarakteristiek kort houden.

Mijn bijzondere dank gaat uit naar:

- Mark en Katja, om mijn *Doktoreltern* te willen zijn. Met jullie beiden is het aanvankelijk allemaal begonnen – of hoe een paar op goed geluk verstuurd mailtjes de volgende jaren van je leven kunnen bepalen. Wat heb ik me al vaak gelukkig geprezen dat ik uitgerekend bij jullie terechtkwam, twee van de meest warme, milde en ruimhartige mensen die ik ken. Jullie niet-aflatende steun en vertrouwen zijn werkelijk van onschatbare waarde geweest voor mij, zowel op professioneel als op persoonlijk vlak.
- Evy, om de rol van reddende engel te vervullen en naar aanleiding van Marks vervroegde emeritaat het promotorschap over te nemen. Een promotorwissel halverwege het doctoraatstraject is voor niemand evident, maar samen hebben we het tot een goed einde gebracht.
- Wouter, voor het openen van de deur naar de psychologie en naar je lab.
- Arnaud, voor je bereidheid om naast het doctoraat van Cathy ook dat van mij op te volgen.
- De talrijke collega's aan de UGent en daarbuiten die zijn ingegaan op al mijn vragen en verzoeken. Zoals bij ieder gedegen wetenschappelijk onderzoek, heb ook ik veelvuldig gebruik gemaakt van de tips, inzichten en materialen van andere onderzoekers, en daarvoor ben ik hen zeer erkentelijk.
- Mijn mede-doctorandi en -doctorandae, voor jullie gezelschap, de praktische en morele ondersteuning, en de aangename gesprekken over werk en over alles behalve werk. Dankzij jullie houd ik aan mijn tijd als doctoraatsstudent vele fijne herinneringen over.
- Lidewij en de andere collega's aan de UvA, om mij met open armen te ontvangen in Amsterdam en bij te dragen aan wat voor mij een onvergetelijk verblijf was.
- Alle leraren, directieleden, andere schoolpersoneelsleden en leerlingen die dit doctoraat mede mogelijk hebben gemaakt. Ik heb veel tijd en inspanning van jullie gevraagd, en zonder jullie medewerking was er eenvoudigweg geen onderzoek geweest.
- De studenten die interesse toonden in mijn onderzoek en mij met enthousiasme hebben geassisteerd tijdens grote dataverzamelingssessies. Jullie aanwezigheid betekende soms een wereld van verschil.
- Mijn voormalige leraren, professoren en promotoren binnen de klassieke talen, die mij ieder op hun manier hebben geholpen om de stappen te zetten die uiteindelijk naar dit doctoraat zouden leiden.

- Mijn ouders, zonder wie ik misschien niet aan een doctoraat was begonnen, en zonder wier economische geesten ik de zin van klassieketalenonderwijs wellicht nooit zozeer in vraag had gesteld.
- Mijn goede vrienden, om er altijd te zijn en mij te aanvaarden zoals ik ben. Ook al zie ik niet iedereen van jullie even vaak, ik heb de afgelopen jaren onbeschrijflijk veel aan jullie gehad. Jullie zijn mijn uitlaatklep, mijn therapeuten, mijn zelfgekozen familie.
- Andy, om mijn wederhelft te zijn, in goede en kwade tijden, εἰς αἰῶνα. Bedankt om steeds weer voor klankbord te willen spelen, mij steeds weer te willen helpen mijn gedachten op een rijtje te zetten, en mij iedere dag opnieuw alle liefde en partnerschap te bieden die ik me kan wensen.

Table of Contents

General Introduction.....	1
<i>Non scholae sed vitae?</i>	3
Dissertation roadmap	6
Works cited in the General Introduction.....	8
 Part I: Status Quaestionis.....	 9
Introductory remarks to Part I.....	11
Cohesion between the chapters	11
The value of classical language education.....	12
Grammar and translation.....	12
 Chapter 1 The Public Debate on Classical Language Education	 17
1.1 Introduction.....	17
1.2 Delineation of topic and method.....	18
1.3 The state of classical language education in Western Europe and the US	20
1.3.1 The bloom of classics and the seeds of doubt	20
1.3.2 From fundament to <i>Fremdkörper</i> in the 20th century	21
1.3.3 Between tradition and opposition in the 21st century	24
1.4 Contra side of the public debate	27
1.4.1 A two-fronted offensive	27
1.4.2 The egalitarian argument	27
1.4.3 The utilitarian argument	31
1.5 Pro side of the public debate.....	33
1.5.1 A kaleidoscopic counteroffensive	33
1.5.2 Rebuttal of the egalitarian argument.....	34
1.5.3 Rebuttal of the utilitarian argument by resisting utilitarianism itself.....	38
1.5.4 Rebuttal of the utilitarian argument by offering counterarguments.....	39
1.5.4.1 The cultural-historical and general-developmental arguments.....	40
1.5.4.2 The linguistic and cognitive arguments.....	42
1.6 Epilogue	45
1.6.1 Some reflections on the value of the arguments discussed.....	45
1.6.2 Conclusions and suggestions for future research	46

Works cited in this chapter	47
Chapter 2 State of the Art.....	57
2.1 Introduction	57
2.2 Characteristics of classical language impact studies.....	58
2.2.1 <i>Quis?</i> The classicist as layman educational measurer	58
2.2.2 <i>Quid?</i> The myth of classics	59
2.2.3 <i>Ubi?</i> An American dream	60
2.2.4 <i>Quibus auxiliis?</i> The tools of systematic oversimplification.....	60
2.2.5 <i>Cur?</i> Research in defence of the discipline.....	62
2.2.6 <i>Quomodo?</i> Divided rather than united.....	63
2.2.7 <i>Quando?</i> Twelve decades of research.....	64
2.3 A history of American research on the cognitive impact of studying classical languages.....	64
2.3.1 The beginnings of the research field.....	64
2.3.2 The Classical Investigation.....	66
2.3.3 Research between the Classical Investigation and the Second World War	72
2.3.4 Midcentury research	74
2.3.5 Research in the seventies and eighties	75
2.3.6 Methodological and other shortcomings	78
2.3.7 Recent research (1990–present).....	81
2.4 Conclusion.....	83
Works cited in this chapter.....	84
Addenda to Chapter 2.....	90
Recent non-American research	90
Mathematics impact studies?	92
Works cited in these addenda	93
Chapter 3 Theoretical Framework.....	97
3.1 Introduction	97
3.2 A definition of cognitive transfer	98
3.3 A concise history of cognitive transfer research.....	99
3.4 Applying the cognitive transfer construct to classical language education	101
3.4.1 The training task.....	102
3.4.2 The content factor	106
3.4.3 The context factor	107
3.4.4 The individual factor	108
3.5 Supporting and related literatures	108
3.5.1 Near transfer / Linguistic benefits	108
3.5.2 Far transfer / General-cognitive benefits.....	110
3.5.3 Related topics	112
3.6 Conclusion.....	112
Works cited in this chapter.....	113
Addenda to Chapter 3.....	117
Current trends in cognitive transfer research	117
Works cited in these addenda	118

Part II: Methodological Preamble	119
Introductory remarks to Part II.....	121
Relation to Parts I and III.....	121
Instrument selection process.....	121
Chapter 4 Dutch Language Test	125
4.1 Introduction.....	125
4.2 Building a test for data collection	125
4.3 Revising the test for data analysis	128
Works cited in this chapter	131
Chapter 5 Validation of Ch-ICAR.....	135
Introductory remarks to Chapter 5.....	135
5.1 Introduction.....	136
5.2 Findings from the initial literature review	139
5.3 Selecting items for a Children's ICAR	140
5.4 Summary of the pilot study	141
5.5 Summary of study 1	142
5.6 Summary of study 2.....	143
5.7 Conclusion	143
Works cited in this chapter	144
Chapter 6 Reflections on School-Based Research.....	151
6.1 Introduction.....	151
6.2 The school as a research setting.....	152
6.2.1 The importance of school-based research.....	152
6.2.2 The difficulty of school-based research	154
6.3 The microcosm that is a school	155
6.3.1 Complexity	156
6.3.2 Task.....	156
6.3.3 Structure	158
6.3.4 Systems	158
6.3.5 People	160
6.4 Working relationships with schools	160
6.4.1 Gatekeepers.....	160
6.4.2 Face-to-face communication	161
6.4.3 Investment.....	162
6.5 Planning research activities in schools	162
6.5.1 Peace, love, and understanding	163
6.5.2 Additional considerations	163
6.6 Participant recruitment.....	164
6.6.1 Informed consent procedures	164
6.6.2 Information and consent documents	167
6.7 Administration of research materials	167
6.8 Test-taking motivation	169
6.9 Research management.....	170
6.10 Conclusion	171

Works cited in this chapter.....	172
Part III: Empirical Research	177
Introductory remarks to Part III	179
Chapter 7 Teaching Method.....	183
Introductory remarks to Chapter 7	183
English abstract.....	185
7.1 Inleiding	185
7.2 Context: het Vlaamse curriculum Latijn	186
7.2.1 Leerlingen, onderwijsnetten en koepels	186
7.2.2 De geschiedenis van de belangrijkste leerplannen Latijn in een notendop.....	187
7.2.2.1 Latijn in het officieel onderwijs: de leerplannen van het Gemeenschapsonderwijs.....	187
7.2.2.2 Latijn in het vrij onderwijs: de leerplannen van Katholiek Onderwijs Vlaanderen	189
7.3 Onderzoeksmethode.....	191
7.3.1 Instrumenten en dataverzameling	191
7.3.2 Steekproefname en deelnemers aan de enquête.....	194
7.4 Bevindingen.....	195
7.4.1 Lectuur centraal in beide onderwijsnetten	195
7.4.2 Grote verschillen tussen individuele leraren	198
7.4.3 Extreme lectuurgerichtheid en de Ørbergmethode.....	200
7.4.4 Welke plaats heeft zuivere grammatica nog?	201
7.4.5 Voorstanders van een meer grammaticagerichte aanpak.....	203
7.4.6 <i>Multum legendum est an multa?</i> De grenzen van de geïntegreerde benadering	205
7.5 Slot.....	208
7.5.1 Beperkingen van onze onderzoeksopzet en wenken voor verder onderzoek	208
7.5.2 Conclusies	209
Werken geciteerd in dit hoofdstuk	210
Corrigendum bij Hoofdstuk 7	212
Chapter 8 Methods and Results of the Longitudinal Study.....	215
8.1 Methods	215
8.1.1 From design to data	215
8.1.2 From participants to cases.....	216
8.1.2.1 Recruitment of schools and pupils	216
8.1.2.2 Composition of the program group and the comparison group....	217
8.1.2.3 Sample size and sample variants	217
8.1.3 From materials to variables	219
8.1.3.1 Time-fixed data components.....	219
8.1.3.2 Time-varying data components.....	222
8.1.4 From hypotheses to analyses	226

8.1.4.1	Operationalising cognitive transfer	226
8.1.4.2	Dealing with confounding influences, notably selection bias.....	227
8.1.4.3	Longitudinal linear mixed modelling.....	229
8.2	Results: Latin, without matching	229
8.2.1	Models and interpretation.....	229
8.2.2	Descriptive statistics	231
8.2.3	Dutch language ability	233
8.2.3.1	Lexicon.....	233
8.2.3.2	Verbs.....	236
8.2.4	General cognitive ability.....	239
8.2.4.1	Verbal intelligence: crystallised intelligence (Gc).....	239
8.2.4.2	Verbal intelligence: short-term memory (Gsm).....	241
8.2.4.3	Non-verbal intelligence: fluid intelligence (Gf)	245
8.2.4.4	Non-verbal intelligence: visuo-spatial processing (Gv).....	247
8.2.4.5	Overall general cognitive ability (G)	249
8.2.5	School achievement	251
8.2.5.1	Achievement for Dutch.....	251
8.2.5.2	Achievement for English.....	252
8.2.5.3	Achievement for French.....	253
8.2.5.4	Achievement for mathematics	254
8.2.5.5	Total achievement.....	255
8.3	Results: Greek.....	256
8.3.1	Models and interpretation.....	256
8.3.2	Descriptive statistics	257
8.3.3	Dutch language ability	260
8.3.3.1	Lexicon.....	260
8.3.3.2	Verbs.....	263
8.3.4	General cognitive ability.....	266
8.3.4.1	Verbal intelligence: crystallised intelligence (Gc).....	266
8.3.4.2	Verbal intelligence: short-term memory (Gsm).....	269
8.3.4.3	Non-verbal intelligence: fluid intelligence (Gf)	271
8.3.4.4	Non-verbal intelligence: visuo-spatial processing (Gv).....	273
8.3.4.5	Overall general cognitive ability (G)	275
8.3.5	School achievement	277
8.3.5.1	Achievement for Dutch.....	277
8.3.5.2	Achievement for English.....	280
8.4	Results: Latin, with matching	283
8.4.1	Matching procedure, models and interpretation.....	283
8.4.2	Matching solution.....	284
8.4.2.1	Balance in terms of sex	284
8.4.2.2	Balance in terms of SES.....	285
8.4.2.3	Balance in terms of native language	286
8.4.2.4	Balance in terms of non-cognitive variables	286
8.4.2.5	Balance in terms of Dutch language ability.....	287
8.4.2.6	Balance in terms of general cognitive ability.....	288
8.4.3	Dutch language ability	289
8.4.3.1	Lexicon.....	289
8.4.3.2	Verbs.....	289

8.4.4	General cognitive ability	290
8.4.4.1	Verbal intelligence: crystallised intelligence (Gc)	290
8.4.4.2	Verbal intelligence: short-term memory (Gsm)	291
8.4.4.3	Non-verbal intelligence: fluid intelligence (Gf)	291
8.4.4.4	Non-verbal intelligence: visuo-spatial processing (Gv)	292
8.4.4.5	Overall general cognitive ability (G)	293
8.4.5	School achievement	293
8.4.5.1	Achievement for Dutch	293
8.4.5.2	Achievement for English	294
8.4.5.3	Achievement for French	294
8.4.5.4	Achievement for mathematics	294
8.4.5.5	Total achievement	294
	Works cited in this chapter	295
Chapter 9	Discussion of Results and Some Additional Analyses	299
9.1	Introduction	299
9.2	Overview of empirical findings	300
9.2.1	Group differences in absolute performance across time: pupil profiles	300
9.2.2	Group differences in evolution over time: to transfer or not to transfer? ...	303
9.2.3	Other observations from the longitudinal analyses	305
9.2.4	Group differences in empathy and reading motivation	307
9.3	Evaluation of the cognitive transfer effects of studying classical languages in Flemish secondary education	308
9.3.1	Potential candidates for cognitive transfer	308
9.3.2	Strength of the evidence	309
9.3.3	Group versus individual	311
9.3.4	Context is everything	313
9.3.5	Synthesis	314
9.4	Limitations and implications for further research	314
	Works cited in this chapter	317
	<i>General Conclusion</i>	<i>319</i>
	Main study on cognitive transfer	321
	Complementary study on teaching method	322
	<i>Appendices</i>	<i>325</i>
Appendix 1	Instrument Selection Process	327
Appendix 2	Try-out Dutch Language Test: Parallel Test no.1	335
Appendix 3	Full List of Items in the Dutch Language Test	339
Appendix 4	Classification of Items in the Dutch Language Test	349

Appendix 5	Scoring Protocol for the Dutch Language Test.....	351
Appendix 6	Factor Loadings of All Dutch Language Test Items	359
Appendix 7	Factor Structure of the Revised Dutch Language Test.....	365
Appendix 8	Difficulty Levels of the Revised Dutch Language Test	371
Appendix 9	Sample of Ch-ICAR Items	375
Appendix 10	Information Letter for Schools	377
Appendix 11	Cover Letter for Pupils & Parents	383
Appendix 12	Informed Consent Documents for Pupils & Parents	385
Appendix 13	Background Data Form for Pupils & Parents	391
Appendix 14	Research Flyer 2023–2024	393
Appendix 15	Individual Feedback Report	395
Appendix 16	Testing Manual, Version October 2023 (Laptop).....	405
Appendix 17	Empathy Scales, Flemish Version	415
Appendix 18	Reading Motivation Scale	417
Appendix 19	Descriptive Statistics of the Non-Cognitive Variables	419
Appendix 20	Data Storage Fact Sheets	421

General Introduction

General Introduction

Non scholae sed vitae?

In the first century AD, the Roman Stoic Seneca the Younger wrote a series of moralising letters to his friend and student Lucilius. In one of these letters, he lamented that philosophical reasoning is often wasted on superfluous matters, such as whether virtue is corporeal or not. By approaching philosophy in this way, we humans may become book-smart, but we will not attain true wisdom about life: *non vitae sed scholae discimus*. Later, Seneca's words were reversed into the now famous saying *non scholae sed vitae discimus*, which literally means “we do not learn for school, but for life”. In this form, the saying became a popular motto for educational institutions keen to assert, in the spirit of Seneca, that they do not merely raise book-smart individuals, but that they prepare students for their future lives. As such, *non scholae sed vitae* seemed a fitting title for a dissertation revolving around classical language education (i.e., the instruction of Latin and Ancient Greek as part of the regular secondary school curriculum) and cognitive transfer (i.e., the cognitive mechanism whereby learning one task also improves performance on another task from a different domain; see Barnett & Ceci, 2002). After all, if a subject is shown to produce cognitive transfer effects, that can be seen as the ultimate proof of its value, not just in school, but in life.

The study of classical languages in particular is expected to bring about cognitive transfer effects. To understand the origins of this extraordinary expectation, we must consider the public debate surrounding the study of classical languages, and the circumstances under which previous research on their transfer effects came about. These issues will be explored in depth in Chapters 1 and 2 respectively. Here in the General Introduction, we will briefly reflect on classical language education in Flanders specifically, which presents an interesting paradox. On

the one hand, Flanders has a strong tradition of teaching classical languages. Despite Dutch being a Germanic language, and despite all secondary school degrees being equal in terms of admission to higher education, Latin and Greek figure prominently as study options within the general track.⁴ Testimonies from pupils saying they enrolled in Latin “because their parents made them” suggest that these study options still enjoy a certain prestige today. On the other hand, enrolment numbers slowly decline year after year. Whereas two decades ago nearly 13% of all Flemish secondary school pupils followed at least one classical language, in recent years this proportion has diminished to just 9% (Vereeck, 2024). There appears to be a widespread sentiment that the study of classical languages is a waste of time, that Latin and Ancient Greek have become useless and obsolete in the twenty-first century.

As is the case abroad, Flemish proponents of classical language learning have reacted to this utilitarian criticism by emphatically praising the many qualities of the classical languages, among which their alleged transfer value. The association between classical languages and cognitive transfer has found its way into official educational documents, in which it is plainly claimed that Latin and Greek enhance linguistic feeling and reading ability in the mother tongue, facilitate modern foreign language learning, and even boost problem-solving skills (CLB, 2025; Katholiek Onderwijs Vlaanderen, 2019). Such claims have become important arguments in defending the value and utility of classical language education. Former classical language pupils, especially those graduated from Greek-mathematics, Latin-mathematics and Greek-Latin, are also known to have the highest study success rate at university (Declercq & Verboven, 2010; Duyck et al., 2017). Furthermore, studies show that youngsters who have taken Latin score higher on standardised tests of Dutch spelling and syntax than their peers coming from a general study option without Latin (Devos & De Vilder, 2019; Vanheule, 2015).

The question is whether these performance differences are created by a cognitive advantage already present in pupils who opt for classical languages, or by a cognitive advantage bestowed upon these pupils through their classical language study. Without the right data, this is a chicken and egg kind of question. As Chapters 2 and 3 will show, the existing research literature is

⁴ With “general” I refer to the type of secondary education that is non-technical or -vocational, but aimed towards preparing pupils for higher education. Latin can be chosen from the first year onwards. It can be studied alone, or, from the second year onwards, in combination with Greek. Usually, in the fifth and sixth year pupils also have the option to drop Latin and continue with Greek. Depending on the year of secondary education and the exact study option, three to five class hours a week are allocated to each language.

unable to provide a definitive answer. The centrepiece of the doctoral project⁵ documented in this dissertation is therefore an empirical study on the cognitive transfer effects of studying classical languages in Flemish secondary education. At the outset of the project, the general goal of determining cognitive impact was concretised into two objectives and their corresponding research questions (quoted from the original fellowship application with minor edits).

Objective 1. The first objective narrows down the broad notion of “impact”. In this research I will study potential beneficial effects on:

- a. language ability in Dutch as an L1 (near transfer);
- b. school achievement in other areas such as modern foreign languages (near transfer) and mathematics (far transfer);
- c. general cognitive ability (far transfer).
 - ⇒ In comparison to pupils with a similar cognitive profile who have not studied a classical language, are pupils who do study a classical language more skilled in Dutch as an L1, and/or do they achieve better grades at school in some area, and/or do they perform better on tests of general cognition?

Objective 2. Secondly, I will look at the part of the specific teaching method therein, namely whether different teaching methods currently employed in Flemish classical language instruction would bring about a difference in outcome.

- ⇒ Does the employed teaching method influence the magnitude of the transfer effects in any of the areas 1a till 1c?

The starting hypotheses were (1) that the study of classical languages does indeed result in demonstrable near and far transfer, and (2) that a more grammar-oriented teaching method is conducive to transfer. There is already a lot of suggestive evidence that learning Latin benefits language ability in English as an L1 (Vereeck et al., 2025), and foreign language learning in general is known to have positive effects on L1 proficiency (e.g., Murphy et al., 2015) and on executive cognitive functions (e.g., Woumans et al., 2016).

A large-scale longitudinal observational study was then set up, in which roughly a thousand Flemish pupils in general secondary education – some of whom studied classical languages and some of whom did not – were followed for three school years and tested yearly in order to

⁵ The project was funded by Research Foundation - Flanders (grant 1110621N, fellowship fundamental research) and Ghent University.

chart their cognitive development. The study ran from 2021 to 2024. This study was the first in Flanders to use a longitudinal design to investigate the cognitive effects of classical language learning over time, and the first internationally to do so on this scale and according to modern methodological standards.

In addition to the main study on cognitive transfer, in the course of the doctoral fellowship a complementary study was designed with the following objective in mind (quoted from the application for renewal of the fellowship).

Objective 3. A third objective has been added, which will be addressed by the complementary study, and that is to map out the educational practices of Latin classes in Flanders.

Dissertation roadmap

Preparation is said to be half the battle. In the case of this dissertation, it is seven ninths: The first seven chapters set the stage in preparation for the report of the main study given in the final two chapters. The structure of the dissertation is 3 x 3, three parts of three chapters each.⁶ Part I forms the status quaestionis, which transcends the boundaries of Flanders. Chapter 1 offers an outline of the international public debate regarding classical language education. This chapter embodies the fact that cognitive transfer is but one narrow account of the value of classical language education, and that transfer arguments are part of a much larger web of arguments pro and contra. Chapter 2 sketches the state of the art and puts classical language impact studies on the proverbial map as a research field. The chief portion of this chapter is taken up by a history of American research on the cognitive impact of studying classical languages, since the majority of previous research was conducted in the United States. Chapter 3 approaches the topic from the perspective of psychological theory: It defines the construct of cognitive transfer, and builds a theoretical framework of findings from various research literatures that support the hypothesis of classical language learning inducing cognitive transfer.

⁶ Every chapter is preceded by a title page with a note about publication status. The four already published articles required for an article-based PhD are included in Chapters 1, 2, 3 (in English, with me being the first author) and 7 (in Dutch, with me being the second author).

Part II is a methodological preamble, singling out some aspects of the main study's methodology in order to avoid cumbersome prolixity of the methods section in Part III. Chapters 4 and 5 pertain to the search for adequate measuring instruments (for a more elaborate clarification of their contents, see the introductory remarks to Part II). Chapter 6 dwells on the intricacies and complications of doing fieldwork in schools to collect data on underage school children. Part III is about the empirical research that was carried out. Chapter 7 is wholly dedicated to teaching method, providing even more context on Flemish classical language education. The chapter reports on the complementary study corresponding to objective 3, while the introductory remarks also go into objective 2. Finally, the last two chapters report on the main study in relation to objective 1. Chapter 8 combines the more technical sections containing the methods and results. Chapter 9 discusses these results, with the aim of answering the central question of whether or not the study of classical languages produces cognitive transfer effects.

The official language of this Latin-titled dissertation is English. However, some parts are phrased in Dutch: the preface, because of its essayistic character; the acknowledgements, because of the intended audience; Chapter 7, because of the local theme of classical language teaching method in Flanders and, accordingly, because it was published in a Dutch journal.⁷ The Appendices are largely in Dutch as well. Thus the present dissertation is a macaronic writing of sorts, encompassing multiple languages as well as multiple research disciplines. May the reader enjoy this *satura lanx*, this varied dish of intellectual fares I have prepared; it is my personal ode and contribution to classics and psychology, an endeavour to which I have gladly devoted the last handful of years of my life.

⁷ The introductory remarks to the chapter are in English. There is also an English abstract of the article in Chapter 7, as well as an English summary of the complementary study in the General Conclusion at the end of this dissertation.

Works cited in the General Introduction

- Barnett, S., & Ceci, S. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637.
- CLB. (2025). Latijn. *Onderwijsskiezer*.
https://www.onderwijsskiezer.be/v2/secundair/sec_detail_2.php?detail=11548
- Declercq, K., & Verboven, F. (2010). *Slaagkansen aan Vlaamse universiteiten: Tijd om het beleid bij te sturen?* Vlaams Instituut voor Economie en Samenleving.
- Devos, F., & De Vilder, J. (2019). Een replica-onderzoek naar de kennis van Nederlandse spelling en grammatica bij laatstejaarsleerlingen aso en vwo. *Tijdschrift Over Taal*, 57(1), 1–9.
- Duyck, W., Janse, M., & Soltic, J. (2017). Stem voor Grieks en Latijn. Wat overleeft, leeft. In J. Strobbe, P. Thoen, & G. Van Coillie (Eds.), *Vorming door onderwijs. Humaniora tussen verleden en toekomst* (pp. 169–184). Acco.
- Katholiek Onderwijs Vlaanderen. (2019). *Leerplan Basisoptie klassieke talen*. Brussel.
<https://pro.katholiekonderwijs.vlaanderen/I-KITa-a>
- Murphy, V., Macaro, E., Alba, S., & Cipolla, C. (2015). The influence of learning a second language in primary school on developing first language literacy skills. *Applied Psycholinguistics*, 36(5), 1133–1153.
- Vanheule, J. (2015). Latijn en grammaticale competentie in het Nederlands. Verslag van een nieuw onderzoek. *Didactica Classica Gandensia*, 51, 89–107.
- Vereeck, A. (2024). Een nieuwe blik op de Vlaamse leerlingenaantallen klassieke talen. *Prora*, 29(1), 7–12.
- Vereeck, A., Janse, M., De Herdt, K., Hauspie, C., Szmalec, A., & Woumans, E. (2025). Does studying Latin make pupils smarter? Presenting the field of classical language impact studies. *Classical Journal*, 120(4), 479–519.
- Woumans, E., Surmont, J., Struys, E., & Duyck, W. (2016). The longitudinal effect of bilingual immersion schooling on cognitive control and intelligence. *Language Learning*, 66(2), 76–91.

plures aliorum eventis docentur

most of us learn from the experiences of others

~ Tacitus, Annals IV.33

Part I: Status Quaestionis

Introductory remarks to Part I

Cohesion between the chapters

This status quaestionis frames the empirical research: why there is a need for research on the cognitive impact of classical language learning in the first place, which research studies have already taken place, and what earlier findings can tell us. Each of the chapters does so from a different angle. Chapter 1 starts from the real-world situation of the controversy around classical language education, Chapter 2 takes stock of previous research on classical languages specifically, while Chapter 3 seeks connection with various other strands of relevant psychological and psycholinguistic research, especially on cognitive transfer.

The mutual relations between these three chapters are as follows. Research on any sort of impact classical language instruction might have on pupils that exceeds the realm of classics, as discussed in Chapter 2, can be viewed as an integral part of the wider public debate, as discussed in Chapter 1. It is no coincidence that the arguments in favour of classical language education (cf. Section 1.5.4) correspond to the categories of classical language impact studies (cf. Section 2.2.4).⁸ The theoretical framework proposed in Chapter 3 can be viewed as a reaction against the lack of theoretical foundation by the research studies included in Chapter 2.

⁸ The attentive reader will not fail to notice that there is actually a subtle difference between the two classifications in Sections 1.5.4 and 2.2.4, which is the consequence of advancing insight. In the public debate article of Chapter 1, I grouped together skills such as historical awareness and critical thinking with non-cognitive characteristics such as ambition and self-confidence under the umbrella of the general-developmental argument, since they are often mentioned in the same breath. When preparing the classical language impact article of Chapter 2, however, I realised that these skills are actually cognitive in nature rather than non-cognitive, hence the distinction between categories 3 and 4. This also means that the cultural-historical argument of Section 1.5.4 is not represented in the research categories of Section 2.2.4. The cultural-historical argument, as defined in Chapter 1, focuses on directly classics-related knowledge (and the intellectual and emotional enrichment this knowledge may provide) rather than on skills derived from this knowledge.

The value of classical language education

Even if this status quaestionis suffices as background to the conducted empirical research, it can hardly be called a complete study of the issue posed by the value of classical language education. I am aware that I have only perused and processed a fraction of the countless writings on this topic, and I continue to encounter new sources daily, both written a few days ago and written a century ago. The Dutch classicist Helena Stellwag (1902–1996), who devoted a monograph to the topic that was published in 1949,⁹ had a similar experience. She gave her 350-page book the telling subtitle “an introduction”, and in the preface she likened her research efforts to fighting the Lernaean Hydra, a mythical many-headed snake that regrew two new heads for every single head that was chopped off. To study every existing publication on the value of classical language education would be a veritable Herculean task indeed.

Stellwag, who in addition to classics had studied pedagogy and psychology as well, also realised the importance of interdisciplinarity:

“Immers, vanuit de studie der klassieken moet dit probleem worden aangevat, — aangezien slechts door iemand, die de Oudheid kent, het volle pond er aan kan worden geschonken — en van uit de studie der Paedagogiek en Psychologie.” (Stellwag, 1949, p. 3)

[After all, this problem must be approached from the study of classics—since only someone who knows Antiquity can give it its full due—as well as from the study of pedagogy and psychology. (Stellwag, 1949, p.3)]

The idea of the necessity and fecundity of marrying classics with psychology / pedagogy in matters of classical language education will recur multiple times throughout Chapters 2 and 3.

Grammar and translation

Something that perhaps may not have received enough attention in the articles that make up the following chapters, is the point of which elements of classical language education were previously assumed to “form” the mind, or in modern terms, to have transfer value. Stellwag’s

⁹ Stellwag, H. (1949). *De waarde der klassieke vorming. Een cultuur-historische, paedagogisch-psychologische en didaktische inleiding*. Wolters.

monographic synthesis is very instructive on this point. Like in the psychological theoretical framework of Chapter 3, Stellwag and others ascribed formative value to the combination of (1) the Latin and Ancient Greek languages themselves, characterised by inflection and morphosyntactic agreement, and (2) the way in which these languages were taught, with Stellwag being inclined to attribute more weight to the latter. Grammar rules and translation exercises, the basic ingredients of the traditional grammar-translation method, were considered particularly powerful educational devices. Due attention to grammar was thought to foster insight in language as system, whilst also inspiring an attitude of diligence and systematic reasoning, and translating was thought to be the mental training *par excellence*, because of the many both substantial and minute problems a translator is constantly faced with.

However, according to Stellwag, the teaching method of her day was already but a watered down version of the true grammar-translation method of old, and therefore less effective. Flemish teaching practices in the 2020s are even further removed from the original grammar-translation method than Dutch teaching practices in the 1940s. As we will see later in Chapter 7, reading, understanding and enjoying texts has become the ultimate goal of classical language education in Flanders. Grammar is no longer as heavily emphasised as it once was, although some level of grammar instruction is still an essential part of the dominant teaching methods. Translating into Dutch remains a common practice to establish the meaning of sentences and texts, having the potential to facilitate linguistic transfer: What begins as superficial comparisons like “Latin *accusativus cum infinitivo* is translated with a *dat*-construction in Dutch” could provide a gateway into a more fundamental understanding of structural differences between the Dutch language and the classical language in question.

Chapter 1: The Public Debate on Classical Language Education

The contents of this chapter are nigh identical to the following article:

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*, 25(50), 101–115.
<https://doi.org/10.1017/S2058631024000424>

Minor edits were made with an eye to harmonising spelling, punctuation, title numbering, cross references, bibliographical style etc. throughout the dissertation.

Chapter 1

The Public Debate on Classical Language Education

1.1 Introduction

Why study classical languages? That perennial question lies at the basis of this contribution. It is also the basis of a dispute that is waged in many classrooms and living rooms, as well as publicly in the media. Passionate pleas why we should or should not study classical languages (hereafter CLs) are published in papers and journals on a regular basis. As former UN ambassador Austin Lashbrook (1969, p. 163) remarked some fifty years ago, “[i]t seems as though men are forever trying to prove or disprove the value of Latin as part of a liberal education”.

The constant debate on the value of CL education forms the context in which classics teachers operate, and therefore it is well-known and often referenced. When reference is made to this debate, either its existence is mentioned in passing as general context, or a specific part of the debate is zoomed in on, such as the outcome of a particular research study or educational project. For instance, research on the impact of studying CLs (e.g., Masciantonio, 1977) can indeed be seen as a response to the criticism that CL education is useless, saying for example that it improves native language ability, just like educational equity projects (e.g., Holmes-Henderson et al., 2018) can be seen as a response to the criticism that CL education is an elitist affair, saying for example that minority groups perhaps benefit the most from it.

The actual debate itself, however, is seldom viewed in total, as a research topic in its own right, and in a way that transcends a mere listing of arguments. This contribution aims to do just that, and look at the internal logic of the debate in its entirety. Much of what is laid out will probably be familiar to the reader. Rather than in the information on its own, the novelty lies in the way it is assembled, arranged and interpreted. The first author’s experience with

presenting this material tells that it can be of use and interest to classicists who feel that they are entangled in a jumble of arguments pro and contra they find hard to oversee. The outline of the debate given here provides a framework for situating and evaluating arguments, and adds a historical dimension to the dispute as well.

1.2 Delineation of topic and method

The topic of the current article is the public debate regarding CL education. Straightforward as this may seem, a more precise delineation is necessary. Here, the phrase “public debate” mainly refers to the discussion in traditional media. The definition of public debate as such is a direct consequence of the article’s genesis: It stems from an inquiry performed in 2020 into Flemish newspaper articles on CL education from the last few years. This inquiry was supplemented by readings of about a century of literature on CLs in other countries, in particular the United Kingdom and the United States. Already in that initial stage, it soon became clear that the same elements returned over and over again, and that the obtained rudimentary analysis of the debate could have wider bearing. From the onset, a key source in this exploration has been *Classics teaching in Europe*, edited by former Euroclassica chairman John Bulwer.

An extensive process of additional research and refinement eventually led to the paper in its present form. The end result is reckoned to be applicable to (Western) Europe and the United States of America as a whole, as sources from various countries within this area have been consulted and incorporated for all of the onward sections. The timeframe has been kept limited to recent and contemporary history. Focus lies on CL education at the secondary level. Unless otherwise indicated, instruction of CLs at the primary and tertiary levels is not taken into account. Neither is this article about shifts and developments within CL courses, larger evolutions in the classical discipline, or the role of classical heritage in society, although pieces of these related themes will inevitably come up along the way. In sum, the topic is delineated as the controversy about the instruction of Latin and/or Ancient Greek at the secondary educational level, as it emerges from Western traditional media in the 20th and 21st centuries.

The objective is to outline said controversy in a coherent historical narrative, and in doing so, to analyse CL education as a social phenomenon. The narrative starts from newspaper articles, opinion pieces and educational documents, and relates what is written there to relevant

academic literature. For the analysis, we will make use of the sociolinguistic theory of language ideology. Language ideologies are understood as conceptualisations about languages, speakers, and discursive practices (Irvine, 2012). These conceptualisations are not purely linguistic, but rather social and cultural givens which can have important material effects (Farr & Song, 2011). Woolard (2020, p. 1) defines them as “morally and politically loaded representations of the nature, structure, and use of languages in a social world”. The adoption of a sociolinguistic approach allows us to view the public debate on CL education from a new, theoretical perspective.¹⁰

The method chosen to study and describe the topic of course has its limitations: The picture painted below will undoubtedly be somewhat coloured by the first author’s background and lived experience as a pupil, student, teacher and researcher of CLs in Flanders. In the words of Dahlberg (2006), as researchers “we are not these objectivistic scientists that distantly register meanings, (...) but immensely involved in the explication of meaning” (Dahlberg, 2006, p. 16). Especially in qualitative and historical research, it is therefore essential to reflect on our social identities and personal biases that come with it, and to recognise their role in our research (Janak, 2018). My own self-examination process as main author of this article, which resembled Janak’s (2018) technique of bracketing, yielded the following. Above all, I am a Flemish classicist who believes in the intrinsic value of the study of CLs. My core understanding of the current topic is indebted to the training and instruction I received in teaching and didactics. Moreover, my main research centres around the linguistic and cognitive effects of studying CLs, which is why I feel most tuned in to that part of the debate.

Before finally turning to the public debate itself, a succinct but indispensable historical overview will be given of the state of CL education (Section 1.3). Next, the most prominent arguments pro and contra the study of CLs will be discussed, elucidating the language ideologies informing them (Sections 1.4 and 1.5). The last section contains concluding remarks on these arguments and on the debate as a whole (Section 1.6).

¹⁰ Vice versa, it may be noted that looking at CLs opens up uncultivated terrain for sociolinguistics. In its roughly sixty years of existence, the field of language ideology has proven to be a fertile one, with contributions from various scientific backgrounds (Irvine, 2012). However, it has invariably brought forth research concentrating on modern language varieties (e.g., Milroy, 2001), whereas Latin and Ancient Greek have largely been overlooked.

1.3 The state of classical language education in Western Europe and the US

1.3.1 The bloom of classics and the seeds of doubt

In order to grasp the situation in the 20th and 21st, a preamble on prior centuries is in place. The Humanist scholars played a pivotal role in bringing classics into blossom as a field of study and instruction. In his insightful essay entitled “A Theory of Classical Education”, written for the inaugural volume of the journal published by the then newly-founded British Joint Association of Classical Teachers, Robert Bolgar (1963, p. 14) explains:

“The origins of that dominance go back to the days of the Humanists who believed that the ancient literatures provided the best key to a proper understanding of life, and persuaded the world that their belief was right. They pleaded their case to such good effect that the pre-eminence of the classical heritage was accepted not only by the educated, but also by many who never saw the inside of a grammar school.”

And thus, schooling remained synonymous to CL study throughout the Modern Era. However, the origins of the debate in question stretch back many centuries as well. After the Humanist surge of enthusiasm to return *ad fontes*, there have always been voices doubting and criticising the CL education of the time. A selective overview can be found in Pire (1971, pp. 1–52). He summarises:

“Les critiques, les attaques essayées par les langues anciennes et leur enseignement depuis Montaigne forment une marée si tumultueuse, si puissante qu’il semble impossible de résister à son courant. Devant tant d’accusations, la défense se sent gagnée par un sentiment d’impuissance; que dis-je, par un sentiment de culpabilité: les langues classiques semblent devoir mourir pour la seconde et la dernière fois.” (Pire, 1971, pp. 24–25)

[The criticisms, the attacks suffered by the classical languages and their instruction since Montaigne form a tide so tumultuous, so powerful that it seems impossible to resist its current. In face of so many accusations, the defence feels overcome by a feeling of impotence; nay, by a feeling of guilt: it seems the classical languages must die for the second and final time. (Pire, 1971, pp. 24–25)]

Nevertheless, 19th-century educators even went a step further. Within the pedagogy of the 1800s, the socialising function of education was rated highly, and there were strongly normative and prescriptive views of how the nature of a child should be developed and perfected (Depaepe, 1999). When general secondary education as we more or less know it took shape, its aim was *Bildung*:¹¹ a Humanistic ideal of general formation which lets the student flourish into a well-mannered and well-thinking human being. The classical languages and literatures were deemed highly suitable to achieve this *Bildung*-ideal, and therefore took centre stage at school (Pfeiffer, 1976). Those were the last palmy days for CLs, because from the late 19th century onwards, a public debate ensued that reached unseen heights.

1.3.2 From fundament to *Fremdkörper* in the 20th century

In Western Europe and the US, the study of Latin and – to a lesser extent – Ancient Greek had been the fundament of the educational system for many centuries (e.g., Wachter, 2008, p. 135). In the 20th century, the state of CL education underwent serious changes. There is seldom a single causal explanation for historical evolutions, nor do they run an identical course everywhere. Likewise, the decline of secondary-level CL instruction was due to a concurrence of factors, which played out differently in different regions and countries. Here we will briefly state six influential factors, being well aware that their importance, timing and even order varies across Europe and the US. This summary is necessarily generalising, and the number of examples is limited.

For a start, there was the wave of criticism from progressive men who had grievances about their classical education. As already mentioned in the previous subsection, critiques of CL education are of all time. Yet it was not until the late 19th, early 20th century that these managed to thoroughly disrupt the *status quo*. In the United States, this dispute is known as the Battle of the Classics (Adler, 2020). In Victorian England as well, schoolmasters and university lecturers revolted against the way school establishments had been run, and attacked the special status which CLs enjoyed.

¹¹ The pair *Bildung* vs. *Ausbildung* was originally coined by the German scholar Wilhelm von Humboldt (1767–1835).

Bolgar (1963) analyses an exemplary article by Cambridge classicist Henry Sidgwick written in this context. One of the main propositions of Sidgwick and associates was that Latin and Greek as they were being taught back then, with heavy emphasis on perfecting translation and composition, but little room for literary appreciation and study of the ancient cultures,¹² did not serve the needs of society. They called for a wider curriculum, in which mathematics and science would hold an equally important position. Inadvertently, their criticisms had much farther-reaching consequences: “The educational reformers of the late 19th century gave Latin a bad name, and we late in the 20th still have to struggle against the views they popularised: that Latin is dreary, that it is useless, that it is remote.” (Bolgar, 1963, p. 11)

As far as the public image of CL education is concerned, the tone was set for the decades to follow. A second factor in close connection to the first, is the scientific progress at this same juncture in time. The natural sciences made great strides forward, making significant discoveries back-to-back, which shaped and adjusted humanity’s outlook on the world. The ancient literatures were not the exceptional source of knowledge anymore that they had been for the Humanists as well as in later centuries; people now became more orientated towards the exact sciences. The *auctoritas* of the classical writers took a severe blow around this point, and with it the rationale to study CLs.

Another punch was thrown on behalf of educational psychology; this is the third factor. Part of the pedagogical philosophy of *Bildung* is the doctrine of formal discipline (Paulsen, 1965). The premises of formal discipline are that the incidental teaching of a complex and demanding subject like a CL serves the general objective of cultivating all-round mental faculties, and that such transfer of learning occurs automatically. The famous American educational psychologist Edward Thorndike discredited this doctrine. From the empirical investigations he conducted in the twenties (Broyler et al., 1927; Thorndike, 1924a, 1924b), he concluded that the effect of one’s high school study programme on cognitive ability was negligible. To current methodological standards Thorndike’s research is questionable (as also pointed out by Gutacker, 1979, p. 10; Haag, 1995, p. 251; Ortner et al., 2008, p. 190), but at the time his findings were revolutionary,

¹² Words of similar tenor are found in the report of the Classical Investigation launched by the American Classical League (Advisory Committee, 1924). Among other things, the goal of this investigation was “to make recommendations as to how Latin could be taught more effectively to achieve the claims made for it” (Douglass & Kittelson, 1935, p. 27). The final report, dating from 1924, stressed the value of culture and general language study in any Latin course (Lucasse, 1929).

and did not pass unheeded. Their impact was felt in the United States and far beyond (Haag & Stern, 2000; Sparks et al., 1995).

Under the influence of factors one to three, CL education no longer held out the prospects of gaining world knowledge and sharpening your wits as it used to do before. Of course, proficiency in at least Latin was a prerequisite for attending university, but that too would change. This brings us to our fourth factor, namely the disappearance of institutional protection. The United States and the United Kingdom are two clear examples of countries where the alteration of higher education policy has had a direct impact on CL education. By the late twenties, Latin ceased to be a compulsory testing area on American college entrance exams;¹³ the rise of mass education and the scientific rivalry accompanying the Cold War then brought enrolment in Latin to its lowest point by the late sixties (Sparks et al., 1995, p. 165). In Great Britain, the universities of Oxford and Cambridge decided to drop Latin as a matriculation requirement in 1960 (e.g., Paul, 2013, p. 145). The aftermath of Oxbridge's decision led to profound reform of the classical subjects (Forrest, 2003; Morwood, 2003). Similar events happened in other European countries,¹⁴ with fewer pupils and class hours for CLs as a logical consequence.

Institutional protection has not vanished completely in all countries, however. Several German universities still demand a *Latinum* (a certificate obtained after passing the state exam in Latin) for certain study fields and levels, as well as a *Graecum* (ditto for Ancient Greek) for study options like classics, ancient history, classical archaeology and philosophy (Glücklich, 2006). This also goes for Switzerland (Schweizerische Maturitätskommission, 2009). The University of Zürich, for instance, has only in recent years relaxed its requirements for prior knowledge of Latin (Universität Zürich, 2022). Austria requires prospective students of about forty university degrees, even in less directly related domains like law and medicine, to have successfully taken their final examinations in Latin (Reitermayer, 2006).

Simultaneously with the opening-up of higher education, a fifth factor was set in motion: the invention of the modern secondary school. In stark contrast with the 19th-century model

¹³ Certain colleges were even earlier to admit freshmen who had never taken Latin (cf. Wier, 1921).

¹⁴ For the Netherlands, see Verhoeven, 1997, pp. 2–3. In Belgium, secondary school certificates were declared partially polyvalent in 1956 and completely omnivalent in 1964 (e.g., Grootaers, 1998). In Italy, the Codignola law of 1969 made it possible for all secondary education graduates to continue their studies at university (e.g., Aloï, 2010).

which revolved around CLs in the name of *Bildung*, new models of general secondary education were being proposed devoid of Latin or Greek. This modern type of education was presented as a viable alternative to a classical education, and was thought to offer an equivalent, if not better preparation for a pupil's further career. As Townsley (1985) eloquently phrased it: "The First and Second World Wars ejected Latin from its exalted seat of predominance and respect, replacing it with modern languages which provided the immediate practicality so urgently needed during those times of international upheaval" (Townsley, 1985, p. 4). The CL-free study options soon gained popularity.

In the second half of the 20th century, a sixth factor emerged in the form of innovative pedagogical theory. Led by the spirit of the sixties and seventies and by contemporary insights from developmental psychology, pedagogical theory of the time prescribed a more comprehensive approach to schooling (Depaepe, 1999; Ellis et al., 1991). It highlighted the need to bolster not just cognitive abilities, but all sorts of competences, including emotional and motorial ones. Moreover, education was supposed to serve the development of all individuals; therefore, it should not act as a mechanism to exclude anyone. Rather, it ought to be communal for as many pupils as possible, regardless of their background or capability.

Latin and Greek, which were notoriously difficult to master to the expected degree of accuracy, and which had functioned as a selection instrument for the intellectually most able pupils (Bolgar, 1963, p. 24), did not fit in this scheme. Though once the fundament of all Western non-vocational education, the study of CLs had become a *Fremdkörper* within in the educational system – more than ever a relic from days long gone.

1.3.3 Between tradition and opposition in the 21st century

Where does that leave us, two decades into the 21st century? The legacy of the two previous centuries is still tangible. Therefore, CL education finds itself in an ambiguous position, hovering between a strong tradition and strong opposition.

"[Es] dürfte kaum ein Schulfach geben, das die Geister so scheidet, bei dem so leidenschaftlich diskutiert wird, bei dem ein Standpunkt *sine ira et studio*, d.h. sachlich, kaum möglich erscheint." (Haag, 2001, p. 31)

[There must hardly be another school subject which is so controversial, which is so passionately discussed, about which a stance *sine ira et studio*, i.e. objective, seems hardly possible. (Haag, 2001, p. 31)]

On the one hand, despite the disruptive factors just discussed, the instruction of Latin and Greek has persisted, and they are not rarely encountered on the timetables of secondary schools (as is also the conclusion of Skurzak, 2013). It is safe to say that in several European countries, among which Austria, Belgium, Germany, Italy and The Netherlands, CLs – all things considered – continue to take up a relatively large portion of the available school curriculum (Facq et al., 2006; Glücklich, 2006; Korn, 2020; Laes, 2019; Perra, 2006; Reitermayer, 2006; Tijsseling, 2006). Some governments even formulate obligations or recommendations regarding the provision of CLs (Turpin, 2020): For example, Latin and/or Greek can be mandatory for pupils in academically oriented streams, such as the *gymnasia* and *licei classici*.

On the other hand, the general public opinion gravitates towards a negative appraisal. The place of CLs in the curriculum is no longer commonly seen as self-evident (e.g., Caerols, 2013 on perceptions in Spain). School reforms have pushed CL programmes in tight corners, and enrolment numbers keep on dwindling (Bulwer, 2018). Since the 20th century, the pedagogical antipodes of CLs have always been the exact sciences. In the past few years this contrast was further intensified, since the economic demand for more technically-schooled employees occasioned heavy promotion of STEM studies (Science, Technology, Engineering and Mathematics), and not without effect.¹⁵

The narrative is not quite complete yet, however. Besides the discourse of CLs slowly losing terrain but holding their own in spite of it all, that dominates publications by classicists (e.g., Bulwer, 2006a; Lister, 2008; Morwood, 2003), there are signs that classics is finding a new elan. A tendency has become perceptible towards renewed appreciation of CLs by policymakers. The French-speaking Community of Belgium has been working on an Excellency Pact which, among other things, will make initiation in Latin a compulsory part of a *truncus communis* in secondary education (Fédération Wallonie-Bruxelles, 2023). In 2021, the British Department for Education announced a £4 million Latin Excellence Programme to bring back Latin to state-maintained schools (Department for Education, 2021). There are also several successful

¹⁵ A note on Flanders specifically: As STEM programmes attracted ever more pupils (Departement Onderwijs and Vorming, 2021), it was felt that this was often at the expense of CL programmes. STEM became seen as a competitor of CLs, vying for pupil enrolment (e.g., Duyck et al., 2017; Laes, 2016).

projects running in Europe and the United States which employ Latin instruction as a social catalyst.¹⁶

It is unclear what all of this means for the future of CL education. Could another Greco-Roman renaissance possibly be in the making? Will pupil enrolment plummet to an all-time low? Will CL education more or less carry on as it is? Or will new functions for CLs complement current educational functions as language ideologies develop into the 21st century? The pacts, programmes and projects of the previous paragraph are very diverse qua content, target group and finality, and by no means a return to the CL instruction of yore. Both the stimulation and denunciation of CL education are driven by an underlying language ideology. Let us therefore take an in-depth look at the arguments that are used pro and contra. Figure 1 shows the categories in which they can be subsumed.

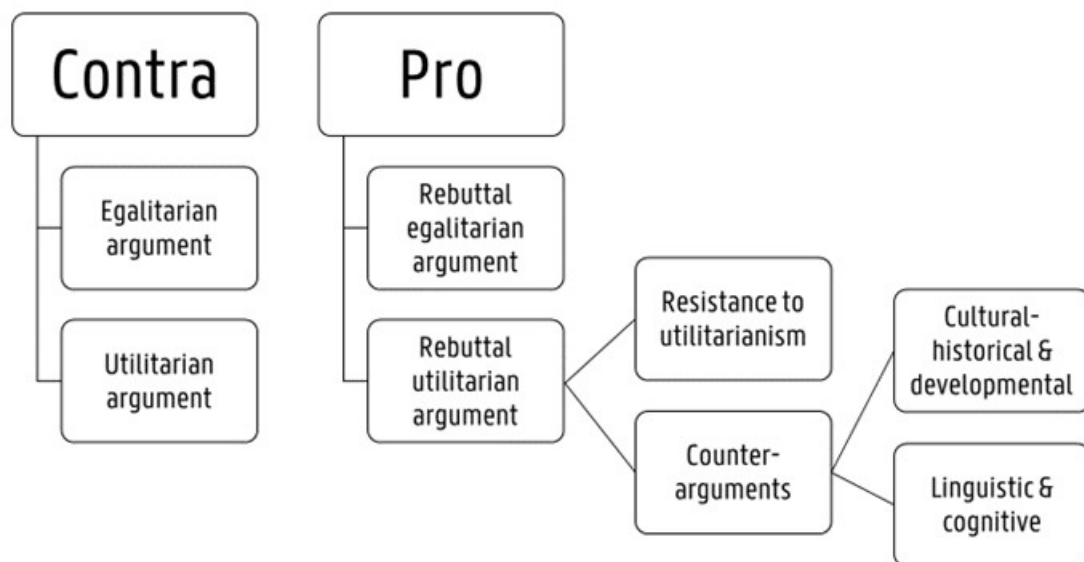


Figure 1 Classification of arguments pro and contra CL education

¹⁶ In German-speaking regions for instance, we have *Latinus Pons* in Germany (Kipf, 2010), and *Lingua Latein* in Switzerland (Wesselmann & Walker, 2021). An American example is *Aequora* (Butterworth, 2017).

1.4 Contra side of the public debate

1.4.1 A two-fronted offensive

As the mixed picture of the historical overview shows, there is plenty of room for debate. Is the study of CLs for everyone or for the happy few? What is the purpose of devoting oneself to ancient languages that are no longer spoken? Can cultural expressions that are thousands of years old still be relevant today? Especially since the post-war period, an international public debate regarding CL education has manifested itself. On the contra side, its role and value are openly and continuously questioned from two main angles: egalitarianism and utilitarianism. In Sections 1.4.2 and 1.4.3, we will discuss for each of the two contra-arguments what the criticism entails, with which elements in the reality of CL education past and present it can be connected, and how it can be understood from a language ideology perspective. These three components will partly be intertwined.

Note straight away how these two lines of opposition imply a different outlook on CL education. Insofar as CL education is seen as a time-honoured tradition, it enjoys a certain prestige which is interpreted negatively by egalitarianists as being elitist and exclusionary. Insofar as CL education is seen as a curiosity in the modern world, it is vulnerable to the utilitarian criticism that CLs are useless and irrelevant today. One of the two can be more prevalent than the other in certain countries. In Flanders for instance, the utilitarian argument dominates the public debate (Spelmans, 2010), whereas in the United Kingdom for instance, the egalitarian argument is of relatively greater significance (e.g., Foster & Wise, 2022; Lanvers, 2017). Concerns can change over time as well, which appears to have happened in the United States: In the 20th century, many efforts were directed at invalidating utilitarian criticism (Vereeck et al., 2025), whereas in the 21st century, equality issues are much more in the foreground (notably also in academia, cf. de Melo, 2023; Libatique & McHardy, 2023). It is not impossible either to combine the two contra arguments in one critique, as done by Merry & Boterman (2020).

1.4.2 The egalitarian argument

The general idea behind the egalitarian argument is that CL education is not inclusive. This implies that the connection between participation in CL education and certain demographic

characteristics is denounced, but not necessarily that the intrinsic value of studying CLs is denied. Alternatively, this contra-argument could also be termed the accessibility argument. The suggested solutions for inequal access range from a turnabout in teaching method to complete abolition of CL education. The egalitarian argument is arguably more complex than the utilitarian argument that schooling should have practical use, in the sense that it touches on fundamental societal and educational issues which transcend CL education as such. Moreover, there are different kinds of egalitarian thinking, with different visions on where, when and between which groups equality should be achieved (Hirtt et al., 2007). Likewise, we can discern multiple variants of the egalitarian argument against the study of CLs.

The oldest kind of egalitarianism is concerned with socio- economic class. Classics and CL education are perceived as elitist. In regions with a long history of instructing Latin and Ancient Greek, this perception is obviously rooted in a historical reality: For centuries, Latin has been the *lingua franca* of the higher classes, and later, CL education continued to be a feature of class distinction. Up to the early 20th century, CLs marked the dividing line between schooling of high and low stature (Herrlitz et al., 1993; Sawert, 2018). When in the 20th century CL-free alternatives arose within general secondary education, the choice for the classical or modern type usually coincided with the social background of the pupil (Gerhards et al., 2019). Does this still hold true today? Research studies from various countries indicate that to a certain extent it does indeed (e.g., Boone, 2011; Gerhards et al., 2021; Hunt & Holmes-Henderson, 2021; Merry & Boterman, 2020). Hunt and Holmes-Henderson (2021) describe the inequality of access to CL study as “classics poverty”.

This accessibility issue extends beyond socio-economic class alone. It is known that social and ethnic background tend to co-vary (Bracke, 2023; Schleicher, 2019; Stevens & Dworkin, 2019). Concerns about the inclusion and representation of diverse ethnic groups in CL education have been expressed since at least the seventies (Barnes, 2018; Nesbitt, 1979). Nowadays social egalitarianism is widened to encompass as many aspects of human identity as possible, with particular focus on gender and ethnicity. Like nearly all areas of society, the study of CLs has not escaped the critical gaze of our time (see also de Melo, 2023 on classics at university in particular). On account of its past, the classical discipline is charged with racism and sexism: It is pre-eminently a discipline of old White men, so it seems, going from the chiefly high-born men that make up the classical writers, to the elite White male scholars who later read and studied them (Butterfield, 2020).

As far as the current situation is concerned, the debate around CL education is often entangled with broader debates about education and equality. What kind of schools offer CLs in the first place depends on how education in general is organised. In some countries, school types marked by CL instruction are criticised for maintaining social disparity (e.g., Merry & Boterman, 2020 on Dutch *gymnasias*). Even if these are accessible to all pupils in theory, for a variety of reasons they may not be so in practice: Sociological research from Europe and the United States has shown that even though democratised education operates on egalitarian assumptions and aims to be merit-based, the actual participation and outcomes tell a different story (e.g., Domina et al., 2017; Hirtt et al., 2007). Ironically, in the UK, opposition against CLs has led to their removal from most state school curricula, cordoning them off to the private school sector – the very situation the Latin Excellence Programme is now seeking to reverse.

Aside from the socio-economic and identitary variants of the argument at hand, another variant stems from an intellectual egalitarianism. The pedagogy of CL education has been frowned upon as well. It must be noted that sometimes the study of CLs is intentionally presented as highbrow.¹⁷ “Several countries hint that Latin is continued in some schools to give them an appearance of high standards of learning and discipline which is attractive to some parents. (...) It is even possible that a country may see nothing wrong with an elitist education.” (Bulwer, 2006b, p. 9). Historically, there was a link between CL education and the selection of an intellectual upper class too. This link lives on in the collective memory to this day. Some CL programmes might deliberately appeal to it and seek to preserve CLs’ function as a selection instrument, while other deliberately do not. Either way it is a persistent conception (Van Damme, 2021, p. 179).

According to some, the traditional teaching paradigm, sprung from the grammar-translation method, is to blame for the systemic inequality (summary of such critical views in Hunt, 2022, 175–176). Suspending a more detailed definition, grammar-translation consists of “presentation and study of grammar rules, which are then practised through translation exercises” (Richards & Rodgers, 2014, pp. 6–7). This method lays a heavy tax on pupils’ declarative memories, and for this reason it is accused of being inherently exclusionary:

¹⁷ To be sure, Latin and Ancient Greek are not the only gatekeepers. Such intellectual elitism is also found outside the context of CL education: many secondary schools award prizes for the best pupils in any subject, and universities grant their excelling students distinctions and honours programmes.

“Since the goal is accuracy, students with good memories who maintain those exceptionally high standards validate the teacher’s practices. Teachers, then, have more control over the direction of their program. In the highest-level classes, students fulfil the teacher’s goal of producing students who meet the high levels of accuracy that has been valued for so long. Such a practice is self-fulfilling, however, since little attention is given to the students without good enough memories to live up to the gold standard of accuracy and excellence. This is exclusivity.” (Piantaggini, 2020, p. 94)

How can the egalitarian argument be understood from a language ideology perspective? Two principles seem to underpin and explain the egalitarian argument, first and foremost the hierarchisation of language. Jaspers (2009) captures this hierarchisation in a prestige binary, distinguishing between “prestige multilingualism”, e.g. learning the dominant Western languages in school, and “plebeian multilingualism”, e.g. speaking a different language at home than at school because of a migrant background. The second principle is the importance of equal access to education in the prestige languages for all, no matter your socio-economic status, declarative memory capacity, gender, ethnicity, or whatever aspect of your identity.

Knowledge and instruction of CLs are regarded as a textbook example of prestige multilingualism. A classical education constitutes symbolic capital in the sense of Bourdieu (1992). Latin and Ancient Greek have long functioned as a cultural code of belonging for the social and intellectual elite, as both a hallmark and tool of privilege (Bracke, 2023). As recent as 2019, an attitude research study in Germany among parents of secondary school pupils found that the majority of these parents rated a person with knowledge of Latin higher on *Allgemeinbildung*, being culturally well-educated, and having high status, than a person with knowledge of French (Gerhards et al., 2019).

Yet it could be argued that a view on the study of CLs as prestigious multilingualism is gradually becoming ever more anachronistic. Increasingly, the opposite side of the binary comes into play as well: Marginalised groups have found their way to classics and used it to offer resistance and develop their own identities, so that we may speak of a democratic turn (Bracke, 2023). In fact, the prestige binary is problematised altogether by the ambiguous position in which CLs find themselves in the 21st century. The egalitarian argument is only tenable insofar as CL education is a solid and well-respected tradition. Realistically speaking, how much prestige do Latin and Greek really have left?

1.4.3 The utilitarian argument

This slightly provocative question leads us seamlessly into the second argument on the contra side of the public debate, known as the utilitarian argument. The gist of it is that CLs are not useful in present-day society, or at least not useful enough to justify the time spent on them; that they are “*onerosi ed inutili*” [onerous and useless], to quote Caerols (2013, p. 322). In this respect it is remarkable how little the objections made today differ from those made half a century or a century ago, as the following two random examples illustrate.

In a French essay that was published in the early seventies, Adrien Cart (1971, pp. 175–77) conveys a vivid impression of a customary critique of Latin. It runs somewhat as follows: Seeing that Latin is a dead language which virtually no one speaks and but few can read, the benefits reaped in adult life of having learned Latin as a pupil are limited to the dubious joys of deciphering stone inscriptions, quickly grasping legal and administrative terms, and understanding the labels on pharmacists’ phials and in botanical gardens; surely those many laborious hours spent on Latin are only worth their while for the fools who dream of becoming a teacher in language, literature, history or philosophy. If we go back another fifty years in time, we find Wier’s delightfully witty account of American college life in the early twenties, of which we will quote the introductory paragraph for the purpose of illustration as well as enjoyment:

“If the value of Latin and Greek were determined by the regard in which the average student claims to hold them, they would have no plea for inclusion in the curriculum of the up-to-date school. For they belong to that definite class of studies considered by the modernist not only useless but to a degree pernicious, in that it delays the pursuit of more important things. Therefore, the student, when he sets out for college, leaves them behind him with little regret; and if he thinks of them further it is with that kindly tolerance that one may well feel for useless things that are over and done.” (Wier, 1921, p. 326)

The utilitarian argument fits in with a broader evolution towards a more instrumental vision on education, which Wier in the quote above refers to as “modernism”.¹⁸ Throughout the 20th

¹⁸ This term takes on a specific meaning in the historical context of the Battle of the Classics: “Traditionalists in the Battle of the Classics hoped to retain the classical languages (...) as the pedagogical core of higher

and 21st century, for reasons sketched out in Section 1.3, the 19th-century ideal of *Bildung* was gradually being abandoned in favour of *Ausbildung*, ‘professional training’ (Masschelein & Simons, 2012; Tippelt, 2009). Public expectations of education shifted from providing the young mind with a general formation, to equipping youth with measurable assets for their future and preparing them to become capable citizens in modern society. Nowadays traditional subjects such as Latin and Greek are under pressure, and have to give way to more modern forms of schooling like project-based learning and cross-subject clusters, above all STEM.

This instrumental vision is apparent in the current educational policy of the European Union. As mentioned above, because of the need for technical and engineering profiles in the labour market, STEM is being actively promoted in secondary education. The European government issues guidelines, which local STEM platforms implement in collaboration with partner foundations (Secretariat of the EU STEM Coalition, 2023). This development aligns with utilitarianist opposition to the study of CLs. In the public debate, the utilitarian argument that CLs are of no use and occupy precious time in the curriculum is often followed by praise of STEM studies, which are put forward as a quintessentially useful and contemporary area of study – the sensible alternative, as it were.

Another aspect of the EU’s educational policy revealing an instrumental vision on education, is the formulation of “key competences”. The European as well as national and regional governments have defined sets of key competences, of which mastery is to be expected of a youngster having graduated from secondary school. These “include knowledge, skills, and attitudes needed by all for personal fulfilment and development, employability, social inclusion and active citizenship”(Council of the European Union, 2018). At first glance, the instruction of ancient languages seems hard to reconcile with the goal to impart 21st century key competences. In Section 1.5.4, however, it will be argued that CL education might have quite some common ground with this pedagogic project after all.

The ideology behind the utilitarian argument can be summarised with the next series of inequalities: Science is more important than language, and modern languages are more relevant than classical ones. What we have here is also – like with the egalitarian argument – a hierarchisation, but based on the widely- endorsed principle of direct usability. In this hierarchy,

education in the US. Their opponents, who were often called “modernists”, by contrast, aimed to end curricular prescription and the dominance of Latin and [A]ncient Greek in the American colleges.” (Adler, 2020, p. 6)

CLs are decidedly situated at the lowest end. The suggestion that CL study could serve as a preparation for modern language study is usually dismissed by critics as an inefficient trajectory. Thus, CLs are forced to compete directly with the sciences as well as with modern foreign languages, in both cases occupying the position of the underdog.

1.5 Pro side of the public debate

1.5.1 A kaleidoscopic counteroffensive

“Classicists have a duty of care to their subject”, says classics lecturer David Butterfield (2020). Whereas other subjects survive on their own, the survival of Latin and Greek in secondary education is left to the loving care of people sympathetic to the classical cause. In the face of incessant attacks of egalitarian and/or utilitarian tenor, and the ever-looming threat of complete abolition, they have continued to praise the qualities of CLs even more vigorously and pugnaciously. As such, proponents of classics have created what could be named an apologetic tradition, typically characterised by lamentation about the downfall of their beloved discipline and determination to convince the general public of the value of studying CLs.

It is true that CL education was already contested and defended centuries ago, but since the public debate gained momentum in the 20th century, the number of apologetic writings appears to have grown profusely. In addition to the occasional newspaper column, full-length articles and even entire books are dedicated to the matter. To cite a couple of examples would be an arbitrary choice; to inventory all of these works would be a research endeavour in its own right.

Another significant source for the pro side of the public debate, and in a way part of the apologetic tradition, are educational documents pertaining to CL programmes. It is telling that authors of attainment targets for classical subjects tend to specify not just what needs to be learned, but also why it needs to be learned. They do not content themselves with listing the grammatical knowledge that is to be expected from a secondary school pupil, for instance, but hasten to add that study of the Latin / Ancient Greek language system stimulates abstract insight in languages (Maes, 2021). The justifications that are given reveal a lot about policymakers’ ideology regarding language and education. Is classics about maximising linguistic talent? Is it an instrument in identity building? Is it both, or something else?

There is one thing about which there is consensus: The engagement with classical languages and cultures is enriching and worth striving for. This belief unites those who find themselves on the pro side of the debate, but there are indeed numerous ideologies that can underpin it. Classicists and other enthusiasts have adduced many and varied points in favour of the study of CLs. This argumentative diversity can be connected to the versatility of the classical subjects themselves. In its broadest interpretation, the nature of CL instruction is threefold: language, history, and literature. To these three core aspects may be added a fourth aspect, namely pupils' general (non-cognitive) development, which is mostly addressed in an indirect manner through the other three. Moreover, there is notable variation between countries in the way CL education is organised, not only in terms of legislation, starting age, duration and intensity, but in terms of content and focus too: It can be more of a linguistic exercise or rather more of a cultural exploration, for instance.¹⁹

To bring some clarity to this kaleidoscopic image, in what follows we will propose a classification that accommodates the myriad of pro-arguments in a logical instead of enumerative fashion. Essentially, all pro arguments can be understood as a direct reaction to either the egalitarian or utilitarian argument against the study of CLs. As Figure 1 indicates, an initial division is therefore imposed depending on which contra-argument is rebutted, egalitarian (Section 1.5.2) or utilitarian. Within the latter category a further division is made, since there are two distinct reactions to the utilitarian argument, albeit they need not automatically exclude one another. The first is to get to the root of the problem, so to speak, and resist utilitarianism itself (Section 1.5.3). The second is to acknowledge the prevailing world view and build a case for CLs within a functionality-oriented world, in other words to offer counterarguments in the strictest sense (Section 1.5.4).

1.5.2 Rebuttal of the egalitarian argument

Pertinent or impertinent as the egalitarian argument may be, it cannot be an excuse to abolish CL education; that is probably a good summary of the rebuttal of the egalitarian argument.

¹⁹ This variation in content is reflected in the variety of names referring to the relevant school subjects: 'classics', 'classical civilisation', 'classical languages' (e.g., Dutch *klassieke talen*), 'classical literature' (e.g., French *lettres classiques*), 'ancient philology' (e.g., German *Altphilologie*), 'Latin and general language studies' (e.g., Swedish *Latin med allmän språkkunskap*), et cetera.

While it is a logical fallacy to assert that the plain removal of Latin and Greek from the timetables would guarantee a more inclusive school experience, as the egalitarian argument is sometimes used to imply (Bulwer, 2006b, p. 8), the concrete allegations of non-inclusivity are taken seriously by proponents of CL education as valid criticism. While certain people might seek to uphold the elitist aura surrounding CL education (as hinted by Merry & Boterman, 2020), a demand for change and innovation is ringing out from the contra side as well as from the pro side. In the Netherlands, for example, classical organisations have cast doubt themselves on the concept of so-called categorical *gymnasia*, where all pupils study CLs and which used to be held in very high esteem (OIKOS, 2021).

In the last couple of decades, a lot has been stirring in this respect. However, the effort of CL teachers and researchers towards greater inclusivity is not very conspicuous in the public sphere. Most of it is occurring either in specialist journals or through educational equity projects that are the exception rather than the norm (cf. *infra*). Although social media have become an important platform for dissemination and interaction, in the traditional media the reaction to the egalitarian argument is not as visible as the accusation of elitism. Nonetheless, let us briefly go over the ways that the egalitarian argument and its variants are rebutted.

Recently, several hands-on CL projects have called social-egalitarianist criticisms into question. Here CLs are being deployed to improve literacy and raise aspirations in deprived areas or with vulnerable groups, both at the primary and the secondary school level. Examples are *Literacy through Latin* in Great Britain (Bracke, 2016), *Ancient Greeks, Young Heroes* in Flanders (Bracke, 2020), *Latinus Pons* in Germany (Kipf, 2010), and *Lingua Latein* in Switzerland (Wesselmann & Walker, 2021), or on a smaller scale, the introduction of a *gymnasium* in the underprivileged neighbourhood of southeast Amsterdam (Vereniging Classici Nederland, 2021). These projects are living proof that classics can be of value for children and youngsters of all backgrounds.

In regard to the pupils in the average Latin or Greek classroom, there may be societal factors (touched upon in Section 1.4.2) as a result of which not all population groups currently partake equally of CL education. Content-wise, however, there is a lot of potential, and a lot has been done as well, to make the literary and cultural-historical aspects of CL education more inclusive and accessible (e.g., Barnes, 2018; Bracke, 2021; Conley, 2019; Fengler, 2000; Rijser, 2018; Sawyer, 2016). There is much to be said on this topic, but this article is not the right place to

delve deeply into CL teaching practices for inclusivity, so the following single paragraph will suffice.

Seeing that the ancient world was vast and pluralistic and witnessed profound cross-cultural influence (Fibiger Bang, 2010; Hall, 2015; Rüpke, 2010), it has a ton of possibilities to offer for intercultural education. Something as simple as paying heed to the geography of the Hellenistic and Roman empires could be a first step in the right direction (as also noted by Sawyer, 2016). The classical discipline itself has evolved, and can no longer be imagined without gender and queer studies (e.g., Budin & Macintosh Turfa, 2016; Churchill, 2006; Corbeill, 2010; Surtees & Dyer, 2020). One of the main opportunities for a CL teacher to cater for a diverse pupil audience and to showcase diversity, is the choice and discussion of texts. To give some examples: Famous Latin and Greek authors abound of what we would now consider non-Western origin, documentary texts lend a voice to the lower ranks of society, and widening the scope beyond Antiquity to post-classical literature opens up options for reading women. According to an emergent body of anecdotic evidence, studying CLs can in fact enrich pupils' perception of diversity and help them express more balanced judgements (e.g., Barnes, 2018; Hunt, 2016; Rabinowitz & McHardy, 2014).

That leaves the intellectual variant of the egalitarian argument. The language aspect of CL instruction actually harbours a lot of potential too. First of all, it has been pointed out many a time (e.g., Bell & Wing-Davey, 2018; Hill, 2006; Van Houdt, 2012) that in a CL class starting positions are the same for all pupils, which can seldom be said of native language or modern foreign language classes. In case of the CLs, no one has any prior knowledge worth mentioning, so everyone starts learning from scratch. While some pupils will already be more familiar with Antiquity and classical culture than others depending on the input they receive at home, the Latin and Ancient Greek languages are usually just as new to every beginning pupil.

Furthermore, research suggests that several linguistic and didactic characteristics of the CLs and the traditional grammar-oriented way of teaching them can contribute to the creation of an inclusive learning environment (Ancona, 1982; Ashmore & Madden, 1990; Block et al., 1995; Hill, 2006; Hill et al., 1995): The linguistic make-up of Latin and Greek, together with the focus on written words and the explicit presentation of grammar, make them attractive to weaker language learners and accommodate for the difficulties they typically encounter when learning

foreign languages.²⁰ The attentive reader will notice that these findings are diametrically opposed to the claims regarding the grammar-translation method mentioned in Section 1.4.2.

Within other teaching paradigms as well, pupil diversity is the order of the day. Textbooks for CLs have evolved and continue to evolve to be more accessible and appeal to wider audiences (e.g., Paul, 2013 on the *Cambridge Latin Course*). Communicative approaches to CL learning and instruction, which have caught on in the US some years ago and are slowly gaining acceptance in Europe, also often present themselves as a possible route towards greater inclusivity (Hunt, 2021). Teachers who have incorporated elements of active Latin in their lessons report an increase in pupil engagement, motivation and enjoyment, which may in turn lead to a more successful learning experience for a larger number of pupils (e.g., Hunt, 2021; Rasmussen, 2015).

In short, teachers and researchers of CLs have managed to mitigate the egalitarian argument in every respect, by proving that the possibility for greater inclusivity is engrained in the very subjects themselves. Latin and Ancient Greek “have been rescued from what sometimes seemed a terminal association with elite power, dead White males, imperial and gender exploitation” (Hardwick, 2013, p. 16). There is reason to speak of a shift in both practice and image: As

²⁰ At least three linguistic characteristics can be identified that meet the needs of weaker language learners who tend to perform poorly in modern language classes. The first one is grapheme-phoneme correspondence: Letters customarily produce one sound when pronounced orally, which make spelling and pronunciation less troublesome to process. The second one is inflection, since it clarifies syntactic and semantic relations. Unlike in English (or French, Dutch, Spanish, Italian...), one does not rely exclusively on word order to obtain meaning; rather, decoding a sentence requires diligent inspection of every word and its components. Breaking down words into morphemes and relying on inflections to find out the function of a constituent enforces understanding of the mechanics of language. The third one has to do with vocabulary. Latin specifically has got a relatively small lexicon and uses few idioms (Hill et al., 1995, p. 49), which simplifies the task of learning vocabulary.

If one moves on to teaching method, besides the fact that low incidence classes in which all pupils take off at the same point in the learning curve are a better fit for pupils who experience severe foreign language learning difficulties, there are two more relevant characteristics. Firstly, the focal point lies on written rather than spoken language, which appeals to learners with poor auditory ability who cannot easily discern the sounds of spoken language. Moreover, the approach is usually multisensory, meaning that sentences can simultaneously be seen and heard when the teacher reads them aloud. Weaker language learners typically fall short in phonological coding (Sparks and Ganschow, 1993), so that in an oral message they fail to distinguish different words from each other. Secondly, grammar tends to be presented explicitly, which is convenient for those who struggle to build internal rules on the implicit basis of contact with authentic language material, another typical obstacle for weaker language learners. A slower, incremental way of language instruction is conducive to the language acquisition of these pupils.

participation is being widened to historically non-traditional target groups, the CLs are increasingly associated with social critique and liberation movements (Bracke, 2023; Hardwick, 2013). It will be clear from the discussion of the egalitarian argument and its rebuttal that the conceivable interpretations and receptions of classical material are manifold, and that these can be prompted by strongly divergent values and ideologies.

1.5.3 Rebuttal of the utilitarian argument by resisting utilitarianism itself

Now, how is the utilitarian argument rebutted in the public debate? One fashionable reaction amongst the proponents of CL education is to launch a counterattack and question the axioms of the utilitarian approach towards education. By means of *exemplum* we quote John Bulwer (2006b, p. 9):

“Parents want to be confident that what their children learn at school will be useful for them in the future. This can lead them to some curious conclusions, as they appear to believe that a lot of material studied at a fairly advanced stage of school will be actually used by the pupils in their future careers. The rapid pace of change in many areas means that almost all of this material will be out of date by the time these pupils enter the world of work. (...) Parents and students are still over-intent on the immediate use of a subject to a possible future career, and need convincing that education at the higher level is more than simply training.”

To put it differently, the very idea that general secondary education is capable of providing *Ausbildung*, is largely an illusion. Bulwer continues (2006b, pp. 9–10):

“Utilitarianism also has its own drawbacks. Its concentration on the greatest good for the greatest number poses problems for minorities and those who have a strong commitment to a particular subject but who cannot be accommodated alongside the greater numbers who opt for the more popular subjects. If a student genuinely does not want to study what everyone else does but opts for something else, his or her wishes are ignored under a utilitarian system which emphasises conformity and coherence. Philosophers argue that utilitarianism can force someone to act in [a] way that compromises their integrity. It could be argued that a strong wish to study Classical languages represents a kind of integrity which our education systems tend to compromise.”

How precisely his words mirror those of Bolgar (1963, p. 11) a good forty years earlier:

“A boy today cannot choose to study Latin as freely as he would choose German or Geography or Mathematics. He has to reckon with a hostile public opinion, and he has to learn to defend his fancy against critical parents, friends and enquirers.”

If you look at it from that angle, a utilitarian approach towards education is not in the best interest of the pupil: It restricts their freedom for self-development, and in a rather naïve way at that, considered the speed at which the sciences and the labour market evolve. CLs, on the contrary, embody the timeless ideal of liberal education, of a broad formation which above all benefits the pupil as a human being.²¹ In regard to education, the principle of direct usability on which the utilitarian ideology is founded is rejected altogether by this strand of the pro side, in favour of *Bildung*.

1.5.4 Rebuttal of the utilitarian argument by offering counter-arguments

Another strategy is to choose pragmatism over idealism, and search for counterarguments within a utilitarian framework. Advocates of CL education who have ventured to ask what roles CLs can fulfil in contemporary education, given that the mainstream vision is highly instrumental, have come up with a plethora of arguments. Here we wish to propose a classification into four overarching categories: the cultural-historical argument, the general-developmental argument, the linguistic argument and the cognitive argument. Figure 2 visualises how these tie in with the four aspects of CL instruction distinguished in Section 1.5.1. The cultural-historical argument groups the history and literature aspects. The general-developmental argument speaks for itself. The linguistic and cognitive arguments are both linked to the language aspect.

What all four counterarguments have in common is that they carry an acknowledgement of educational utilitarianism. The goal of this strand of the pro side is to get CLs higher up within the hierarchy of usability. The first two arguments attempt to do so by propagating the

²¹ In Flanders, the old name for general secondary education, which is used in colloquial speech to this day, is *humaniora*, “that which makes you more human”.

Primärnutzen or “primary use” of CLs (Section 1.5.4.1), the latter two their *Sekundärnutzen* or “secondary use” (Section 1.5.4.2). As a note in advance, different individuals and organisations situated on the pro side of the debate do not necessarily agree on the significance of each of these counterarguments, just like they might disagree on the weight that should be attached to each of the instructional aspects. For 20th-century examples in various languages of statements representative of each of these arguments, we refer to Stellwag (1949, pp. 165–168).

Aspect of CL instruction	Counterargument
history	cultural-historical argument
literature	
general development	general-developmental argument
language	linguistic argument
	cognitive argument

Figure 2 Classification of counterarguments in relation to aspects of CL instruction

1.5.4.1 The cultural-historical and general-developmental arguments

The cultural-historical and general-developmental arguments counter the utilitarian argument by stating the merit of CLs in their own right. Granted, they may not be all that useful for daily-life communication. Their *Primärnutzen* does not lie in the conveyance of practical knowledge, but in the conveyance of culture and of soft skills, which are equally valuable assets.²²

The cultural-historical argument emphasises the rich historical and literary heritage that pupils are brought into contact with, which first of all is intrinsically worthwhile, and second of all has a special relation to Western society. It is a truism that what we consider to be Western culture can be traced back to Greco-Roman Antiquity and Christianity. Examples of Greek and Roman influences are legion: in law, politics, philosophy, oratory... A sprinkle of

²² As far as the *Primärnutzen* is concerned, there is a notable difference between the situation in the 20th and 21st centuries and the situation in previous centuries. For as long as Latin was a customary language of science and international communication, its instruction was defensible with an eye to Latin literacy in itself. By the 20th century, the Latin literacy argument *stricto sensu* had gone out of fashion for obvious reasons, and was completely superseded by the cultural-historical and general-developmental approaches towards reading Latin presented in this subsection.

pathos is not alien to this argument. Just think of common expressions like “the cradle of Western civilisation”, or the dictum by classical philologist Karl Büchner (1978), “die lateinische Sprache ist es gewesen, die Europa geschaffen hat” [it was the Latin language that has created Europe].

The cultural-historical branch of argumentation can be pursued further in various directions (Maes, 2021; Spelmans, 2010, pp. 176–179). It can take a purely aesthetic turn, focusing on the quality of classical literature and other cultural expressions, and on the educational profit of engaging with beautiful texts and artefacts. By virtue of their literary and artistic education, Latin and Greek distinguish themselves from other subjects. It can also take an ethic turn: Qualitative literature often treats general existential themes, from which we can draw lessons even today. Furthermore, the foreignness of the texts, written in a context wildly remote from that of the 21st-century reader, presumably heightens tolerance towards others. Finally, it can take a sociocultural turn, which is related more to the Greco-Roman inheritance of Western culture. The study of CLs is supposed to foster pupils’ awareness of the specific and relative value of Western cultural expressions as well as their reflective attitude towards contemporary norms, values and beliefs.

Typical for the cultural-historical argument is the exceptional status that is assigned to Classical Greece and Rome. The idolisation of Classical Antiquity as the cradle of Western civilisation really blossomed during the Enlightenment (Verbaal, 2021), although its roots go back much further: In Late Antiquity there already was a well-developed sense of what counted as classical literature, and obviously Classical Antiquity had already known a great revival in the Renaissance. It says a lot that there has not been a single point in time, after it was written, that Virgil’s *Aeneid* has not been on the curriculum. Even so, the Latin and Greek literary traditions did not break off in the first century AD, but rather continued deep into the Modern Era; on that basis, the opinion has been voiced that CL instruction ought to take later periods into account as well, if it truly wants to lay claim to a cultural-historical argument (e.g., Verbaal, 2021).

The general-developmental argument showcases the broad range of soft skills that pupils acquire by studying a CL. These have been documented by reports of CL programmes, academic articles and opinion pieces (e.g., Bracke, 2016; DeVane, 1997; Harrington- Lueker, 1992; Katz et al., 2020; Sussman, 1978; Van Damme, 2021). A strong cultural and historical

awareness is one of these skills. Somewhat related are a critical mind, independent judgement,²³ and the capacity to deal with ambiguity. Metacognitive and affective abilities are thought to be honed as well, such as the application of learning strategies, intellectual ambition, discipline, confidence and positive self-image. Personality traits that are associated with CL learning are grit, conscientiousness, self-control, motivation and inquisitiveness.²⁴ Last but not least, studying a CL is likely to increase appreciation of linguistics, arts and humanities.²⁵

Interestingly, the tenor of the cultural-historical and general- developmental arguments is perfectly compatible with the European key competences for the 21st century. If we zoom in on the history and literature aspects of Latin and Greek, it is evident they are easily connected with several key competences (Council of the European Union, 2018). The most self-explanatory one of all, is “cultural awareness and expression”. “Personal competence”, like skills related to personal mental and emotional health, is addressed nowhere more explicitly in the general secondary curriculum than in classes on ancient philosophy, which was after all about the art of living. “Citizenship competence” is developed through the confrontation with the society of the Greeks and Romans, and the comparison with our own. More specifically, argumentative skill receives ample treatment in classes on rhetoric. So, without having to look hard or interpret in overly creative ways, it soon becomes clear that CL education can significantly contribute to the realisation of the key competences.

1.5.4.2 The linguistic and cognitive arguments

The remaining two counterarguments are especially fascinating from an ideological perspective. The linguistic and cognitive arguments go along with the utilitarian way of thinking, only to seek

²³ See Crump 2008 for how CL education in the Netherlands ought to be more committed to cultivating independent judgement.

²⁴ In March 2023, however, the IZA Institute of Labor Economics, an independent economic research institute, released a report of an Italian study on personality traits with a rather surprising conclusion. The report states that CL study does not affect conscientiousness and openness, but does increase neuroticism and self-reported unhappiness (Brunello et al., 2023). This finding circulated both in the traditional media and on social media for a while, despite the research being methodologically flawed. As McDaniel (2023) aptly analyses, the authors overstepped the actual scope of their research and drew a too far- reaching conclusion with respect to their limited data on *licei classici* and *licei scientifici*.

²⁵ Testimonies of (former) CL pupils often state that they find learning Latin/ Greek fun, and choose to study it for the fun (see also Butterworth, 2017; Katz et al., 2020; Taylor et al., 2023). Pleasure might not be an argument for the preservation of a school subject *per se*, but at the very least this observation deserved a footnote.

to defeat the utilitarian opposition on its own terrain, by asserting that the study of CLs *is* directly useful for something else. In other words, it is not just about Latin and Ancient Greek themselves, but there is a *Sekundärnutzen*. On a linguistic level, this may be called language awareness or proficiency or the like; on a cognitive level, it may be called problem-solving, analytical reasoning or something of that purport. In both academic and everyday speech, these secondary uses are referred to as transfer value.

Some of the most rapidly mentioned reasons to study CLs fall under the category of the linguistic argument or of the cognitive argument (illustrated for example in the research of Smith, 2007 and Taylor et al., 2023). When someone with a favourable disposition towards Latin is asked to advocate for the subject, it is not implausible the advocacy goes like this: Latin enhances mastery of one's native tongue, opens up the realm of scientific terminology, facilitates the acquisition of modern and in particular Romance languages, and teaches you how to study, reason and think. It is a well-known cliché that Latin is the ultimate “brain gymnastics”.

Although there are national differences and nuances, these common-sense reasons are popular across countries. The German attitude research study, already cited in Section 1.4.2, also found 80% of its participants held the view that Latin is of greater use than any modern language with an eye to logical thinking, command of German, and foreign language aptitude (Gerhards et al., 2019). The linguistic argument in favour of CL learning is all the more poignant in countries where the official language is considered to be closely connected to its classical predecessor, like in Greece.

The *in se* rather bold claims of linguistic and cognitive benefits are supported by a mountain of anecdotic evidence (see references and quotes in Holmes-Henderson & Kelly, 2022), by people confirming them to be true going on their perceptions of their own experiences with CL education. The arguments are sometimes further backed up with examples of higher achievement by (former) CL pupils – who do indeed tend to outperform their peers on a variety of measures (Bracke & Bradshaw, 2020), from standardised language tests, over medical entrance exams, to success rates and study efficiency in higher education. In the United States, for instance, Latin pupils consistently achieve high scores on the verbal portion of the SAT,²⁶ a standardised test widely used for college admissions (Cooper et al., 2008; LaFleur, 1981, 1982).

²⁶ Conversely, CL programmes have been designed with the express purpose of coaching pupils for the SAT (Holmes & Keffer, 1995).

The linguistic and cognitive arguments really are at the heart of the public debate regarding CL education. Official educational documents are often keen to praise the alleged linguistic and cognitive benefits of CL study, and use them as justification for new or existing educational practices. The abovementioned Walloon Excellency Pact, for instance, imposes Latin in order to uplift the command of French (Fédération Wallonie-Bruxelles, 2023). In France too, Latin is taught with the explicit goal of supporting native language ability (Denis-Laroque, 2006). The Swedish attainment targets for Latin literally include scientific terminology and modern foreign language aptitude (Skolverket, 2023). In Germany there even exists a thriving programme, known under the name of *Latein plus* or *Englisch- und Lateinunterricht in Kooperation* depending on the *Bundesland*, where Latin and English are taught together in parallel for the sake of language-specific and language-general linguistic insight (Thies, 2004; Van Houdt, 2012).

Provided that the study of Latin and Ancient Greek boosts linguistic and cognitive abilities, CL education can add yet more European key competences to their record, namely the “literacy”, “multilingual” and “learning to learn” competences (Council of the European Union, 2018). Hence Figure 2 can be expanded with the corresponding European key competences, shown in Figure 3 below.²⁷

Aspect of CL instruction	Counterargument	European key competence
history	cultural-historical argument	cultural awareness and expression
literature		
general development	general-developmental argument	personal, social, and citizenship competence
language	linguistic argument	literacy and multilingual competence
	cognitive argument	learning to learn competence

Figure 3 European key competences corresponding to the counterarguments

²⁷ The discussion of CL education and key competences in this article is of a theoretical nature. For case studies in Italy and Spain, see Canfarotta et al., 2022.

1.6 Epilogue

1.6.1 Some reflections on the value of the arguments discussed

CL education is a controversial topic. Unlike in a fair trial, in a public debate like the one outlined above, the burden of proof lies with the defence. How do they perform? Starting with the issue of accessibility and inclusivity, one could say that CL teachers and researchers have gone beyond a mere refutation of the egalitarian argument and have actually turned the tables in their favour through their rebuttal. As things stand though, CL education in general still has a long way to go if it wants to earn the reputation of open-minded and intercultural education. A great deal more time and effort will be required to inflict lasting changes on the public image of CL education.

In response to the criticism of uselessness, the resistance to utilitarianism sketched above will perhaps resonate with many a classicist. Considering strike power in the debate, however, surely it is not the optimal strategy to reject the utilitarian stance altogether. Due to the immediate confrontation of opposite ideologies, such a reaction to the utilitarian argument is more likely to serve as preaching to the converted than to making any new converts.

What is then the value of the counterarguments? The cultural- historical and general-developmental arguments are powerful, and arguably the most compelling (Adler, 2020), but they lose strength owing to woolly and esoteric parlance (Van Damme, 2021) and their disconnect from modern psychology (Vereeck et al., 2025). The linguistic and cognitive arguments have been repeated so many times that they almost have become facts in the minds of those disposed to believe them; for a more sceptical audience, they fail to convince because they are based on perceptions and correlations at best, not on causal empirical evidence. As long as the transfer value of CL education is not submitted to rigorous scientific scrutiny,²⁸ the linguistic and cognitive arguments amount to little more than an apologetic *topos*.

The crux here remains whether any extraordinary achievements by (former) CL pupils are chiefly explained by the transfer value of CL education, or by preselectivity, i.e. by the innate

²⁸ Under the auspices of Research Foundation – Flanders, such innovative research on the transfer effects of CL learning is currently being performed at Ghent University by PhD students Alexandra Vereeck and Cathy Hauspie. The empirical findings can be expected in the following years.

capacities of pupils who choose to study Latin and/or Ancient Greek. Research so far has not been able to solve this conundrum. There is nonetheless a substantial research literature on the linguistic and general-cognitive impact of studying CLs (Vereeck et al., 2025), which was already alluded to in the introduction. Particularly in the United States, dozens of studies were conducted in the course of the 20th century, with predominantly positive findings regarding CL impact. Despite the large number of studies, they do not seem to have been disseminated very widely. Moreover, due to methodological issues, this body of research has not managed to rule out the preselectivity explanation. Thus, the door is left wide open for sceptics to brush aside the intellectual successes of CL pupils as mere selection bias, to which there is no sound reply on hand.

Finally, the objection has been raised that none of the arguments on the pro side are exclusive to Latin and Greek. Some might indeed apply to other school subjects,²⁹ but it is doubtful whether there is another single subject to which all arguments apply. Maybe the profound entanglement of language and culture is what truly makes Latin and Greek unique, albeit that there are large differences in this regard between teaching paradigms.

1.6.2 Conclusions and suggestions for future research

We can conclude that this debate is fully rooted in language ideology. More particularly, it is a matter of perceived prestige and utility. Where on these spectra CLs belong and what the implications are thereof, that's always the question. Whichever arguments are put in the forefront depends on the image CL education has in that particular time and place. Future research could pay more attention to differences in the debate between countries, or give a more detailed account of one particular region or country.³⁰ Other options for future research include studying social media instead of traditional media, or outlining the debate in earlier time periods predating the late 19th century.³¹

²⁹ Proposals exist in abundance for other school subjects that ought to replace CLs. We have chosen not to elaborate this point, seeing how this would necessitate a discussion of to what extent the classical subjects live up to the arguments pro, which surpasses the scope of the current article.

³⁰ For Flanders, there has previously been Spelmans (2010), who took a more philosophical approach.

³¹ Steps have been taken in that direction by Pire (1971) and Drago (2008).

One thing is certain anyway: The public debate on the role and value of CL education is going round in circles. Very little has changed over time, as demonstrated by multiple quotes in this article. The final word is for the author of *Max Havelaar*, best known under his *nom de plume* Multatuli, who in 1873 penned an aphorism not any less apt at the time of writing in 2023, exactly one and a half centuries later:

“Het komt me voor, dat de telkens hernieuwde strijd over het nut der klassieken bij het onderwijs slecht gevoerd wordt. De wederzijdse argumenten geven meer blijk van voor- en tegeningenomenheid, dan van onpartijdig onderzoek en zaakkennis.” (Multatuli, 1873, no. 1105)

[It appears to me, that the continually renewed fight about the usefulness of classics in education is waged poorly. The arguments on both sides mostly reflect biases pro and contra, rather than impartial research and expertise. (Multatuli, 1873, no. 1105)]

Works cited in this chapter

- Adler, E. (2020). *The Battle of the Classics: How a nineteenth-century debate can save the humanities today*. Oxford University Press.
- Advisory Committee. (1924). *Part one, general report*. Princeton University Press.
- Aloi, N. (2010). *La scuola in Parlamento*. Pellegrini.
- Ancona, R. (1982). Latin and a dyslexic student: An experience in teaching. *The Classical World*, 76(1), 33–36.
- Ashmore, R., & Madden, J. (1990). Literacy via Latin: A case study. *Journal of College Reading and Learning*, 23(1), 63–70.
- Barnes, J. (2018). Developing students’ ideas of diversity in the ancient and modern worlds through the topic of Alexandria in the *Cambridge Latin Course, Book II*. *Journal of Classics Teaching*, 19(37), 1–9.
- Bell, B., & Wing-Davey, Z. (2018). Delivering Latin in primary schools. In A. Holmes-Henderson, S. Hunt, & M. Musié (Eds.), *Forward with Classics. Classical languages in schools and communities* (pp. 111–128). Bloomsbury Academic.
- Block, L., Brinckerhoff, L., & Trueba, C. (1995). Options and accommodations in mathematics and foreign language for college students with learning disabilities. *Higher Education and the Handicapped*, 14(2), 1–5.
- Bolgar, R. (1963). A theory of classical education. *Didaskalos. The journal of the Joint Association of Classical Teachers*, 1(1), 5–26.
- Boone, S. (2011). *Sociale ongelijkheid bij de overgang van basis- naar secundair onderwijs: Een onderzoek naar de oriënteringspraktijk*. Vlaams Ministerie van Onderwijs en Vorming.
- Bourdieu, P. (1992). *Language and symbolic power*. Polity Press.
- Bracke, E. (2016). The role of university student teachers in increasing Widening Participation to Classics. *Journal of Widening Participation and Lifelong Learning*, 18(2), 111–129.

- Bracke, E. (2020). De impact van Oudgriekse studie op kansarme leerlingen. Participatieverruiming in het lagere scholenproject Oude Grieken—Jonge Helden. *Lampas*, 53(3), 90–103.
- Bracke, E. (2021). 'Iedereen Latijn' van theorie naar praktijk: Een diversiteitstoolkit. In T. Velle & Y. Maes (Eds.), *Latijn anders? Waarom er nood is aan ander Latijn in de klas* (pp. 49–74). Skribis.
- Bracke, E. (2023). Connecting social justice and classical languages. In *Classics at primary school. A tool for social justice* (pp. 1–21). Routledge.
- 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.
- Broyler, C., Thorndike, E., & Woodyard, E. (1927). A second study of mental discipline in high school studies. *Journal of Educational Psychology*, 18(6), 377–404.
- Brunello, G., Esposito, P., Rocco, L., & Scicchitano, S. (2023). *Do classical studies open your mind?* [Discussion paper series]. IZA Institute of Labor Economics.
- Büchner, K. (1978). *Latein und Europa. Traditionen und Renaissance*. Reclam.
- Budin, S., & Macintosh Turfa, J. (Eds.). (2016). *Women in Antiquity. Real women across the ancient world*. Routledge.
- Bulwer, J. (Ed.). (2006a). *Classics teaching in Europe*. Duckworth.
- Bulwer, J. (2006b). Introduction. In *Classics teaching in Europe* (pp. 1–18). Duckworth.
- Bulwer, J. (2018). Changing priorities in classics education in mainland Europe. In A. Holmes-Henderson, S. Hunt, & M. Musić (Eds.), *Forward with Classics. Classical languages in schools and communities* (pp. 67–88). Bloomsbury Academic.
- Butterfield, D. (2020, July 18). What would it mean to 'decolonise' the Classics? *The Spectator*. <https://www.spectator.co.uk/article/what-would-it-mean-to-decolonise-the-Classics>
- Butterworth, E. (2017). Latin in the community: The Paideia Institute's Aequora program. *The Classical Outlook*, 92(1), 2–8.
- Caerols, J. (2013). Lo studio delle lingue classiche nel sistema educativo spagnolo. In L. Canfora & U. Cardinale (Eds.), *Disegnare il futuro con intelligenza antica. Insegnamento del latino e del greco antico in Italia e nel mondo* (pp. 317–337). Il Mulino.
- Canfarotta, D., Tosto, C., & Casado-Muñoz, R. (2022). Development of key competences through Latin and Greek in secondary school in Italy and Spain. *Journal of Classics Teaching*, 23(45), 13–21.
- Cart, A. (1971). Les langues anciennes. *Bulletin de l'Association Guillaume Budé*, 4(2), 175–183.
- Churchill, L. (2006). Is there a woman in this textbook? Feminist pedagogy and elementary Latin. In J. Gruber-Miller (Ed.), *When dead tongues speak. Teaching beginning Greek and Latin* (pp. 86–109). Oxford University Press.
- Conley, B. (2019). Non-elite Latin for the classroom. *Sententiae Antiquae*. <https://sententiaeantiquae.com/2019/02/01/non-elite-latin-for-the-classroom/>
- Cooper, T., Yanosky, D., & Wisenbaker, J. (2008). Foreign language learning and SAT Verbal scores revisited. *Foreign Language Annals*, 41(2), 200–217.
- Corbeill, A. (2010). Gender studies. In A. Barchiesi & W. Scheidel (Eds.), *The Oxford handbook of Roman studies* (pp. 220–233). Oxford University Press.
- Council of the European Union. (2018). Council Recommendation of 22 May 2018 on key competences for lifelong learning. *Official Journal of the European Union*, C 189, 13.
- Dahlberg, K. (2006). The essence of essences—The search for meaning structures in phenomenological analysis of lifeworld phenomena. *International Journal of Qualitative Studies on Health and Well-Being*, 1(1), 11–19.
- de Melo, W. (2023). Classics in translation? A personal angle (part II). Translation in Anglophone universities. *Antigone*. <https://antigonejournal.com/2023/03/classics-translation-ii/>
- Denis-Laroque, O. (2006). France. In J. Bulwer (Ed.), *Classics teaching in Europe* (pp. 52–57). Duckworth.
- Depaepe, M. (1999). *Orde in vooruitgang*. University Press Louvain.
- Departement Onderwijs & Vorming. (2021). *STEM-monitor*. Vlaamse Overheid. <https://onderwijs.vlaanderen.be/sites/default/files/2021-07/STEM%20MONITOR%202021.pdf>

- Department for Education. (2021). *Thousands more students to learn ancient and modern languages. £4 million new scheme to give opportunity for all to study Latin*. GOV.UK.
<https://www.gov.uk/government/news/thousands-more-students-to-learn-ancient-and-modern-languages>
- DeVane, A. (1997). *Efficacy of Latin studies in the information age* [Paper submitted for PSY 702: Educational Psychology, Valdosta State University].
<http://www.edpsycinteractive.org/files/latin.html>
- Domina, T., Penner, A., & Penner, E. (2017). Categorical inequality: Schools as sorting machines. *Annual Review of Sociology*, 43(1), 311–330.
- Douglass, H., & Kittelson, C. (1935). The transfer of training in high school Latin to English grammar, spelling, and vocabulary. *Journal of Experimental Education*, 4(1), 26–33.
- Drago, R. (Ed.). (2008). Opinioni illustri dal XVIII al XX secolo. In *Latino perché? Latino per chi?* (pp. 139–163). Associazione TreeLLLe.
- Duyck, W., Janse, M., & Soltic, J. (2017). Stem voor Grieks en Latijn. Wat overleeft, leeft. In J. Strobbe, P. Thoen, & G. Van Coillie (Eds.), *Vorming door onderwijs. Humaniora tussen verleden en toekomst* (pp. 169–184). Acco.
- Ellis, A., Cogan, J., & Howey, K. (1991). Historical perspectives. Education in America. In *Introduction to the foundations of education* (3rd edition, pp. 77–96). Allyn and Bacon.
- Facq, J., Waumans, L., Paternotte, F., Goedert, C., Brux, C., & Hanegreefs, N. (2006). Belgium. In J. Bulwer (Ed.), *Classics teaching in Europe* (pp. 22–30). Duckworth.
- Farr, M., & Song, J. (2011). Language ideologies and policies: Multilingualism and education. *Language and Linguistics Compass*, 5(9), 650–665.
- Fédération Wallonie-Bruxelles. (2023). *Pacte pour un enseignement d'excellence*. Enseignement.be.
<https://pactepourunenseignementdexcellence.cfwb.be/>
- Fengler, C. (2000). Lateinunterricht und ausländische Schüler—Ein Erfahrungsbericht. *Pegasus Online-Zeitschrift*, 1(1).
http://www.pegasus-onlinezeitschrift.de/alte_seite/erga12000fengler.htm#5
- Fibiger Bang, P. (2010). Imperial ecumene and polyethnicity. In A. Barchiesi & W. Scheidel (Eds.), *The Oxford handbook of Roman studies* (pp. 671–684). Oxford University Press.
- Forrest, M. (2003). The abolition of compulsory Latin and its consequences. *Greece & Rome*, 50(2), 42–66.
- Foster, F., & Wise, J. (2022). Social tensions in studying ancient history. *The Curriculum Journal*, 33(4), 536–552.
- Gerhards, J., Kohler, U., & Sawert, T. (2021). Educational expansion, social class, and choosing Latin as a strategy of distinction. *Zeitschrift Für Soziologie*, 50(5), 306–321.
- Gerhards, J., Sawert, T., & Kohler, U. (2019). Des Kaisers alte Kleider: Fiktion und Wirklichkeit des Nutzens von Lateinkenntnissen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 71(2), 309–326.
- Glücklich, H.-J. (2006). Germany. In J. Bulwer (Ed.), *Classics teaching in Europe* (pp. 58–69). Duckworth.
- Grootaers, D. (1998). *Histoire de l'enseignement en Belgique*. CRISP.
- Gutacker, B. (1979). Fördert Lateinunterricht sprachliche Fertigkeiten im Deutschen? In K. J. von Klauer & H.-J. Konradt (Eds.), *Jahrbuch für Empirische Erziehungswissenschaft 1979* (pp. 9–32). Schwann.
- Haag, L. (1995). Auswirkungen von Lateinunterricht. *Psychologie in Erziehung und Unterricht*, 42, 245–254.
- Haag, L. (2001). Auswirkungen von Lateinunterricht – Ergebnisse nach zwei Lernjahren. *Psychologie in Erziehung und Unterricht*, 48, 30–37.
- Haag, L., & Stern, E. (2000). Non scholae sed vitae discimus? Auf der Suche nach globalen und spezifischen Transfereffekten des Lateinunterrichts. *Zeitschrift für Pädagogische Psychologie*, 14(2/3), 146–157.
- Hall, E. (2015). *Introducing the Ancient Greeks*. Bodley Head.
- Hardwick, L. (2013). Against the 'democratic turn'. Counter-texts; counter-contexts; counter-arguments. In L. Hardwick & S. Harrison (Eds.), *Classics in the modern world. A 'democratic turn'?* (pp. 15–32). Oxford University Press.

- Harrington-Lueker, D. (1992). Latin redux. *The Executive Educator*, 14(8), 21–25.
- Herrlitz, H., Hopf, W., & Titze, H. (1993). *Deutsche Schulgeschichte von 1800 bis zur Gegenwart. Eine Einführung*. Juventa.
- Hill, B. (2006). Latin for students with severe foreign language learning difficulties. In J. Gruber-Miller (Ed.), *When dead tongues speak. Teaching beginning Greek and Latin* (pp. 50–67). Oxford University Press.
- Hill, B., Downey, D., Sheppard, M., & Williamson, V. (1995). Accommodating the needs of students with severe language learning difficulties in modified foreign language classes. In G. Crouse, P. Campana, & M. Harmon Rosenbusch (Eds.), *Broadening the frontiers of foreign language education. Selected papers from the 1995 Central States Conference* (pp. 46–56). National Textbook.
- Hirtt, N., Nicaise, I., & De Zutter, D. (2007). *De school van de ongelijkheid*. Epo.
- Holmes, T., & Keffer, R. (1995). A computerized method to teach Latin and Greek root words: Effect on Verbal SAT scores. *Journal of Educational Research*, 89(1), 47–50.
- Holmes-Henderson, A., Hunt, S., & Musić, M. (Eds.). (2018). *Forward with Classics. Classical languages in schools and communities*. Bloomsbury Academic.
- Holmes-Henderson, A., & Kelly, K. (2022). *Ancient languages in primary schools in England: A literature review*. Department for Education.
<https://www.gov.uk/government/publications/ancient-languages-in-primary-schools-in-england>
- Hunt, S. (2016). Teaching sensitive topics in the secondary classics classroom. *Journal of Classics Teaching*, 17(34), 31–43.
- Hunt, S. (2021). Active Latin teaching for the inclusive classroom. In M. Lloyd & S. Hunt (Eds.), *Communicative approaches for ancient languages* (pp. 55–64). Bloomsbury Academic.
- Hunt, S. (2022). *Teaching Latin. Contexts, theories, practices*. Bloomsbury Academic.
- Hunt, S., & Holmes-Henderson, A. (2021). A level Classics poverty. Classical subjects in schools in England: Access, attainment and progression. *Council of University Classical Departments Bulletin*, 50.
<https://cucd.blogs.sas.ac.uk/bulletin/>
- Irvine, J. (2012). Language ideology. In *Oxford Bibliographies*.
<https://www.oxfordbibliographies.com/view/document/obo-9780199766567/obo-9780199766567-0012.xml>
- Janak, E. (2018). Bracketing and bridling: Using narrative reflexivity to confront researcher bias and the impact of social identity in a historical study. *Philanthropy & Education*, 1(2), 82–93.
- Jaspers, J. (Ed.). (2009). *De klank van de stad. Stedelijke meertaligheid en interculturele communicatie*. Acco.
- Katz, J., Noels, K., & Fitzner, A. (2020). Why learn Latin? Motivation for learning a classical language. *Teaching Classical Languages*, 11(1), 89–129.
- Kipf, S. (2010). Integration durch Bildung. Schülerinnen und Schüler nichtdeutscher Herkunftssprache lernen Latein. *Forum Classicum*, 14(3), 181–197.
- Korn, M. (2020). Die Sicht eines Fachdidaktikers auf die aktuelle Stellung des Unterrichtsfachs Latein in Deutschland. *Cursor Latein4EU. Zeitschrift für Freunde der lateinischen Sprache und europäischen Kultur*, 16, 16–17.
- Laes, C. (2016). Waarom STEM het nieuwe Latijn niet is (en dat ook niet hoeft te zijn). *Prora*, 21(4), 14–17.
- Laes, C. (2019). Quo vadimus? Latin and Greek in Flemish secondary schools. *Cursor Latein4EU. Zeitschrift für Freunde der lateinischen Sprache und europäischen Kultur*, 15, 58–60.
- LaFleur, R. (1981). Latin students score high on SAT and achievement tests. *The Classical Journal*, 76(3), 254.
- LaFleur, R. (1982). 1981 SAT and Latin achievement test results and enrollment data. *The Classical Journal*, 77(4), 343.
- Lanvers, U. (2017). Elitism in language learning in the UK. In D. Rivers & K. Zotzmann (Eds.), *Isms in language education. Oppression, intersectionality and emancipation*. De Gruyter Mouton.
- Lashbrook, A. (1969). Latin: The basic language. *The Classical Journal*, 64(4), 163–166.

- Libatique, D., & McHardy, F. (Eds.). (2023). *Diversity and the study of Antiquity in higher education. Perspectives from North America and Europe*. Routledge.
- Lister, B. (Ed.). (2008). *Meeting the challenge. International perspectives on the teaching of Latin*. Cambridge University Press.
- Lucasse, F. (1929). What the Classical Investigation has accomplished. *Journal of Educational Research*, 20(1), 15–21.
- Maes, Y. (2021). China, de toekomst voor het Latijn. In T. Velle & Y. Maes (Eds.), *Latijn anders? Waarom er nood is aan ander Latijn in de klas* (pp. 31–48). Skribis.
- Masciantonio, R. (1977). Tangible benefits of the study of Latin. A review of research. *Foreign Language Annals*, 10(4), 375–382.
- Masschelein, J., & Simons, M. (2012). *Apologie van de school. Een publieke zaak*. Acco.
- McDaniel, S. (2023). Does classical studies really make people neurotic and unhappy? *Tales of Times Forgotten*.
<https://talesoftimesforgotten.com/2023/04/09/does-classical-studies-really-make-people-neurotic-and-unhappy/>
- Merry, M., & Boterman, W. (2020). Educational inequality and state-sponsored elite education: The case of the Dutch gymnasium. *Comparative Education*, 56(4), 522–546.
- Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics*, 5(4), 530–555.
- Morwood, J. (2003). *The teaching of classics*. Cambridge University Press.
- Multatuli. (1873). *Ideeën VI*. Funke.
- Nesbitt, E. (1979). Classics and multicultural education. *Hesperiam*, 2, 78–95.
- OIKOS. (2021). OIKOS conference ‘Shared Antiquity’—Pt. 1. *Rijksuniversiteit Groningen > OIKOS*.
<https://www.rug.nl/research/research-let/onderzoekscholen/oikos/calendar/2021/0212-shared-ant>
- Ortner, T., Asanger, M., Kubinger, K., & Proyer, R. (2008). Zur Frage nach Auswirkungen von Lateinunterricht auf die kognitive Fähigkeit ‘Reasoning’. *Psychologie in Erziehung und Unterricht*, 55, 189–195.
- Paul, J. (2013). The democratic turn in (and through) pedagogy. In L. Hardwick & S. Harrison (Eds.), *Classics in the modern world. A ‘democratic turn’?* (pp. 143–156). Oxford University Press.
- Paulsen, F. (1965). *Geschichte des gelehrten Unterrichts auf den deutschen Schulen und Universitäten vom Ausgang des Mittelalters bis zur Gegenwart. Mit besonderer Rücksicht auf den klassischen Unterricht* (R. Lehmann, Ed.; 3rd edition, Vol. 2). De Gruyter.
- Perra, A. (2006). Italy. In J. Bulwer (Ed.), *Classics teaching in Europe* (pp. 70–80). Duckworth.
- Pfeiffer, R. (1976). *History of classical scholarship* (Vol. 2). Pergamon Press.
- Piantaggini, L. (2020). Grammar-translation: What is it—really—for students? *Journal of Classics Teaching*, 21(42), 92–94.
- Pire, G. (1971). *Le latin en question. Vues modernes sur l’étude des langues anciennes*. Dessain.
- Rabinowitz, N., & McHardy, F. (2014). *From abortion to pederasty. Addressing difficult topics in the classics classroom*. Ohio State University Press.
- Rasmussen, S. (2015). Why oral Latin? *Teaching Classical Languages*, 6(1), 37–45.
- Reitermayer, A. (2006). Austria. In J. Bulwer (Ed.), *Classics teaching in Europe* (pp. 19–21). Duckworth.
- Richards, J., & Rodgers, T. (2014). *Approaches and methods in language teaching. A description and analysis* (3rd revised edition). Cambridge University Press.
- Rijser, D. (2018). Open het gymnasium! *De Nederlandse Boekengids*, 3(2), 3.
- Rüpke, J. (2010). Religious pluralism. In A. Barchiesi & W. Scheidel (Eds.), *The Oxford handbook of Roman studies* (pp. 748–766). Oxford University Press.
- Sawert, T. (2018). *Latente Mechanismen sozialer Hierarchisierung. Die Wahl alter Sprachen als Reproduktionsmechanismus des Bildungsbürgertums*. VS Verlag für Sozialwissenschaften.
- Sawyer, B. (2016). Latin for all identities. *Journal of Classics Teaching*, 17(33), 35–39.
- Schleicher, A. (2019). *PISA 2018. Insights and interpretations*. OECD.

- Schweizerische Maturitätskommission. (2009). *Latinum Helveticum*. Eidgenössisches Department für Wirtschaft, Bildung und Forschung.
<https://www.sbf.admin.ch/sbf/de/home/bildung/maturitaet/gymnasiale-maturitaet/latinum-helveticum.html>
- Secretariat of the EU STEM Coalition. (2023). *EU STEM Coalition*.
<https://www.stemcoalition.eu/>
- Skolverket. (2023). *Latin – språk och kultur*.
<https://www.skolverket.se/om-oss/var-verksamhet/regeringsuppdrag-och-remisser/skolverkets-remisser/remisser/2023-01-23-remiss-gallande-amnesplaner/remiss-for-nationella-hogskoleforberedande-program>
- Skurzak, T. (2013). Panoramica sull'insegnamento delle lingue classiche nei paesi europei. In L. Canfora & U. Cardinale (Eds.), *Disegnare il futuro con intelligenza antica. Insegnamento del latino e del greco antico in Italia e nel mondo* (pp. 369–384). Il Mulino.
- Smith, J. (2007). *The correlation of Latin, Greek and the Classical Education model with learning other subjects* [Doctoral dissertation]. Capella University.
- Sparks, R., & Ganschow, L. (1993). Searching for the cognitive locus of foreign language learning difficulties: Linking first and second language learning. *The Modern Language Journal*, 77(3), 289–302.
- Sparks, R., 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.
- Spelmans, I. (2010). Het schoolvak Latijn: Naar een tweede Sanskriet? *Kleio. Tijdschrift voor oude talen en antieke cultuur*, 39(4), 166–188.
- Stellwag, H. (1949). *De waarde der klassieke vorming. Een cultuur-historische, paedagogisch-psychologische en didaktische inleiding*. Wolters.
- Stevens, P., & Dworkin, A. (Eds.). (2019). *The Palgrave handbook of race and ethnic inequalities in education*. Palgrave Macmillan.
- Surtees, A., & Dyer, J. (2020). *Exploring gender diversity in the ancient world*. Edinburgh University Press.
- Sussman, L. (1978). The decline of basic skills. A suggestion so old that it's new. *The Classical Journal*, 73(4), 346–352.
- Taylor, A., Holmes-Henderson, A., & Jones, S. (2023). Classics education in Northern Irish primary schools; curriculum policy and classroom practice. *Journal of Classics Teaching*, 24(47), 52–58.
- Thies, S. (2004). "Latein-Plus". Latein und Englisch in den ersten Jahren des Gymnasiums in Baden-Württemberg und Nordrhein-Westfalen. *Der Altsprachliche Unterricht*, 47(1), 54–59.
- Thorndike, E. (1924a). Mental discipline in high school studies. *Journal of Educational Psychology*, 25(1), 1–22.
- Thorndike, E. (1924b). Mental discipline in high school studies. *Journal of Educational Psychology*, 25(2), 83–98.
- Tijsseling, E. (2006). The Netherlands. In J. Bulwer (Ed.), *Classics teaching in Europe* (pp. 90–96). Duckworth.
- Tippelt, R. (Ed.). (2009). *Steuerung durch Indikatoren. Methodologische und theoretische Reflektionen zur deutschen und internationalen Bildungsberichterstattung*. Barbara Budrich.
- Townsley, L. (1985). Latin as a vocabulary builder for hearing-impaired and second-language students of English. *Teaching English to Deaf and Second-Language Students*, 3(1), 4–8.
- Turpin, F. (2020). The situation of classics in Europe. Evaluation of the Euroclassica questionnaire 2019/2020. *Cursor Latein4EU. Zeitschrift für Freunde der lateinischen Sprache und europäischen Kultur*, 16, 20–22.
- Universität Zürich. (2022). Latin and Greek. *University of Zürich, Faculty of Arts and Social Sciences*.
<https://www.phil.uzh.ch/en/studium/studentservices/lateingriechisch.html>
- Van Damme, D. (2021). Wat kunnen jongeren vandaag in hemelsnaam nog leren van Homeros of Tacitus? *Kleio. Tijdschrift voor oude talen en antieke cultuur*, 50(3–4), 179–182.
- Van Houdt, T. (2012). Iedereen Latijn. *Streven*, 79, 738–742.

- Verbaal, W. (2021). De oudheid en de bunker. In T. Velle & Y. Maes (Eds.), *Latijn anders? Waarom er nood is aan ander Latijn in de klas* (pp. 15–30). Skribis.
- Vereeck, A., Janse, M., De Herdt, K., Hauspie, C., Szmalec, A., & Woumans, E. (2025). Does studying Latin make pupils smarter? Presenting the field of classical language impact studies. *The Classical Journal* 120(4), 479–519.
- Vereniging Classici Nederland. (2021). Klassieke talen dichten kloven in Amsterdam Zuidoost. *Klassieken.nu*.
<https://klassieken.nu/2021/03/24/niemand-snapt-de-abl-abs-klassieke-talen-dichten-kloven-in-amsterdam-zuidoost/>
- Verhoeven, P. (1997). *Tekstbegrip in het onderwijs Klassieke Talen. Een onderzoek naar de verwezenlijking van nieuwe doelstellingen in het schoolonderzoek Grieks en Latijn*. Universiteit Leiden.
- Wachter, R. (2008). Latin and European language history. In B. Lister (Ed.), *Meeting the challenge. International perspectives on the teaching of Latin* (pp. 135–149). Cambridge University Press.
- Wesselmann, K., & Walker, M. (2021). Chance Mehrsprachigkeit: Latein als sozialer Katalysator. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 26(2), 251–270.
- Wier, M. (1921). Latin and Greek as aids to English composition. *The Classical Journal*, 16(6), 326–338.
- Woolard, K. (2020). Language Ideology. In J. Stanlaw (Ed.), *The International Encyclopedia of Linguistic Anthropology* (pp. 1–21). Wiley Blackwell.

Chapter 2: State of the Art

The contents of this chapter are nigh identical to the following article:

Vereeck, A., Janse, M., De Herdt, K., Hauspie, C., Szmalec, A., & Woumans, E. (2025). Does studying Latin make pupils smarter? Presenting the field of classical language impact studies. *Classical Journal*, 120(4), 479–519.

Minor edits were made with an eye to harmonising spelling, punctuation, title numbering, cross references, bibliographical style etc. throughout the dissertation. One remaining artefact of the paper's original bibliographical style is that in-text author-date references are not necessarily listed in alphabetical order, as should be the case in the rest of the dissertation (in accordance with the APA manual of style).

Chapter 2

State of the Art

2.1 Introduction

Classicists are used to being asked regularly about the whys and wherefores of their trade. You could even say it is engrained in what it means to be a classicist to fight for classics *tout court*, and in particular for the presence of classical languages (henceforth CLs) in the secondary school curriculum. Continuous and explicit support for Latin and Greek is necessary to secure their future as school subjects. One particular argument that is often put forward by way of justification, is that CL education has a transdisciplinary function: Among other things, it is thought to sharpen linguistic and other cognitive abilities. Just think of many popular statements in the vein of “if you study a CL, you will have a better command of your native tongue”, or “CLs are great to learn how to think analytically”, and so on.

Some accept it as common sense that CL learning stimulates various language and reasoning skills; others dismiss it as superstition. “Undoubtedly these proclaimed benefits of CL education have been subjected to scientific research?”, you might ask. In fact, a rather astonishing amount of research has indeed been carried out, but the results must be approached with caution, as will become clear. We present: the field of CL impact studies, whereby we understand “impact” as everything that transcends the boundaries of classics subject matter in the strictest sense (the Latin and Ancient Greek languages and literatures, as well as their cultural-historical context). While this article is by no means the first on this topic or even on this body of research literature, it is the first of its kind, in the sense that CL impact studies are named and framed as a field and examined from a history of science perspective.

Furthermore, our literature review distinguishes itself from other fairly recent literature reviews (Kennedy, 2006; Smith, 2007; Holliday, 2012; Bracke & Bradshaw, 2020; Holmes-Henderson & Kelly, 2022) in both selectiveness and comprehensiveness. We employ a clear-cut inclusion criterion: For a work to be part of CL impact studies, it must contain empirical research, not just anecdotic evidence. Such a distinction was not observed by Kennedy (2006), Smith (2007) or Holmes-Henderson and Kelly (2022). Even so, we found the corpus to be a lot larger than the fifty studies mentioned in the most complete review to date by Bracke and Bradshaw (2020).

In Section 2.2, we define and describe the field of CL impact studies in more detail, paying attention to its place within the history of classical scholarship. For structure, we turn to ancient rhetoric and make use of the famous hexametrical verse “*Quis, quid, ubi, quibus auxiliis, cur, quomodo, quando?*”.³² Along the way, we also clarify the choice of content and approach for Section 2.3, wherein we sketch a history of a very substantial part of CL impact studies, namely the American research on cognitive impact, by going over it in chronological fashion. In Section 2.4, an attempt is made at drawing some conclusions about the field as a whole and its present state of the art.

2.2 Characteristics of classical language impact studies

2.2.1 *Quis?* The classicist as layman educational measurer

Over time, a great many people have taken an interest in unravelling the impact of studying CLs. Not by chance, these people often were classicists who were looking to prove a point. Feeling the direct and pressing need to defend their discipline before parents, policymakers and the public at large, they had no choice but to take matters in their own hands and become a “layman educational measurer” themselves (Carr, 1919, p. 20). Thus originated a peculiar branch of classical scholarship, which has hitherto been given little attention. In her survey of research on Latin learning and instruction, Adema (2019) does refer to a couple of studies on

³² Literally translated: “Who, what, where, by which means, why, how, when?”.

potential crosslinguistic advantages of learning Latin. She subsumes these under “second language acquisition”, mixed together with research on competence development in Latin itself. CL impact studies as such, however, have been neglected as part of the history of classical scholarship. This article aims to change that, and therefore our primary approach is historical in nature.

It is worth noting that not all research has been conducted by CL instructors. Part of it took place at the hands of other educational practitioners and professional educational measurers (such as psychologists, pedagogues and social scientists), as we will see in the remainder of this article. The field of CL impact studies is inherently interdisciplinary: The topic and research questions spring from classics (see also Section 2.2.5), but the procedures and solutions must come from the behavioural sciences (see also Section 2.2.4). Moreover, the actuality of a CL effect has consequences for other subjects and for schools at large. That explains why researchers from various backgrounds have been involved. Regardless of who is collecting and analysing the data, the research rationale remains inextricably tied to the discipline of classics.

2.2.2 *Quid?* The myth of classics

Could there really be more than what meets the eye – more than the Greek declensions, the Latin texts, the Greco-Roman culture, et cetera? CL education provides an extraordinary mental training, so goes “the myth of classics”, as Cardinale (2011) calls it. The perpetual quest for the mythical classics-transcending impact of classics is what constitutes the field of CL impact studies, which seeks to answer the question whether CL education yields any additional benefits. In accordance with common contemporary psychological practice (e.g., Richardson et al., 2012), these benefits can be divided into cognitive traits, such as literacy, mathematical ability or critical thinking, and non-cognitive traits, such as personality and motivation. We will elaborate on the various cognitive and non-cognitive benefits in Section 2.2.4.

As the overview in Section 2.3 will illustrate, the vast majority of studies is actually about Latin alone. The impact of Ancient Greek has been blatantly disregarded in the research efforts. Two studies that did look at Greek specifically, are the ones by Chanock (2006) and Vanheule (2015). To a certain extent the choice – be it conscious or unconscious – to concentrate on Latin is justifiable, since it has always been taught more frequently than Greek. Despite this uneven distribution in both education and research, we decided to stick with the term “classical languages”, being aware that it usually represents just “Latin”.

2.2.3 *Ubi?* An American dream

Taking stock of the existing body of research, the search for the impact of CLs is predominantly an American pursuit, it seems. The most and the earliest research has been conducted in the United States. Multiple studies have been performed in Europe as well, however, for instance in Belgium (e.g., Pyck, 1947; Pire, 1971; Vanheule, 2015; Hauspie et al., 2024). German-speaking Europe has something of a research tradition of its own, featuring more social scientists (e.g., Castiello, 1934; Gutacker, 1979; Haag & Stern, 2003; Ortner et al., 2008; Goik, 2021). There have also been studies in Australia (e.g., Chanock, 2006). For reasons of feasibility, Section 2.3 will be limited to the United States.

2.2.4 *Quibus auxiliis?* The tools of systematic oversimplification

The methods of CL impact studies are not those of philology, but rather of what we would now label as psychology. It is important to realise in this respect that the methodological toolbox of experimental cognitive psychology was not yet a given from the start, but that the development of this latter discipline coincided in time with the appearance of CL impact studies. In addition to our primary historical approach, we therefore also take a cognitive-psychological stance, and try to give some indications as to the scientific quality of the works discussed. Section 2.3.6 in particular touches upon the methodological shortcomings that unfortunately characterise the field.

Roughly speaking, there are two empirical paradigms to grasp the world around us and – remembering the winged words of Karl Popper – to simplify its mechanisms artfully and systematically: quantitative research on the one hand, and qualitative research on the other. In the case of CL impact studies, the methods happen to be closely connected with the outcome variables and with the frequency of the research. On the basis of these three parameters, we can split up the field in four to five categories.

- (1) **Effects on language ability**, which are studied quantitatively and very frequently. Quantitative language assessment is probably about as old as language education itself (Farhady, 2018), so it did not pose any conceptual problem for the CL impact measurers. The bulk of all research has focused on the supposed linguistic benefits of CL study (i.c., Latin), mostly for the first language (i.c., English) but also for second languages.

- (2) **Effects on general cognitive ability**, which are studied quantitatively and less frequently compared to language ability. Nowadays we think of general cognitive ability as an obvious quantifiable, but in the previous century, when intelligence testing was still in its infancy, that used to be a very different matter. None of the American studies had general cognitive ability as an outcome variable; only two included mathematical ability (Smith et al., 1927; Sheridan, 1976), which could be seen as a proxy for non-verbal intelligence. Several studies did involve academic achievement and SAT performance.
- (3) **Effects on less prototypical cognitive traits** like cultural awareness, historical awareness or critical thinking, which are studied infrequently. In contrast to language and general cognitive ability, which are base constructs in modern psychology, the quantification of these traits is ongoing at present (e.g., Kumlien et al., 2020; Aisiah et al., 2016; Tiruneh et al., 2018).
 - (a) Nevertheless, one American study from the twenties ventured a quantitative measurement of reasoning ability in societal topics (Smith et al., 1927).
 - (b) Otherwise there is little or no CL impact research targeting these outcomes specifically. Usually these are studied qualitatively, bracketed together with the variables in the next category.
- (4) **Effects on non-cognitive traits** like ambition, discipline or self-confidence, which are studied qualitatively and infrequently. Without wanting to detract from the value of qualitative research, the lack of quantitative research on these traits is a notable deficiency of the field. There is no reason why there should not be any such research, in view of the availability of valid and reliable non-cognitive scales these days (e.g., Duckworth and Quinn, 2009; Tangney et al., 2004; Vansteenkiste et al., 2009). Moreover, aligning this strand of CL impact research with established psychological constructs could remedy the usage of somewhat woolly terms (criticised by e.g., Van Damme, 2021) and lead to theoretical refinement regarding the non-cognitive benefits of coming into contact with classics.

Section 2.3 will be wholly dedicated to categories 1 to 3a, in other words, to the quantitatively measured³³ cognitive benefits. For the qualitatively examined benefits, we hereby direct the reader to a few empirical studies, both American and European (Advisory Committee, 1924; Sussman, 1978; Bracke, 2016; 2020; Holmes-Henderson, 2021; Manolidou et al., 2023).

2.2.5 *Cur?* Research in defence of the discipline

The classical discipline distinguishes itself from other academic disciplines in the sense that the justification of its existence forms an integral part of it. As Kitchell (2008, p. 150) points out:

“We classicists are rather a unique kind, when all is said and done. Teachers of history, mathematics, chemistry or physics rarely have to justify their profession. Neither must they engage in promotional activities to fill their classrooms or to ensure the future of their discipline. Yet no matter the country, and no matter the level at which we teach, the classics profession is filled with people dedicated to just these sorts of endeavours, and such activities will continue to absorb our energies for the foreseeable future.”

And so they have, for many decades already, in the United States and Western Europe alike. After secondary-level CL instruction as we roughly know it had lived its heyday in the nineteenth century (Pfeiffer, 1976), the first serious doubts about this educational practice arose towards the turn of the next. Against the backdrop of revolutionary scientific discoveries and unbridled faith in progress, the opinion that CLs are dull and unworldly began to gain increasing acceptance (Bolgar, 1963; Adler, 2020). Later, the changing visions on cognition, pedagogy and society in the course of the twentieth century jeopardised the privileged educational position of CLs even further (Sparks et al., 1995; Haag & Stern, 2000; Vereeck et al., 2024). Their position has been precarious ever since, and their educational value a matter of ceaseless debate and discussion.

The public debate regarding CL education falls into two main lines of criticism (Vereeck et al., 2024). The first is concerned with the social equity and accessibility of classics, which we

³³ Note how using qualitative methods to gauge cognitive outcomes, for instance through self-reports by pupils and/or teachers (e.g., Smith, 2007; Taylor et al., 2023), inevitably shifts the research topic from CL impact to attitudes about CL education. According to our definition, *perceived* evolutions of language or general cognitive ability – be it recorded through qualitative or quantitative interviews – fall outside the realm of CL impact studies.

leave aside for now. The second line of criticism is concerned with the practical use of classics: In a society where Latin or Ancient Greek are no longer actively spoken, from a strictly functional point of view they are useless and irrelevant. Sure enough, for each attack on CL education, a eulogy in favour has been written. Classicists and other proponents of CL education have wielded every thinkable counterargument to affirm the validity, meaningfulness and even utility of their beloved subjects, claiming all sorts of advantages that allegedly result from studying CLs.³⁴

In the modern world, such a dispute cannot be settled by verisimilar words skilfully arranged according to the art of rhetoric, but only by persuasive numbers competently calculated according to the rules of statistics. This is where CL impact studies come in, to strengthen the counterarguments against utilitarian criticism. Again, this kind of research in defence of a discipline is highly typical of classics, and unseen for other school subjects. While it can easily be agreed upon that CL education is a cultural enrichment of some sorts and that it builds good study habits (e.g., Smith, 2007; Taylor et al., 2023), the reality of the linguistic and general-cognitive benefits is much more contested. Hence, the field of CL impact studies has chiefly been preoccupied with proving transfer from the study of CLs to these areas.

2.2.6 *Quomodo?* Divided rather than united

One could contend that the amalgam of CL impact studies does not really make up a field. That is to say, it is by no means a unified or incremental field, but rather a highly shattered one with little to no self-awareness. Researchers display only limited knowledge of the work already carried out by their predecessors. The academic literature seems to exist in separate bubbles, so to speak: In two different articles, even from the same era, references to previous research can vary completely. Consequently, the heuristics, especially of the initial period, fan out further with almost every new discovery. Presenting a comprehensive inventory is therefore beyond the scope of the current article. Moreover, some of the oldest work can be hard to track down, and it is beyond doubt that several treasures lie hidden in unpublished student papers. Instead of exhaustivity, representativeness is what Section 2.3 aims for.

³⁴ In turn, learners often mention the same advantages as motivation for studying Latin (see Katz et al., 2020; Smith, 2007).

2.2.7 *Quando?* Twelve decades of research

The history of the field of CL impact studies begins at the daybreak of the twentieth century. One of the earliest academic references to CLs in relation to cognition is found in the work of Charles Spearman (1863–1945), an Englishman and trailblazer in experimental psychology. In his seminal paper on intelligence (1904), Spearman theorised the existence of a “common intellective function” underlying all forms of mental activity, in other words, of a general intelligence determining a person’s performance on just about anything from a maths test to a pitch discrimination test. Per Spearman’s theory, different mental activities “load” to a different degree onto the intellective function, and the one he found to have the highest correlation with general intelligence was none other than classics. To paraphrase, at a time when there were no IQ tests yet, the simplest way to estimate someone’s general cognitive ability was apparently to ask for their CL marks. After Spearman’s 1904 paper, another ten years went by until the first proper research into the effects of studying Latin. CL impact research has continued ever since, spanning twelve decades so far.

2.3 A history of American research on the cognitive impact of studying classical languages

2.3.1 The beginnings of the research field

The oldest known instances of people checking, calculating and reporting the impact of studying Latin on language ability originate from the United States in the mid-nineteen-tens (e.g., Perkins, 1914; Harris, 1915). It was no coincidence that these first tentative tests and experiments of Latin effects cropped up at a time of great change in both education and industry: as high schools emerged across the country and vocational subjects were introduced *en masse* in their curricula, practical use became of increasing concern and Latin was forced onto the defensive (Barber, 1985). Moreover, empirical scientific measurement was becoming the epistemological norm. Carr (1919, p. 20) vividly sketched the scene as follows:

“[The value of Latin as an aid in improving a pupil’s English] is the one value which the Latinist never fails to claim for his subject, whether he puts it first or last in order of importance. If he can only *prove* the validity of this claim in the language of, and to the complete satisfaction of, the controllers of educational policies in our schools (...). ‘But,’ the Latinist insists, ‘this claim, above all others, has been proven valid again and again.’ Well, not so the professional measurer could notice it! It is true that the Latinist can produce pages and pages of what would seem to be unimpeachable testimony from reliable sources, and he has again and again summoned statistics to back up the testimony. Still his opponents continue to say in print and from the platform that no convincing evidence has been offered (...).”

The decade of the nineteen-tens harbours a treasure-chest of small research studies, which apparently have known little dissemination, or in any case have not been systematically reviewed by contemporary or later authors. When Latin and non-Latin pupils were compared, the predominant result was that the former did better on various measures of English vocabulary, spelling, grammar and writing (e.g., Perkins, 1914; Barber, 1985 [1914]; Harris, 1915; Foster, 1917; Dallam, 1917; Wilcox, 1917; Carr, 1919, 1921)³⁵. Some thought was also given to possible factors that interacted with or confounded the observed effect, such as number of years of CL instruction, employed CL teaching method, pupil characteristics, and instrument bias.

Regarding the first factor, years of instruction, different subpopulations of CL pupils were investigated. After one year of Latin, pupils already exhibited a growth in passive English vocabulary multiple times larger than that of their peers in the Latin-less track (Carr, 1919, 1921). The more the better, though: Among high school graduates, only the scores of those with three to four years of Latin experience really stood out (Harris, 1915; Foster, 1917). Regarding teaching method, the amount of attention paid to English derivation turned out to be directly proportional to test performance by the Latin pupils (Carr, 1921). Perkins (1914), who organised a Latin class in his school wholly centred around English derivation, was able to bring the lexical prowess of his pupils to unseen heights after just two years: Compared to

³⁵ Some clarification of these references is called for. Barber 1985 is a posthumously published abbreviation of a 1914 contribution by professor Grove Barber to his university newspaper, in which he gave a summary of Perkins 1914. Carr 1921 was published in the twenties, to be granted, but reports on tests during the school year 1919–1920, well before the start of the Classical Investigation (cf. Section 2.3.2).

pupils of equal mean ability³⁶ who had chosen a modern foreign language class, the Latin pupils fared on average 29 percent better on English tests.

According to some, however, pupil characteristics like initial mental ability and gender had more explanatory power than having a curriculum with or without CL (Foster, 1917; Wilcox, 1917). On that note, it is worth adding that many studies are unclear about how they control for potentially confounding pupil characteristics, if at all (more on this in Section 2.3.6). Finally, apropos the last factor, most vocabulary and spelling tests used in this early research contained almost exclusively words of Latin origin. Then again, for commonly known historical reasons, this is almost unavoidable when the higher registers of the English lexicon are entered. Carr (1919) recognised this issue of instrument bias, and in later work (1921) deliberately adjusted his language test to incorporate both words of Latin and of Anglo-Saxon origin. He concluded that consistently large vocabulary gains by Latin pupils only occurred on the Latin-derived words (e.g., ‘filial’, ‘vulnerable’, ‘veracious’).

2.3.2 The Classical Investigation

The research endeavour that had been commenced here and there in the early twentieth century soon culminated into one large, systematical enterprise: the Classical Investigation³⁷, which was launched by the American Classical League and ran between 1921 and 1924 (Wraga, 2009). The sample sizes of the previous pioneering work were small, and its results were felt to be inconclusive. In the face of widened access to secondary schooling and the concomitant educational debates, the need for a more thorough inquiry on a larger scale imposed itself. The Classical Investigation aspired to be just that: a grand scheme “to determine objectively the value of Latin as it was being taught in the average high school, and, through experimental

³⁶ There were six measurements in total. For tests 1 to 5, 21 Latin pupils and 21 modern foreign language pupils were selected, “such that each group had virtually the same average mark in Latin, on the one hand, and modern language, on the other, and also in English” (Perkins 1914, 11). For test no. 6, 17 Latin pupils and 17 modern foreign language pupils were selected, considering “not only foreign language and English (...) but also all studies the pupils had taken during the year” (Perkins 1914, 12).

³⁷ As the largest study on CL education to date, the Classical Investigation deserves due consideration in this research history. Unless other sources are referred to, Section 2.3.2 is summarizing the facts laid out by Wraga (2009), who devoted a full-length article to the Classical Investigation alone.

procedures, to make recommendations as to how Latin could be taught more effectively to achieve the claims made for it” (Douglass & Kittelson, 1935, p. 27).

Although the absolute number of Latin pupils multiplied around this time because of the surge of youngsters attending high school, the relative proportion of enrolment in Latin declined significantly because of the prevailing idea that classics had little to offer in modern life (Wraga, 2009, pp. 82–84; Holliday, 2012, p. 3). An especially harsh and high-profile attack came in the form of the reformatory proposal *A Modern School* by progressive educator Abraham Flexner (1866–1959), which would indirectly lead to the Classical Investigation. Flexner (1916) proposed that Latin and Greek be removed from the school curriculum entirely. His paper provoked a flood of reactions from classicists, brimming with indignation and contempt (Wraga, 2009, pp. 84–85).

One man in particular stepped to the fore as the preeminent advocate of CLs, namely Princeton professor Andrew West (1853–1943). Amongst other initiatives, he was the driving force behind the foundation of the American Classical League in 1919, the intentions of which were clear enough from the very outset: to roll out a publicity campaign on behalf of CLs and to combat all educational theories hostile to CL teaching and learning (Phinney, 1994). In March 1921, the American Classical League was awarded a generous grant by the General Education Board, a Rockefeller philanthropy that existed between 1903 and 1964 (Rockefeller Archive Center, 2022), to conduct what later became known as the Classical Investigation (Wraga, 2009, p. 85).

The investigation consisted of two parts (Wraga, 2009, p. 86). In the first, quantitatively-oriented part, data on the status and effects of CL education³⁸ were garnered through the administration of tests and surveys during the school years 1921–1922 and 1922–1923. After all, Flexner (1916) had claimed that no evidence existed for the acclaimed beneficial impact of

³⁸ The scope was much wider than the assessment of cognitive, specifically linguistic impact that is of interest here. It entailed studies on secondary-school and college enrolment in CLs, on the number of years college matriculants had taken Latin and Greek, on the supply of CL and teacher training courses, on teacher requirements and other matters of educational policy, on the extent to which language graduates made use of their studied languages and the extent to which they showed interest in the literature of these languages, on the etymology of all words included in the Oxford Dictionary, on the most frequently occurring words of the English language, on the frequency and range of Latin words and phrases as well as of classical allusions found in contemporary English, on the grammatical principles common to Latin and English, and many, many more (West et al., 1923).

CLs on English, and admonished that schooling include “nothing for which an affirmative case cannot now be made out”. In the second, qualitatively-oriented part, the most important objectives, content, teaching methods and teacher qualifications were established by means of surveys, score-cards and symposia. The scale of this operation required an extensive team in charge (Wraga, 2009, pp. 85–86):

“A nationwide study, the investigation was organised in three tiers: an Advisory Committee of 15 members, all teachers of the classics on the secondary or post-secondary levels and chaired by Andrew West; a Special Investigating Committee (...), which managed the logistics of the entire undertaking; and eight Regional Committees whose total of 55 members would coordinate regional activities, mainly testing.”

On the grounds, about seven thousand school teachers volunteered to cooperate (West et al., 1923, p. 549). The American Classical League also entered partnerships with the national Bureau of Education and College Entrance Examination Board (West et al., 1923), state departments of education (e.g., Coxe, 1924), as well as universities³⁹. Multiple graduate and doctoral dissertations were completed under the co-auspices of the Special Investigating Committee (e.g., Heald, 1923; Haskell, 1923; Bates, 1924; Hill, 1924; Lawler, 1925; Hamblen, 1925). Additionally, the organisational team sought the cooperation and advice of numerous university professors in the psychological and pedagogical sciences. There are reasons to suspect, however, that in practice this interdisciplinary collaboration was altogether limited.⁴⁰ For that matter, only one of these external professors, renowned psychologist Edward Thorndike (1874–1949), appears to have been remunerated by the League for his services (Wraga, 2009, pp. 86–87).

Thorndike⁴¹ actually supervised several research studies under the flag of the Classical Investigation. More particularly, he looked into effects on reading comprehension and

³⁹ Notably the University of Iowa, to which the following of the references in this section can be traced back: Kirby, 1923; Heald, 1923; Bates, 1924; Hill, 1924; Lawler, 1925.

⁴⁰ West and his colleagues (1923, p. 557) give us an inkling of what transpired: “They were almost unanimously affirmative with regard to the possibility of transfer, and with regard to the probability that experiments such as those suggested in the letter would yield fruitful results. They were, however, almost unanimously negative on the crucial question of co-operation in such experiments.”

⁴¹ Thorndike’s involvement is hardly a surprise: outside of the Classical Investigation, he repeatedly researched mental training in general (Thorndike & Woodworth 1901a, 1901b, 1901c) as well as the

vocabulary range (Thorndike, 1923a, 1923b, 1923c; Thorndike & Ruger, 1923). Upon analysis of the reading scores of several thousands of high school pupils, he found that after one year of differentiated instruction, that is to say a curriculum with or without Latin, “the gain made by the Latins is about one and one half times as great as the gain made by the non-Latins” (Thorndike, 1923b, p. 167). Strangely enough, this gain was already observed at the interim test halfway through the school year, and did not increase between the middle and the end of the year. As for vocabulary, Latin pupils showed superior half-year gains and year gains (Thorndike, 1923a). When year gains were computed separately according to the origin of the words, Latin pupils had a distinct edge on knowledge of Latin-derived words, and a slight edge on knowledge of words of Anglo-Saxon and Greek origin (Thorndike & Ruger, 1923). In all studies some allowance was made for differences in initial scores, but other a priori pupil characteristics were not taken into account.

Thorndike’s research associate, Edith Newcomb, wrote two succinct commentaries (1923a, 1923b) on the Classical Investigation, adding a couple of *caveats* to the then present and future findings. The first commented on the “mental make-up of the Latin pupils”: The reading ability, knowledge of fundamental grammar, and general intelligence of American pupils starting Latin were found to vary immensely.⁴² Newcomb warned this complicating factor could conceal the true effect of Latin study. The second commented directly on the results of Thorndike’s recent research. Newcomb rejoiced in the vocabulary study with a hearty “I told you so!” and attempted to explain the surprising development of reading gains throughout the first year of Latin as “a temporary condition of intellectual excitement, which would fade away after the first few months”, caused by “the increased interest in language in general, due to the introduction to a tongue differing so markedly from English”.

Of course the Thorndike studies are but a small, albeit perhaps better-known portion of what has come down to us. Educational journals around this time were flooded with papers pertaining to the Classical Investigation (e.g., Gilliland, 1922; Otis, 1922; Kirby, 1923; MacPhail, 1924; Cole, 1924; Coxe, 1924). *School and Society*, a weekly periodical with usually short

influence of various high school subjects on intelligence (Thorndike 1924a, 1924b; Broyler et al., 1927). His findings rendered him skeptical about the possibility of any particular type of training or schooling having such broad effects on the mind.

⁴² This need not astonish, since the wider availability of secondary schooling had led to more heterogeneous class groups (Wraga, 2009, p. 84; Angus & Mirel, 1999).

contributions edited by James Cattell (1860–1944), yet another famous name in psychology (Webster, 1985, p. 107), was a recurrent choice for publication. A great deal more studies was never put out in print,⁴³ but only obliquely referenced in publications by the investigative committees (e.g., West et al., 1923; Advisory Committee, 1924), and we should not forget the abovementioned dissertations which also remained unpublished.

In the cognitive research on studying classical languages of the early twenties, the previous issues of English etymology, instruction time, and teaching method were revisited. Otis (1922) secured a Latin and a non-Latin group of exactly equal intelligence scores on the Terman Mental Ability Test, counting 42 pupils each. The participants are said to have frequented high school for on average two and a half years, and to have received the same amount of Latin and modern foreign language instruction respectively. Otis administered tests in Latin-derived English vocabulary as well as in short composition. The Latin pupils excelled in vocabulary by 33 percent and in composition by 6.5 percent.

Gilliland (1922) tested knowledge of English words with Latin, Anglo-Saxon and Greek roots, this time in 115 randomly selected college freshmen. The ability to define words of any origin was dependent upon the number of years spent studying Latin, a pronounced advantage emerging from three years onwards. According to various researchers (Haskell, 1923; Bates, 1924; Coxe, 1924; Hamblen, 1925), however, the amount of automatic transferal from Latin to English was fairly limited. The effects were reinforced if part of the Latin class had been deliberately dedicated to spelling, derivation, and analogy,⁴⁴ and even greater effects were achieved when rules and principles had been developed in this regard.

New lines of research were the effects of Latin on the ability to master the Romance languages and on academic achievement in higher education. Kirby (1923) calculated the correlations between the number of years of Latin in high school, the intelligence scores, and the first-year French grades of 268 college freshmen. When the intelligence variable was rendered constant, he found a partial correlation between Latin years and French grades of 0.22. Cole (1924) obtained similar results for both French and Spanish, nonetheless signalling this might be explained by Latin pupils' "superior capacity for work" rather than by Latin-

⁴³ Some undisclosed data resurfaced in later publications by former Committee members (e.g., Carr, 1942).

⁴⁴ This finding would later be repeated by Berelson and Steiner (1964, p. 162) in their impressive work inventorying the state of the art in the behavioural sciences, as part of a section on transfer of training.

Romance transfer. Heald (1923) and Hill (1924) came to the conclusion that Latin study was a vital factor for success in French even apart from intelligence.

After the massive data collection and analysis had been concluded, the Advisory Committee prepared a report in six volumes, of which only volume one and three survive in print. The others were never made public; they were last declared to be seen together in 1929 (Lucasse, 1929b, p. 15) by a Harvard graduate who had just written a thesis about the Classical Investigation (Lucasse, 1929a), but any modern-day efforts to track them down have failed (Wraga, 2009, p. 88). At the time, the results of the Classical Investigation gave cause to some disappointment: The advantage of Latin pupils over non-Latin pupils was not nearly as large as had been generally supposed (Wraga, 2009, pp. 86–90; Advisory Committee, 1924). It is imperative to keep in mind the scientific-historical context here: Whereas nowadays we know that effect sizes in educational research are nearly always small (Evans & Yuan, 2022), back then expectations were wound up to a higher pitch.

In the *General Report*, the first volume of the Classical Investigation report, an overview is given of the objectives on which CL instruction – or rather, Latin instruction, given the negligible enrolments in Greek – could lay claim. The “increased ability to read and understand Latin” was reaffirmed as “the indispensable primary immediate objective” (Advisory Committee, 1924, p. 32). In an era when the everyday professional use of Latin literacy had been replaced by its utility in improving English skills as the number-one rationale for taking on this ancient language (Wraga, 2009, p. 79), it was a remarkable move on the part of the American Classical League, and one likely reflective of the dissatisfying results, to make Latin for the sake of Latin such a prominent argument in the defence.

The additional benefits of studying Latin were not discarded, however. On the contrary, the Committee (1924, p. 78) drew a double card and continued the overview of objectives with the following both cognitive and non-cognitive benefits: “increased understanding of those elements in English which are related to Latin; increased ability to read, speak and write English; increased ability to learn other foreign languages; development of correct mental habits; development of a historical and cultural background; development of the right attitudes toward social situations; development of literary appreciation; elementary knowledge of the simpler general principles of language structure”.

2.3.3 Research between the Classical Investigation and the Second World War

However impressive the Classical Investigation may be, it was not the be-all and end-all apology of CL education that West and his many supporters had undoubtedly aimed for. Although it did spark substantial didactic innovations (Lucasse, 1929b; Phinney, 1994), the issue of the cognitive and in particular crosslinguistic impact of the study of CLs was far from resolved. The *General Report* had a mixed reception: In the main, it was praised by classicists and condemned by educationalists (Lucasse, 1929b). Consequently, in the years and decades after the Classical Investigation, various new research studies were set up, mostly by “educational people” (if we may borrow this rather vague expression from Lucasse, 1929b, p. 16 as an umbrella term).

Arthur Mead, professor of Education at Ohio Wesleyan University, and his student David Smith tried a completely new tack (Mead, 1923, p. 113; Smith et al., 1927).⁴⁵ They raised the following question: “Does Latin translation, on the part of the high-school student, lead to a transfer of training in reasoning ability to other fields or subjects?” (Smith et al., 1927, p. 381). More specifically, reasoning problems from the fields of arithmetic, history, sociology and economics were administered at two points during the school year to a paired⁴⁶ sample of Latin pupils and non-Latin pupils with similar curricula. Latin pupils who showed little to no progress in Latin translation, as measured by an additional test, were discarded from the analysis together with their respective non-Latin counterparts. In just more than half of all 214 so created pairs, the gain in average grade on the reasoning test of the Latin pupil was greater than that of their counterpart.

⁴⁵ A description of the study was only published in 1927, hence its inclusion in Section 2.3.3. In the article from 1927, it is revealed that Smith carried out this research for his M.A. thesis in 1922 already. In the opening chapter of his thesis, of which a staff member of the Ohio Wesleyan University Library was kind enough to provide us with a scan, Smith made mention of an ongoing investigation by the American Classical League. As far as we can tell, however, Smith’s study under the supervision of professor Mead was not associated with the Classical Investigation.

⁴⁶ As many “true pairs” as possible were formed: pupils from the same class in the same school with the same intelligence quotient. 161 such pairs were found of a proficient latinist and a non-latinist to match. To raise participant numbers, “approximate pairs” were allowed as well: pupils from the same class in the same school with an IQ difference of no more than 4 points. 53 extra pairs were secured in this manner.

Another professor of Education, Harl Douglass, together with high school teacher Clifford Kittelson reopened the question whether the study of CLs is beneficial for native language ability in English (Douglass & Kittelson, 1935). To verify whether Latin pupils had greater command of English than non-Latin pupils of equal mental ability, they handed out tests of vocabulary, spelling and grammar to matched pairs⁴⁷ of high school seniors who differed in the number of years of Latin under their belts. Pupils with three to four years of Latin performed better than those with two years, who in turn performed better than those without Latin. Douglass and Kittelson were critical of the results, pointing out that differences were “small” (i.e., in fact, smaller than 10 percent on the raw scores). They hinted that the teaching method could be further improved “as to contribute most effectively to developing ability in English” (Douglass & Kittelson, 1935, p. 33).

Frederick Pond (1938), a high school principal, approached the issue of English ability from a different angle: He sought to determine the relative contribution to English word knowledge of general intelligence, reading experience, number of semesters of Latin, and achievement in Latin. Regression analysis elucidated that pupils’ general intelligence, as reflected by the Otis Test of Mental Ability, and how many semesters they took Latin made the greatest contribution, the other two factors being of meagre weight. However, in the original sample of 208 pupils the average intelligence quotient of the Latin group was slightly higher than that of the non-Latin group, and after assembling 30 matched pairs⁴⁸ the difference in vocabulary score between the two groups dropped from almost six to just two percent. This led Pond to conclude, quoting Wagener (1937, p. 459), that the non-cognitive effects are the primary reason for and justification of studying CLs: “The study of language, literature, and the artistic side of classical culture must be used to foster the aesthetic and intellectual development of the individual as an individual and to create a pleasure in living that is highly personal.”

⁴⁷ The pupils were paired on the basis of gender, age, intelligence, economic status, number of years spent on modern foreign language study, and first year English mark. It is doubtful how we are to understand this “pairing”: the article merely reports the average values, which are roughly the same for all groups. There were 112 duos of pupils with no Latin and with two years of Latin, having a mean intelligence quotient around 107; there were also 29 trios of pupils with no Latin, two years of Latin and three/four years of Latin, having a mean intelligence quotient of around 116.

⁴⁸ Besides the intelligence quotient, the bases for matching were sex, age, school achievement and semesters of school attendance.

Furthermore, there were several studies on academic achievement and modern foreign language ability in the late twenties and thirties. Positive relations were found between CL study in high school and college marks, even when allowance was made for differences in general intelligence (e.g., Sorenson, 1929; Yates, 1929; Douglass, 1931; Clark, 1932). A large-scale study of the records of 1025 college freshmen from the academic year 1933–1934 confirmed the earlier results (Smith & Douglass, 1937). French composition in particular seemed to profit from training in Latin (Whelden, 1933). On the secondary school level as well, it was discovered that a background in Latin was a benefit when entering a modern foreign language course (Richardson, 1933).

2.3.4 Midcentury research

From the forties through the sixties, there appears to have been less research on the cognitive impact of the study of CLs; no more than a handful of studies are known to exist. A conspicuous one is the study by Margarete Altenhein (1950) from Hunter College: For the first and only time in American CL impact research, the topic of interest was achievement in German, which shares with the CLs its use of cases. Altenhein examined the grades in a German course of a cohort of 183 students, and noticed these went up progressively with the number of years of Latin study prior to the German course. When making pairs with similar intelligence scores, as measured by a scholastic aptitude test mandatory for all incoming freshmen, the differences in German grades only stood out more. Altenhein (1950, p. 328) remarked that “[t]his explodes the idea that greater success in German grades on the part of the Latin students is brought about by initial intellectual superiority.” The difference between the paired groups of students with no Latin background and students with three years of high school Latin and a semester of college Latin even reached statistical significance, although this is hardly meaningful, given there were scarcely five such pairs.

Inquiries into French, Spanish and English also went on in much the same vein as before (e.g., Morton-Finney, 1941; Parounagian, 1942; Sister Josephina, 1963), with positive results regarding the influence of Latin study on youngsters’ foreign language achievement and English word knowledge. The middle of the twentieth century was marked by a development in CL

impact research, however, and that is the appearance of extensive literature studies,⁴⁹ not entirely unlike the current article. A remarkable master's thesis was written at Indiana State Teachers College by Samuel Nisbet Junior (1951). Nisbet started from the premise that the study of Latin has practical and cultural value, and then substantiated it through library and documentary research. His goal was not so much to write a chronological history of a field, as it is here, but to gather and thematically arrange separate studies that could corroborate existing arguments in favour of studying Latin: a prime example of research in defence of the discipline, in other words.⁵⁰

2.3.5 Research in the seventies and eighties

The turbulent late sixties ushered in a new era for CL education in the United States. After enrolment numbers had hit a national all-time low, they slowly began to climb again (Sparks et al. 1995, p. 165). Projects and programmes of varying scale cropped up left and right, devoted to “the resurrection of Latin” (e.g., LeBovit, 1967; Cederstrom, 1974; Offenbergh et al., 1982; Polsky, 1986). The steady national decline in SAT Verbal scores likely had something to do with this evolution (Sussman, 1978, p. 346; Holliday, 2012, p. 4). Thanks to increasing awareness of the importance of foreign language learning (National Council of State Supervisors of Foreign Languages, 2003, p. 147) and the rising popularity of FLES, which stands for Foreign Language in Elementary School, Latin now found its way to primary education (Sparks et al., 1995, pp. 165–167).

The classics, traditionally a middle- and upper-class affair, were deliberately being brought to a broader public, particularly to underprivileged pupils attending inner-city schools (Bracke & Bradshaw, 2020, pp. 5–6). The Latin programmes of the time were not only characterised by new audiences, but by a new pedagogy as well. In brief, “[t]he Latin curricula employed in the test situations were not a revival of the traditional methods with their heavy emphasis on

⁴⁹ Outside of the United States, Nisbet was preceded in this endeavour by Castiello (1934) and Stellwag (1949), who combined an in-depth literature survey with experiments of their own.

⁵⁰ Nisbet ventured into comparative and historical linguistics as well: in the vein of the Classical Investigation (cf. footnote 38), clearly a source of inspiration for this thesis, he dedicated chapters to the relationship between Latin and English, principles of word formation in both Latin and English, and derivation and borrowing from Latin in English as well as Romance languages. Related topics like spelling, technical terms, phrases and other influences of Latin on the English language received ample treatment too.

grammar, but, rather, ones based on modern linguistic theory and typified by the oral-aural approach with special attention paid to word derivations and cultural background” (Sussman, 1978, p. 346). Exemplary for this pedagogic movement are Rudolph Masciantonio’s writings (1968a, 1968b, 1968c). Moreover, several courses were developed specifically tailored to students with special needs, both those with learning difficulties (see below) and the most gifted ones (e.g., Reinhardt, 1983; Bassman & Ironsmith, 1984; VanTassel-Baska, 1987).

As a result of the regenerated interest in CL learning, research on its cognitive impact took flight again in the second half of the seventies and in the eighties. At this point in the history of the field of CL impact studies, a yet unseen type of publication also appeared, which was neither literature study nor record of original research. Instead, these articles discussed multiple novel research studies on a local level – the findings of which were not always disseminated via academic journals – in some detail. Such overviews were compiled by a trio of classicists, the already mentioned Rudolph Masciantonio (1977), Nancy Mavrogenes (1977) and Lewis Sussman (1978). The focus of the research continued to lie on English language ability in its various aspects, since most, if not all Latin projects called into being during this era were intentionally devised, implemented and evaluated on the basis of measurable outcomes for English skills.

As far as the preserved reports and articles tell us, the desired outcomes were generally achieved too. In numerous cases it was found that regular Latin instruction decidedly improved pupils’ word knowledge and reading ability, as demonstrated by higher gains on standardised vocabulary and reading comprehension tests than those of fellow pupils who did not receive Latin instruction (see also e.g., Scott & LeBovit, 1971; Bowker, 1975; Sheridan, 1976; Baca et al., 1979; VanTassel-Baska, 1982; Offenberget al., 1982; Bassman & Ironsmith, 1984; Fromchuck, 1984; Polsky, 1986). Some studies tested for spelling, and time and again Latin pupils proved superior to their Latin-less peers (see also e.g., Carlisle & Liberman, 1989). There were also indications that Latin instruction aided grammar skills (see also e.g., Sheridan, 1976; VanTassel-Baska, 1982). However, the sample sizes of these studies were often small.

Worth noting for its spectacular result is the evaluative study of FLES education in the public schools of the District of Columbia (Scott & LeBovit, 1971; discussed by Masciantonio, 1977, p. 378; Mavrogenes, 1977, p. 270; and Sussman, 1978, pp. 347–348). The progress in reading achievement in the course of one school year was measured in hundreds of sixth grade pupils, consisting of a no-foreign-language group, a group in their fourth year of French or Spanish, and a group in their first year of Latin. Though the no-foreign-language group ranked

highest at the pre-test, the Latin group came out on top at the post-test, thereby making the most progress of all and surpassing the French/Spanish group who had had not one, but four years of foreign language instruction. No precautions were taken to control for other pupil characteristics.

In underperforming populations, the effects of Latin were perhaps even more striking. An exceptionally successful experiment took place in a secondary school in Worcester, Massachusetts (discussed by Masciantonio, 1977, pp. 380–381; Sussman, 1978, p. 350). Two groups were selected at random from incoming pupils who scored multiple grades below the average reading level for their age. The pilot group participated in a Latin programme, while the control group continued the regular school curriculum without Latin. The first group exhibited incredible improvement in English, outstripping the second group in tests at the end of the school year. Their vocabulary was increased by fourteen months and their reading comprehension by nineteen months, whilst pupils this far behind were expected to advance only about six months. Furthermore, Latin classes have been fruitfully deployed to enlarge the English vocabulary of hearing-impaired pupils whose first language was American Sign Language (Townsley, 1985).

Aside from the linguistic benefits, for which evidence accumulated, one study included general-cognitive benefits in its design (Sheridan, 1976; discussed by Masciantonio, 1977, pp. 377–378; Mavrogenes, 1977, p. 270; Sussman, 1978, pp. 349–350). In order to evaluate the FLES Latin programme in Indianapolis, for three consecutive school years sixth graders were subjected to the intermediate Metropolitan Achievement Test twice a year. The experimental groups counted roughly 300 participants, and the control groups were composed of roughly 100 participants with similar socio-economic and academic profiles. The pretests revealed they were “quite reasonably well matched”⁵¹; nevertheless the existing differences may have distorted the results. Overall, the Latin pupils made the greatest gains and outperformed the group without Latin, not only in the field of English, but in that of mathematics as well. Sussman (1978, p. 350) concluded with respect to the Indianapolis study that “[it] is therefore

⁵¹ For the first year (1973–74), results are not available from either the pre-test or the post-test. At the pre-test of the second year (1974–75), there was a difference in spelling and math concepts significant at the 5%-level in favour of the experimental group. Minor, statistically insignificant differences on other subtests existed in both directions. At the pre-test of the third year (1975–76), the only difference significant at the 5%-level was in spelling, this time in favour of the control group.

important since it demonstrates transfer not only in the specific areas of language skills, but also in logical, precise thinking, as exemplified by the mathematics achievement scores, a heretofore unexpected result.”

Furthermore, in the eighties the relationship of high school foreign language study with SAT scores and with college success was scrutinised. LaFleur (1981, 1982) remarked that on the national level, Latin pupils scored miles above the average on the SAT Verbal as well as Math, and higher than all other language students with the exception of pupils taking Russian⁵², who had the upper hand on the SAT Math. Eddy (1981) conducted a regional study, in which he proved the favourable effect of foreign language study on the verbal section of the SAT, but concluded that it did not matter which language was opted for. This research was later criticised for its restricted size and distribution of sample population (Davis-Wiley, 1985, p. 36). A replication study was carried out by Cooper (1987), who did find an effect of language; not Latin, but German was linked to the highest SAT Verbal scores⁵³ in his sample of 1778 pupils. Regarding college cumulative GPA (grade point average) and freshman English GPA, however, in student groups nearly homogenous as for high school achievement, Davis-Wiley (1985) recorded the strongest correlation with high school Latin.

2.3.6 Methodological and other shortcomings

Before delving into CL impact research from the nineties onwards, we first reflect on the methodological and other shortcomings of research from prior decades. All of the studies reviewed so far make for an impressive pile of research. In general, the findings tend to be positive, but how much value can be attached to them? This is where history, psychology, and history of science meet. In the previous sections, we have mainly been historiographers of American CL impact studies, embedding research efforts in their historical background, conveying the findings at that time, and as far as possible taking heed of design and methods.

⁵² Like Latin, Russian is of course a highly inflected and linguistically complex language.

⁵³ Much to Cooper’s own surprise, as we read in a note to the article: “The finding that German students scored the highest of all language groups is surprising. Classics teachers have long been very persuasive about the linguistic benefits of Latin study. (...) Results of these studies seem to substantiate this claim. Thus, one would expect Latin students to perform best on the SAT–Verbal (...). Such is not the case here, though. Perhaps the superior showing of students of German is due to the fact that this language attracts more serious students than other languages do.”

In this section, we let the cognitive-psychological approach prevail, without losing sight of the scientific-historical context.

The studies discussed happened in different times, with indeed very different standards, paradigms, tools and devices. The available statistical apparatus used to be way less elaborate than it is now. For example, the Classical Investigation was concluded years before significance testing became common practice (Moore & McCabe, 2005, p. 303). Also, it is easy to forget in an age of convenient software packages, but empirical research used to involve a lot more calculating by hand. That being said, from a modern perspective, the majority of studies fall “short in academic rigor”, as Kennedy (2006, p. 3) puts it. Prudence is thus called for, and some critical comments are in order. More specifically, the corpus suffers from five interrelated problems, having to do with methodology, theory, reporting style, contextuality and scope.

- (1) Methodologically, the principal shortcoming is the inadequate handling of potentially confounding variables such as pupil characteristics; this applies as much to CL impact research from the nineteen-eighties as from the nineteen-tens. In the worst case, pupil characteristics are not taken into account at all. Oftentimes it is promptly assumed that confounding influences are nil because the participant groups appear to be similar enough – an assumption which mostly seems untenable. Many studies attempt matching groups, but the matching procedures used are sub-optimal, superseded, or even suspicious from a modern point of view (see e.g., Ho et al., 2007 on matching). Statistical modelling is conspicuously absent, which would be unthinkable today. Pond (1938) and Offenbergs team (1982) are some of the exceptions who actually build a regression model, including a limited selection of control variables.

Sample sizes vary widely. Several studies have an acceptable sample of some hundreds of participants, but many do not have enough subjects for the results to be reliable (see e.g., Brysbaert & Stevens, 2018 on sample size and psychology’s replication crisis). The fact that sampling from a narrow number of schools can create cluster effects in the data is utterly overlooked (see e.g., Hox et al., 2018 on multilevel analysis and how people belonging to the same social group, for instance a school, tend to be more alike). Range restriction is also a regular issue, because of the need to find matched pairs, or as in Smith, Mead and Peters’ research (1927), because only proficient latinists were deemed eligible. Another grave shortcoming is the lack of longitudinal research spanning multiple years, which is a must to chart cognitive development over time and to investigate causality. And so the list goes on.

- (2) Of course the abovementioned methodological flaws can be explained by the times, as already pointed out, as well as by feasibility, coincidence, and the theoretical deficit. The research designs of the CL impact studies are not theory-driven. Instead of starting from a theoretical framework about human cognition and language learning (like the one proposed by Vereeck et al., 2023), the studies are very much rooted in and dictated by everyday practice. Conversely, the analysis of the data does not lead to any real generation of theory. For want of a theoretical basis that can account for the supposed CL effects and the learning mechanisms underlying them, the premises of any hypotheses about the impact of studying CLs remain implicit.
- (3) The style of research reporting, then, tends to be on the vaguer side.⁵⁴ Many details of the design are not always disclosed. What exactly are pupils of “similar intellectual profiles” and the sort? What exactly does it mean when they are “paired”? It is hardly unusual that background data on participants are omitted. The psychometric qualities of the measuring instruments are categorically kept silent about, which is especially problematic in those cases when materials of own produce were used, rather than existing validated tests (see e.g., Reeves & Marbach-Ad, 2016 on test validity). Furthermore, as can be expected, graphical representations of data distributions are very rare.
- (4) In short, the body of research in question is outdated in many ways. Not only do its methods and reports not quite live up to current standards for good scientific practice, its age has yet another consequence. Regardless of scientific quality, this research is simply not that relevant anymore in the twenty-first century, seeing how drastically education has changed. The geographical and linguistic setting of the United States poses additional restrictions. In other words, the results are highly contextual, and generalisability to other times or areas is low (see e.g., Hanushek, 2021 on cross-national generalisability of educational research).

⁵⁴ That is to say nothing of researcher bias. Especially for the Classical Investigation, with its unmistakable political agenda, one cannot shake off the feeling that the content of the report was predetermined. Such criticisms were already voiced at the time, including by classicists like Arthur Tappan Walker (1929, p. 85), who wrote that “the influence of the tests on the recommendations was no greater than if they had been an afterthought”.

- (5) That leaves the matter of scope. Besides the limitation to speakers of English, there are various gaps in the literature. The relative effects of variables like age, socio-economic status, mother tongue, duration and method of CL instruction remain understudied, and so is the impact of learning Ancient Greek on top of Latin. Among the outcome variables, native language ability has stood centre stage. The impact of studying CLs on foreign language aptitude and proficiency, as well as on non-linguistic cognitive abilities have received considerably less attention. The available research on CL impact leaves much to be desired, both formally and substantively.

On the ground of the methodological and other related problems, we cannot consider the alternative hypothesis convincingly discredited, the alternative hypothesis being that the high performance associated with CL study cannot be attributed to the CLs themselves, but is in fact nothing more than preselectivity. If CL pupils are smarter, more motivated, and/or more resilient to begin with, if the programme preselects a certain type of somehow stronger pupils, so to speak, it is only natural that these pupils achieve better scores. That is the downside of the myth of classics: It has always been followed by suspicions of preselectivity. Since the earliest days of CL impact studies there have been skeptics (Foster, 1917; Wilcox, 1917), and subsequent research was not successful in bringing about change.

2.3.7 Recent research (1990–present)

Seeing the abundance of studies throughout the twentieth century, there is a noticeable lack of recent research in the United States. One can merely speculate as to why this is the case. Is it a matter of changing needs, perceptions, or research climates? Is it a matter of chance? While the cognitive impact of studying CLs has attracted some research interest in other countries (e.g., Vanheule, 2015; Hauspie et al., 2024; Haag, 1995, 2001; Haag & Stern, 2000, 2003; Ortner et al., 2008; Goik, 2021; Kipf & Frings, 2021; Chanock, 2006; Holmes-Henderson, 2023), to our knowledge the American research of the past thirty years is limited to the following studies.

In the nineties, Sparks, Ganschow and associates (1995) explored the effects of CL education for pupils with learning difficulties. Later that decade, DeVane (1997) composed a selective and rather associative literature review in which she made a case for the efficacy of Latin studies in the Information Age. Since the turn of the millennium, one doctoral dissertation has been devoted to quantitative measurement of cognitive CL impact (Kennedy, 2006). Though more than fifteen years old already, the methodology and statistical analysis can

still be considered up to date. This makes Kennedy's research an extremely valuable source, albeit that it was never peer-reviewed and published in scientific journals. Cooper and colleagues (2008) replicated and upscaled his 1987 study on SAT Verbal scores: It was confirmed that foreign language study was a significant predictor on top of a range of demographic characteristics, but this time Latin and German pupils were found to obtain equally high scores. Finally, there have been some studies not so much about CL instruction, but about English programmes that taught Latin and/or Greek roots in order to boost word knowledge, and with success (Holmes & Keffer, 1995; Crosson et al., 2019).

Let us zoom in a little more on the two most important studies by Sparks et al. (1995) and Kennedy (2006). Sparks, Ganschow and associates conducted an exploratory study with both normally developing and learning disabled high school pupils, and their findings suggested that the benefits of Latin might be even greater for those who normally struggle with language. The pupils diagnosed with language learning disabilities made considerable progress in phonological/orthographical and semantical/lexical measures of native language proficiency as well as in foreign language aptitude after taking Latin for a year, provided that it was taught in a highly structured, step-by-step fashion with sufficient attention to language sounds and their graphic representations.

Kennedy's dissertation gives an account of a cross-sectional study, using data extracted from the educational records of 227 people graduated from the same private high school in North Carolina between 1996 and 2005. Gender, length of foreign language study and foreign language grades were included as covariates; the sample was considered homogeneous⁵⁵ for the rest. In sharp contrast to contemporary skepticism about the possibility of transfer of learning (e.g., De Bruyckere et al., 2020), the results of Kennedy's study are actually very promising. In summary, he obtained significant results for all of his research questions, bearing upon English ability, standardised assessment, overall school achievement, and math achievement: In all of these areas, graduates who had elected Latin reached higher levels than graduates who had

⁵⁵ Because of the high financial (sizeable tuition fee, no tuition assistance) and intellectual (only pupils who had performed excellently in elementary school) entrance requirements of this school, the sample was assumed to be sufficiently homogeneous in terms of socio-economic background and cognitive ability. This assumption was not empirically verified. Moreover, when true, the range of this research is substantially restricted.

elected a modern foreign language. Thus twenty-first-century research would seem – as Kennedy avidly pointed out himself – to corroborate previous research.

2.4 Conclusion

In this contribution we have provided access and drawn attention to the field of CL impact studies. Despite having more than a century-old history, this research field has remained largely under the radar. Many people from various backgrounds, including countless classicists, have nonetheless taken an interest in the issue of whether the study of CLs transfers to other, not classics-related domains. Consequently, a lot of work has been done, but it is ill-known among the researchers themselves, as proven by their widely differing bibliographies, let alone by a broader public. It is our hope that the current article can serve in establishing the field of CL impact studies as such: a small but definite research field in its own right, with strong ties to classics as well as to cognitive psychology.

Finally, what is the present state of the art regarding the hotly debated cognitive effects of studying CLs, more specifically in the United States? It is far from clear what these effects are, much less so what their effect sizes are. Whilst existing results are auspicious, they supply suggestive evidence at best for the hypotheses that CL study has linguistic and general-cognitive benefits. Despite many research efforts, the classical discipline has been unable to rid itself of the suspicion that preselectivity is the paramount explanation for each and any observation to the advantage of CL pupils. Notwithstanding the methodological and other shortcomings of the older research, we are inclined to say that the sheer amount of evidence does justify at least a strong presumption with regard to possible linguistic benefits. Does studying Latin make pupils smarter? Additional, carefully designed longitudinal research will be needed to make a start on truly answering that question.

Works cited in this chapter

- Adema, S. (2019). Latin learning and instruction as a research field. *Journal of Latin Linguistics*, 18(1–2), 35–59.
- Adler, E. (2020). *The Battle of the Classics: How a nineteenth-century debate can save the humanities today*. Oxford University Press.
- Advisory Committee. (1924). *Part one, general report*. Princeton University Press.
- Aisiah, Suhartono, & Sumarno. (2016). The measurement model of historical awareness. *Research and Evaluation in Education*, 2(2), 108–121.
- Altenhein, M. (1950). The relationship of Latin to achievement in German. *School and Society*, 72(1874), 326–329.
- Angus, D., & Mirel, J. (1999). *The failed promise of the American high school, 1890-1995*. Teachers College Press.
- Baca, A., Mandel, J., & Neshat, K. (1979). Language transfer project of the Los Angeles unified school district. *The Classical Outlook*, 56(4), 74–80.
- Barber, G. (1985). Latin as a practical study. *The Classical Journal*, 81(2), 158–160.
- Bassman, M., & Ironsmith, M. (1984). An experimental FLES program in Latin. *ADFL Bulletin*, 16(1), 39–41.
- Bates, F. (1924). *A controlled experiment in the teaching of English grammar through Latin* [Master's dissertation]. University of Iowa.
- Berelson, B., & Steiner, G. (1964). *Human behavior. An inventory of scientific findings*. Harcourt, Brace & World, Inc.
- Bolgar, R. (1963). A theory of classical education. *Didaskalos. The journal of the Joint Association of Classical Teachers*, 1(1), 5–26.
- Bowker, R. (1975). *English vocabulary comparison of Latin and non-Latin students*. Boston: Human Engineering Laboratory.
- Bracke, E. (2016). The role of university student teachers in increasing Widening Participation to Classics. *Journal of Widening Participation and Lifelong Learning*, 18(2), 111–129.
- Bracke, E. (2020). De impact van Oudgriekse studie op kansarme leerlingen. Participatieverruiming in het lagere scholenproject Oude Grieken—Jonge Helden. *Lampas*, 53(3), 90–103.
- 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.
- Broyler, C., Thorndike, E., & Woodyard, E. (1927). A second study of mental discipline in high school studies. *Journal of Educational Psychology*, 18(6), 377–404.
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1), no. 9.
- Cardinale, U. (2011). Linguistics and the teaching of classical languages. In R. Oniga, R. Iovino, & G. Giusti (Eds.), *Formal linguistics and the teaching of Latin. Theoretical and applied perspectives in comparative grammar*. Cambridge Scholars Publishing.
- Carlisle, J., & Liberman, I. (1989). Does the study of Latin affect spelling proficiency? *Reading and Writing*, 1(2), 179–191.
- Carr, W. (1919). The English vocabulary of the high-school freshman. *The Classical Journal*, 15(1), 20–29.
- Carr, W. (1921). First-year Latin and growth in English vocabulary. *School and Society*, 14(351), 192–198.
- Carr, W. (1942). By their fruits. *The Classical Journal*, 37(6), 334–350.
- Castiello, J. (1934). *Geistesformung. Beiträge zur experimentellen Erforschung der formalen Bildung*. Dümmler.
- Cederstrom, E. (1974). Quid agunt discipuli? Latin in Philadelphia. *The Independent School Bulletin*, 33, 56–57.
- Chanock, K. (2006). Help for a dyslexic learner from an unlikely source: The study of Ancient Greek. *Literacy*, 40(3), 164–170.

- Clark, E. (1932). Amount of high school Latin as an indicator of success in college work. *School and Society*, 35(893), 189–190.
- Cole, L. (1924). Latin as a preparation for French and Spanish. *School and Society*, 19(491), 618–622.
- Cooper, T. (1987). Language study and SAT-Verbal scores. *The Modern Language Journal*, 71(4), 381–387.
- Cooper, T., Yanosky, D., & Wisenbaker, J. (2008). Foreign language learning and SAT Verbal scores revisited. *Foreign Language Annals*, 41(2), 200–217.
- Coxe, W. (1924). The influence of Latin on the spelling of English words. *Journal of Educational Research*, 9(3), 223–233.
- Crosson, A., McKeown, M., Moore, D., & Ye, F. (2019). Extending the bounds of morphology instruction: Teaching Latin roots facilitates academic word learning for English Learner adolescents. *Reading and Writing*, 32(3), 689–727.
- Dallam, T. (1917). Is the study of Latin advantageous to the study of English? *Educational Review*, 54, 500–503.
- Davis-Wiley, P. (1985). High school foreign language study and college academic performance. *The Classical Outlook*, 62(2), 33–36.
- De Bruyckere, P., Kirschner, P., & Hulshof, C. (2020). If you learn A, will you be better able to learn B? Understanding transfer of learning. *American Educator*, Spring, 30–34.
- DeVane, A. (1997). *Efficacy of Latin studies in the information age* [Paper submitted for PSY 702: Educational Psychology, Valdosta State University].
<http://www.edpsycinteractive.org/files/latin.html>
- Douglass, H. (1931). *The relation of high school preparation and certain other factors to academic success at the University of Oregon* (Vol. 1). University of Oregon Press.
- Douglass, H., & Kittelson, C. (1935). The transfer of training in high school Latin to English grammar, spelling, and vocabulary. *Journal of Experimental Education*, 4(1), 26–33.
- Duckworth, A., & Quinn, P. (2009). Development and validation of the Short Grit Scale (grit-s). *Journal of Personality Assessment*, 91(2), 166–174.
- Eddy, P. (1981). *The effect of foreign language study in high school on verbal ability as measured by the Scholastic Aptitude Test – Verbal*. Center for Applied Linguistics.
- Evans, D., & Yuan, F. (2022). How big are effect sizes in international education studies? *Educational Evaluation and Policy Analysis*, 44(3), 532–540.
- Farhady, H. (2018). History of language testing and assessment. In J. Liantas & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching*. John Wiley & Sons.
<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118784235.eelt0343>
- Flexner, A. (1916). A Modern School. *The American Review of Reviews*, 53(4), 465–474.
- Foster, F. (1917). The results of a recent spelling test at the University of Iowa. *School and Society*, 5(122), 506–508.
- Fromchuck, A. (1984). The measurable benefits of teaching English through Latin in elementary school. *The Classical World*, 78(1), 25–29.
- Gilliland, A. (1922). The effect of the study of Latin on ability to define words. *Journal of Educational Psychology*, 13(8), 501–505.
- Goik, L.-S. (2021). *Zur Wirkung des Lateinunterrichts. Ergebnisse einer Längsschnittstudie* (U. Röhr-Sendlmeier, Ed.). Logos Verlag.
- Gutacker, B. (1979). Fördert Lateinunterricht sprachliche Fertigkeiten im Deutschen? In K. J. von Klauer & H.-J. Konradt (Eds.), *Jahrbuch für Empirische Erziehungswissenschaft 1979* (pp. 9–32). Schwann.
- Haag, L. (1995). Auswirkungen von Lateinunterricht. *Psychologie in Erziehung und Unterricht*, 42, 245–254.
- Haag, L. (2001). Auswirkungen von Lateinunterricht – Ergebnisse nach zwei Lernjahren. *Psychologie in Erziehung und Unterricht*, 48, 30–37.
- Haag, L., & Stern, E. (2000). Non scholae sed vitae discimus? Auf der Suche nach globalen und spezifischen Transfereffekten des Lateinunterrichts. *Zeitschrift für Pädagogische Psychologie*, 14(2/3), 146–157.
- Haag, L., & Stern, E. (2003). In search of the benefits of learning Latin. *Journal of Educational Psychology*, 95(1), 174–178.

- Hamblen, A. (1925). *An investigation to determine the extent to which the effect of the study of Latin upon a knowledge of English derivatives can be increased by conscious adaptation of content and method to the attainment of this objective* [Doctoral dissertation]. University of Pennsylvania.
- Hanushek, E. (2021). Addressing cross-national generalizability in educational impact evaluation. *International Journal of Educational Development*, 80, no. 102318.
- Harris, L. (1915). A study in the relation of Latin to English composition. *School and Society*, 2(33), 251–252.
- Haskell, R. (1923). *A statistical study of the comparative results produced by teaching derivation in the ninth grade Latin classes and in the ninth grade English classes of non-Latin pupils in four Philadelphia high schools* [Doctoral dissertation]. University of Pennsylvania.
- Hauspie, C., Duyck, W., Schelfhout, S., Vereeck, A., Janse, M., & Szmalec, A. (2024). Is studying Latin associated with (non-)linguistic cognitive transfer? A large-scale cross-sectional study. *Contemporary Educational Psychology*, 79, no. 102325. <https://doi.org/10.1016/j.cedpsych.2024.102325>
- Heald, I. (1923). *Relation between the study of Latin in high school and first-year college French* [Master's dissertation]. University of Iowa.
- Hill, J. (1924). *The relation of the amount of Latin pursued in high school to success in first-semester French in the University of Iowa* [Master's dissertation]. University of Iowa.
- Ho, D., Imai, K., King, G., & Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3), 199–236.
- Holliday, L. (2012). The benefits of Latin? *Educational Research Quarterly*, 36(1), 3–12.
- Holmes, T., & Keffer, R. (1995). A computerized method to teach Latin and Greek root words: Effect on Verbal SAT scores. *Journal of Educational Research*, 89(1), 47–50.
- Holmes-Henderson, A. (2021). Developing multiliteracies through classical mythology in British classrooms. In L. Maurice (Ed.), *Our mythical education* (pp. 139–151). Warsaw University Press.
- Holmes-Henderson, A. (2023). Ancient languages for 6- to 11-year olds. Exploring three pedagogical approaches via a longitudinal study. In A. Holmes-Henderson (Ed.), *Expanding Classics. Practitioner perspectives from museums and schools* (pp. 9–29). Routledge.
- Holmes-Henderson, A., & Kelly, K. (2022). *Ancient languages in primary schools in England: A literature review*. Department for Education. <https://www.gov.uk/government/publications/ancient-languages-in-primary-schools-in-england>
- Hox, J., Moerbeek, M., & Van de Schoot, R. (2018). Introduction to multilevel analysis. In *Multilevel analysis. Techniques and applications* (3rd edition, pp. 1–7). Routledge.
- Katz, J., Noels, K., & Fitzner, A. (2020). Why learn Latin? Motivation for learning a classical language. *Teaching Classical Languages*, 11(1), 89–129.
- Kennedy, C. (2006). *The study of Latin and its correlation with improved English abilities* [Doctoral dissertation]. Touro University.
- Kipf, S., & Frings, K. (2021). Latein als Brückensprache. In *Integration durch Sprache: Schüler nichtdeutscher Herkunftssprache lernen Latein* (pp. 22–42). Propylaeum.
- Kirby, T. (1923). Latin as a preparation for French. *School and Society*, 18(463), 563–569.
- Kitchell, K. (2008). Promotion of the classics in the United States: New initiatives for a new millennium. In B. Lister (Ed.), *Meeting the challenge. International perspectives on the teaching of Latin* (pp. 150–164). Cambridge University Press.
- Kumlien, C., Bish, M., Chan, E., Rew, L., Chan, P., Leung, D., & Carlson, E. (2020). Psychometric properties of a modified cultural awareness scale for use in higher education within the health and social care fields. *BMC Medical Education*, 20, no. 406.
- LaFleur, R. (1981). Latin students score high on SAT and achievement tests. *The Classical Journal*, 76(3), 254.
- LaFleur, R. (1982). 1981 SAT and Latin achievement test results and enrollment data. *The Classical Journal*, 77(4), 343.
- Lawler, L. (1925). *The potential remediability of errors in English spelling through the study of high school Latin* [Doctoral dissertation]. University of Iowa.

- LeBovit, J. (1967). Qui timide rogat, docet negare. *The Classical World*, 61(2), 37–40.
- Lucasse, F. (1929a). *A critical study of the report of the classical investigation* [Master's dissertation]. Harvard University.
- Lucasse, F. (1929b). What the Classical Investigation has accomplished. *Journal of Educational Research*, 20(1), 15–21.
- MacPhail, A. (1924). Freshman academic achievement in college of students presenting four years of Latin and those presenting no Latin. *School and Society*, 19(479), 261–262.
- Manolidou, E., Goula, S., & Sakka, V. (2023). Ancient Greek for kids: From theory to praxis. *Journal of Classics Teaching*, 24(47), 3–11.
- Masciantonio, R. (1977). Tangible benefits of the study of Latin. A review of research. *Foreign Language Annals*, 10(4), 375–382.
- Masciantonio, R. (1968c). *How the Romans lived and spoke (Romani viventes et dicentes). A humanistic approach to Latin for children in the fifth grade*. Instructional services of the school district of Philadelphia.
- Masciantonio, R. (1968a). Latin materials for the inner city public school. *The Classical Outlook*, 45(6), 61–62.
- Masciantonio, R. (1968b). Latin's needs in the big city schools. *The Classical World*, 61(9), 390–393.
- Mavrogenes, N. (1977). The effect of elementary Latin instruction on language arts performance. *The Elementary School Journal*, 77(4), 268–273.
- Mead, A. (1923). *Learning and teaching. Psychological foundations of educational technique*. J.B. Lippincott Company.
- Moore, D., & McCabe, G. (2005). *Introduction to the practice of statistics* (5th edition). W.H. Freeman and Company.
- Morton-Finney, P. (1941). Latin, a basis for French and Spanish study as evidenced by teacher's marks. *Modern Language Journal*, 25(11), 873–880.
- National Council of State Supervisors of Foreign Languages. (2003). The role of Latin in American education. *The Classical Outlook*, 80(4), 147–149.
- Newcomb, E. (1923a). Latin in the high school. *School and Society*, 17(423), 129–130.
- Newcomb, E. (1923b). The testing program of the Classical League. *School and Society*, 17(442), 665.
- Nisbet, S. (1951). *The practicability of Latin in the secondary school* [Master's dissertation]. Indiana State Teachers College.
- Offenberg, R., Epstein, B., & Rodriguez-Acosta, C. (1982). *Language arts through Latin 1980-1981*. The school district of Philadelphia office of research, planning and evaluation.
- Ortner, T., Asanger, M., Kubinger, K., & Proyer, R. (2008). Zur Frage nach Auswirkungen von Lateinunterricht auf die kognitive Fähigkeit 'Reasoning'. *Psychologie in Erziehung und Unterricht*, 55, 189–195.
- Otis, A. (1922). The relation of Latin to the study of English vocabulary and composition. *School Review*, 30, 45–50.
- Parounagian, M. (1942). The Portland derivatives test. *The Classical Outlook*, 19(6), 54–55.
- Perkins, A. (1914). Latin as a vocational study in the commercial course. *The Classical Journal*, 10(1), 7–16.
- Pfeiffer, R. (1976). *History of classical scholarship* (Vol. 2). Pergamon Press.
- Phinney, E. (1994). The beginning. In E. Phinney (Ed.), *The history of the American Classical League, 1919-1994* (pp. 23–40). American Classical League.
- Pire, G. (1971). *Le latin en question. Vues modernes sur l'étude des langues anciennes*. Dessain.
- Polksy, M. (1986). The NEH/Brooklyn College Latin Cornerstone Project 1982-84: Genesis, implementation, evaluation. *The Classical Outlook*, 63(3), 77–83.
- Pond, F. (1938). Influence of the study of Latin on word knowledge. *The School Review*, 46(8), 611–618.
- Pyck, J. (1947). *Prof. Dr. Jaime Castiello, Geestesvorming: Bijdrage tot het experimenteel onderzoek van de formele vorming. Vertaling en nota's omtrent middelbaar en normaal onderwijs*. De Kinkhoren.
- Reeves, T., & Marbach-Ad, G. (2016). Contemporary test validity in theory and practice: A primer for discipline-based education researchers. *CBE Life Sciences Education*, 15(1).
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4803101/>
- Reinhardt, E. (1983). They don't call this class 'bor-ing'. *The English Journal*, 72(1), 37.

- Richardson, H. (1933). Discovering aptitude for the modern languages. *Modern Language Journal*, 18(3), 160–170.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138, 353–387.
- Rockefeller Archive Center. (2022). "Without distinction of race, sex, or creed": The General Education Board, 1903–1964. RE:Source. <https://resource.rockarch.org/story/the-general-education-board-1903-1964/>
- Scott, H., & LeBovitz, J. (1971). *A study of the effect of Latin instruction on English reading skills of sixth grade students in the public schools of the district of Columbia, school year, 1970-71*. District of Columbia public schools, Division of Planning, Research and Evaluation.
- Sheridan, R. (1976). *Augmenting reading skills through language learning transfer. FLES Latin program evaluation reports, 1973-74, 1974-75, 1975-76*. Indianapolis public schools.
- Sister Josephina. (1963). Latin: An effective word builder. *The Classical Journal*, 58(8), 354–355.
- Smith, D., Mead, A., & Peters, C. (1927). The transfer of 'translation thinking'. *School and Society*, 25(639), 380–384.
- Smith, J. (2007). *The correlation of Latin, Greek and the Classical Education model with learning other subjects* [Doctoral dissertation]. Capella University.
- Smith, M., & Douglass, H. (1937). The relation of high-school Latin to marks in the first year of arts college. *School Review*, 45(9), 695–701.
- Sorenson, H. (1929). High-school subjects as conditioners of college success. *Journal of Educational Research*, 19, 237–254.
- Sparks, R., 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.
- Spearman, C. (1904). 'General intelligence', objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–292.
- Stellwag, H. (1949). *De waarde der klassieke vorming. Een cultuur-historische, paedagogisch-psychologische en didaktische inleiding*. Wolters.
- Sussman, L. (1978). The decline of basic skills. A suggestion so old that it's new. *The Classical Journal*, 73(4), 346–352.
- Tangney, J., Baumeister, R., & Luzzo Boone, A. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2), 271–324.
- Taylor, A., Holmes-Henderson, A., & Jones, S. (2023). Classics education in Northern Irish primary schools; curriculum policy and classroom practice. *Journal of Classics Teaching*, 24(47), 52–58.
- Thorndike, E. (1923a). The influence of first-year Latin upon range in English vocabulary. *School and Society*, 27(421), 82–84.
- Thorndike, E. (1923b). The influence of first-year Latin upon ability to read English. *School and Society*, 27(424), 165–168.
- Thorndike, E. (1923c). The gains made in ability in English by pupils who study Latin and by pupils who do not. *School and Society*, 28(467), 690.
- Thorndike, E. (1924a). Mental discipline in high school studies. *Journal of Educational Psychology*, 25(1), 1–22.
- Thorndike, E. (1924b). Mental discipline in high school studies. *Journal of Educational Psychology*, 25(2), 83–98.
- Thorndike, E., & Ruger, G. (1923). The effect of first-year Latin upon knowledge of English words of Latin derivation. *School and Society*, 28(453), 260–270.
- Thorndike, E., & Woodworth, R. (1901a). The influence of improvement in one mental function upon the efficiency of other functions (I.). *Psychological Review*, 8(3), 247–261.
- Thorndike, E., & Woodworth, R. (1901b). The influence of improvement in one mental function upon the efficiency of other functions II. The estimation of magnitudes. *Psychological Review*, 8(4), 384–395.

- Thorndike, E., & Woodworth, R. (1901c). The influence of improvement in one mental function upon the efficiency of other functions III. Functions involving attention, observation and discrimination. *Psychological Review*, 8(6), 553–564.
- Tirunch, D., Gu, X., De Cock, M., & Elen, J. (2018). Systematic design of domain-specific instruction on near and far transfer of critical thinking skills. *International Journal of Educational Research*, 87, 1–11.
- Townsley, L. (1985). Latin as a vocabulary builder for hearing-impaired and second-language students of English. *Teaching English to Deaf and Second-Language Students*, 3(1), 4–8.
- Van Damme, D. (2021). Wat kunnen jongeren vandaag in hemelsnaam nog leren van Homeros of Tacitus? *Kleio. Tijdschrift voor oude talen en antieke cultuur*, 50(3–4), 179–182.
- Vanheule, J. (2015). Latijn en grammaticale competentie in het Nederlands. Verslag van een nieuw onderzoek. *Didactica Classica Gandensia*, 51, 89–107.
- Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology*, 101(3), 671–688.
- VanTassel-Baska, J. (1982). Results of a Latin-based experimental program for the verbally precocious. *Roeper Review*, 4(4), 48–50.
- VanTassel-Baska, J. (1987). A case for the teaching of Latin to the verbally talented. *Roeper Review*, 9(3), 159–161.
- 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*, 25(50), 101–115.
- 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. *Horizons of Psychology*, 32, 121–130.
- Wagener, A. (1937). The adaptation of objectives in ancient language teaching to present educational practices. *The Classical Journal*, 32(8), 455–466.
- Walker, A. (1929). The report of the Classical Investigation. A criticism. *The Classical Journal*, 25(2), 83–92.
- Webster, D. (1985). James McKeen Cattell and the invention of academic quality ratings, 1903–1910. *Review of Higher Education*, 8(2), 107–121.
- West, A., Carr, W., Gray, M., & McDuffee, W. (1923). The Classical Investigation: The work of the first two years. *The Classical Journal*, 18(9), 548–568.
- Whelden, C. (1933). Training in Latin and quality of other academic work. *Journal of Educational Psychology*, 24(7), 481–497.
- Wilcox, M. (1917). Does the study of high-school Latin improve high-school English? *School and Society*, 6(133), 58–60.
- Wraga, W. (2009). Latin literacy redux: The classical investigation in the United States, 1921–1924. *History of Education*, 38(1), 79–98.
- Yates, J. (1929). *The type of high school curriculum which gives the best preparation for college* (Vol. 1). College of Education, University of Kentucky.

Addenda to Chapter 2

Recent non-American research

At first, the intention was for the above article to cover more or less all previous research. As I began to unravel the history of the field I would dub classical language impact studies, it turned out that the American research was much more abundant and went much farther back in time than initially thought, so the article's scope was adjusted accordingly. To round out this chapter on the state of the art, I will now briefly discuss some recent non-American research efforts that must not go unmentioned. This section includes all other published empirical studies that I am currently aware of on the impact of studying classical languages in pre-tertiary education.

Around the turn of the century, two studies were undertaken by the German classicist and educational psychologist Ludwig Haag, together with fellow psychologist Elsbeth Stern (Haag, 2001; Haag & Stern, 2000, 2003). One study tested fifty university students taking a beginners' course in Spanish. At the end of the course, the students received a translation assignment from German to Spanish. Their scores were compared depending on whether they had previously studied French or Latin as their second foreign language (the first foreign language being English for all of them). Those who had previously studied French made fewer grammar and vocabulary errors, which led Haag and Stern to declare that, for native speakers of German, French was a better preparation for learning Spanish than Latin.

The other study was a longitudinal study taking place in secondary education, in which a program group of 115 Gymnasium pupils who had studied Latin for four years was set against a comparison group of 93 Gymnasium pupils of which 26 had never studied Latin and 67 had studied Latin for only two years, which Haag and Stern apparently deemed negligible. Within this set-up, they arrived at the following findings. Regarding native language ability, transfer effects were found for error detection, construction of compound sentences, and school achievement in German. Regarding general cognitive ability, no transfer effects of Latin were found whatsoever. Furthermore, a difference in educational climate was observed: The four-year Latin pupils experienced more pressure to achieve, spent more time on homework, were more competition-minded, displayed more discipline, and judged their teachers as more competent.

These two studies were quite impactful, inside as well as outside academia. The negative findings in the areas of foreign language ability and general cognitive ability were framed as proof for the uselessness of classical language study, both by the media and by the researchers themselves (Haag, 2001; Haag & Stern, 2003; Wirth, 2011). Haag and Stern's methodology and argumentation would later be heavily criticised, however (e.g., Ortner et al., 2008; Wirth, 2011). Ortner and colleagues (2008) zeroed in on the fact that the intelligence tests Haag and Stern used were already outdated and inadequate at the time. In reaction, they conducted an exploratory cross-sectional study, administering an up-to-date battery of five reasoning ability tests to 137 ninth-graders, some of which had just started Latin and some of which did not take Latin, and 112 eleventh-graders, some of which had taken Latin for two years and some of which had never taken Latin. The ninth-grade sample showed a reversed preselectivity effect: The group who had chosen not to study Latin performed better on four of the tests, though only one of these four group differences was statistically significant. In the eleventh-grade sample, the group who had been studying Latin scored significantly higher on the Symbol Calculations test.

A few years ago, a longitudinal study was conducted in Bonn (Goik, 2021), albeit on a somewhat smaller scale than the study recorded in this dissertation, and with a different target population as well. During school years 2017–2018 and 2018–2019, Goik followed a sample of pupils aged ten to twelve as they transitioned from the fifth to the sixth grade. In Flanders, these grades are still part of primary education, but for the participants from Bonn, the fifth and sixth grades represented the first two years of Gymnasium. By the end of the study, the final sample fell into three different (non-overlapping) groups: 70 pupils who had started learning Latin from grade 5 onwards, 107 pupils who had started learning Latin from grade 6 onwards, and 88 pupils who had started learning French from grade 6 onwards.

Goik investigated transfer from Latin to various language-related and non-language related domains. Despite some issues with statistical power and the quality of the measuring instruments, she came to the conclusion that Latin had a beneficial impact on both native language ability, in particular on making lexical associations, and foreign language ability, in particular on recognising Latin / French cognates in English (a language which none of the pupils had previously received any formal instruction in). She also noted that the pupils' German vocabulary widened as a result of learning a second language, but that the duration of second language study (i.e., two years for the pupils beginning in grade 5 and one year for the pupils beginning in grade 6) mattered more than which language they had chosen (i.e., Latin or

French). Furthermore, Goik asked the pupils about their enjoyment of various school subjects, and Latin by far received the highest ratings.

In the United Kingdom, there has been a lot of interest in the potential of Latin to improve English literacy, especially for pupil populations facing educational and/or socio-economic barriers, such as children who learn English as an Additional Language, who have Special Educational Needs, who qualify for Free School Meals, or who come from ethnic-minority backgrounds (review in Holmes-Henderson & Kelly, 2022). Many classics initiatives have been set up with this idea in mind, and in general they also seem to achieve their goal of supporting literacy. However, it appears there has been little to no empirical research that adheres to a scientific hypothesis-testing paradigm. Most studies narrating the effects of classical language instruction on English language ability relied exclusively on descriptive statistics and intuitive observations (e.g., Holmes-Henderson, 2023; Wing-Davey, 2018; Wright, 2023).

Elsewhere in Europe, Ghent University in Flanders has emerged as somewhat of a centre for classical language impact studies. Prior to the completion of the present dissertation, there have already been a study on the correlation between Latin study in secondary education and academic achievement in higher education (Hauspie, Schelfhout, et al., 2024), a preliminary cross-sectional study on linguistic and general-cognitive transfer (Hauspie, Duyck, et al., 2024, mentioned again in Section 9.2.1), and another tentative study on linguistic transfer that used retrospective grouping (Vanheule, 2015). Vanheule recruited 280 pupils in their final year of general secondary education, all from the same school, and administered a Dutch language test consisting of various syntactical exercises. Including achievement for English, mathematics and history as covariates, she analysed the relation between the test scores and the number of years of Latin study (between 0 and 6) and found that having studied Latin for three or more years was a significant predictor. Vanheule also took into account whether the pupils were studying Greek, and this variable had additional explanatory value on top of years of Latin. In sum, her study indicated that studying classical languages benefitted Dutch syntactic competence.

Mathematics impact studies?

One might ask how unique this kind of impact research is to classical languages. The situation of the classical languages is unique in the sense that other school subjects like modern languages, mathematics, sciences and history are not disputed in a similar way, even though not everything that is taught there has direct practical use either. For those other subjects, there is

no need for “research in defence of the discipline” (cf. Section 2.2.5). Nevertheless, researchers have wondered whether mathematics possessed any transfer value. What classical languages (as they are typically instructed) and mathematics have in common, is that these subjects are systematic, that they are organised in a cumulative way, and that they require a certain degree of abstract and analytical thinking (DeVane, 1997; VanTassel-Baska, 1987). Latin has even earned the nickname of “maths of the humanities” (Pelling & Morgan, 2010, p. 6).

Based on a cursory search, mathematics impact research appears to be less abundant, but it does appear to span roughly the same time period as classical language impact research. In 1917, the American philosopher Ernest Moore (1871–1955) wrote an animated article for *School and Society* which could be characterised as a literature review about the theory of formal discipline, from Socrates to Spearman, with a focus on the role of mathematics in education. Among other things, Moore (1917, p. 489) described an “experiment” by one Lewis of Dartmouth. In this early example of mathematics impact research, Lewis contrasted college students’ performance on a geometry test with their academic achievement in a law class and their performance on an essay task. He noted that high scores on the geometry test did often not go hand in hand with high scores in the other two areas, which was inconsistent with transfer. Recent research, on the other hand, has looked at transfer from one mathematical domain to another (e.g., Jerrim et al., 2025). The effects of spatial training in particular have been a topic of interest. A recent meta-analysis concluded that spatial training enhanced mathematics understanding and performance in general (Hawes et al., 2022).

Beyond regular school subjects, other pursuits that have frequently been linked to cognitive transfer are music, chess, and computer programming. These research fields are touched upon in the theoretical framework, in both Section 3.5.2 and the addenda to Chapter 3.

Works cited in these addenda

- DeVane, A. (1997). *Efficacy of Latin studies in the information age* [Paper submitted for PSY 702: Educational Psychology, Valdosta State University].
<http://www.edpsycinteractive.org/files/latin.html>
- Goik, L.-S. (2021). *Zur Wirkung des Lateinunterrichts. Ergebnisse einer Längsschnittstudie* (U. Röhr-Sendlmeier, Ed.). Logos Verlag.
- Haag, L. (2001). Auswirkungen von Lateinunterricht – Ergebnisse nach zwei Lernjahren. *Psychologie in Erziehung und Unterricht*, 48, 30–37.
- Haag, L., & Stern, E. (2000). Non scholae sed vitae discimus? Auf der Suche nach globalen und spezifischen Transfereffekten des Lateinunterrichts. *Zeitschrift für Pädagogische Psychologie*, 14(2/3), 146–157.

- Haag, L., & Stern, E. (2003). In search of the benefits of learning Latin. *Journal of Educational Psychology*, 95(1), 174–178.
- Hauspie, C., Duyck, W., Schelfhout, S., Vereeck, A., Janse, M., & Szmalec, A. (2024). Is studying Latin associated with (non-)linguistic cognitive transfer? A large-scale cross-sectional study. *Contemporary Educational Psychology*, 79, no. 102325.
- 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*, 74(4), 853–883.
- Hawes, Z., Gilligan-Lee, K., & Mix, K. (2022). Effects of spatial training on mathematics performance: A meta-analysis. *Developmental Psychology*, 58(1), 112–137.
- Holmes-Henderson, A. (2023). Ancient languages for 6- to 11-year olds. Exploring three pedagogical approaches via a longitudinal study. In A. Holmes-Henderson (Ed.), *Expanding Classics. Practitioner perspectives from museums and schools* (pp. 8–29). Routledge.
- Holmes-Henderson, A., & Kelly, K. (2022). *Ancient languages in primary schools in England: A literature review*. Department for Education.
<https://www.gov.uk/government/publications/ancient-languages-in-primary-schools-in-england>
- Jerrim, J., Lopez-Agudo, L., Sims, S., & Marcenaro-Gutierrez, O. (2025). Teaching for near transfer: Is maths instruction aimed at schema formation and abstraction associated with pupils' ability to answer unfamiliar maths questions? *Learning and Individual Differences*, 118, no. 102609.
- Moore, E. (1917). Does the study of mathematics train the mind specifically or universally? *School and Society*, 6(148), 481–491.
- Ortner, T., Asanger, M., Kubinger, K., & Proyer, R. (2008). Zur Frage nach Auswirkungen von Lateinunterricht auf die kognitive Fähigkeit 'Reasoning'. *Psychologie in Erziehung und Unterricht*, 55, 189–195.
- Pelling, C., & Morgan, L. (2010). *Latin for language learners. Opening opportunities for primary pupils*. Politeia.
- Vanheule, J. (2015). Latijn en grammaticale competentie in het Nederlands. Verslag van een nieuw onderzoek. *Didactica Classica Gandensia*, 51, 89–107.
- VanTassel-Baska, J. (1987). A case for the teaching of Latin to the verbally talented. *Roeper Review*, 9(3), 159–161.
- Wing-Davey, Z. (2018). Delivering Latin in primary schools. In A. Holmes-Henderson, S. Hunt, & M. Musié (Eds.), *Forward with Classics. Classical languages in schools and communities*. (pp. 117–127). Bloomsbury Academic.
- Wirth, T. (2011). Unfaire Latein-Kritik. *Pegasus Online-Zeitschrift*, 11(1), 133–140.
- Wright, P. (2023). Including the excluded. Teaching Latin in an area of high socio-economic disadvantage. In A. Holmes-Henderson (Ed.), *Expanding Classics. Practitioner perspectives from museums and schools* (pp. 30–41). Routledge.

Chapter 3: Theoretical Framework

The contents of this chapter are nigh identical to the following article:

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. *Horizons of Psychology*, 32, 121–130.
<https://doi.org/10.20419/2023.32.580>

Minor edits were made with an eye to harmonising spelling, punctuation, title numbering, cross references, bibliographical style etc. throughout the dissertation.

Chapter 3

Theoretical Framework

3.1 Introduction

During the bustling and buzzing 17th edition of the European Congress of Psychology in Ljubljana, the first author had the pleasure of presenting a poster entitled “A debate in need of data. Report of an empirical study on the cognitive transfer effects of studying classical languages” (Vereeck et al., 2022). With this article we would like to expand on one aspect of that presentation. In particular, we want to give a more complete and detailed account of the underlying theoretical framework, because we believe it nicely exemplifies how psychology functions as a hub science today. The central construct stemming from cognitive psychology, this theoretical framework combines insights from various psychological and non-psychological literatures in order to shed light on a problem originating in an entirely different discipline, namely classics.

First, we shall briefly reiterate the research topic: the cognitive impact of classical language education. In Europe, the instruction of Latin and Ancient Greek has a longstanding tradition and is still an important part of most educational systems. Nevertheless, across decades and countries, the value of studying these ancient languages has been the subject of fierce public debate. In reaction to criticisms about their perceived futility and irrelevance in today’s world, classicists and other proponents have suggested that the study of classical languages transfers to other cognitive domains, resulting in improved native as well as foreign language abilities, analytical reasoning skills, and so on. To this day, however, the alleged transfer benefit amounts to little more than an apologetic *topos*, not (yet) proven by scientific evidence.

As may already be inferred from the title and the description of the research topic, the central construct here is cognitive transfer. Although there has been some research into the linguistic and other benefits of studying classical languages (Bracke & Bradshaw, 2020; Vereeck et al., 2025) and these have even been connected with cognitive transfer before (e.g., De Bruyckere et al., 2020), this article will present what is to our knowledge the first proposal of a fully elaborated theoretical framework for the issue at hand. Given the age-long history of the debate on classical language education (Vereeck, 2024), such a theoretical framework is definitely a desideratum, as it is a necessary condition for sound and rigorous research.

3.2 A definition of cognitive transfer

Let us begin by defining the central construct. Cognitive transfer is the mechanism where learning one task, say Latin grammar, has a favourable impact on performance on another task, say native language ability. These tasks are respectively called the training task and the transfer task. There are two types of cognitive transfer. On the one hand, transfer is said to be “near” if the tasks are similar, that is if they fall within the same cognitive modality – for example linguistic competence. On the other hand, transfer is said to be “far” if the tasks are dissimilar, that is if they fall within different cognitive modalities. In many studies on cognitive transfer one will find a definition along these lines (e.g., Harrison et al., 2013), whereby similarity is conventionally interpreted as the extent to which the task domains share common features.

There is more to cognitive transfer than is captured by this intuitive definition, however, which is why Barnett and Ceci (2002) have devised a transfer taxonomy. In their taxonomy the multifaceted conditions of transfer are broken down into two global factors, the content and the context. Both of these global factors are broken down further into several independent dimensions. The content factor is subdivided into the specificity of the learned skill, the memory demands, and the nature of the performance change. To paraphrase, a skill that is learned during the training task, specific or generic, is remembered and activated during the transfer task, be it prompted or unprompted, and thus affects the speed, accuracy, or approach of performance on the transfer task. The context factor is subdivided into the knowledge domain, the physical context, the temporal context, the functional context (e.g.,

academic or informal), the social context (e.g., individual or in group), and the modality (e.g., written or oral). Importantly, the training and the transfer task can be less or more similar along each of the six dimensions of the context factor. Instead of just two different types of cognitive transfer, near and far, theoretically there are now sixty-four (2⁶).

But it does not end there. It is not enough to take into account the transfer conditions in all its facets; transfer effects ought to be studied in relation to the participants' characteristics as well. As adequately discussed by von Bastian and Oberauer (2014) in their review of working memory training, research indicates that cognitive performance is affected by an array of individual characteristics. These include age, initial cognitive ability, genetic predispositions, motivation, and personality traits such as neuroticism and conscientiousness. Additionally, people's beliefs about or awareness of the different conditions in a research design can also influence their performance (Green et al., 2019). In an attempt to give due attention to inter-individual differences and participant expectations, whilst matching the terminology of the transfer taxonomy, in what follows this third factor will be termed the individual factor.

3.3 A concise history of cognitive transfer research

Transfer of learning has fascinated people as long as learning itself. In Classical Greece, Plato, inspired by Pythagoras, already voiced the idea that musical instruction ought to be the basis of all education, for mastering musical harmony would lead to a more harmonious character. He also recommended thorough mathematical instruction: A mind well versed in mathematical problems can solve any problem at all, he believed (Stellwag, 1949). From the Renaissance onwards, this special educational position was bestowed upon the study of classics. Historically, there is therefore a strong link between the concept of cognitive transfer and classical language education.

It was not until the Enlightenment that the idea really gained ground that mental capacities could be trained in the same way as muscles, by applying them to subject matter with a formative value. In Locke's and Kant's writings, for example, we clearly find this theory of mental discipline (Stellwag, 1949). While the term "mental discipline" may be old-fashioned and outdated, it essentially describes nothing else than transfer of learning. In the nineteenth

century, a lively scholarly debate developed itself about the possibility of a formal training which would have broad transfer effects (Castiello, 1934).

When psychology tore itself loose from philosophy and adopted the scientific method, cognitive transfer could no longer be dogmatically accepted or rejected, but its existence had to be empirically verified. Charles Spearman (1927, p. 163) even referred to it as “the most vital of all educational problems”. His theory of intelligence is reconcilable with the possibility of transfer. According to Spearman (1904), general intelligence underlies all branches of intellectual activity, so provided it can be trained, transfer to a great many specific activities is conceivable. His contemporary Edward Thorndike was of a different opinion. After running several experiments (notably Thorndike & Woodworth, 1901a, 1901b, 1901c) in which participants fulfilled tasks under slightly different training and transfer conditions, he formulated the theory of identical elements (Thorndike, 1914), which is still invoked today (e.g., De Bruyckere et al., 2020). This theory states that transfer is directly proportional to the number of elements in common between two tasks, which strictly interpreted equals a negation of the existence of proper cognitive transfer.

Many other psychologists also took an interest in this vital educational problem, so that by the time the nineteen- thirties came round, over a hundred transfer experiments were known (Castiello, 1934). Methods gradually improved, yet results and conclusions greatly differed. Thus the field of cognitive transfer research grew into one of extreme positions and little consensus. For all the progress we have made in our understanding of human cognition, even today the likelihood of transfer is still a matter of scholarly debate. It ranges from hopeful optimism (e.g., Barnett & Ceci, 2002; Green et al., 2019; Jaeggi et al., 2014) to uncompromising skepticism (e.g., Gobet & Sala, 2022; Harrison et al., 2013; Shipstead et al., 2010). A prominent subset of modern cognitive transfer research has come to focus on working memory in particular (von Bastian & Oberauer, 2014).

In short, cognitive transfer is a hot topic in psychology, past and present. A few years ago, Sala, Gobet and colleagues (2019) performed a second-order meta-analysis of recent research on cognitive training programs. They found unequivocal evidence for the occurrence of near transfer from working memory training to memory tasks, especially in normally developing children. The existence of far transfer could not be confirmed, partly due to methodological problems in far transfer studies like placebo effects, publication bias and lack of statistical power. The explanation Gobet and Sala (2022) proposed for the pattern of results across studies, is that the between-study variance mainly springs from sampling error and that the

true effect size of far transfer, when corrected for placebo effects and publication bias, is actually zero.

According to an article composed by forty-eight leading scientists in the field, however, “the absence of clear methodological standards has made it difficult-to-impossible to easily and directly compare results across studies (either via side-by-side contrasts or in broader meta-analyses)” (Green et al., 2019, p. 3). They also argued that each type of behavioural intervention for cognitive enhancement is too different to study them all together under one overarching category. Indeed, lumping large amounts of research together, which is of course the very point of a meta-analysis, contradicts the philosophy of Barnett and Ceci (2002) and is bound to lead to negative conclusions, as far as cognitive transfer is concerned. As highlighted by the transfer taxonomy, there are always several dimensions to be taken into account: Transfer might occur in some configurations of these dimensions and not in others. Therefore, a lack of significant and replicable far transfer effects in a certain type of situation does not mean that the possibility of far transfer is to be dismissed altogether.

3.4 Applying the cognitive transfer construct to classical language education

Is the study of classical languages useful, in the (narrow) sense that pupils become better readers, writers and thinkers than they otherwise would have been? This question can readily be rephrased as a cognitive transfer issue. The training task consists of classical language education, as offered by the local educational system. The transfer task can be anything which is mentioned as a likely benefit of classical language education. Language-related effects may be considered near transfer, whereas effects on general cognitive abilities and not language-related achievement may be considered far transfer. As simple and obvious as the preceding perhaps sounds, the impact of classical language education has yet to be truly viewed from this perspective and its potential cognitive transfer effects have yet to be verified.

3.4.1 The training task

Now, let us try to grasp the training task in more detail. To this end, it is necessary to introduce another core construct, being language complexity. Linguists use this term in multiple ways.⁵⁶ The relevant approach here is the psycholinguistic one (Housen, 2020): Complexity then refers to the cognitive effort and resources deployed to process language features in second language learning and use. The complexity a learner experiences emanates from the language features themselves, from a multitude of learner characteristics, among which age, motivation and the distance between the first and the second language, and from the learning context, where second language acquisition theory makes a distinction between explicit and implicit settings.

The inconstant sum of feature-related, learner-related and context-related complexity is sometimes called relative complexity, as opposed to absolute or structural complexity, which is an unchanging property of individual language features. Linguistically speaking (Housen, 2020), phonological, lexical, and grammatical features are complex or not, depending on the internal structuring of linguistic units and their interconnections with other linguistic units. Structural complexity can be of a purely formal nature or rather of a functional nature, the latter being determined by how many functions a linguistic unit has and how transparent the form-function mappings are.

For decades, it was common knowledge in the field of linguistics that the overall absolute complexity of every language was the same, an axiom known as the principle of invariance of language complexity, the ALEC-statement (an acronym for “all languages are equally complex”), or the equi-complexity dogma (Becerra-Bonache & Jiménez- López, 2015; Sampson, 2009). This axiom has been refuted and is no longer universally accepted by present-day linguists (e.g., Joseph & Newmeyer, 2012; Gil et al., 2009). Researchers have in

⁵⁶ The literature reveals at least four approaches to defining complexity (Housen, 2020; Housen & Simoens, 2016). Some define it in terms of developmental timing (1), a central concept in the Processability Theory on second language acquisition (see for example Pienemann & Lenzing, 2015). According to this theory, a linguistic feature is more complex than another feature if it takes more time to emerge or to be fully mastered. The pedagogic approach (2) focuses on the complexity of descriptions and explanations provided to learners to help them acquire or use a second language feature: their elaborateness, the amount of meta-language, the degree of conceptual clarity or abstractness of the terminology, etc. These two approaches have no bearing on the current theoretical framework. Then there is the linguistic approach (3), which can in itself be seen as one aspect of the psycholinguistic approach (4).

fact detected differences in total complexity between languages (e.g., Dahl, 2009; McWhorter, 2001). The assumption behind complexity invariance, namely that complexity or a lack thereof in one linguistic area is compensated for in another, does not hold up either. Morphological and syntactical complexity do not by default balance each other out, but tend to go hand in hand, especially when a language has case marking: As Dahl (2009) remarks, morphological distinctions between cases are only viable if there are syntactic rules fixing their distribution.

So, where do the classical languages stand in all this? In brief, Latin and Ancient Greek are without a doubt pretty complex. The number-one characteristic causing complexity is their extensive inflectional morphology – those familiar with either language need but think of the many noun and verb classes. By means of argumentation, the interested reader will find more information as well as other characteristics in Table 1. The overall complexity of Ancient Greek has actually been shown to be very significantly higher than that of modern Indo-European languages (Bentz et al., 2022, p. 13). The features that make the classical languages complex in the absolute sense simultaneously make them complex in the relative sense for learners, compared with Western first languages and with modern second languages commonly studied in Western secondary school systems.

Still to be discussed is the learning context. Besides their linguistic make-up, the way classical languages are taught sets them apart from their modern counterparts.⁵⁷ Pertinently, the main objective of Latin and Greek classes is not being able to communicate in everyday situations, but being able to read and analyse sophisticated literary texts. The combination of high structural complexity and the singular goal of building a reading knowledge explains why –

⁵⁷ In the controversy around the usefulness of classical language education, a recurring suggestion is whether youngsters would not be better off studying languages like Russian, Arab or Chinese instead. These languages can also be regarded as complex, both absolutely and relatively, so that instruction therein could hypothetically have comparable cognitive effects. Of all three the most eligible candidate for near transfer would probably be Russian, which as a Slavonic Indo-European language is still sufficiently similar to promote direct insight in Germanic and Romance languages and thus produce transfer effects to pupils' native language ability. In practice, however, these modern languages would never be taught in the same way as the classical languages; the focus would rightly lie on communication. With Latin and Ancient Greek it is a different story. Through the lens of the cognitive transfer framework, they are therefore not interchangeable. Furthermore, there are obvious cultural-historical arguments that can be adduced.

unlike in modern foreign language classes – the predominant teaching paradigm⁵⁸ is explicit instruction aimed at a passive knowledge of vocabulary and grammar.

In this particular learning context, pupils probably draw more heavily upon declarative rather than procedural memory. Thanks to neurocognitive research on second language acquisition (overview by Ullman, 2015), it is known that whereas implicit learning opportunities such as language immersion lead to more language learning in procedural memory, explicit classroom instruction encourages language learning in declarative memory. Classical language education as customarily practiced represents the explicit classroom setting *par excellence*: Pupils are confronted with no natural oral language use whatsoever, and are trained exclusively in the skill of reading. The exposure to classical texts, however, is too limited to allow for solid proceduralisation of linguistic knowledge or reading ability. Low second language exposure has indeed been found to correlate with learning in declarative memory as well (Ullman, 2015).

When studying classical languages, a lot of time and effort is directed towards the language system: rote learning of large amounts of vocabulary, memorizing many declensions and conjugations including exceptions, and getting a grip of countless grammatical rules and irregularities. Declarative memory is crucial for absorbing idiosyncratic information such as vocabulary or irregular morphological forms. In addition, and in competition with procedural memory, declarative memory is also available for learning grammatical rules and rule-governed forms. Coincidentally, for this it is chiefly used in the event of explicit classroom instruction or low exposure (Ullman, 2013; Ullman, 2015).

To summarise the training task, classical language education teaches pupils structurally complex languages (feature-related) with a certain distance to their mother tongue (learner-related) by means of a teaching method that appeals to acquisition in declarative memory (context-related). Unlike procedural knowledge, which is hyper-specialised, declarative knowledge is flexible, broadly available, consciously accessible, and utilisable in circumstances that only indirectly resemble the original conditions of learning (Squire & Wixted, 2011). Hence, declarative acquisition of a complex language system constitutes a learning task that is not unlikely to result in cognitive transfer; of course, at present, this is but a hypothesis

⁵⁸ Some teachers diverge from this paradigm and try to model classical language learning as much as possible after natural language acquisition. Although the so-called communicative method (Lloyd & Hunt, 2021) has gained popularity in recent years, in the whole of Europe it is quantitatively of minor significance.

regarding the underlying learning mechanisms.⁵⁹ The first point on the research agenda is to empirically establish whether any transfer effects of classical language learning do in fact occur.

Table 1 Structural complexity of the classical languages

Functional structural complexity	
▪ multiple form-function mappings	⇒ For example, inflectional morphology, where a single ending conveys multiple grammatical categories like case, gender and number at once. These fusional endings are more complex than features with a one-to-one mapping.
▪ opacity of a form-function mapping	⇒ For example, the variable functions of cases. The accusative for instance is ambiguous in that it can indicate both an object or an adverbial.
▪ communicative redundancy	⇒ For example, the Greek augment in particular and tense marking in general. Tense marking is redundant when other cues to the time frame are present.
▪ optionality of a feature	⇒ For example, null subjects. Subject expression is optional in Latin and Greek.
Formal structural complexity	
a. Morphological complexity	
▪ declensional and conjugational allomorphy (i.e. the encoding of distinctions between noun and verb classes)	⇒ For example, the numerous declension and conjugation types. Encoded allomorphy is a complexifying by-product of inflectional morphology.
▪ arbitrary allomorphy	⇒ For example, variation in the way the principal parts of the verb are formed. To illustrate, the Latin verb <i>monere</i> has perfect <i>monui</i> , whereas <i>maneo</i> has <i>mansi</i> . Arbitrary allomorphy adds another layer of complexity within the noun and verb classes.
▪ additional morphophonemics	⇒ For example, positional variants of stems. To illustrate, the Latin word for ‘name’ is <i>nomen</i> in the nominative, but uses the stem <i>nomin-</i> in other cases. Such morphophonological processes are likely to develop in inflecting languages, which creates an extra component of grammar to be learned.
▪ marking of agreement	⇒ For example, agreement of case, gender, number, and so on. Agreement can get particularly complex with embedded relative clauses or with noun phrase hyperbata,

⁵⁹ Some preliminary confirmation for this hypothesis can for example be seen in educational research on generic skills like critical thinking, which has shown that transfer of critical thinking skills is only achieved if they are taught explicitly within a meaningful domain-specific context (e.g., Tiruneh et al., 2018). The fact that teaching method is decisive for learning outcomes *tout court* is well-documented in both cognitive transfer research (e.g., Karbach & Kray, 2009) and second language acquisition research (e.g., Bagheri et al., 2019).

since a hyperbatic placement increases the number of intercurrent constituents between a linguistic unit and its nearest head or dependent, in itself a hallmark of complexity. Generally, agreement-marking inflection renders a grammar inherently more complex.

- number of constituent components of a feature
 - ⇒ For example, compound or circumscribed forms like *amatus sum*, *pugnaturus est*, *λελυκώς ὦ* (*lelukōs ō*) et cetera. They require multiple components for one paradigm slot, which is more complex than a single component.
- b. Syntactic complexity
 - asymmetries between matrix and subordinate clauses
 - ⇒ For example, the usage of mood and tense, notably the relative usage in Latin and the aspect in Greek. These characteristics augment the syntactic complexity by increasing the number of rules to process.
- c. Semantic complexity
 - fine subdivisions requiring overt specification
 - ⇒ For example, the obligatory expression of modality by verbal mood and tense. Overt and grammaticalized expression of such fine-grained semantic distinctions makes the grammar more complex.
 - semantically opaque or arbitrary case government
 - ⇒ For example, verbs governing different cases for their first object. Several possibilities mean more processing and thus more complexity.

Note. The categories are adapted from Housen, 2020; Housen & Simoens 2016; and McWhorter, 2001. The examples are our own.

3.4.2 The content factor

Following the transfer taxonomy (Barnett & Ceci, 2002), it must be specified what is transferred – also known as the content factor – and when and where it is transferred from and to – also known as the context factor. To recapitulate, the content factor encompasses the specificity of the skill learned during the training task, the memory demands of the transfer task, and the nature of the change in performance on the transfer task. The application of these last two dimensions is straightforward. Under the hypothesis that studying classical languages produces cognitive transfer effects, we expect that after having studied classical languages pupils will perform better, that is more accurate, on a set measurement (see paragraph on the context factor), without being at all prompted that the skills learned in Latin or Greek class could be of any help. Since subjects have to select and execute a task approach by themselves, the memory demands are recall, recognition and execution. The assessed performance change is accuracy, rather than speed or something else.

As for the skills learned by engaging in the study of classical languages, they can be formulated at different levels of specificity. An example of a language-specific transferrable skill

is deconstructing complex words and sentences to discover their meaning, which could come in handy while performing modern language tasks. The language-specific skills can also be reformulated and used at a more generic level: Examples are pattern recognition, learned for instance by looking for regularities in inflection, and analytical reasoning, learned for instance by parsing long sentences, which could be used while performing a wide range of problem-solving tasks. In case of a broadly defined training task like classical language education, the activated skill and its specificity will vary according to the transfer task.

3.4.3 The context factor

The dimensions of the context factor can be filled in as follows, starting with the knowledge domain. With regard to the training task, the knowledge domain is classics, in particular the Latin and/or Ancient Greek language. With regard to the transfer task, knowledge domains that have aroused interest as possible areas of transfer are: native language ability, modern foreign language ability, general cognitive ability, and school achievement. The modality is, primarily, written. The physical, temporal, functional, and social contexts are undetermined, in that they depend on the research study. In previous research (see Bracke & Bradshaw, 2020; Vereeck et al., 2025), the tasks mostly take place in the pupils' school environment during school hours. If the transfer task consists for instance of a standardised language test administered by a researcher or educator, it resembles regular school tests in its academic function and individual execution, the only difference being that the stakes are lower if the pupils' performance on this test does not count towards their grades.

Barnett and Ceci (2002) pointed out that the training and the transfer task can be near to each other or far apart along each of the dimensions of the context factor. When it comes to measuring the cognitive impact of classical language education, the largest differences exist on the knowledge domain dimension. Therefore, it makes sense to operationalise near and far transfer in relation hereto. Potential linguistic benefits of the study of classical languages – or in other words, crosslinguistic transfer to native language ability, modern foreign language ability, or achievement in language courses – should be labelled near transfer. Potential general-cognitive benefits – or in other words, transfer to general cognitive ability or achievement in other courses like mathematics and sciences – should be labelled far transfer.

3.4.4 The individual factor

Ultimately, cognitive transfer is a matter of human behaviour, hence the importance of the individual factor. For classical language pupils no less than for other research participants, characteristics that may have a confounding influence on the cognitive transfer process must be taken into account. Besides age, gender, socio-economic status, motivation, and personality, a key variable to consider is initial cognitive ability. Transfer can only be meaningfully evaluated if any additional effect from the training task can be isolated and distinguished from the effect of initial cognitive ability. Study options are not assigned randomly, so they are usually correlated with demographic and other individual-difference variables (Green et al., 2019). In other terms, there is a high risk of selection bias. Study options are also liable to prejudices and stereotypes, above all Latin and Greek, which is why measures are needed to minimise any Pygmalion effects and stereotype threat effects.

3.5 Supporting and related literatures

So far we have reframed one of the pivotal moot points in the classical language debate as a psychological issue of cognitive transfer. Given that the sheer existence of cognitive transfer is not a foregone conclusion, except perhaps for very near transfer as a result of working memory training, the theoretical framework is not quite complete yet. What basis is there to expect transfer effects of classical language learning? As laid out in this section, there is a number of research findings from diverse scientific fields on the basis of which such transfer effects seem, at the very least, plausible.

3.5.1 Near transfer / Linguistic benefits

The linguistic benefits have been extensively studied in the United States in the course of the twentieth century. Without going into more detail here, this research is outdated and has many methodological shortcomings and generalisability problems (Bracke & Bradshaw, 2020; Vereeck et al., 2025). It is a substantial body of research nonetheless, which may be taken as suggestive evidence. Numerous American studies found a positive effect of Latin instruction

on English word knowledge (see for example Masciantonio, 1977), with both Latin-derived words and words of Anglo-Saxon or other origin (e.g., Gilliland, 1922). English reading ability was found to be positively impacted by Latin instruction as well: One report on a second language programme even noted more progress and higher absolute levels after one year of Latin than after four years of French or Spanish (Mavrogenes, 1977, p. 270). The improvement in reading ability might in part be a consequence of the improvement in word knowledge, for it is a well-known fact in the fields of first language acquisition (e.g., Hacquebord, 2006) as well as second language acquisition (e.g., Hu & Nation, 2000) that vocabulary size and text comprehension are closely connected.

Furthermore, the possibility of near transfer from the classical languages to the native language is supported by second language acquisition research on bidirectional crosslinguistic transfer. While since its early days the field of second language acquisition has traditionally focused on unidirectional transfer from the first language to the second, in later decades attention increased for the influence of second language learning on first language proficiency. In her acclaimed multicompetence framework, Cook (1991) urged that people who know more than one language have a compound state of mind, distinct from the sum of two or more monolingual states.

Later, models of multilingualism have been proposed inspired by Complexity Theory (Larsen-Freeman, 2015, 2020), a modern transdisciplinary theory concerned with complex systems that is well-applicable to language and language acquisition. Their fundamental principle is that, just as so much in this world, language and language acquisition need to be seen as complex and dynamic, rather than static and stable. Being preoccupied with change and adaptation, Complexity Theory “inspires the thinking that a multilingual system expands the language resources from which a multilingual may draw, the use of one language affects the use of another, and thus the influence between languages is bidirectional” (Larsen-Freeman, 2015, p. 230).

Research on bidirectional crosslinguistic transfer endorses the idea that second language learning is a dynamic action with potential consequences for first language proficiency. In bilingual adults, transfer has been observed across all areas of their native language independently of context factors, which points towards a formal restructuring of linguistic competence (Pavlenko & Jarvis, 2002). Studies with small children show that learning a second language with a transparent correspondence between oral sounds and written letters, such as Italian, aids the development of native literacy in a language where this correspondence

is less transparent, such as English. Murphy et al. (2015, p. 1151) draw the conclusion that second language learning “can also be viewed as language awareness training: an appreciation for, and understanding of language as a system of sounds, words, and structure that can be manipulated in different ways”.

When reviewing the above results of second language acquisition research, we must keep in mind the differences and similarities with classical language learning. The main differences are that Latin and Ancient Greek are not used for oral or written production in secondary education,⁶⁰ and that pupils’ active knowledge is minimal. A similarity is that these languages do also have a high correspondence between oral sounds and written letters. Moreover, classical language education in particular qualifies as language awareness training: Its transfer value may very well lie in general language principles and the approach of language as a system. At any rate, it has been known for some time now that second language instruction is beneficial for first language performance (as already noted by Ganschow & Sparks, 1995).

3.5.2 Far transfer / General-cognitive benefits

Prior research on the general-cognitive benefits of studying classical languages is virtually limited to one experiment, also conducted in the United States some decades ago. In this experiment, Latin pupils made considerably more progress in not only English but in mathematics as well, compared to pupils with similar socio-economic and academic profiles who did not take Latin (Masciantonio, 1977, pp. 377–378; Mavrogenes, 1977, p. 270). Within the classic lab-based research on cognitive transfer, far transfer does not have a great track record and, as noted above, is still surrounded by scepticism. Even so, the suggestion that the study of classical languages could elevate general cognitive ability is supported by two other research lines. The broad notion that cognitive ability is susceptible to environmental factors is corroborated by psychological research on intelligence, and the more specific notion that second language learning boosts intelligence is corroborated by psycholinguistic research on the cognitive effects of multilingualism and second language acquisition, above all the bilingual advantage.

⁶⁰ That is, not taking into account the communicative method already mentioned in footnote 58.

Research on the malleability of intelligence has amassed a large amount of evidence that intelligence, as gauged by an IQ score, can and does change considerably depending on environmental factors like the opportunity to acquire or practice certain mental capabilities. This evidence comes from various sources: adoption studies, intervention studies, studies on schooling effects, studies on nutritional effects, studies on generational changes, and so on (Sauce & Matzel, 2018). A person's IQ scores may be relatively stable from one year to the next, but that does not mean intelligence is an entirely inherent and unalterable quality. In a meta-analysis of the effect of education on intelligence, Ritchie and Tucker-Drob (2018) discovered that IQ consistently increases with approximately one to five points per additional year of schooling. From there it is not a far stretch to hypothesise that particular types of education could have a more pronounced effect on cognitive ability than other types.

The bilingual advantage in cognitive functioning has been an avidly studied topic in psycholinguistics since in the sixties⁶¹ research demonstrated that bilingualism was not detrimental to the mind, as contemporary consensus had it, but rather an advantage in intellectual reasoning (Peal & Lambert, 1962). Modern research has frequently associated bilingualism with improved working memory and especially improved executive control (e.g., Bialystok et al., 2005; Bialystok & Barac, 2012; Carlson & Meltzoff, 2008). It must be said that, like with cognitive transfer, not all research supports the existence of a bilingual advantage in executive functioning (e.g., Dick et al., 2019; Paap, 2023). The first longitudinal study assessing the cognitive effect of becoming a bilingual within participants was conducted by Woumans et al. (2016). Using a matched comparison group design, they noted significant gains in non-verbal intelligence scores for the bilingual group, thus making a solid case for far transfer as a result of second language learning.

⁶¹ Long before that, the Russian child psychologist Lev Vygotsky (1896-1934) already proclaimed the beneficent nature of foreign language instruction (Fredericks, 1974). During his life Vygotsky's writings were not influential in the West, but later he became a famous name in the field of educational psychology for, among other things, his theory of the zone of proximal development.

3.5.3 Related topics

Besides classical language instruction, there are other kinds of instruction commonly thought to have a beneficial impact that stretches beyond the training domain, in particular instruction in music, chess, and programming. Despite the resemblance, only the latter three seem to occupy an established place within the psychological cognitive transfer literature, whereas the classical languages do not. On a side note, Green et al. (2019) in fact see a trichotomy in the literature: (1) cognitive training in the strictest sense, (2) related behavioural interventions (among which musical interventions and the like) that are on the radar of cognitive enhancement research, and (3) other relevant contexts that could come up in educational or clinical research.

Music training has been associated with enhanced general cognitive ability, spatial ability, executive control and various language skills, and chess training has been linked to improvement on general-cognitive measures among which fluid intelligence, processing speed and working memory (Hille et al., 2011; Janus et al., 2016; Sala & Gobet, 2017). Just like research on the classical languages, however, research on music and chess has suffered from methodological shortcomings such as the lack of control for selection bias and the absence of active comparison groups, which prevent a definitive conclusion about their true transfer value (Sala & Gobet, 2017). Computer programming, on the other hand, has proven transfer effects to creative thinking, mathematical skills, metacognition, spatial skills and reasoning (Scherer et al., 2019).

3.6 Conclusion

One could say psychology is deemed a hub science because of its high transfer value: The theories, methods, and findings have a profound impact outside of the realm of psychology itself. As this article illustrates, the adoption of a psychological approach can shed new light on scholarly problems originating from different disciplines. Moreover, it can even be the driving force towards resolving them. The old American research on classical language education lacked a theoretical framework. In general, the supposed secondary benefits of studying classical languages have not been approached as fundamentally a phenomenon of

human cognition and language learning, and the influence of participant characteristics and expectancy effects has largely been overlooked. The framework presented here defined classical language learning as a specific type of cognitive training, akin to other research topics such as crosslinguistic transfer, the bilingual advantage, and music education. We also discussed the various factors that can affect the possible transfer resulting from this training. These insights ought to be translated into the designs of future empirical research on the topic.⁶²

The exploration of new topics can also yield worthwhile expansions of existing psychological literatures. Just like psychology contributes to classical language impact research, the reverse is true too. In their methodological manifesto, Green and his assembly of scientists (2019) remark that the link between lab tests and real-world outcomes is not always clear, and that studies on the effectiveness of real-world behavioural interventions for cognitive enhancement are rare. Since so much of the classic research on cognitive transfer has relied solely on lab-based experiments, field data from an already existing and heavily discussed educational situation will be an invaluable contribution to this branch of research. Furthermore, the focus on classical languages will be an interesting addition to the literature concerning the bilingual advantage, as well to the field of second language acquisition.

Works cited in this chapter

- Bagheri, M., Hadian, B., & Vaez-Dalili, M. (2019). Effects of the Vaughan method in comparison with the Audiolingual method and the communicative language teaching on Iranian advanced EFL learners' speaking skill. *International Journal of Instruction*, 12(2), 81–99.
- Barnett, S., & Ceci, S. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637.
- Becerra-Bonache, L., & Jiménez-López, D. (2015). A grammatical inference model for measuring language complexity. In I. Rojas, G. Joya, & A. Catala (Eds.), *Advances in computational intelligence*. 13th International Work-Conference on Artificial Neural Networks, IWANN 2015, Palma de Mallorca, Spain, June 10-12, 2015 (Vol. 1). Springer.

⁶² For the first time in the history of the educational debate about the cognitive impact of classical language education, we are applying psychological research paradigms regarding the cognitive effects of multilingualism and second language acquisition to classical language learning in our ongoing research at Ghent University. This research is funded by The Research Foundation – Flanders (FWO), grants 1110621N and G049821N.

- Bentz, C., Gutierrez-Vasques, X., Sozinova, O., & Samardžić, T. (2022). Complexity trade-offs and equi-complexity in natural languages: A meta-analysis. *Linguistics Vanguard: A Multimodal Journal for the Language Sciences*, 14(9), 9–25.
- Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, 122(1), 67–73.
- Bialystok, E., Martin, M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, 9(1), 103–119.
- 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.
- Carlson, S., & Meltzoff, A. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, 11(2), 282–298.
- Castiello, J. (1934). *Geistesformung. Beiträge zur experimentellen erforschung der formalen Bildung*. Dümmler.
- Cook, V. (1991). *Second language learning and language learning*. Edward Arnold.
- Dahl, Ö. (2009). Testing the assumption of complexity invariance: the case of Elfdalian and Swedish. In D. Gil, G. Sampson, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 50–63). Oxford University Press.
- De Bruyckere, P., Kirschner, P., & Hulshof, C. (2020). If you learn A, will you be better able to learn B? Understanding transfer of learning. *American Educator*, 44(1), 30–34.
- Dick, A., Garcia, N., Pruden, S., Thompson, W., Hawes, S., Sutherland, M., Riedel, M., Laird, A., & Gonzalez, R. (2019). No evidence for a bilingual executive function advantage in the ABCD study. *Nature Human Behaviour*, 3(7), 692–701.
- Fredericks, S. (1974). Vygotsky on language skills. *The Classical World*, 67, 283–290.
- Ganschow, L., & Sparks, R. (1995). Effects of direct instruction in Spanish phonology on the native-language skills and foreign-language aptitude of at-risk foreign language learners. *Journal of Learning Disabilities*, 28(2), 107–120.
- Gil, D., Sampson, G., & Trudgill, P. (Eds.). (2009). *Language complexity as an evolving variable. Studies in the evolution of language*. Oxford University Press.
- Gilliland, A. (1922). The effect of the study of Latin on ability to define words. *Journal of Educational Psychology*, 13(8), 501–505.
- Gobet, F., & Sala, G. (2022). Cognitive training: A field in search of a phenomenon. *Perspectives on Psychological Science*, 20(10), 1–17.
- Green, C., Bavelier, D., Kramer, A., Vinogradov, S., Ansorge, U., Ball, K., Bingel, U., Chein, J., Colzato, L., Edwards, J., Facoetti, A., Gazzaley, A., Gathercole, S., Ghisletta, P., Gori, S., Granic, I., Hillman, C., Hommel, B., Jaeggi, S., ... Witt, C. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, 3(1), 2–29.
- Hacquebord, H. (2006). Woordkennis als onderdeel van taaldiagnostisch onderzoek. *Levende talen*, 7(1), 15–22.
- Harrison, T., Shipstead, Z., Hicks, K., Hambrick, D., Redick, T., & Engle, R. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, 24(12), 2409–2419.
- Hille, K., Gust, K., Bitz, U., & Kammer, T. (2011). Associations between music education, intelligence, and spelling ability in elementary school. *Advances in Cognitive Psychology*, 7, 1–6.
- Housen, A. (2020). Difficulty and complexity of language features and second language instruction. In C. Chapelle (Ed.), *The concise encyclopedia of applied linguistics* (pp. 389–395). Wiley Blackwell.
- Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, 38, 163–175.
- Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430.
- Jaeggi, S., Buschkuhl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, 42(3), 464–480.
- Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *Journal of Experimental Child Psychology*, 144, 84–97.

- Joseph, J., & Newmeyer, F. (2012). All languages are equally complex. The rise and fall of a consensus. *Historiographia Linguistica*, 39(2–3), 341–368.
- Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, 12(6), 978–990.
- Larsen-Freeman, D. (2015). Complexity theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 227–244). Routledge.
- Larsen-Freeman, D. (2020). Chaos/complexity theory for second language acquisition/development. In C. Chapelle (Ed.), *The concise encyclopedia of applied linguistics* (pp. 181–188). Wiley Blackwell.
- Lloyd, M. & Hunt, S. (2021). *Communicative approaches for ancient languages*. Bloomsbury Academic.
- Masciantonio, R. (1977). Tangible benefits of the study of Latin: A review of research. *Foreign Language Annals*, 10(4), 375–382.
- Mavrogenes, N. (1977). The effect of elementary latin instruction on language arts performance. *The Elementary School Journal*, 77(4), 268–73.
- McWhorter, J. (2001). The world simplest grammars are creole grammars. *Linguistic Typology*, 5(2–3), 125–166.
- Murphy, V., Macaro, E., Alba, S., & Cipolla, C. (2015). The influence of learning a second language in primary school on developing first language literacy skills. *Applied Psycholinguistics*, 36(5), 1133–1153.
- Paap, K. (2023). *The bilingual advantage in executive functioning hypothesis: How the debate provides insight into psychology's replication crisis*. Routledge.
- Pavlenko, A., & Jarvis, S. (2002). Bidirectional transfer. *Applied Linguistics*, 23(2), 190–214.
- Peal, E., & Lambert, W. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied*, 76, 1–23.
- Pienemann, M., & Lenzing, A. (2015). Processability theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 159–179). Routledge.
- Ritchie, S., & Tucker-Drob, E. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, 29(8), 1358–1369.
- Sala, G., Aksayli, N., Tatlidil, S., Tatsumi, T., Gondo, Y., & Gobet, F. (2019). Near and far transfer in cognitive training: A second-order meta-analysis. *Collabra: Psychology*, 5(1), no. 18.
- Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science*, 26(5), 515–520.
- Sampson, G. (2009). A linguistic axiom challenged. In D. Gil, G. Sampson, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 1–18). Oxford University Press.
- Sauce, B., & Matzel, L. (2018). The paradox of intelligence: Heritability and malleability coexist in hidden gene- environment interplay. *Psychological Bulletin*, 144(1), 26–47.
- Scherer, R., Siddiq, F., & Sánchez Viveros, B. (2019). The cognitive benefits of learning computer programming: A meta-analysis of transfer effects. *Journal of Educational Psychology*, 111(5), 764–792.
- Shipstead, Z., Redick, T., & Engle, R. (2010). Does working memory training generalize? *Psychologica Belgica*, 50(3–4), 245–276.
- Spearman, C. (1904). 'General intelligence', objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–292.
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. The Macmillan Company.
- Squire, L., & Wixted, J. (2011). The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience*, 34, 259–288.
- Stellwag, H. (1949). *De waarde der klassieke vorming. Een cultuur-historische, paedagogisch-psychologische en didaktische inleiding*. Wolters.
- Thorndike, E. (1914). *The psychology of learning*. Teachers College.
- Thorndike, E., & Woodworth, R. (1901a). The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review*, 8(3), 247–261.
- Thorndike, E., & Woodworth, R. (1901b). The influence of improvement in one mental function upon the efficiency of other functions. II. The estimation of magnitudes. *Psychological Review*, 8(4), 384–395.

- Thorndike, E., & Woodworth, R. (1901c). The influence of improvement in one mental function upon the efficiency of other functions: III. Functions involving attention, observation and discrimination. *Psychological Review*, 8(6), 553–564.
- Tiruneh, D., Gu, X., De Cock, M., & Elen, J. (2018). Systematic design of domain-specific instruction on near and far transfer of critical thinking skills. *International Journal of Educational Research*, 87, 1–11.
- Ullman, M. (2013). Declarative/procedural model of language. In H. Pashler (Ed.), *Encyclopedia of the mind* (Vol. 1, pp. 224–226). Sage Publications.
- Ullman, M. (2015). The declarative/procedural model. A neurobiologically motivated theory of first and second language. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 135–158). Routledge.
- 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*, 25(50), 101–115.
- Vereeck, A., Janse, M., De Herdt, K., Hauspie, C., Szmalec, A., & Woumans, E. (2025). Does studying Latin make pupils smarter? Presenting the field of classical language impact studies. *Classical Journal*, 120(4), 479–519.
- Vereeck, A., Janse, M., De Herdt, K., Szmalec, A., Hauspie, C., & Duyck, W. (2022). A debate in need of data: Report of an empirical study on the cognitive transfer effects of studying classical languages [Conference presentation abstract]. In M. Jurišević (Ed.), *Book of Abstracts: 17th European Congress of Psychology. Horizons of Psychology*, 31, 59–447.
- von Bastian, C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: A review. *Psychological Research*, 78(6), 803–820.
- Woumans, E., Surmont, J., Struys, E., & Duyck, W. (2016). The longitudinal effect of bilingual immersion schooling on cognitive control and intelligence. *Language Learning*, 66(2), 76–91.

Addenda to Chapter 3

Current trends in cognitive transfer research

Research on cognitive transfer has continued to develop and expand since the conceptualisation and the publication of the article above. These are some current relevant trends.

- (1) There has been a growing interest in individual differences. Two theories compete to explain the role of baseline cognition: the magnification theory, according to which higher ability individuals get the most benefit from cognitive training (e.g., Ørskov et al., 2021), and the compensation theory, according to which lower ability individuals get the most benefit from cognitive training (e.g., Traut et al., 2021).
- (2) Advances have also been made in our understanding of the neural correlates of cognitive transfer (e.g., Ericson et al., 2024 on near transfer from one visuo-spatial working memory task to another).
- (3) Further issues with existing meta-analyses have been brought to light by Sandoval-Lentisco and colleagues (2024). In their meta-review of 97 meta-analyses on cognitive training, they criticise various aspects of methodology and reporting style, among which the lack of assessment of bias risk, the inadequate handling of effect sizes, the non-disclosure of deviations from a pre-registered protocol, and the intransparency about key meta-analytical decisions in general.
- (4) The existence of a bilingual advantage is increasingly being called into question. A systematic review by Planckaert and colleagues (2023) highlights the many conflicting results that have been produced over the years. The authors suggest that learning a second language might be advantageous for cognitive control, but only during the sensitive periods of cognitive development between the age of three and six, and of cognitive decline later in life. Moreover, after the age of six, monolingual children tend to catch up again with bilingual children.

Bialystok (2024) has proposed an alternative explanation for bilingual effects on executive functioning: Linguistic cognitive processes do not really “transfer” to non-linguistic cognitive processes, but the constant demand on attention by multiple

languages being jointly activated in the brain causes bilinguals to apply their attention more effectively.

- (5) The field examining the cognitive impact of music training has evolved in the opposite direction: There is now increasing support for far transfer effects. The data of Sala and Gobet's (2020) meta-analysis that concentrated on music training have been re-analysed by Bigand and Tillmann (2022). The latter addressed some methodological issues in the original meta-analysis and found evidence for a small but robust effect of music training to general cognitive ability. Subsequent studies have come to similar conclusions (e.g., Jamey et al., 2024 on inhibition control).

Works cited in these addenda

- Bialystok, E. (2024). Bilingualism modifies cognition through adaptation, not transfer. *Trends in Cognitive Sciences*, 28(11), 987–997.
- Bigand, E., & Tillmann, B. (2022). Near and far transfer: Is music special? *Memory & Cognition*, 50, 339–347.
- Ericson, J., Palva, S., Palva, M., & Klingberg, T. (2024). Strengthening of alpha synchronization is a neural correlate of cognitive transfer. *Cerebral Cortex*, 34(2), no. bhad527.
- Jamey, K., Foster, N., Hyde, K., & Dalla Bella, S. (2024). Does music training improve inhibition control in children? A systematic review and meta-analysis. *Cognition*, 252, no. 105913.
- Ørskov, P., Norup, A., Beatty, E., & Jaeggi, S. (2021). Exploring individual differences as predictors of performance change during Dual-N-Back training. *Journal of Cognitive Enhancement: Towards the Integration of Theory and Practice*, 5(4), 480–498.
- Planckaert, N., Duyck, W., & Woumans, E. (2023). Is there a cognitive advantage in inhibition and switching for bilingual children? A systematic review. *Frontiers in Psychology*, 14, no. 1191816.
- Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, 48(8), 1429–1441.
- Sandoval-Lentisco, A., López-Nicolás, R., Tortajada, M., López-López, J., & Sánchez-Meca, J. (2024). Transparency in cognitive training meta-analyses: A meta-review. *Neuropsychology Review*.
- Traut, H., Guild, R., & Munakata, Y. (2021). Why does cognitive training yield inconsistent benefits? A meta-analysis of individual differences in baseline cognitive abilities and training outcomes. *Frontiers in Psychology*, 12, no. 662139.

life must be understood backwards
~ Søren Kierkegaard, Journals JJ:167

Part II: Methodological Preamble

Introductory remarks to Part II

Relation to Parts I and III

In the first part of the dissertation, the status quaestionis, the topic of the cognitive impact of studying classical languages was theoretically and historically examined from various angles. In the third part of the dissertation, the research report, the phenomenon of the cognitive impact of studying classical languages will be empirically studied through newly gathered field data. The function of the second part is to elaborate on some aspects of the methodology of the longitudinal study that warrant a detailed discussion. This part takes a practical approach to the matter at hand, showcasing how three central concepts from the research objectives and questions (cf. General Introduction), namely Dutch language ability, general cognitive ability and school pupils, are translated into research realities.

Dutch language ability and general cognitive ability, on the one hand, are abstract, so-called latent constructs that have to be estimated through observed test scores, hence requiring suitable measuring instruments. School pupils, on the other hand, are very concrete, but engaging them as research participants is not as straightforward as it may seem at first glance; successful data collection in schools is necessarily preceded by an intensive communicative and logistical process. To phrase it differently, Part II concerns the choice of linguistic and general-cognitive measuring instruments for the longitudinal study, and the underestimated art of doing school-based research. The section hereafter briefly recounts the instrument selection process. Chapter 4 zooms in on the Dutch language test. Chapter 5 is about the validation of a general-cognitive test that ultimately did not make it into the test battery. Finally, Chapter 6 is the product of my methodological reflections on school-based research.

Instrument selection process

Suitable measuring instruments were unfortunately not readily available. The composition of the test battery therefore took up the majority of the first year of the doctoral fellowship; an internal report of this process is given in Appendix 1. No instrument could be found that

measured Dutch language ability broadly in the study's target population. Consequently, we developed a language test ourselves, by borrowing items from several sources. Test development ran largely parallel to the research on classical languages. The scope and structure of this test are discussed in full in Chapter 4.

For the measurement of general cognitive ability, the initial choice fell on the latest edition of the Wechsler Intelligence Scale for Children, or WISC for short. The WISC had the advantage of being an internationally well-known and often-used IQ-test with positive reviews in Flanders and abroad, but it also presented a number of disadvantages, above all the high price of materials and the need for individual administration. Two other options surfaced, namely ICAR and COVAT-3. All three options were thoroughly explored (see also Appendix 1), in light of which preparations were made for a pilot study with ICAR items. Having considered the alternatives, eventually the COVAT test was preferred for the following reasons. Contrary to the WISC, the COVAT-3 allowed for digital group administration and there was no financial cost. Contrary to the ICAR, the COVAT-3 was to a much larger degree a finished product, so that it would not be necessary to fit in a pilot study before the start of the longitudinal study, and it was also founded in intelligence theory and tailored specifically to youngsters.

The COVAT-3 is part of a family of COVAT tests that were developed at Thomas More University of Applied Sciences, inspired by the authoritative CHC model of intelligence.⁶³ The original CoVaT-CHC Basic Version is a paper & pencil test, on the basis of which the digital COVAT-3 was later created. The COVAT-3 is still partially in progress, for example, norm scores are not available yet. In exchange for free access to the COVAT-3 testing platform, it was formally agreed that I would share my pseudonymised language and COVAT data with the Thomas More team with an eye to further refinement of their test. The ICAR, CoVaT-CHC Basic Version, and CHC model of intelligence all feature in Chapter 5, which is dedicated to the ICAR test for children that was conceived during the abovementioned ICAR pilot study. To the COVAT-3 we return in Section 8.1.3.2 of Part III.

⁶³ In a nutshell, this model discerns several broad cognitive abilities, among which fluid intelligence (reasoning, problem solving) and crystallised intelligence (learned procedures and knowledge); these abilities are distinct from each other to some extent, but at the same time they are all influenced by an overarching general intelligence factor or g-factor.

Chapter 4: Dutch Language Test

This chapter as such is not intended for publication.

Chapter 4

Dutch Language Test

4.1 Introduction

The Dutch language test was a continuous work in progress from 2020 to 2024. Developing an instrument whilst already employing it to collect substantive data is far from ideal, but it was the only option, as there did not seem to exist a Dutch language ability test for non-clinical populations between the ages of 12 and 16. The selection and scoring of items as well as their division in subtests were changed and refined many times over the years. This chapter provides some background information about the final instrument, with the goal of clarifying the linguistic variables that feature in the analyses of Part III. Section 4.2 gives a concise account of the creation process of the test as it was used for data collection in the longitudinal study, and Section 4.3 describes the principal adaptations that were made later, using these three years of data. Many more details can be found in Appendices 1 to 8. Consultation of these appendices is recommended, as they might make the information in this chapter more meaningful to the reader, especially Appendix 3 with the items and Appendix 4 about the grouping of the items.

4.2 Building a test for data collection

After extensive heuristics (the findings of which are summarised in Appendix 1), five promising-sounding tests were identified. At least parts of these five tests seemed potentially

suitable for digital group administration in general secondary education, and together, they covered a nice variety of language aspects.

- (1) A receptive vocabulary test developed at the Ghent Center for Reading Research (Doreleijers & van der Sijs, 2019; Vander Beken et al., 2018). Although the full test is aimed at advanced language users such as university students, it also contains a dozen of easier items of which the difficulty levels (calculated as the percentage of correct answers) lay between 0.9 and 1 in the original student sample of Vander Beken and colleagues (2018).
- (2) A Dutch language test developed at Taalcentrum-VU (van Beersum & van Straalen, 2010), which VU Amsterdam gives to incoming students in order to detect students with insufficient language skills. The item pool is updated regularly, among other reasons to avoid questions leaking out, and so I was able to obtain some ninety recently removed items from diverse categories.
- (3) The LUCI-test (Leuven Universitair Competentie Instrument) developed at the Leuven Language Institute as part of the university's TaalVaST project (De Wachter & Heeren, 2013). Inspired by the policy of VU Amsterdam, KU Leuven wished to introduce their own Dutch language test for incoming students. However, the resulting LUCI-test also puts a lot of emphasis on specifically academic skills, such as reading and analysing scientific texts or adopting the right study habits.
- (4) The *Spellingtest Secundair Onderwijs* developed at Artevelde University of Applied Sciences (Van Vreckem & Desoete, 2023b). Around the start of my longitudinal study, the test had not yet been finalised, but the authors were kind enough to send me their provisional testing manual as well as additional psychometric analyses. Since the test is primarily intended as a diagnostic instrument to screen pupils in lower secondary education for language deficiencies, however, the difficulty level of most of these materials is much too low to be discriminating at the higher end of performance, which is needed to register possible transfer effects of studying classical languages.
- (5) The *Begrijpend Leestest Secundair Onderwijs* developed at Artevelde University of Applied Sciences (Van Vreckem & Desoete, 2023a). Idem as the previous test, but for reading comprehension instead of spelling.

From the sources above, ca. 100 items from 13 different subtests were selected for further consideration, on the basis of their estimated suitability for the target audience and representativeness of general language ability. The selected items fell into five language aspects:

Vocabulary, Grammar, Spelling, Text structure and Reading comprehension. Items were divided over five parallel tests in a way that every test contained several language aspects; test no.1 is given in Appendix 2 as an example. A pencil & paper version of each of these parallel tests was piloted in a school that was not part of the main sample for the longitudinal study.⁶⁴ On average, there were 85 participants per school and hence per test, consisting of latinists and non-latinists from the second and fourth year of general secondary education. This set-up allowed to assess the functioning of the items in both second-year and fourth-year pupils, which would be the age range of the longitudinal study. Particular attention was paid to the items' difficulty levels per language aspect, because in order to create a sensitive instrument it is desirable to have a good distribution of easy, medium and hard items for each construct (Brysbaert, 2024).

Following analysis of the try-out data, 9 subtests were retained and the total number of items was reduced to 69. Appendix 3 lists all of the test items, including the instructions, answer options and correct answers. To enable digital group administration, the test was programmed in LimeSurvey, a web-based survey platform. No time restrictions or answering obligations were imposed. The LimeSurvey version counted 10 exercises (due to slight variations in the instructions of one of the subtests) and 62 questions (due to some of the items being programmed as subquestions). Appendix 4 offers a diagram of how exactly the items were grouped into questions, exercises, subtests, language aspects, and LimeSurvey screens.⁶⁵ In this way, the test was built that would subsequently be administered three years in a row to collect data on Dutch language ability as part of the main study on the cognitive effects of studying classical languages. It should be stressed once more that I do not own any of the test items. Via personal communication, the authors of the items gave me permission to incorporate them in this research.

⁶⁴ A few months into the longitudinal study, one school from the main sample resigned. This school was eventually replaced by one of the schools that helped pilot the language test. However, there was no overlap in participants, since at the time of the try-out the cohort that was to be followed up longitudinally was still in the first year of secondary education.

⁶⁵ Note that in Appendices 3 and 4 the sentence spelling items have been split up (as explained in Section 4.3), so that there are 84 items listed instead of 69.

4.3 Revising the test for data analysis

In the course of processing the language data that came in during the various data collection rounds, three types of problems surfaced.

- (1) There were multiple issues with the automatic scoring of answers as it had originally been programmed into LimeSurvey. Among other things, missing values were incorrectly scored as 0, and open-ended questions (where participants had to respond by writing words rather than selecting their answer from a multiple-choice list) regularly gave rise to inaccurate scores.
- (2) The 3 sentence spelling items deviated from the rest, in the sense that multiple points were awarded depending on the number of words in the sentence that were evaluated (between 5 and 7), whereas the 66 other items were binary (0 = false, 1 = correct).
- (3) The small number of items per subtest resulted in the internal scale consistencies being subpar.

The first two problems were eliminated by drawing up an entire new scoring protocol, which was applied to the raw data using a combination of SPSS Syntax, R code, and manual scoring. The sentence spelling items were split up, so that each evaluated word now became a separate item, bringing the total number of items to 84. The scoring protocol is given in Appendix 5.

To solve the third problem, a more thorough revision of the test itself was needed. Its original structure of 9 subtests / variables looked like this:

- Vocabulary without context (15 items);
- Vocabulary in context (5 items);
- Prepositions (5 items);
- Word transformations (6 items);
- Spelling of verbs (17 items);
- Spelling of sentences (3 → 18 items);
- Paragraph structure (3 items);
- Paragraph boundaries (3 items);
- Reading (12 items).

Item Response Theory analyses were carried out on the data of the full samples of each of the three measuring moments (cf. Section 8.1.2.3), using the *mirt* package in R. In the words of Hedge, Powell and Sumner (2018), “[t]he goal of IRT is to characterize the relationship

between typically a single latent trait (...) and the probability of a binary response (...) on individual test items”. Solutions with one to five factors were explored. A three-factor solution produced the best interpretable factor scores, and aligned reasonably well with the scree plots (see Figure 4). While the *AIC* fit measure preferred to have more factors, the *BIC* fit measure also favoured a model with three factors (see Table 2). Factor 2 coincided with the verb spelling subtest, and factor 3 with the sentence spelling subtest. Items from the other subtests predominantly loaded onto factor 1. All factor loadings are given in Appendix 6.

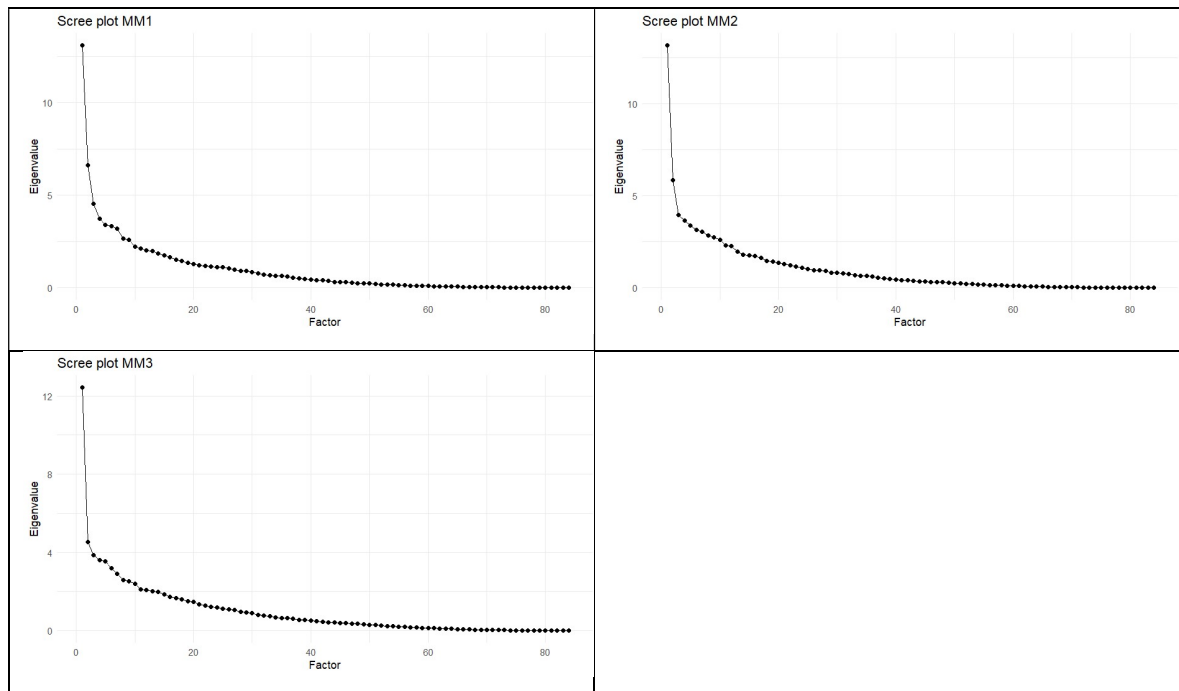


Figure 4 Scree plots for the language data of measuring moments 1 (upper left), 2 (upper right) and 3 (below)

Table 2 Fit measures of different IRT models for the language data of all three measuring moments

Factors	T1		T2		T3	
	<i>AIC</i>	<i>BIC</i>	<i>AIC</i>	<i>BIC</i>	<i>AIC</i>	<i>BIC</i>
1	87006.78	87817.46	58405.27	59155.06	45409.73	46123.72
2	85660.20	86871.39	57369.77	58489.99	44869.37	45936.11
3	85256.84	86863.71	57156.20	58642.39	44672.29	46087.53
4	85203.70	87201.44	57087.27	58934.96	44614.84	46374.33
5	85165.46	87549.24	57063.35	59268.09	44631.59	46731.08

Once the test's overall structure had been simplified from 9 subtests to 3 factors, the next step was to decide which items to preserve and which to discard. Weighing up the items' factor loadings, the items' contributions to the internal consistency of the factor scale, and the distribution of difficulty levels per factor scale, it was eventually decided to discard the following 24 items: Q1, Q8, Q13, Q14, Q20, Q24, Q30_SQ301, Q34_SQ342, Q35_SQ352, Q37_SQ371, Q39_SQ391, Q40_SQ401, Q41_SQ411, Q42_SQ423, Q42_SQ424, Q43_SQ432, Q44_SQ443, Q45, Q46, Q48, Q50, Q54, Q57, Q61. The factor structure (calculated using the *mirt* package in R) and difficulty levels (calculated in SPSS) of the preserved items are given in Appendix 7 and Appendix 8 respectively. The end result of these revisions were the following three scales.

- (1) **The lexical scale.** The first scale combines items from the Vocabulary subtests (i.e. Vocabulary without context and Vocabulary in context), the Grammar subtests (i.e. Prepositions and Word transformations), the Text structure subtests (i.e. Paragraph structure and Paragraph boundaries), and the Reading comprehension subtest (i.e. Reading). Upon closer inspection, this is not at all surprising. The so-called Grammar subtests in fact depended heavily on word knowledge. The Text structure subtests bear affinity to the Reading comprehension subtest, and the connection between vocabulary and reading skills is well-established (e.g., Hacquebord, 2006; Wagner & Meros, 2010). 36 out of 49 items were preserved. For the three measurements, Cronbach's α respectively equals .77, .81, and .82. There is a good distribution of difficulty levels across the spectrum of 0 to 1.
- (2) **The verbal scale.** The second scale is derived from the Spelling of verbs subtest, in which verbs have to be conjugated so that they fit the sentence context. 10 out of 17 items were preserved. For the three measurements, Cronbach's α respectively equals .79, .81, and .78. The distribution of difficulty levels is rather limited, ranging only between .44 and .75 for the first measurement and between .61 and .82 for the third measurement.
- (3) **The orthographical scale.** The third scale is derived from the Spelling of sentences subtest, in which aurally presented sentences have to be written down. 14 out of 18 items were preserved. For the three measurements, Cronbach's α respectively equals .73, .69, and .69. The distribution of difficulty levels is problematic, seeing that pronounced ceiling effects emerge. To illustrate; during the third measuring moment

13 of the 14 items were answered correctly by more than 70% of participants, of which 5 by even more than 85%.

In the analyses of Part III, orthography was dropped because of the ceiling effects observed with the third scale. The other two scales are represented by the variables Lexicon and Verbs, which express the percentage of correct answers as a number between 0 and 100.

Works cited in this chapter

- Brysbaert, M. (2024). Designing and evaluating tasks to measure individual differences in experimental psychology: A tutorial. *Cognitive Research: Principles and Implications*, 9, no. 11.
- De Wachter, L., & Heeren, J. (2013). Een taaltest als signaal. De ontwikkeling en implementatie van een strategische taalvaardigheidstoets. *Levende talen*, 14(1), 19–27.
- Doreleijers, K., & van der Sijs, N. (2019). *Vertrokken Nederlands. Pilotonderzoek naar de Nederlandse taal en cultuur in den vreemde*. Research report under the authority of the Taalunie and the Meertens Instituut.
- Hacquebord, H. (2006). Woordkennis als onderdeel van taaldiagnostisch onderzoek. *Levende talen*, 7(1), 15–22.
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166–1186.
- van Beersum, M., & van Straalen, E. (2010). Verplicht toetsen en bijspijkeren of eigen verantwoordelijkheid? De basisvaardigheden Nederlands van eerstejaars VU-studenten. In S. Vanhooren & A. Mottart (Eds.), *Vierentwintigste conferentie Het Schoolvak Nederlands* (pp. 211–216). Academia Press.
- Van Vreckem, C., & Desoete, A. (2023a). *Test: Begrijpend lezen. Voor het eerste tot het vierde middelbaar*. Academia Press.
- Van Vreckem, C., & Desoete, A. (2023b). *Test: Spellen. Voor het eerste tot het vierde middelbaar*. Academia Press.
- 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.
- Wagner, R., & Meros, D. (2010). Vocabulary and reading comprehension: Direct, indirect, and reciprocal influences. *Focus on Exceptional Children*, no. 1G1-245473138.

Chapter 5: Validation of Ch-ICAR

This chapter is based on:

Dutry, M., Vereeck, A., Duyck, W., Deros, E., Schelfhout, S., Szmalec, A., Woumans, E., Schittekatte, M., Debeer, D., & Dirix, N. (2025). Validation of the Children's International Cognitive Ability Resource (Ch-ICAR). *Behavior Research Methods*, 57(2), no. 66.
<https://doi.org/10.3758/s13428-024-02591-1>

Chapter 5

Validation of Ch-ICAR

Introductory remarks to Chapter 5

Admittedly, this chapter on the validation of the Children's ICAR is an odd one out, as there is no direct link with the central theme of the cognitive effects of studying classical languages. I preferred to include this chapter nonetheless, since the work on ICAR formed an integral part of my PhD, and I envisioned this dissertation as an accurate reflection of the entire research process, rather than merely the purified and polished end product of this process.

My involvement in this test development project was as follows. Fellow PhD student Merel Dutry and I both needed a general-cognitive measuring instrument for our respective school-based research studies. We were introduced to ICAR, which stands for International Cognitive Ability Resource: an online item database that researchers can freely make use of to design tests according to their own needs. This resource counters some of the practical problems associated with proprietary licensed intelligence tests like Wechsler's, such as administration duration and financial cost. However, there were some issues that needed clarification, in particular the appropriateness of ICAR items for use in non-adult populations.

Dividing the workload equally among us, Merel and I conducted a literature review, selected test items, and ran a pilot study. Whereas Merel eventually continued with ICAR in her main research, my supervisors and I decided it would be more opportune for me not to use another self-designed test in my study in addition to the Dutch language test, but to opt for the already existing and validated test that was the COVAT-3. Merel then led the following validation studies. Study 1 assessed the basic psychometric qualities of the instrument we would later name Ch-ICAR (for Children's ICAR), and study 2 examined its construct validity through

cross validation with two other cognitive tests, the CoVaT-CHC Basic Version and Raven's 2 Progressive Matrices. I withdrew from further data collection and analysis, apart from some assistance with scoring and processing the paper & pencil test forms of the CoVaT-CHC Basic Version.

Meanwhile, I did keep abreast of her results, and we started co-writing the paper that recently appeared in *Behavior Research Methods*. Merel drafted the sections on the pilot study, study 1 and study 2. I took responsibility for the global introduction and the precursory sections on the literature review and the initial selection of items; additionally, I reviewed and edited the entire manuscript multiple times throughout the writing and publication process. Up until the first round of peer-review, we went over all feedback we received together and consulted about how to incorporate it. The precursory sections ended up having to be drastically abridged in the final draft due to space limitations. In this chapter, I will present a more elaborate version of the first part of the article, based on earlier drafts of mine, followed by summaries of the pilot study, study 1 and study 2. The conclusion is copied from the published article.

5.1 Introduction

“[L]ife is an intelligence test, and the importance of measuring intellectual ability as people respond to life's challenges cannot be overstated.” (Dworak et al., 2021, p. 1).

Cognitive ability refers to the capacity to acquire, process, manipulate, organise, and apply knowledge effectively (Gottfredson, 1997). It is widely recognised to reliably predict important outcomes such as academic achievement (Roth et al., 2015; Tikhomirova et al., 2020; Vilia et al., 2017) and work performance (Sackett et al., 2017). Furthermore, cognitive ability is significantly associated with happiness (Ali et al., 2013; Kanazawa, 2013) and overall health (Wrulich et al., 2014). Unsurprisingly, cognitive ability and its assessment therefore belong to the most discussed and central topics within psychology (Gottfredson & Saklofske, 2009).

For children, the most commonly used test worldwide to measure cognitive ability is the Wechsler Intelligence Scale for Children (WISC; Evers et al., 2012; Flanagan & Alfonso, 2017; Weiss et al., 2019). The WISC is a time- and field-tested instrument that has strong psychometric features (Oakland et al., 2016) and boasts international positive reviews (e.g., Na & Burns, 2016, for the English version; Kwaliteitscentrum voor Diagnostiek, 2023, for the

Dutch version). However, the WISC is primarily designed for clinical contexts, and less attuned to the needs of many research settings. Because this test can only be administered individually, the workload soon becomes unmanageable for studies aiming at larger samples, even when using abbreviated scales like the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011) or other short forms containing a reduced number of subtests (Aubry & Bourdin, 2018). Financially as well, the cost of administering the WISC often exceeds budgets of (junior) researchers.

To alleviate the practical problems that researchers encounter with Wechsler scales and other proprietary licensed intelligence tests, and “to address the need for psychometrically valid tools that are well-suited for large-scale, remote data collection” (Dworak et al., 2021, p. 3), an interuniversity team of cognitive ability experts developed the International Cognitive Ability Resource (ICAR; Revelle et al., 2020), a public-domain measure of cognitive ability designed for online administration. Public-domain measures are valuable tools in a research context. Open-access instruments enable more researchers to conduct studies, share findings, and ultimately advance the field (Condon & Revelle, 2014). The public nature of these instruments encourages collaboration within the research community regarding test development, refinement, and validation (Goldberg, 1999). On the downside, publicly available measures of cognitive ability sometimes either lack sufficient psychometric rigor (e.g., Cog15; Kajonius, 2014; Kristjánsdóttir & Zaiter, 2023) or are too narrow in focus, concentrating exclusively on a particular type of item (e.g., only matrix reasoning; Zorowitz et al., 2024).

The ICAR, on the other hand, is psychometrically sound and offers a wide variety of subtests (Revelle et al., 2017) including items of high and low difficulty. The items can be programmed into any software of choice, and the extensive database (The International Cognitive Ability Resource Team, 2014) makes it possible to tailor the instrument to the intended audience or to switch up the items and minimise test–retest effects. In the earliest days of the project, the ICAR consisted of just 60 items, divided over four subtests: Verbal Reasoning, Three-dimensional Rotation, Matrix Reasoning, and Letter-Number Series (Condon & Revelle, 2014). This original ICAR60, as well as a 16-item subset known as the ICAR16, was validated in a sample of 96,958 participants ranging in age from 14 to 90 years (Condon & Revelle, 2014). Since its initial validation, the ICAR has established itself as a cognitive ability measure for research on adolescents and adults (used in 79 published studies according to Dworak et al., 2021, and more since), the majority of studies employing the ICAR16.

The appropriateness of an ICAR for children, however, is still unknown. Considering the dynamic nature of brain development across the lifespan, with rapid changes in early childhood and more gradual changes in adulthood, it is essential to acknowledge that the ICAR items designed for adults may not be adequate for children (Flanagan & McDonough, 2018). For instance, through neurodevelopment, brains become more organised, efficient, and faster at processing information during childhood and adolescence (Lenroot & Giedd, 2006; Low & Cheng, 2006). Moreover, research suggests that age and frontal lobe development play pivotal roles in shaping cognitive ability (Shaw et al., 2006), highlighting the need for assessment tools tailored to the cognitive abilities of children.

Therefore, we aimed to compile an instrument based on ICAR items that unites the existing practical advantages of the ICAR with an adequate measurement of the cognitive ability of children between the ages of 11 and 14, more specifically for use in a Flemish school context. As the ICAR project is not based a priori on a particular theoretical framework such as the Cattell-Horn-Carroll (or CHC) model of intelligence (Schneider & McGrew, 2012), we also looked into the question of which broad cognitive abilities are measured by which ICAR subtest. The compilation of a children's ICAR, or Ch-ICAR in brief, not only broadens the scope of the ICAR project; it also addresses a notable gap in the literature, for at the moment there is a lack of psychometrically sound measures of cognitive ability that are freely available, limited in administration time, and suited for children.

After conducting a literature review (see Section 5.2), we tried out a selection of test items on some members of the target audience and compiled a preliminary instrument (see Section 5.3). With this preliminary instrument we carried out a pilot study, which occasioned further adaptations (see Section 5.4). Finally, with this adapted instrument, we conducted two studies assessing psychometrics and convergent validity (see Sections 5.5 and 5.6). Study 1 investigated the basic psychometric qualities of the Ch-ICAR. Study 2 cross-validated the Ch-ICAR with two commercial tests, on the one hand Raven's 2 Progressive Matrices (henceforth RPM), and on the other hand the Flemish CoVaT-CHC Basic Version (henceforth Covat). The former is a measure of nonverbal intelligence, which provides an effective estimate of cognitive ability (McLeod & McCrimmon, 2021) and is commonly used and suitable for group administration (Oakland et al., 2016). The latter is a comprehensive measure of intelligence, whose composition reflects the CHC model of intelligence (Magez et al., 2015). In Flanders, the Covat offers a valid alternative to other intelligence tests, since—unlike RPM—it encompasses multiple broad cognitive abilities (Flanagan & McDonough, 2018). Cross-validation of the Ch-

ICAR using both the RPM and Covat instruments provides us with empirical verification as to whether the Ch-ICAR is indeed a valid measure for children's cognitive ability.

5.2 Findings from the initial literature review

First of all, we reviewed the literature with the following questions in mind: (1) Has the ICAR been successfully used with children before, (2) has it already been used in Dutch-speaking regions, (3) which subtests are employed, (4) what constructs do the subtests measure, and (5) how much time is required to administer the ICAR? Dworak and colleagues (2021) provide a list of 79 publications incorporating the ICAR so far. Consultation of these publications yielded the following findings.

In line with the original validation, most studies involved adolescents and adults. Of those reporting a central tendency of the participants' age, the lowest was a median of 19 (Zhang & Goffin, 2018) and the highest a mean of around 47 (Goda et al., 2019). Only four studies exclusively tested youngsters below the age of 20 (Jankovski et al., 2017; Kajonius, 2016; Kirkegaard & Bjerrekær, 2016; Subotić et al., 2019), nevertheless signalling problems with too high difficulty levels. Save some, the ICAR studies predominantly took place in English-speaking regions, often relying on the Amazon Mechanical Turk. One study that was conducted in the Netherlands (Van Der Krieke et al., 2016) limited itself to items from Matrix Reasoning and Three-dimensional Rotation, for which no translation was needed.

The majority of studies employed the ICAR16, also known as the ICAR Sample Test (Condon & Revelle, 2014). Eight employed a different set of items from the same four subtests (Choma & Hanoch, 2017; Hood, 2015; Karwowski et al., 2017, 2020; Karwowski & Brzeski, 2017; Kirkegaard & Bjerrekær, 2016; Vermunt et al., 2018; Womick et al., 2019). Regarding the constructs measured, we looked for any explicit link to theory such as the CHC model of intelligence (Schneider & McGrew, 2012). The authors of the 79 studies mostly spoke in general terms of “intelligence” or “cognitive ability”. Some specifically mentioned “fluid intelligence”. A proper theoretical foundation, however, tying the ICAR subtests to broad cognitive abilities in the CHC model for instance, we did not encounter.

The administration time varied across studies. Frick (2017, p. 27) suggested that 20 minutes is the recommended amount of time for completing the ICAR16. By contrast, Thurston (2016)

reported that less than 20 minutes are required, and Zabelina and colleagues (2019) imposed a time limit of 18 minutes with the instruction to fill out as many of the 16 items as possible within the allotted frame. The bulk of studies giving any clue with regard to timing presented the ICAR as an untimed test. Interestingly, of the studies that administered just 5 items, the duration differed as much as 2,5 minutes (Gillen et al., 2019) and 15 minutes (Kirkegaard & Bjerrekær, 2016).

5.3 Selecting items for a Children's ICAR

Subsequently, we solicited feedback from experts in the field and from children of the target age to assess the suitability of various ICAR item types. With the consent of their respective parents, a total of eight children were tested in this phase, all Flemish and their age ranging from 11 to 14. Given our expectation that the test would be too difficult for children, we started by selecting roughly the 10 easiest items according to the ICAR database (The International Cognitive Ability Resource Team, 2014) from the original subtests Verbal Reasoning, Three-dimensional Rotation, Matrix Reasoning, and Letter-Number Series. Upon inspection of the other available subtests, we added Number Series and Figural Analogies to this list. For the Verbal Reasoning items we procured the Dutch translation by Fontaine (Nelissen et al., manuscript in preparation). With regard to the instructions, we provided each subtest with a written Dutch explanation ourselves.

During the first couple of administrations, participants performed better than expected. For the following administrations, we therefore reinserted the remaining items of Verbal Reasoning and Figural Analogies, and expanded Number Series with some more difficult items as well. The items were administered on paper during one-on-one sessions, either in person or via videocall. Subtests were broken off after three consecutive wrong answers. After examining the test scores, we decided not to include Three-dimensional Rotation and Letter-Number Series in the preliminary instrument. The former proved unsuitable for the target audience, with half of the children unable to solve any item correctly. The latter raised concerns about test conditions: If participants are allowed to write down the alphabet (or rather not prevented from doing so), it becomes substantially easier to complete the letter series.

In sum, we decided to continue with the four subtests Verbal Reasoning, Matrix Reasoning, Number Series and Figural Analogies, with 16, 11, 23 and 20 items, respectively. Verbal Reasoning contains questions on vocabulary and general knowledge as well as mathematical problems.⁶⁶ Matrix Reasoning items consist of eight geometric shapes, placed in a 3-by-3 grid, with one shape missing; participants have to choose the response option that represents the missing ninth shape. Number Series items show a short sequence of numbers; participants have to complete the sequence by entering the next number of the sequence. Figural Analogies items consist of three coloured geometric shapes, with the first two shapes standing in a certain relation to each other; participants have to choose the response option that represents the missing fourth shape, which stands in the same relation to the third as the second does to the first. A specimen of each item type can be found in Appendix 9.⁶⁷

5.4 Summary of the pilot study

With the goal in mind of putting minimal strain on researchers' time and participants' cognitive load, whilst still obtaining an accurate estimate of cognitive ability, it seemed desirable to reduce the number of items per subtest. A pilot study was carried out among five class groups of three different Flemish primary schools ($N_{\text{Pilot}} = 99$; male 41%; age 11 to 12). On the basis of this pilot study, a total of 31 items was preserved in the final Ch-ICAR: eight items each for Verbal Reasoning, Number Series, and Matrix Reasoning, and seven for Figural Analogies. We prioritised items with stronger item–subscale correlations, while simultaneously maintaining a balanced range of difficulty levels. We also ensured that selected items showed minimal sex differences, and avoided disproportionate response times. Additionally, we added a self-designed example item to the subtests Matrix Reasoning and Figural Analogies, in order to clarify instructions for the test-taker.

⁶⁶ As such, in our opinion, the name of the subtest does not quite cover the content, but for matters of consistency and comparability, we chose to retain the original subtest name.

⁶⁷ The final compilation of the Ch-ICAR is available on the ICAR website, via: <https://icar-project.com/projects/childrens-icar>. It is necessary to create a user account first.

5.5 Summary⁶⁸ of study 1

In a large sample of pupils recruited from 23 Flemish schools ($N_1 = 820$; male 53.5%; $M_{age} = 12.02$), study 1 evaluated four aspects of the psychometric qualities of the Ch-ICAR. Firstly, the internal structure of the test and the psychometric properties of the individual items were analysed, using hierarchical IRT (item response theory) analysis. The test structure was found to align with the two-level structure of intelligence posited by the CHC model, with several broad cognitive abilities at the lowest level (represented by the four Ch-ICAR subtests) and one general cognitive ability at the highest level. Four items that exhibited suboptimal fit were removed, lowering the total number of items in the Ch-ICAR from 31 to 27. Secondly, the test's internal consistency was determined by calculating Cronbach's alpha and McDonald's omega. The full Ch-ICAR demonstrated good internal consistency ($\alpha = .82$; $\omega = .82$). At subtest level, however, results varied: While reliability estimates for Number Series and Figural Analogies were favourable, those for Verbal Reasoning and Matrix Reasoning were on the lower side.

Thirdly, the distribution of test scores was examined, specifically in respect to sex. T-tests revealed no significant differences, in line with previous research showing that overall cognitive ability is generally equal across sexes. Fourthly, the relationship between the Ch-ICAR and academic performance was measured. The total Ch-ICAR score uniquely explained significant portions of the within-school variance in GPA (18%), mathematics score (26%), and score on a standardised mathematical skill test (35%). Overall, the Ch-ICAR seems to be a reliable and valid measure of cognitive ability in children, providing a quick, cost-efficient, and useful tool for researchers.

⁶⁸ This summary was created with limited assistance from Microsoft Copilot.

5.6 Summary⁶⁹ of study 2

In two different samples of respectively 91 secondary school pupils ($N_{\text{RPM}} = 91$; male 40%; $M_{\text{age}} = 12.99$) and 96 primary school pupils ($N_{\text{Covat}} = 96$; male 50%; $M_{\text{age}} = 11.05$), study 2 cross-validated the Ch-ICAR with the RPM and the Covat, looking at relations on both test level (RPM and Covat) and subtest level (only Covat). The included Covat subtests measured three broad cognitive abilities: fluid intelligence (Gf), crystallised intelligence (Gc), and visuospatial processing (Gv). Total Ch-ICAR scores correlated strongly with both RPM ($r = .67$) and total Covat scores ($r = .62$). These correlations indicate that the Ch-ICAR effectively measures cognitive abilities in a comparable way to the commercial tests. The Verbal Reasoning subtest of the Ch-ICAR was found to predominantly assess Gf and, to a somewhat lesser extent, Gc. The Figural Analogies subtest mainly tapped into Gv, and the other subtests Number Series and Matrix Reasoning mainly tapped into Gf.

However, some issues surfaced with two of the subtests. Figural Analogies had a non-normal score distribution, which suggests that difficulty levels may not be well-calibrated for the target age group. Matrix Reasoning displayed low reliability, which suggests that the items may not consistently measure the intended construct. The study nevertheless concluded that the total Ch-ICAR score is a valid measure of children's cognitive ability, establishing itself as a robust alternative to proprietary intelligence tests for research purposes. At this stage, the independent use of individual subtests is not recommended due to fluctuating reliabilities. Further research is needed to improve the reliability and validity of individual subtests.

5.7 Conclusion

In developing the Ch-ICAR, our aim was to expand the scope of the ICAR project, considering that its suitability for children had not been previously assessed (Dworak et al., 2021; The International Cognitive Ability Resource Team, 2014). Additionally, we sought to address the

⁶⁹ This summary was created with limited assistance from Microsoft Copilot.

dearth of psychometrically robust cognitive ability measures that are freely accessible, brief in administration time, and tailored to children.

The results of studies 1 and 2 provide support for the utility of the full Ch-ICAR as a measure for children's cognitive abilities within a research context. This was demonstrated by the observed correlations between the Ch-ICAR and the RPM and Covat, two established intelligence tests (Magez et al., 2015; McLeod & McCrimmon, 2021), and by the established relationship with academic performance. The magnitude of the correlation coefficients is in line with prior validation research of the ICAR16 (Condon & Revelle, 2014; Young & Keith, 2020). At this point, however, we advise against the use of the Ch-ICAR subtests in isolation, in contrast to the use of the battery as a whole. Further research should ascertain whether the subtests can reliably stand alone and yield meaningful interpretations. Therefore, replications of the relations with the broad cognitive abilities are necessary, as well as improvements in the Matrix Reasoning and Figural Analogies subtests.

The primary constraint of this investigation pertains to the restricted scope of the Ch-ICAR, which solely assessed a limited subset of ICAR items in children. Consequently, the possibility cannot be dismissed that an alternative or more extensive item selection may yield superior performance as a child-adapted version of the ICAR. Another consideration involves the absence of established norms for the Ch-ICAR. Given the dynamic developmental changes in cognitive processing efficiency during childhood and adolescence (Flanagan & McDonough, 2018), it is imperative to consider factors such as age or grade level when interpreting and comparing Ch-ICAR results. Therefore, we underscore the significance of future research endeavours aimed at establishing age- or grade-based norms to ensure precise interpretation and comparison of Ch-ICAR outcomes.

Works cited in this chapter

- Ali, A., Ambler, G., Strydom, A., Rai, D., Cooper, C., McManus, S., ... Hassiotis, A. (2013). The relationship between happiness and intelligent quotient: The contribution of socio-economic and clinical factors. *Psychological Medicine*, 43(6), 1303–1312.
- Aubry, A., & Bourdin, B. (2018). Short forms of Wechsler scales assessing the intellectually gifted children using simulation data. *Frontiers in Psychology*, 28(9), no. 830.
- Choma, B. L., & Hanoch, Y. (2017). Cognitive ability and authoritarianism: Understanding support for Trump and Clinton. *Personality and Individual Differences*, 106, 287–291.

- Condon, D., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43, 52–64.
- Dworak, E., Revelle, W., Doebler, P., & Condon, D. (2021). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences*, 169, no. 109906.
- Evers, A., Muniz, J., Bartram, D., Boben, D., Egeland, J., Fernandez- Hermida, J. R., ... Urbanek, T. (2012). Testing practices in the 21st century: Developments and European psychologists' opinions. *European Psychologist*, 17(4), 300–319.
- Flanagan, D., & Alfonso, V. (2017). *Essentials of WISC-V Assessment*. John Wiley & Sons.
- Flanagan, D., & McDonough, E. (Eds.). (2018). *Contemporary intellectual assessment: Theories, tests, and issues (Fourth Edition)*. Guilford Press.
- Frick, S. (2017). *Leaders on their best behavior: Leader behaviors resulting in effective virtual teams* [Doctoral dissertation]. University of South Florida.
- Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the Caltech cohort study. *Journal of Political Economy*, 127(4), 1826–1863.
- Goda, G., Levy, M., Manchester, C., Sojourner, A., & Tasoff, J. (2019). Predicting retirement savings using survey measures of exponential-growth bias and present bias. *Economic Inquiry*, 57(3), 1636–1658.
- Goldberg, L. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (pp. 1–7). Tilburg University Press.
- Gottfredson, L. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24(1), 79–132.
- Gottfredson, L., & Saklofske, D. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology / Psychologie Canadienne*, 50(3), 183–195.
- Hood, J. (2015). *Deconstructing the unskilled-and-unaware problem: Examining the effect of feedback on misestimation while disentangling cognitive bias from statistical artifact* [Doctoral dissertation]. Appalachian State University.
- Jankovski, J., Zečević, I., & Subotić, S. (2017). A preliminary examination of the ICAR progressive matrices test of intelligence. *Empirical Studies in Psychology XXIII*, 152–157.
- Kajonius, P. (2016). Honesty-humility predicting self-estimated academic performance. *International Journal of Personality Psychology*, 2, 1–6.
- Kanazawa, S. (2013). Why is intelligence associated with stability of happiness? *British Journal of Psychology*, 105(3), 316–337.
- Karwowski, M., & Brzeski, A. (2017). Selfies and the (creative) self: A diary study. *Frontiers in Psychology*, 8, no. 172.
- Karwowski, M., Czerwinka, M., & Kaufman, J. (2020). Does intelligence strengthen creative metacognition? *Psychology of Aesthetics, Creativity, and the Arts*, 14(3), 353–360.
- Karwowski, M., Lebeda, I., Szumski, G., & Firkowska-Mankiewicz, A. (2017). From moment-to-moment to day-to-day: Experience sampling and diary investigations in adults' everyday creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 11(3), 309–324.
- Kirkegaard, E., & Bjerreker, J. (2016). ICAR5: Design and validation of a 5-item public domain cognitive ability test. *Open Differential Psychology*.
<https://openpsych.net/paper/5/>
- Kristjánisdóttir, D., & Zaiter, A. (2023). *Public domain intelligence tests: Psychometric properties of the Cog15 and ICAR16 cognitive ability scales* [Master's dissertation]. Lund University.
- Kwaliteitscentrum voor Diagnostiek. (2023). *Beoordeling van de WISC-V-NL voor gebruik in Vlaanderen*.
<https://portaal.kwaliteitscentrumdiagnostiek.be/actueel/beoordeling-van-de-wisc-v-nl-voor-gebruik-in-vlaanderen/>
- Lenroot, R., & Giedd, J. (2006). Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neuroscience and Biobehavioral Reviews*, 30(6), 718–729.
- Low, L., & Cheng, H. (2006). Axon pruning: An essential step underlying the developmental plasticity of neuronal connections. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 361(1473), 1531–1544.

- Magez, W., Tierens, M., Van Huynegem, J., Van Parijs, K., Decaluwé, V., & Bos, A. (2015). *CoVaT-CHC Basisversie*. Psychodiagnostisch Centrum Thomas More.
- McLeod, J., & McCrimmon, A. (2021). Test review: Raven's 2 Progressive Matrices, clinical edition (Raven's 2). *Journal of Psychoeducational Assessment*, 39(3), 388–392.
- Na, S., & Burns, T. (2016). Wechsler Intelligence Scale for Children- V: Test review. *Applied Neuropsychology: Child*, 5(2), 156–160.
- Nelissen, W., Schittekatte, M., Fontaine, J., & Rigolle, F. (Manuscript in preparation). *Validation of the International Cognitive Ability Resource (ICAR) in a Flemish sample*.
- Oakland, T., Douglas, S., & Kane, H. (2016). Top ten standardized tests used internationally with children and youth by school psychologists in 64 countries: A 24-year follow-up study. *Journal of Psychoeducational Assessment*, 34(2), 166–176.
- Revelle, W., Condon, D., Wilt, J., French, J., Brown, A., & Elleman, L. (2017). Web- and phone-based data collection using planned missing designs. In N. Fielding, R. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (2nd ed., pp. 578– 595). Sage Publishing.
- Revelle, W., Dworak, E., & Condon, D. (2020). Cognitive ability in everyday life: The utility of open-source measures. *Current Directions in Psychological Science*, 29(4), 358–363.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118–137.
- Sackett, P., Lievens, F., Van Iddekinge, C., & Kuncel, N. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology*, 102(3), 254–273.
- Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). The Guilford Press.
- Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., ... Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature*, 440(7084), 676–679.
- Subotić, S., Lovrić, S., Gajić, V., Golubović, A., & Sibinčić, D. (2019). Relationship between life stressors, depressiveness, and school achievement in late elementary and late high school students. *Banja Luka November Encounters*, 61–64.
- The International Cognitive Ability Resource Team. (2014). *International cognitive ability resource*. <https://icar-project.com>
- Thurston, A. (2016). *An integrative model of situation awareness* [Master's dissertation]. University of South Florida.
- Tikhomirova, T., Malykh, A., & Malykh, S. (2020). Predicting academic achievement with cognitive abilities: Cross-sectional study across school education. *Behavioral Sciences*, 10(10), no. 158.
- Van Der Krieke, L., Jeronimus, B., Blaauw, F., Wanders, R., Emerencia, A., Schenk, H., De Vos, S., Snippe, E., Wichers, M., Wigman, J., Bos, E., Wardenaar, K., & De Jonge, P. (2016). HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research*, 25(2), 123–144.
- Vermunt, J., Ilie, S., & Vignoles, A. (2018). Building the foundations for measuring learning gain in higher education: A conceptual framework and measurement instrument. *Higher Education Pedagogies*, 3(1), 266–301.
- Vilia, P., Candeias, A., Neto, A., Franco, M., & Melo, M. (2017). Academic achievement in physics-chemistry: The predictive effect of attitudes and reasoning abilities. *Frontiers in Psychology*, 8, no. 1064.
- Wechsler, D. (2011). Wechsler Abbreviated Scale of Intelligence Second Edition (WASI-II) [Database record]. *APA PsycTests*.
- Weiss, L., Saklofske, D., Holdnack, J., & Prifitera, A. (Eds.). (2019). *WISC-V: Clinical use and interpretation*. Elsevier.
- Womick, J., Ward, S., Heintzelman, S., Woody, B., & King, L. (2019). The existential function of right-wing authoritarianism. *Journal of Personality*, 87(5), 1056–1073.
- Wrulich, M., Brunner, M., Stadler, G., Schalke, D., Keller, U., & Martin, R. (2014). Forty years on: Childhood intelligence predicts health in middle adulthood. *Health Psychology*, 33(3), 292–296.

- Young, S., & Keith, T. (2020). An examination of the convergent validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment*, 38(8), 1052–1059.
- Zabelina, D., Friedman, N., & Andrews-Hanna, J. (2019). Unity and diversity of executive functions in creativity. *Consciousness and Cognition*, 68, 47–56.
- Zhang, I., & Goffin, R. (2018). *Evil geniuses at work: Does intelligence interact with the dark triad to predict workplace deviance?* [Undergraduate Honors Thesis]. University of Western Ontario.
- Zorowitz, S., Chierchia, G., Blakemore, S., & Daw, N. (2024). An item response theory analysis of the matrix reasoning item bank (MaRs-IB). *Behavior Research Methods* 56(3), 1104–1122.

Chapter 6: Reflections on School-Based Research

This chapter has yet to be submitted for publication.

Chapter 6

Reflections on School-Based Research

6.1 Introduction

Enlisting school pupils as research participants involves a lot more than simply going to schools to collect data. Rushing into a school-based study unprepared and with unrealistic expectations is a recipe for failure. But then, what is the right preparation and what can realistically be expected? For a junior researcher, or in fact any researcher inexperienced in school-based research, this can be hard to assess, as the typical research article describing a school-based study does not give any pointers in that regard. Therefore, the current paper offers some methodological reflections that will help guide novel school-based researchers. The implied research design is one in which multiple tests and surveys (requiring no specialised equipment) are administered to a large number of pupils during regular teaching hours, but *mutatis mutandis* the recommendations given here should apply to other types of studies in schools as well, such as intervention research for example.

In contrast to many other meritorious articles on how to conduct school-based research, I will do more than just share my experiences and lessons learned. Instead, this paper was conceived as a thematic literature study, with a personal take. After all, sharing experiences is great, but combining and classifying them is even better. By reviewing a sizable number of articles, I sought to identify the best practices as well as the aspects of school-based research that remained neglected. The literature search started from articles citing Alibali and Nathan's 2010 guide (see Section 6.2.2), and continued organically from there. Where necessary, I supplemented the existing literature with insights from other methodological literatures (see particularly Sections 6.4.2, 6.6 and 6.8). Throughout, I have used my own field notes and post-

hoc reflections to provide details, examples, and additional advice not encountered during heuristics.

Section 6.2 introduces the school as a research setting, by giving a nuanced view of both the importance and difficulty of school-based research. Section 6.3 approaches the school as a microcosm: a complex system of interwoven elements that researchers must try to grapple with. Here and elsewhere, it is the researcher's perspective that prevails, although the value of taking the school's perspective is continuously emphasised. The subsequent sections are arranged in accordance with the stages of the research process. Sections 6.4 and 6.5 are on how to build working relationships with schools and how to plan research activities in schools, respectively. Section 6.6 delves into participant recruitment and informed consent. Section 6.7 discusses the administration of research materials from a highly practical point of view. Section 6.8 is about test-taking motivation: a topic that has not nearly received enough attention in the context of school-based research, given that it is one of the backbones of research validity. Section 6.9 brings together some further thoughts on research management. Together, these eight sections aim to give a comprehensive overview of those little important things that do not belong to “core” research design (like variables, conditions, and statistical analysis), but that every school-based researcher ought to think about nonetheless.

6.2 The school as a research setting

6.2.1 The importance of school-based research

The school is a difficult research setting, but an important one. Indeed, it has been claimed that “the importance of schools as a context for scientific research cannot be overstated” (Plummer et al., 2014, p. 15). First of all, schools allow researchers to reach child and adolescent populations, and vice versa: In bringing large numbers of youngsters together, schools provide a unique opportunity in terms of both population accessibility and sample representativity compared to lab-based research, which requires participants to be able to travel to the lab (e.g., Alibali & Nathan, 2010; Bartlett et al., 2017; Haley & Brackney, 2022; Higgins et al., 2019; March et al., 2022; Plummer et al., 2014; Taylor & Owen, 2021). Moreover, the school environment is “familiar, safe, and appropriate for data collection” (Haley & Brackney, 2022,

p. 105), and it is relatively easy to work with long follow-up spans of several months or even years (Higgins et al., 2019). Dey and colleagues (2022, pp. 171–172) also note that response rates are typically much higher in school-based surveys than in surveys distributed otherwise, “at least when no active parental consent is needed” (see also Section 6.6).

Aside from being a convenient option for the recruitment of young research participants, another advantage of the school as a research setting is that it offers real-world data. Depending on the research topic, school-based research may either be a valuable complement to controlled lab experiments, or even the only way of tackling a certain research question. Researchers from various fields within the behavioural sciences have stressed the importance of heading into the classroom and gathering “*in vivo* data” (Plummer et al., 2014, p. 15). Higgins and colleagues (2019, p. 1), for example, point out that there are “mutual benefits of collaborations at the intersection of cognitive science and education practice”. One such research opportunity is putting a lab-developed learning theory to the test in much more complex circumstances to gauge its generalisability. The great potential of schools as a site for studying health, prevention and wellbeing interventions has also been recognised (e.g., Bartlett et al., 2017; March et al., 2022; Santangelo White, 2012; Taylor & Owen, 2021).

It is clear that school-based studies can be highly valuable for various scientific domains, such as education and instruction, human cognition, child development, healthcare, or social functioning. Notwithstanding the value of school-based research, it always remains crucial to consider data quality and watch out for validity threats. Reflecting the variety of research topics, there is also a wide variety in possible research designs and data collection methods, which include focus groups, classroom observation, and test administration. Something all types of research designs must take into account is sample representativity; however, securing a representative sample can be complicated by informed consent procedures, as we will see in Section 6.6. Other validity threats are linked to a specific type of design, and some research designs are more vulnerable in that respect than others. One thing Covid-19 has taught us, is that the mode of supervision can affect survey data quality (Stefes, 2024). For test administration, a typical threat is test-taking motivation, or rather a lack thereof; this issue will be explored further in Section 6.8.

6.2.2 The difficulty of school-based research

The school is an important research setting, but a difficult one. Indeed, this setting has been characterised as “dynamic and unpredictable”, and fraught with challenges that make research there completely unlike any experimental study in a lab (Taylor & Owen, 2021, p. 107). Just how unpredictable and overwhelming the experience can be, becomes evident from the emotive and expressive language that is regularly used: Lonergan and Cumming (2017), for example, liken their experience to single-handedly riding a canoe down fast-flowing rapids full of rocks and other obstacles. In other words, school-based research is not the same on paper and in reality: It is one thing to name school pupils as participants in a research proposal, and another to actually collaborate with schools and their staff to make data collection happen. The task is all the more difficult for researchers who are unfamiliar with this kind of fieldwork, in particular early-career researchers.

At the heart of the numerous “logistical and communication challenges” (Higgins et al., 2019, p. 1) is a mutual lack of procedural knowledge (e.g., Lonergan & Cumming, 2017; March et al., 2022; Plummer et al., 2014). For schools, on the one hand, serving as a research setting is not their primary function (more on that in Section 6.3.2), and therefore they may perceive research participation as a bothersome burden above all else (e.g., Haley & Brackney, 2022; March et al., 2022). Researchers, on the other hand, are all too often ill-prepared when they first set out to conduct school-based research, and may not fully grasp what the schools’ concerns are (for a brief testimony, see e.g., Plummer et al., 2014, pp. 15–16). Additionally, they have the disadvantage of being an outsider to the school (a broader issue within community-based research, cf. Demers et al., 2022). Furthermore, academic colleagues might not be understanding with regard to the difficulty of school-based research. The idea still appears prevalent that data collection in schools is or should be something straightforward. In any case, courses on school-based research are a lot harder to come by than courses on experimental methods like, for instance, eye-tracking.

As a researcher, you can only bring as much experience to the table as you have. Existing research competences can be augmented either by an “inefficient process of trial and error” (Plummer et al., 2014, p. 15), or by learning from the experience of others. Fortunately, there is a growing, be it still small, body of literature on the ins and outs of school-based research. The development of this literature seems to fit in with a much wider tendency of an increasingly open, candid and hands-on approach to research methodology. In different fields and

paradigms, several recent examples can be found of evolutions towards discussing methodology in greater detail: even when it means getting really nitty-gritty, as in the case of the newly introduced reporting guideline for eye-tracking research (Dunn et al., 2024), and even when it means admitting that things sometimes go haywire, for which the introduction of the term Linguistic Non-Understanding in linguistic ethnography is a case in point (van Hest & Jacobs, 2022). Within school-based prevention research, Edwards and colleagues refer to (the need for) such a mentality shift in the title of their 2023 article “What is not in the methods section: Challenges, successes, and lessons learned”.

A pivotal publication on school-based research, although certainly not the first ever piece of its kind, is an article from 2010 by Alibali and Nathan entitled “Conducting research in schools: A practical guide”. Later similar publications cite Alibali and Nathan as a point of reference, while simultaneously underlining the necessity of expanding the literature (e.g., Higgins et al., 2019; Lonergan & Cumming, 2017; Plummer et al., 2014; Santangelo White, 2012; Taylor & Owen, 2021). Some of these methodological articles take a more generic point of view (such as the guides by Alibali & Nathan, 2010 and Plummer et al., 2014), while others are tailored to a clear-cut audience, for example Physical Activity researchers in the United Kingdom (Taylor & Owen, 2021). As Lonergan and Cumming (2017) have noted before, healthcare is particularly well-represented in the methodological literature on school-based research. Not only do healthcare-related articles make up a relatively large portion of the publications (a couple of examples being Bartlett et al., 2017; Butterfield et al., 2003; Dey et al., 2022; Hatch et al., 2023; March et al., 2022; Santangelo White, 2012), there also seems to be a longer tradition of practice-oriented commentary papers in this area (some older examples being Dent et al., 1993; Farrington et al., 2000; Harrington et al., 1997; Olds & Symons, 1990).

6.3 The microcosm that is a school

Now, what exactly makes school-based research so challenging? The answer is that a school is a world of its own, the inner workings of which place limits on what is possible in terms of research. Often, there is a mismatch between the needs and agenda of the researcher and the needs and agenda of the school (Mix & Kalish, 2019, p. 107). A thorough understanding of this microcosm is therefore indispensable, for which we turn to theory in this section.

6.3.1 Complexity

In line with contemporary trends in science and scholarship, several authors have defined and approached the school as a complex system (e.g., Lonergan & Cumming, 2017; March et al., 2022). Here I will build on the theory developed by Lonergan and Cumming (2017), which is itself based on Handy and Aitken's (1986) organisational model.⁷⁰ According to this model, an organisation is characterised by the interplay of four elements: task, structure, systems, and people. By combining Handy and Aitken's model with Robinson's (1993) problem-based methodology and applying it to the school, Lonergan and Cumming arrive at a list of constraints acting upon school-based research. Said list provides a handy overview for aspiring school-based researchers of the various levels where problems potentially can arise when taking a research proposal into a school. In what follows, I will go over the four organisational elements and paraphrase the constraints identified by Lonergan and Cumming, infusing the discussion with insights from others as well as my own.

6.3.2 Task

First of all, the task or collective purpose of a school is to educate its pupils; crucially, its task is not to deliver participants for scientific research. While this observation might seem almost too obvious to mention, that is not necessarily so. Historically speaking, it is not entirely unthinkable that a school would view research participation as one of its core tasks. Of particular interest is the tradition of the laboratory school, first envisioned and realised by the prolific American philosopher John Dewey (1859 – 1952). Glennon et al. (2013, p. 30) explain that “[a]s a laboratory, the school had the same relationship to education and psychology as a teaching hospital has to biology and chemistry – it was a place for putting theory into practice in an experimental setting.” Or in the words of Dewey (1899) himself, his school was “a laboratory of applied psychology”. Ghent University used to have an “experimenting school” too for a large part of the twentieth century, but the school had to close for financial reasons in 1991 (Danniau, 2015).

⁷⁰ Handy and Aitken's model is in turn an adapted version of the original organisational model known as Leavitt's Diamond (Leavitt, 1964).

In the present-day methodological literature on school-based research, we find many calls for institutionalised collaboration. The Switzerland-based research ensemble of Dey and colleagues strongly urges for an ongoing public “dialogue between schools and scholars to better coordinate future school-based research agendas and quality standards”, for instance in the form of “a joint regional survey coordination system” (Dey et al., 2022, p. 179). The creation of such coordinating bodies is not the only possible route to durable collaboration, however. Institutionalised relationships between the world of academic research and the world of education can take various forms and exist at various levels, from local to governmental. At the lowest level, there are long-term partnerships between individual universities and schools. A successful example documented in the literature is the so-called “research school movement” in the United States (Hinton & Fischer, 2008; Kuriloff et al., 2009), which is heavily inspired by the work and legacy of the previously mentioned John Dewey. The aim of the movement, as Glennon and colleagues (2013, p. 30) describe it, is “to reconnect with Dewey’s vision of conducting research in a living school, and extend this work to create an infrastructure that supports sustainable collaboration among researchers and practitioners”.

Other authors focus on the intermediate level, just above the individual school. They advise researchers to strike up relationships with educational stakeholders such as multi-school trusts or school partner agencies, of which the mission aligns with the research topic; in this way, researchers gain access to a pool of schools through an organisation the schools are already familiar with (e.g., Blom-Hoffman et al., 2009; Higgins et al., 2019; March et al., 2022; Taylor & Owen, 2021). Finally, there are also relevant institutions operating at a higher level. In Europe, several countries have national or regional organisations that aim to bridge the gap between research and education by funding research projects, pairing schools and universities, and/or translating research findings to help teachers work in an evidence-informed way. Some examples are the Education Endowment Foundation in the United Kingdom, *Leerpunt* (literally “learning point” or “learning hub”) in Flanders, and the *Nationaal Regieorgaan Onderwijsonderzoek* (Netherlands Initiative for Education Research) in the Netherlands.

If possible, relying on an established relationship of any kind is beneficial, as this can considerably facilitate the entire process from obtaining permission from the school board to cooperating with school staff on data collection days. This being said, the viability as well as longevity of a school-based research project ultimately depends on how well the project can be aligned with the task of the school. There has to be some common ground, a shared goal between researcher and school, like understanding how children learn, or developing more

effective teaching methods (Mix & Kalish, 2019). The research topic needs to be perceived as relevant, important, and somewhat urgent, and the envisaged outcomes must have a tangible educational value (Lonergan & Cumming, 2017; Santangelo White, 2012). In some of the most fruitful partnerships, educators have even helped steer research in new directions (Mix & Kalish, 2019).

6.3.3 Structure

Structure, in Lonergan and Cumming's theory, entails physical and administrative organisation: buildings, rooms, class sizes, calendars, timetables et cetera. When working out a study's methodology, it is imperative to remember that school-based research can only happen on the terms of the schools: The duration of tests and surveys must fit the length of teaching periods, the number of available participants will be contingent on how pupils are grouped into classes, and so on (e.g., Alibali & Nathan, 2010; Bartlett et al., 2017; Taylor & Owen, 2021). The circumstances of administration are also influenced by the school's structure in the literal sense: For example, is there enough space to separate pupils from one another during test administration, or do they have to sit close to each other?

If there is a digital component to the research, for instance if testing takes place through an online platform, that too is likely to depend on the equipment present in the school. For maximum consistency, it would be better if researchers could either use a testing facility or provide all devices themselves, but these situations are rare. In Flanders, all pupils in secondary and upper primary education now have their own devices (Directorate-General for Communication, 2022). Despite opening up possibilities for digital administration, personal devices simultaneously hinder administration: Firstly, because the type of device varies between schools, multiple different versions of the same tests and surveys may have to be programmed to ensure compatibility (e.g., laptop vs. tablet), and secondly, the researcher has very limited control over device settings, which can interfere with testing. Conversely, an "old-fashioned" computer classroom offers the possibility of preparing the devices before pupils come in.

6.3.4 Systems

While task and structure can largely be accounted for with proper forethought, both systems and people are more unpredictable and must be navigated as one moves forward. The term

“systems” is used in this theoretical framework to refer to any policies, procedures and routines that are in place to help reach the organisational goals. Alternatively, this aspect could be referred to as “school culture”.⁷¹ Based entirely on my personal experience with school-based research in the Low Countries, it would seem that the orderliness of the corridors is a reliable proxy for school culture. If a school makes a relatively peaceful impression, it is more likely to be internally well-organised as well; if you are nearly run over and pushed off the stairs by a horde of screaming children as soon as you set foot inside, that is not the best sign – but again, this is merely anecdotal.

It is beyond doubt, however, that school culture and internal systems make all the difference for the logistics of a school-based research project. Since research activities usually involve some sort of disruption of the regular programme (e.g., changes in the timetable and classroom allocation) and require the use of school infrastructure (e.g., rooms, laptops, headphones), a large number of school staff tends to be involved in or affected by the execution of the study (e.g., Bartlett et al., 2017; Plummer et al., 2014). This means that, being an outsider, the researcher has to rely to a certain extent on the internal communication flow of the partner school (March et al., 2022). In other words, no matter how perfect the research plan, it is beyond the researcher’s control exactly how well-informed all members of staff will be on the day of data collection. A suboptimal school-internal communication flow can potentially affect data completeness and quality. This is just one reason why it is always advisable to oversample schools (March et al., 2022).

The dream situation for the school-based researcher is that a member of staff is available who has sufficient leeway within their job role and workload to take on research coordination. In Flanders, I noticed that some schools have a team of coordinators who oversee all special events, so to say, like field trips or research studies. Joining forces with such a coordinator, who is not only committed to the research project but actually has time to take care of the practical organisation, can make data collection proceed much more smoothly. Plummer and colleagues (2014) describe another arrangement, in which the main school contact was paid a special honorarium, covered by project funding. More often, however, research coordination simply

⁷¹ In some extended versions of Leavitt’s Diamond, organisational culture is instead added as a separate fifth element (Nograšek & Vintar, 2014).

has to be taken up by teachers on top of their already overburdened workload without extra compensation (March et al., 2022).

6.3.5 People

That leads to the fourth element: people. In the end, the key is to strike a balance between on the one hand staying true to your own research agenda and scientific non-negotiables, and on the other hand taking the viewpoint of the people you collaborate with in a school context. These people include not just the school staff, but of course pupils and their parents as well. The following five sections of this paper are essentially each about this researcher – people dynamic: Sections 6.4 to 6.8 discuss different aspects of dealing with staff members, participants and parents in detail. As a final remark to this section, I will merely say that school-based researchers should be mentally prepared to receive a range of different reactions to their research study, from indifferent to interested, from defiant to cooperative, and everything in between.

6.4 Working relationships with schools

In the realisation of a school-based study, a first and vital step is to establish and maintain good working relationships with the partner schools. Below, three general recommendations will be discussed: gatekeepers (Section 6.4.1), face-to-face communication (Section 6.4.2), and investment (Section 6.4.3).

6.4.1 Gatekeepers

An oft-recurring recommendation is to seek out and appoint a primary contact (e.g., Bartlett et al., 2017; Hatch et al., 2023; March et al., 2022; Plummer et al., 2014; Santangelo White, 2012; Taylor & Owen, 2021). Ideally, this contact person is someone who is kindly disposed towards the researcher and their research study, because of a personal connection, mutual acquaintances, and/or a strong affinity with the topic. They can function as a facilitator in

several areas, by advocating for the execution of the study, acting as an intermediary between the researcher and the school's headmaster, and assisting the researcher with logistics. The primary contact can change throughout the course of a project. Especially in longitudinal studies spanning multiple years, it can be a challenge to keep identifying the right person to contact every time.

A term that is sometimes used in the methodological literature, is “gatekeeper”. In the definition of Keesling (2008, p. 300), “[a] gatekeeper is a person who stands between the data collector and a potential respondent. Gatekeepers, by virtue of their personal or work relationship to a respondent, are able to control who has access, and when, to the respondent.” This definition can be interpreted in different ways, and indeed one can argue that there are at least two broad types of gatekeepers. On the one hand, there are the abovementioned primary contacts, who are allies and who can figuratively set the school gate ajar for the researcher. On the other hand, there is the school board, who can literally grant or deny access to the school and whose approval needs to be won. Of course, that is not to say the two cannot overlap in practice. In any case, engaging with gatekeepers is “a complex ongoing process requiring a clear understanding of the role [of gatekeeping], strong interpersonal skills, a sound understanding of ethical principles and knowledge of who can be approached for advice and when to do so” (McFadyen & Rankin, 2016, p. 85).

6.4.2 Face-to-face communication

Whenever possible, face-to-face communication is preferable. A study by Roghanizad and Bohns (2017) confirmed that compliance rates are a lot higher when requests are made in person versus when they are made electronically. Interestingly, the people in the study tended to underestimate their persuasiveness in person, and to grossly overestimate their persuasiveness via email. It is not just about initially persuading a school to partake in a study, however: “Access needs not only approval but also cooperation”, as McFayden and Rankin (2016, p. 86) put it. School-based researchers have repeatedly reported that meeting face-to-face with school staff helped build rapport and positive relationships (e.g., March et al., 2022; Plummer et al., 2014). In my own research, I have also found that face-to-face communication increased goodwill as well as adherence to study guidelines, compared to digital communication only. In the same vein, I adamantly advise against having school staff carry out data collection tasks without some in-person briefing beforehand.

6.4.3 Investment

A good partnership works both ways, and it is recommended to invest the necessary time and effort into these working relationships with schools: Not only do schools deserve to be treated well in return for their cooperation, cordial relations will also benefit the research project. This investment takes various forms. Again and again in the school-based research literature, authors emphasize expressing gratitude verbally as well as by offering small gifts, both in acknowledgement of the staff's sacrifices and as incentives for pupils' participation (e.g., Alibali & Nathan, 2010; Higgins et al., 2019; Loneragan & Cumming, 2017; March et al., 2022; Plummer et al., 2014; Rice et al., 2019; Santangelo White, 2012; Taylor & Owen, 2021). Another way of giving back to partner schools is to share the outcomes and findings directly with them, once the study has been completed. Taylor and Owen (2021) particularly recommend the use of infographics in this context.

Furthermore, thoughtful and strategic communication before and during the study is of the essence. It is advised to involve school staff at an early stage (e.g., Alibali & Nathan, 2010; Feldman, 2022; Okello et al., 2013). Communication must be as brief and clear as possible and aimed at supporting research literacy, since staff members might not have any prior experience with a research project in action (March et al., 2022). What may seem a self-evident design choice to a researcher may not be so from the perspective of a teacher. Devoting attention to research literacy, for example by explaining why a certain sample size is an absolute requirement, improves understanding and motivation for the study (March et al., 2022).

6.5 Planning research activities in schools

Once a school has given the green light to a research study, the next step is to plan the research activities. Both relationship-building and planning take up a substantial amount of time, and yet these stages of the research process are rarely mentioned in articles dedicated to the results of a school-based study. It is no wonder, then, that junior researchers and educators sometimes find themselves confronted with unpleasant surprises in regard to the project timeline. This can be helped by adopting the right attitude as a researcher (Section 6.5.1) and keeping in mind some additional considerations (Section 6.5.2).

6.5.1 Peace, love, and understanding

When we look at the literature on school-based research, the number-one recommendation given there is to show patience, flexibility, and understanding towards the school. Because of structural constraints and unexpected factors, the data collection period may be prolonged, the duration of data collection sessions may vary, and in some cases the research design may even have to be adjusted (e.g., Alibali & Nathan, 2010; Bartlett et al., 2017; Hatch et al., 2023; Higgins et al., 2019; Lonergan & Cumming, 2017; March et al., 2022; Plummer et al., 2014; Rutter et al., 2020; Santangelo White, 2012; Taylor & Owen, 2021). Furthermore, to compensate for issues like pupil absence or internet breakdowns, it can be an option to organise make-up sessions at a later date (March et al., 2022; Plummer et al., 2014).

6.5.2 Additional considerations

While being flexible and adaptable are indeed fundamental qualities for a school-based researcher, I believe there is more to be said on the topic of planning which is hardly mentioned in the existing literature. It cannot be stressed enough that planning is a separate and time-consuming phase which precedes the data collection phase.

- (1) To make things a bit more concrete, I recommend factoring in several weeks to a month, just for the schools to go through the initial process of figuring out where they can make space in the curriculum for research participation. During this time, the researcher can expect both a storm of follow-up questions and long stretches of not hearing back from their contact persons.
- (2) Since schools do not necessarily have much experience with research, it can be helpful to be proactive in pointing out that preparations such as scheduling research activities and gathering informed consent forms (cf. Section 6.6) take time, and need to be completed well ahead of the intended period of data collection.
- (3) It is better to overestimate than to underestimate the time that will be needed for testing. A data collection session takes longer than the duration of the tests and surveys (more on that in Section 6.7).
- (4) Different scheduling approaches each have their advantages and disadvantages. Spreading out data collection by dividing pupils into small groups makes each session less consequential for the success of the study, but increases the total time needed for

data collection. A more compact planning where larger groups of pupils are tested at once is more efficient, but raises the stakes: If something goes wrong, a considerable amount of data may be compromised. Pupils' age or other aspects of the design can pose additional constraints on the size of the testing groups.

- (5) There must be a plan in place for pupils who do not participate in the research study (also mentioned by Alibali & Nathan, 2010 and Santangelo White, 2012). Moreover, there must be clear agreements on what pupils are to do as soon as they have completed the research activities (e.g., stay in silence, go home, go to another classroom...).
- (6) Researchers should not forget to schedule in enough time themselves before and after every data collection session, to prepare (e.g., reviewing participants, generating pupil accounts for online testing platforms) and finish up (e.g., checking the data for anomalies, especially when they were collected by somebody else like a student-assistant or a teacher).

6.6 Participant recruitment

Next, we move on to participant recruitment. This section deals with the procedures (Section 6.6.1) and documents (Section 6.6.2) for informed consent in general. If studying minorities or vulnerable populations, a tailored approach and extra recruitment measures may be needed (see for instance Bartlett et al., 2017; Gaias et al., 2020).

6.6.1 Informed consent procedures

Prospective participants must be informed about the details of the research study, and provide their consent for participation; when they are minors, this responsibility lies with their parents or guardians. As a researcher, it is good to be thoroughly aware of the different approaches to informed consent. Since over the last few decades numerous publications have been devoted to the subject, I will limit myself to formulating a definition and summarising the pros and cons described in the informed consent literature. Thereafter, I will go into a couple of practical considerations for school-based research.

The two main categories of informed consent procedures are “active” versus “passive”. Active informed consent (AIC) can also be referred to as opt-in or documented consent, since participants or their parents or guardians have to provide a written response to signal consent. Non-response is consequently interpreted as latent refusal. Passive informed consent (PIC) can also be referred to as opt-out or undocumented consent, since participants or their parents or guardians only need to respond in case they do not agree with research participation. Non-response is consequently interpreted as latent consent.

AIC and PIC differ along three axes: ethics, compliance, and feasibility. AIC is often preferred, because it is viewed as the most ethical option (e.g., Jason et al., 2001; Range et al., 2001). On the downside, participation rates tend to be lower compared to PIC (e.g., Chartier et al., 2008; Courser et al., 2009; Jason et al., 2001; Spence et al., 2015; White et al., 2004). This poses not only a quantitative problem in terms of sample size but also a qualitative problem, in the sense that the complying sample may be less representative. A lot of publications have centred around this very issue, and it has repeatedly been found that there are demographic differences between children whose parents accept and children whose parents refuse in an AIC procedure, with a biased and artificially homogenous study sample as a result (e.g., Bergstrom et al., 2009; Chartier et al., 2008; Courser et al., 2009; Dent et al., 1993; Jason et al., 2001; Spence et al., 2015; Unger et al., 2004; White et al., 2004). Moreover, the administrative load that an AIC procedure imposes on all parties involved should not be underestimated, especially because extensive follow-up measures are usually needed to boost participation rates (e.g., Wolfenden et al., 2009).

PIC has the advantages of being much easier to implement and of boasting high, near-total participation rates (e.g., Chartier et al., 2008; Courser et al., 2009; Jason et al., 2001; Spence et al., 2015; Unger et al., 2004; White et al., 2004), which in turn ensures a more representative sample. According to a recent qualitative study in England, schools also prefer the simplicity of opt-out consent (Hatch et al., 2023). However, PIC may be ethically problematic (e.g., Range et al., 2001), since in reality non-response could signify a number of things, including never having received the information. PIC can therefore only be justified if research participation is minimal-risk (e.g., Jason et al., 2001; Plummer et al., 2014; Unger et al., 2004). Even then, PIC may not be allowed by the university’s ethical committee or by law: For example, the European General Data Protection Regulation (European Union, 2016) requires documented consent when collecting personal data. Indeed, most school-based researchers will be faced with the task of implementing AIC (Wolfenden et al., 2009).

The informed consent phase can partially overlap with the planning phase discussed previously in Section 6.5. Again, it is of the utmost importance to be flexible and allot enough time: Obtaining AIC in a school context can easily take months (e.g., March et al., 2022; Santangelo White, 2012; Taylor & Owen, 2021). It is common practice to have the pupils take home an information letter and consent form and then return the signed form, but as Taylor and Owen (2021, p. 108) put it, “[t]here are numerous opportunities for breakdowns to occur in this ‘school-to-home-to-school’ consent form process”. One of their recommendations is therefore to align the consent procedure as closely as possible with the regular communication between school and parents. Sending out a reminder or two has also proven to be effective (e.g., Ji et al., 2004; Taylor & Owen, 2021; Wolfenden et al., 2009). The opposite is sometimes done as well, that is pre-notifying parents well in advance about the upcoming research study and consent procedure via a school newsletter or the like (e.g., Harrington et al., 1997; Wolfenden et al., 2009).

An alternative strategy to bypass the cumbersome school-to-home-to-school process is to make use of events such as parent-teacher meetings to get informed consent from the parents or guardians directly (e.g., Ji et al., 2004; Taylor & Owen, 2021; Wolfenden et al., 2009). For this strategy to be viable, the timing of these events must be compatible with the research timeline. Another powerhouse strategy which has been shown to significantly influence consent form return rates, is to provide incentives: For example, the class gets a pizza party if everyone returns their form (Wolfenden et al., 2009). Of course, incentives require resources, which may not always be available. In addition to that, policy at either the research institution or the school might not allow giving gifts as an incentive.

Finally, return rates are likely to vary between schools, depending on how much support they can offer the researcher with the consent procedure (Ji et al., 2004; Lueptow et al., 1977). This has everything to do with the existence of the right systems (cf. Section 6.3.4) and the cooperativeness of the people (cf. Section 6.3.5) in them. It can be helpful if staff members are proactive in collecting consent forms, and if they assist the researcher in conveying the importance of the research (Santangelo White, 2012). In my own study, there were a few teachers who overemphasised the fact that participation was non-obligatory, with extremely low participation rates in these particular class groups as a result. While Ji and colleagues (2004) argue that adequate school support is paramount, Wolfenden and colleagues (2009) advise researchers to carry out and oversee recruitment themselves insofar possible, or better still, to have a dedicated recruitment coordinator as part of the research team (see also Section 6.9).

6.6.2 Information and consent documents

Whatever the informed consent procedure looks like, the information letter and consent form ought to be clear and concise (Santangelo White, 2012). This is much easier said than done, since ethical guidelines often call for elaborate and detailed information about the study (as also recognised by Taylor & Owen, 2021). When these documents are too text-heavy, however, they may come across as daunting or evoke apprehensiveness, and hence have the opposite effect of what was intended. For instance, a lengthy technical explanation of how the data will be pseudonymised and protected might actually raise parents' concern about their child's privacy. Some possible ways to inspire confidence and encourage participation, on the other hand, are using the school's rather than the research institution's lay-out, adding a cover letter and/or signature of the principal, and coupling informed consent to another school form that needs to be returned (Ji et al., 2004; White et al., 2004; Wolfenden et al., 2009).

The solicitation of informed consent is sometimes combined with the collection of subject variables (e.g., age, gender, socioeconomic status). When referencing the pupils' home situation on such a subject data form or other document, it is commendable to be mindful of diversity. The formulations and checkboxes should not assume the presence of a mother and a father, but allow for guardians, one-parent families, or families with two same-sex parents. Small adjustments, like consistently writing "parent or guardian" or including a "not applicable" checkbox, can accommodate other family structures besides the traditional nuclear family.

6.7 Administration of research materials

Once research activities have been planned and participants have been secured, then data collection can start. Most articles on school-based research seem to focus on the collaboration process between researcher and school in the run-up to data collection (covered in Sections 6.3 to 6.6), and devote less attention to the actual execution of the measurements itself. This section will share eight practical tips pertaining directly to the administration of research materials. As indicated in the introduction, the imagined situation is that of tests and surveys administered in a classroom setting.

- (1) Provide a manual with all of the administration details, from the oral introduction to troubleshooting guidelines. Not only can it serve as a prompt for the researcher, it is particularly useful to distribute beforehand to anyone assisting with data collection, such as students or teachers. Having a manual to refer to benefits fidelity of implementation (see also Higgins et al., 2019; March et al., 2022; Plummer et al., 2014).
- (2) Bring backup, both an extra test leader and extra materials, e.g., paper-and-pencil versions of questionnaires in case there is no Internet connection, or printed cards in case audio or video does not function properly (see also Plummer et al., 2014).
- (3) Prepare thoroughly for the technical aspect of test administration: Work closely with the school's IT service, do multiple test runs in different circumstances, and have extra devices, batteries and headphones on hand during data collection sessions if possible (see also March et al., 2022; Plummer et al., 2014).
- (4) Make sure there is enough buffer time to get started, all the more when working with large groups. Before all pupils are seated, have their devices ready and are logged in to the right platforms, several minutes are sure to have passed in the best-case scenario, or up to about half an hour if any problems or disruptions occur (see also Plummer et al., 2014).
- (5) Do not overlook pupil seating, and preferably enlist a teacher's help to arrange it. Allowing the pupils to choose themselves where to sit can lead to chaos along the way. Having the pupils sit in alphabetical order or such can cause a delay at the start, but this is likely to be compensated by a more orderly proceeding of the administration process.
- (6) Carefully plan the instructions for the pupils: Ensure that all the tools needed are present (e.g., school board, chalk or markers, projector, hard copies) and do not give instructions in advance they are not allowed to follow until later on.
- (7) Expect the unexpected. Sudden fire drills and field trips appear to be a constant in school-based research (see also Alibali & Nathan, 2010; Plummer et al., 2014; Santangelo White, 2012).
- (8) If any pupils have to stay behind for some minutes to finish up, keep a close eye on them, as they might try to slip away before they are done with their tests.

6.8 Test-taking motivation

Data collected in schools are inevitably noisy. There are all sorts of confounding variables, some of which may be anticipated and incorporated into the research design. Administration circumstances are never as consistent and closely monitored as in a lab, even with optimal organisation. Another major limitation of school-based research of the test administration variety has to do with test-taking motivation, which we will ponder in this section.

It is often implicitly assumed that ability tests measure ability, while in reality they measure both “skill and will” (Eklöf, 2010). With low-stakes assessment in particular there is a real risk of low motivation (e.g., Wise & DeMars, 2005; Wolf & Smith, 1995), especially in males (e.g., Finn, 2015; Schlosser et al., 2019). If pupils are not motivated to do well on a test, they put in less effort, which is in turn reflected in their performance; as such, test-taking motivation introduces construct-irrelevant variance which threatens research validity (e.g., Attali, 2016; Finn, 2015; Ulitzsch et al., 2024). When left unaddressed, scores from unmotivated participants can distort both the psychometric properties of the test and the overall research findings.

The difference in performance between high- and low-stakes assessment – the so-called “stakes effect” (Attali, 2016) – and the validity concerns it raises have prompted quite a bit of research. Unfortunately, as briefly hinted at in the introduction, the stakes literature has not been integrated with the more practice-oriented school-based research literature. Part of the explanation is perhaps that research on the stakes effect has concentrated more on higher education than on primary or secondary school (Finn, 2015). Nonetheless, managing pupils’ test-taking motivation in order to mitigate the stakes effect is a practical aspect of school-based research that needs to be considered and planned, just as much as the communication strategy towards the schools or the type of informed consent procedure.

Various measures can be taken to manage motivation before, during, and after testing. First of all, talking to participants and framing the test as important has been shown to increase motivation and hence performance (Finn, 2015). Motivational framing can be applied to the institutional level, by telling test takers that they are representatives of their educational institution and that their scores will be used to evaluate this institution. Motivational framing can also be applied to the personal level, by telling test takers that there will be personal consequences, for instance that their scores will count towards grades or that they will be released to superiors. The concept of consequentiality is obviously closely related to that of

stakes. Too heavy consequences, however, may induce test anxiety and negatively impact performance. In a study with 757 college students, Liu and colleagues (2015) concluded that there was no motivation or performance difference between the institutional and personal conditions, and that both groups outperformed the control condition in which test takers were simply told that their scores were “for research purposes”.

Other common approaches are providing feedback on performance and handing out incentives such as gift cards (on incentives see also Section 6.6.1). Although their effects on test-taking effort are still unclear (Finn, 2015), these courtesies may improve working relationships (cf. Section 6.4.3) and diminish attrition (March et al., 2022). As far as the test itself is concerned, the ideal test from a motivational point of view is neither too taxing nor too easy, does not contain large chunks of text, and does include graphics (Finn, 2015). Moreover, the smoothness of test administration is of importance as well, as problems and mistakes during this process also lower test-taking motivation (Wise & DeMars, 2005). After data has been collected, motivation can be controlled for through various techniques of sample filtering (e.g., Attali, 2016; Ulitzsch et al., 2024), based on working time, self-reported motivation or effort, performance, person-fit measures or individual differences (Finn, 2015). Since data analysis falls outside the scope of this paper, I will merely note here that some of the more advanced techniques of filtering on the basis of working time require fine-grained response times at the screen or item level.

6.9 Research management

As the previous sections have revealed, there are a lot of moving parts to school-based research. I would like to end with some other recommendable practices which do not have so much to do with the external collaboration with schools, but rather with the challenges of internally managing a school-based research project.

Setting up a school-based study is a huge logistical and communicative operation, in which many different tasks must be fulfilled. It is generally not recommended to have all of these tasks taken up by a single researcher. Instead, there should be a research team, in which everyone has their designated role, and in which preferably at least one person has prior hands-on experience with school-based research. Plummer and colleagues (2014) describe a core team

of four, consisting of a principal investigator, a day-to-day research manager, a technology coordinator, and a school liaison. During the initial stages of a school-based study, the administrative load often becomes quite heavy: The countless visits, emails and phone calls required to get the study up and running across all schools can readily turn into a full-time job of its own. Having a separate school liaison or hiring temporary administration support during busy periods makes communications proceed more efficiently, and frees time for the main researcher(s) for other preparations (March et al., 2022; Plummer et al., 2014; Wolfenden et al., 2009). Data collection also typically calls for multiple people (Plummer et al., 2014; Taylor & Owen, 2021), particularly when dealing with large participant groups or overlapping testing times.

Furthermore, just like schools (cf. Section 6.3.4), researchers do well to have the right systems and workflows in place. One can hardly record too much in logs and roadmaps, or in a “codebook” as Plummer and colleagues (2014) call it. During data collection, field notes are a must (Lonergan & Cumming, 2017; Plummer et al., 2014). Jotting down all technical anomalies that occur, like pupils switching accounts or skipping parts of the test and so on, is an invaluable step with an eye to data processing and analysis afterwards. Personally, I have also found it interesting to make notes from time to time about my general impressions of how test administration went, many of which found their way into this paper. Finally, it is advisable to keep one central overview of participants: who has returned the consent form, who has returned the subject data form, who has participated in which test, what has become of them in between data collection rounds (e.g., a change of school), et cetera. This sounds self-evident, but in practice it is all too easy to get lost in a sea of class lists, participant lists, and testing lists in various different formats. The same goes for contact details of school staff members.

6.10 Conclusion

Every research project is unique and has its individual challenges, but that does not mean all general tips are in vain (a similar sentiment is expressed in the conclusions of March et al., 2022). As this paper hopefully has shown, there is a lot to learn from the experiences of other school-based researchers, and many solid recommendations have been put forward which are broadly valid. By taking all of the above into account, and considering these difficulties in

advance, the process of school-based research can be optimised. There are always aspects that are outside the researcher's control, such as how supportive the school staff is or how large the classrooms are, but with proper preparation and enough resources a school-based study does not have to be difficult or overwhelming. Thanks to field guides such as this one, future school-based researchers no longer need to feel like they are evading rocks in a swirling river, or like they have to keep reinventing the wheel to overcome the challenges they face.

Works cited in this chapter

- Alibali, M., & Nathan, M. (2010). Conducting research in schools: A practical guide. *Journal of Cognition and Development, 11*(4), 397–407.
- Attali, Y. (2016). Effort in low-stakes assessments: What does it take to perform as well as in a high-stakes setting? *Educational and Psychological Measurement, 76*(6), 1045–1058.
- Bartlett, R., Wright, T., Olarinde, T., Holmes, T., Beamon, E., & Wallace, D. (2017). Schools as sites for recruiting participants and implementing research. *Journal of Community Health Nursing, 34*(2), 80–88.
- Bergstrom, J., Partington, S., Murphy, K., Galvao, L., Fayram, E., & Cisler, R. (2009). Active consent in urban elementary schools: An examination of demographic differences in consent rates. *Evaluation Review, 33*(5), 481–496.
- Blom-Hoffman, J., Leff, S., Franko, D., Weinstein, E., Beakley, K., & Power, T. (2009). Consent procedures and participation rates in school-based intervention and prevention research: Using a multi-component, partnership-based approach to recruit participants. *School Mental Health, 1*(1), 3–15.
- Butterfield, P., Yates, S., Rogers, B., & Healow, J. (2003). Overcoming subject recruitment challenges: Strategies for successful collaboration with novice research agencies. *Applied Nursing Research, 16*(1), 46–52.
- Chartier, M., Stoep, A. V., McCauley, E., Herting, J., Tracy, M., & Lymp, J. (2008). Passive versus active parental consent: Implications for the ability of school-based depression screening to reach youth at risk. *The Journal of School Health, 78*(3), 157–186.
- Courser, M., Shamblen, S., Lavrakas, P., Collins, D., & Ditterline, P. (2009). The impact of active consent procedures on nonresponse and nonresponse error in youth survey data: Evidence from a new experiment. *Evaluation Review, 33*(4), 370–395.
- Danniau, F. (2015). *Experimenteerschool*. UGentMemorie.
<https://www.ugentmemorie.be/gebeurtenissen/1948-experimenteerschool>
- Demers, J., Gregus, S., & Petts, R. (2022). New communities, new relationships: Reflections from junior faculty engaging in community-based research. *Journal of Interpersonal Violence, 37*(17–18), NP14938–NP14969.
- Dent, C., Galaif, J., Sussman, S., Stacy, A., Burton, D., & Flay, B. (1993). Demographic, psychological and behavioral differences in samples of actively and passively consented adolescents. *Addictive Behaviors, 18*(1), 51–56.
- Dewey, J. (1899). *The school and society*. University of Chicago Press.
- Dey, M., Marti, L., Helbling, L., & Jorm, A. (2022). Non-participation in epidemiological school-based surveys: Using mixed methods to study predictors and modes of justification for (non-)

- participation at the school level. *International Journal of Social Research Methodology*, 25(2), 171–182.
- Directorate-General for Communication. (2022). ‘Digisprong’ of the Flemish Community. *European Commission*.
https://commission.europa.eu/projects/digisprong-flemish-community_en
- Dunn, M., Alexander, R., Amiebenomo, O., Arblaster, G., Atan, D., Erichsen, J., Ettinger, U., Giardini, M., Gilchrist, I., Hamilton, R., Hessels, R., Hodgins, S., Hooge, I., Jackson, B., Lee, H., Macknik, S., Martinez-Conde, S., Mcilreavy, L., Muratori, L., ... Sprenger, A. (2024). Minimal reporting guideline for research involving eye tracking. *Behavior Research Methods*, 56, 4351–4357.
- Edwards, K., Orchowski, L., Espelage, D., & Temple, J. (2023). What is not in the methods section: Challenges, successes, and lessons learned from conducting school-based interpersonal violence prevention research. *Journal of Interpersonal Violence*, 38(3–4), 4507–4532.
- Eklöf, H. (2010). Skill and will: Test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice*, 17(4), 345–356.
- European Union. (2016). *Document 32016R0679*. EUR-Lex.
<https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
- Farrington, F., McBride, N., & Midford, R. (2000). Conducting research in schools: Lessons learnt from experience. *Health Promotion Journal of Australia*, 10(1), 63–68.
- Feldman, H. (2022). Who’s striking, and who’s not? Avoiding and acknowledging bias in youth climate activism research. *Australian Journal of Environmental Education*, 38(1), 112–118.
- Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Research Report Series*, 2015, no. RR-15-19.
- Gaias, L., Duong, M., Pullmann, M., Brewer, S., Smilansky, M., Halbert, M., Carey, C., & Jones, J. (2020). Race and ethnicity in educational intervention research: A systematic review and recommendations for sampling, reporting, and analysis. *Educational Research Review*, 31, no. 100356.
- Glennon, C., Hinton, C., Callahan, T., & Fischer, K. (2013). School-based research. *Mind, Brain, and Education*, 7(1), 30–34.
- Haley, T., & Brackney, D. (2022). Conducting nursing research in the school setting: Lessons learned. *Creative Nursing*, 28(2), 102–108.
- Handy, C., & Aitken, R. (1986). *Understanding schools as organizations*. Penguin.
- Harrington, K., Binkley, D., Reynolds, K., Duvall, R., Copeland, J., Franklin, F., & Raczynski, J. (1997). Recruitment issues in school-based research: Lessons learned from the High 5 Alabama Project. *The Journal of School Health*, 67(10), 415–421.
- Hatch, L., Widnall, E., Albers, P., Hopkins, G., Kidger, J., de Vocht, F., Kaner, E., van Sluijs, E., Fairbrother, H., Jago, R., & Campbell, R. (2023). Conducting school-based health surveys with secondary schools in England: Advice and recommendations from school staff, local authority professionals, and wider key stakeholders, a qualitative study. *BMC Medical Research Methodology*, 23, no. 142.
- Higgins, E., Dettmer, A., & Albro, E. (2019). Looking back to move forward: A retrospective examination of research at the intersection of cognitive science and education and what it means for the future. *Journal of Cognition and Development*, 20(2), 1–20.
- Hinton, C., & Fischer, K. (2008). Research schools: Grounding research in practice. *Mind, Brain, and Education*, 2(4), 157–160.
- Jason, L., Pokorny, S., & Katz, R. (2001). Passive versus active consent: A case study in school settings. *Journal of Community Psychology*, 29(1), 53–68.
- Ji, P., Pokorny, S., & Jason, L. (2004). Factors influencing middle and high schools’ active parental consent return rates. *Evaluation Review*, 28(6), 578–591.
- Keesling, R. (2008). Gatekeeper. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods* (p. 300). Sage Publications.
- Kuriloff, P., Reichert, M., Stoudt, B., & Ravitch, S. (2009). Building research collaboratives among schools and universities: Lessons from the field. *Mind, Brain, and Education*, 3(1), 34–44.

- Leavitt, H. (1964). Applied organization change in industry: Structural, technical, and human approaches. In W. Cooper, H. Leavitt, & M. Shelly (Eds.), *New perspectives in organization research* (pp. 55–71). John Wiley & Sons.
- Liu, O. L., Rios, J., & Borden, V. (2015). The effects of motivational instruction on college students' performance on low-stakes assessment. *Educational Assessment*, 20(2), 79–94.
- Lonergan, R., & Cumming, T. (2017). Riding the rapids of classroom-based research. *The Australian Educational Researcher*, 44(2), 141–160.
- Lueptow, L., Mueller, S., Hammes, R., & Master, L. (1977). The impact of informed consent regulations on response rate and response bias. *Sociological Methods & Research*, 6(2), 183–204.
- March, A., Ashworth, E., Mason, C., Santos, J., Mansfield, R., Stapley, E., Deighton, J., Humphrey, N., Tait, N., & Hayes, D. (2022). 'Shall we send a panda?' A practical guide to engaging schools in research: Learning from large-scale mental health intervention trials. *International Journal of Environmental Research and Public Health*, 19(6), no. 3367.
- McFadyen, J., & Rankin, J. (2016). The role of gatekeepers in research: Learning from reflexivity and reflection. *GSTF Journal of Nursing and Health Care*, 4(1), 82–88.
- Mix, K., & Kalish, C. (2019). Foreword to the special issue: Found in translation. *Journal of Cognition and Development*, 20(2), 107–109.
- Nograšek, J., & Vintar, M. (2014). E-government and organisational transformation of government: Black box revisited? *Government Information Quarterly*, 31(1), 108–118.
- Okello, G., Jones, C., Bonareri, M., Ndegwa, S., Mcharo, C., Kengo, J., Kinyua, K., Dubeck, M., Halliday, K., Jukes, M., Molyneux, S., & Brooker, S. (2013). Challenges for consent and community engagement in the conduct of cluster randomized trial among school children in low income settings: Experiences from Kenya. *Trials*, 14(1), no. 142.
- Olds, R., & Symons, C. (1990). Recommendations for obtaining cooperation to conduct school-based research. *The Journal of School Health*, 60(3), 96–98.
- Plummer, B., Galla, B., Finn, A., Patrick, S., Meketon, D., Leonard, J., Goetz, C., Fernandez-Vina, E., Bartolino, S., White, R., & Duckworth, A. (2014). A behind-the-scenes guide to school-based research. *Mind, Brain, and Education*, 8(1), 15–20.
- Range, L., Embry, T., & MacLeod, T. (2001). Active and passive consent: A comparison of actual research with children. *Ethical Human Sciences and Services*, 3(1), 23–31.
- Rice, M., Soistmann, H., Ejem, D., Hammack Johnson, A., Turner-Henson, A., Davis, S., & Gray, L. (2019). This work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors—Recruitment of hard to contact, hard to engage child populations in clinics and schools. *Applied Nursing Research*, 46, 72–77.
- Robinson, V. (1993). *Problem-based methodology: Research for the improvement of practice*. Pergamon.
- Roghanizad, M., & Bohns, V. (2017). Ask in person: You're less persuasive than you think over email. *Journal of Experimental Social Psychology*, 69, 223–226.
- Rutter, S., Stones, C., Wood, J., Macduff, C., & Gomez-Escalada, M. (2020). Effectiveness and efficiency of persuasive space graphics (PSG) in motivating UK primary school children's hand hygiene. *International Journal of Environmental Research and Public Health*, 17(7), no. 2351.
- Santangelo White, L. (2012). Conducting school-based research: Lessons learned from an after-school intervention study. *Journal of Pediatric Health Care*, 26(1), 75–79.
- Schlosser, A., Neeman, Z., & Attali, Y. (2019). Differential performance in high versus low stakes tests: evidence from the Gre test. *The Economic Journal*, 129(623), 2916–2948.
- Spence, S., White, M., Adamson, A., & Matthews, J. (2015). Does the use of passive or active consent affect consent or completion rates, or dietary data quality? Repeat cross-sectional survey among school children aged 11–12 years. *BMJ Open*, 5(1), no. e006457.
- Stefes, T. (2024). Data quality of different modes of supervision in classroom surveys. *Education Sciences*, 14(3), no. 299.
- Taylor, S., & Owen, M. (2021). Challenges to school-based physical activity data collection: Reflections from English primary and secondary schools. *Health Education Journal*, 80(1), 106–118.

- Ulitzsch, E., Shin, H. J., & Lüdtke, O. (2024). Accounting for careless and insufficient effort responding in large-scale survey data—Development, evaluation, and application of a screen-time-based weighting procedure. *Behavior Research Methods*, 56(2), 804–825.
- Unger, J., Gallaher, P., Palmer, P., Baezconde-Garbanati, L., Trinidad, D., Cen, S., & Johnson, C. (2004). No news is bad news. Characteristics of adolescents who provide neither parental consent nor refusal for participation in school-based survey research. *Evaluation Review*, 28(1), 52–63.
- van Hest, E., & Jacobs, M. (2022). Spaces of linguistic non-understanding in linguistic ethnography (and beyond). In *Methodological issues and challenges in researching transculturally* (pp. 14–38). Cambridge Scholars Publishing.
- White, V., Hill, D., & Effendi, Y. (2004). How does active parental consent influence the findings of drug-use surveys in schools? *Evaluation Review*, 28(3), 246–260.
- Wise, S., & DeMars, C. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1–17.
- Wolf, L., & Smith, J. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, 8(3), 227–242.
- Wolfenden, L., Kypri, K., Freund, M., & Hodder, R. (2009). Obtaining active parental consent for school-based research: A guide for researchers. *Australian and New Zealand Journal of Public Health*, 33(3), 270–275.

to err is human, to forgive divine;
statistical is to include errors in your design
~ after Alexander Pope and Leslie Kish

Part III: Empirical Research

Introductory remarks to Part III

This final and principal part reports on the empirical research that was conducted into classical language education in Flanders, with the three objectives in view that were laid out in the General Introduction. Chapter 7 corresponds to objectives 2 and 3: to investigate the relation between cognitive transfer and teaching method, and to investigate which methods of teaching Latin are in use in Flanders. The introductory remarks to Chapter 7 explain how the complementary study on teaching method was conceived, how it is connected with the main study on cognitive transfer, and why in the end the sample of schools from the main study did not permit to meet objective 2 and relate teaching method to cognitive transfer. The chapter itself contains the article that resulted from the complementary study, which was published in the Dutch classics journal *Lampas*.

Chapters 8 and 9 correspond to objective 1: to investigate the cognitive transfer effects of studying classical languages in Flemish secondary education by means of a longitudinal study. Insofar possible, I would like to invite the reader of these chapters to keep in mind previous units of this dissertation, especially Chapters 1 and 2, as they elucidate the rationale of this research. Methods and results are taken together in Chapter 8. Here all methodological and statistical details of the longitudinal study can be found. Given that the results sections contain more tables and graphs than text, I have deliberately chosen not to assign numbers to these tables and graphs, feeling that the addition of more titles and numbers would not be conducive to the chapter's surveyability. Chapter 9 links statistics to substance, by thoroughly discussing the results and adding a few extra analyses. With regard to terminology, in previous chapters I have generally preferred to speak about "classical language pupils", but in the last two chapters the terms "latinists" and "hellenists" will figure more prominently as a reflection of the grouping variables Latinist and Hellenist in the dataset (cf. Section 8.1.3.1).

Chapter 7: Teaching Method

The contents of this chapter are nigh identical to the following article:

De Herdt, K., & Vereeck, A. (2024). Van grammaticaal naar genietend lezen? Recente tendensen in de Vlaamse lespraktijk Latijn. *Lampas*, 57(2), 125–151.
<https://doi.org/10.5117/LAM2024.2.002.HERD>

Minor edits were made with an eye to harmonising punctuation, title numbering, cross references, bibliographical style etc. throughout the dissertation, and a correctional note was added at the very end.

Authorship roles are as follows. Conceptualisation: A. Vereeck. Study design and research methodology: A. Vereeck. Instrument development: A. Vereeck and K. De Herdt. Data collection: A. Vereeck and K. De Herdt. Data processing and formal analysis: A. Vereeck. Results and figures: A. Vereeck. Interpretation, discussion and contextualisation of results: K. De Herdt. Draft preparation: K. De Herdt and A. Vereeck. Editing and revision of final manuscript: A. Vereeck and K. De Herdt.

Chapter 7

Teaching Method

Introductory remarks to Chapter 7

One of the original two project objectives was to assess whether any potential transfer effects are influenced by the method that a school employs to teach classical languages. Specifically, we wondered whether the transfer value of classical languages differs depending on whether grammar is taught as an independent component versus as a means to an end (the end being reading). In order to operationalise teaching method, and in particular its grammar-orientedness, as a school-level variable in the longitudinal study, it was quantified through a questionnaire. The questionnaire contained a fixed set of statements and response options, which were assigned numerical values so that a total score could be calculated. As soon as we started developing the questionnaire that was to be used yearly as one of the measures in the main study on cognitive transfer, the idea arose to present a more elaborate version of this questionnaire to a larger audience of Flemish classical language teachers. Accordingly, the complementary study on teaching method came into being, and the third objective of mapping out educational practices was added to the project. For feasibility, it was decided to focus on educational practices for Latin.

This additional study was meant to complement the main study by generating rich and at that point unavailable background information, against which the questionnaire results from the longitudinal study could be more effectively interpreted. After the first year of the longitudinal study, however, it already became clear that it would not be possible to compare the partaking schools on the basis of teaching method: In my sample of schools, recruited on the basis of their enrolment numbers for classical languages (cf. Section 8.1.2.1) the teaching

method differences between teachers were many times larger than those between schools. On the one hand, the total questionnaire scores of teachers from the same school were not at all close together as expected, but sometimes even spanned opposite sides of the spectrum. On the other hand, the total questionnaire scores per school (averaged across all classical language teachers working there) differed only slightly from each other. The schools' scores thus seemed to be determined more so by the coincidental combination of individual teachers than by a clear common policy. Questionnaire results from the following two years consolidated this finding. Therefore, the plan to consider teaching method as a school-level variable in the longitudinal study was abandoned.⁷²

The complementary study could therefore not really fulfil its initially intended role within the cognitive transfer research either. Nevertheless, on its own, the study makes an important and novel contribution to the field of classical language didactics, as explained in the article abstract below. As such, this chapter doubles up as a more in-depth description of Flemish classical language education. The article focuses on the elaborate questionnaire that was distributed among Flemish classical language teachers, but also discusses findings from a series of exploratory interviews that were held during the questionnaire development phase (*passim*), and mentions the findings from the shorter questionnaire used in the longitudinal study as well (Section 7.4.2). Another particularly interesting section within the context of cognitive transfer research is Section 7.4.5 on teachers who favour a more strongly grammar-oriented approach, as it appears they often justify their didactic preferences by referring to the transfer arguments.

⁷² Including a variable at teacher level was not an option. The number of pupils per classical language teacher is too small to be statistically meaningful, and teachers are not a constant factor: It can change from year to year which pupils are taught by which classical language teacher.

English abstract

Until now, a broad overview of Latin didactics in Flanders as a whole was missing from the literature. For the first time, we submitted the Flemish practices of teaching Latin to a large-scale survey, in which over 200 teachers took part. The focal point of our survey was the position of grammar in present-day instruction. In order to quantify this instructional aspect and be able to make detailed comparisons between groups of teachers, we developed a scale measuring “grammar-orientedness”. For the sake of thorough interpretation and contextualisation, we supplemented the quantitative data with qualitative data from teacher comments and interviews, and also related the results to the current educational guidelines for classical languages. Our main finding is that on the whole, Flanders has left the grammar-translation method behind and has embraced the so-called reading method prescribed by the educational guidelines, which puts texts centre stage and integrates the teaching of grammar, reading and cultural-historical content. Nonetheless, more grammar-oriented as well as more extremely reading-oriented practices also exist, and there remains plenty of institutional, interindividual and intra-individual variation.

7.1 Inleiding

Op welke manier wordt Latijn onderwezen in Vlaanderen? Hoewel de door de overheid opgelegde minimumdoelen (zie ook Sectie 7.2) in principe bepalen wat leerlingen moeten leren in het vak Latijn, betekent dit geenszins dat het er in heel Vlaanderen in iedere klas hetzelfde aan toegaat. Leraren hebben immers verschillende opvattingen en hanteren verschillende methodes, zoveel is allang duidelijk. Hoe die verschillen er precies uitzien en hoe ze zich vertalen in dagelijkse leeractiviteiten, dat was vooralsnog minder duidelijk. Voor het eerst is nu de Vlaamse onderwijspraktijk Latijn op grote schaal onderzocht, door middel van onder meer een breed afgenomen enquête.

Meer bepaald hebben we ons toegespitst op de kwestie hoeveel belang en gewicht leraren Latijn toekennen aan grammaticaonderricht. In welke mate focussen ze op lectuur dan wel op grammatica? Strookt hun visie over de ideale verhouding tussen taal, lectuur en cultuur en de manier waarop ze die in hun lessen gestalte geven met de officiële visie van hun onderwijsnet? We richtten ons hiervoor op de zogenaamde eerste graad (leerjaren 1 en 2) en tweede graad (leerjaren 3 en 4), aangezien dit de jaren zijn waarin de Latijnse taal wordt aangeleerd. In de derde graad (leerjaren 5 en 6) worden de aangeleerde grammatica en woordenschat alleen nog toegepast voor het lezen van authentieke teksten: nieuwe taalelementen komen niet meer expliciet aan bod.

Alvorens over te gaan tot de rapportering van het onderzoek, bieden we eerst de nodige context over het Vlaamse curriculum Latijn. In Sectie 7.2 overlopen we kort de relevante elementen van het Vlaamse onderwijs in het algemeen en beschrijven we ook enkele historische evoluties binnen het klassieketalenonderwijs. In Sectie 7.3 gaan we dieper in op de onderzoeksmethode. In Sectie 7.4 behandelen we onze voornaamste bevindingen. Eerder dan een systematisch en exhaustief overzicht te geven van alle data die we met de enquête hebben verzameld, hebben we ervoor gekozen om de meest interessante en opvallende resultaten grondig te bespreken en te interpreteren.

7.2 Context: het Vlaamse curriculum Latijn⁷³

7.2.1 Leerlingen, onderwijsnetten en koepels

In het voorjaar van 2022, wanneer onze enquête liep, volgde 8,8% van de Vlaamse leerlingpopulatie (leerjaren 2-6)⁷⁴ het vak Latijn (Classica Vlaanderen, 2023). Hierbij dienen

⁷³ Zie Laes (2019) voor een situering van het actuele onderwijsaanbod klassieke talen (inclusief lestijden en leerlingenaantallen in de verschillende onderwijsnetten) en Claus (2021) voor een historisch overzicht van de leerplannen van Katholiek Onderwijs Vlaanderen. Het artikel van Claus en alle leerplannen Latijn van deze koepel zijn verzameld op een website: bit.ly/475b3NM.

⁷⁴ Voor het eerste leerjaar zijn de leerlingenaantallen niet bekend, omdat het vak niet tot de basisvorming behoort.

we aan te stippen dat het aandeel in de eerste graad heel wat hoger ligt dan in de tweede en vooral de derde graad. Zelfstandige gymnasia zijn in Vlaanderen zo goed als onbestaande. Latijn en – het veel minder gekozen – Grieks zijn optievakken die worden aangeboden in het secundair onderwijs met “doorstroomfinaliteit”, dat vergelijkbaar is met het Nederlandse vwo.

In Vlaanderen wordt onderwijs aangeboden door twee onderwijsnetten. Er is het officiële net (equivalent aan het openbaar onderwijs in Nederland) en het vrij gesubsidieerd onderwijs (equivalent aan het bijzonder onderwijs in Nederland). Het vrije onderwijsnet is de grootste speler, en dat geldt bij uitstek voor klassieke talen: volgens de meest recente gegevens liep 85,8% van de Vlaamse leerlingen Latijn school in het vrije onderwijs (Classica Vlaanderen, 2023).

Belangrijk om te weten is dat nagenoeg alle scholen van dit vrije onderwijsnet die het vak Latijn aanbieden, deel uitmaken van de koepel “Katholiek Onderwijs Vlaanderen” (KathOndVla) en de leerplannen van deze koepel volgen. In het officiële onderwijsnet behoren de meeste scholen tot het Gemeenschapsonderwijs (GO!). Verder zijn er nog de scholen van het stedelijk en gemeentelijk onderwijs (OVSG) en de scholen van het provinciaal onderwijs (POV). Omdat het GO! richtinggevend is voor het onderwijs klassieke talen in het officiële onderwijsnet, laten we de leerplannen van laatstgenoemde koepels buiten beschouwing. Ook een exhaustieve beschrijving en vergelijking van de leerplannen van de twee grootste onderwijskoepels, KathOndVla en GO!, zou het bestek van dit artikel te buiten gaan. In de volgende subsectie beperken we ons daarom tot enkele historische tendensen die nodig zijn voor een goed begrip van de resultaten van de enquête.

7.2.2 De geschiedenis van de belangrijkste leerplannen Latijn in een notendop

7.2.2.1 Latijn in het officieel onderwijs: de leerplannen van het Gemeenschapsonderwijs

Leerplannen – of “didactische fiches”, zoals ze vandaag in het GO! worden genoemd – omvatten telkens de leerinhouden en leerdoelen per graad. Zo vormen ze een concretisering van de minimumdoelen die door de overheid zijn vastgelegd. Minimumdoelen en leerplannen hebben een verplichtend karakter omdat ze een belangrijk, zij het niet alleenzalmakend instrument vormen van de onderwijsinspectie. In Vlaanderen bestaan immers geen centrale

eindexamens. Het maken en afnemen van examens is dus volledig in handen van vakgroepen, of zelfs individuele leraren. In het officiële onderwijsnet vormen scholen met slechts één leraar Latijn voor alle leerjaren geen uitzondering.

In Vlaanderen is het GO! in de jaren 1960 voorloper geweest in het omhelzen van een leatuurgerichte en inductieve aanpak. De voorschriften van het vernieuwde leerplan van 1963 waren behoorlijk revolutionair (Bosteels, 1997).⁷⁵ De meeste principes van de GO!-leerplannen voor de eerste graad stemmen overeen met die van de *reading method* die internationaal vooral bekend is geworden door latere Britse leergangen als *Oxford Latin Course*. Volgens dit paradigma is lezen niet enkel het doel maar ook het middel: leerlingen krijgen van meet af aan aantrekkelijke Latijnse verhalen voorgeschoteld die hen in contact brengen met de antieke cultuur, en al lezend ontdekken ze de voornaamste grammaticale regels (Gay, 2003; Hunt, 2016).

Een opmerkelijk verschilpunt tussen de visie van het GO! en de hierboven geschetste, gangbare invulling van de *reading method* is het voorschrift van het GO! om vanaf het begin “primaire tekstbronnen” te lezen, hoogstens licht aangepaste originele Latijnse teksten dus. Van dit principe zijn de GO!-leerplannen voor de eerste graad nooit meer afgeweken. De leergang *Tolle Lege* (Coghe et al., 2006) die ook in Nederland niet geheel onbekend is (De Snoo, 2011), past dit principe uiterst consequent toe.⁷⁶ Het meest recente GO!-leerplan definieert de gewenste tekstsoort als volgt: “Het doel is dat leerlingen via (indien nodig) vereenvoudigd (zo dicht mogelijk bij de Latijnse tekst) tekstmateriaal kennis maken met de periode⁷⁷ waarin Latijn actief werd gebruikt” (GO!, 2020) .

Een tweede kenmerk dat de leerplannen van het GO! van oudsher typeert, is dat ze naast een visie en een doelenkader ook een didactiek opleggen en dit niet alleen met betrekking tot het grammaticaonderwijs (inductie) en de authenticiteit van de teksten, maar eveneens wat

⁷⁵ Bosteels (1997, p. 33) twijfelt er niet aan dat de Vlaamse leerplanmakers (mede) geïnspireerd waren door de baanbrekende opvattingen van de Nederlandse classicus A.G. de Man. Regtuit (2015, p. 288) stipt aan dat De Mans aanbevelingen in het Nederlandse onderwijs daarentegen weinig weerklank hebben gevonden.

⁷⁶ In Vlaanderen worden deze lesboeken enkel nog fragmentarisch gebruikt omdat ze de grammaticale minimumdoelen van het actuele GO!-leerplan voor de 1^e graad (2020) ver overschrijden. De grote meerderheid van de leraren in het officieel onderwijs werkt met zelfontworpen syllabi.

⁷⁷ Zoals uit het tweede deel van de definitie blijkt, is er in de actuele Vlaamse minimumdoelen en leerplannen ook aandacht voor Latijn uit andere periodes dan de Oudheid. Een uitgebreide motivering en verkenning van de didactische mogelijkheden van postklassieke, middeleeuwse, vroegmoderne etc. Latijnse teksten is te vinden in Velle en Maes (2021).

betreft de leesmethode. Verplicht is de zogenaamde functionele of lineaire lectuur, Latijn lezen van links naar rechts dus.

7.2.2.2 Latijn in het vrij onderwijs: de leerplannen van Katholiek Onderwijs Vlaanderen

De leerplannen van het veel grotere vrije onderwijsnet, daarentegen, hebben leraren altijd meer vrijheid gelaten. Bovendien hebben ze heel wat langer de sporen gedragen van de traditionele grammatica-vertaalmethode (*grammar-translation method*). Oudere onderbouwleerplannen van Katholiek Onderwijs Vlaanderen mogen dan wel al stellen dat “een tekst begrijpen naar vorm en inhoud” het eerste doel is van een studie klassieke talen en dat “taalkennis beperkt moet blijven tot het minimum dat nodig is om een tekst te verstaan” (NVKSO, 1989, p. 5), in de praktijk is altijd een aanzienlijk deel van de lestijd in de eerste drie jaren naar grammaticaonderwijs blijven gaan. Daarbij werd veel aandacht besteed aan het inoefenen van losse vormen zonder zins- of tekstverband. Net zoals in Nederland bleken vele leerlingen ondanks deze intensieve taalstudie (te) weinig toegerust om in de bovenbouw vlot authentieke teksten te kunnen lezen. Deze discrepantie, het “al decennialang onderkende transferprobleem”, om het met de woorden van *de blauwe broer* te zeggen (Kroon & Sluiter, 2010, p. 38), heeft in de katholieke onderwijskoepel geleid tot een radicale vernieuwing vanaf 2011.

In het vernieuwde leerplan voor de eerste graad werden de doelstellingen in het hoofdstuk lectuur geordend volgens taalreflectie, lectuurreflectie en cultuurreflectie, en was cultuur als afzonderlijk onderdeel verdwenen. Deze indeling wilde leraren stimuleren om vanaf het eerste begin tekstpassages “niet enkel als ‘verhaaltjes’ maar ook als (min of meer) literaire teksten te benaderen” (VVKSO, 2011, p. 37), zelfs al zijn ze niet-authentiek – in het vrije onderwijsnet zijn leraren immers niet verplicht om vanaf de eerste graad authentieke teksten te lezen. Van de nieuwe indeling en de oproep om aan lectuur een centrale plaats toe te kennen, lagen wellicht niet veel leraren wakker. Wat paradoxaal genoeg voor het leesonderwijs veel meer gevolgen had, was een wijziging op grammaticaal vlak.

Wat was de essentie van deze wijziging? Om het grammaticaonderwijs van de klassieke talen meer te laten aansluiten bij dat van de moderne talen, werden leraren verplicht om enkele basisprincipes uit de valentiegrammatica te hanteren, met als belangrijkste het systematisch onderscheiden van morfologische vorm, syntactische functie en semantische rol (Vereeck, 2018). Gedaan met classicistische benamingen zoals “genitief van hoedanigheid”, welkom aan

het inzicht dat in *magnae sapientiae puella* het zinsdeelstuk *magnae sapientiae* een bijvoeglijke bepaling is die de kern *puella* nader omschrijft. Deze terminologische ommekeer noopte lesboekenmakers tot zo'n grote make-over dat ze ook op het vlak van lectuur en oefeningen met een schone lei zijn herbegonnen.

Vanaf 2011 is het landschap van Vlaamse leergangen Latijn voor de onderbouw (jaren 1 tot 3) van het katholiek onderwijs fundamenteel veranderd.⁷⁸ Lees: een ruim aanbod Latijnse teksten, een sterke link tussen de teksten en de cultuurinhouden, en een verschuiving in de taalopdrachten. De focus lag minder op losse vormen, en meer op “contextuele oefeningen”: oefeningen in de vorm van Latijnse zinnen die samen een verhaal vormen dat aansluit bij het betreffende cultuurthema.

Opmerkelijk is dat de leerplanvoorschriften met betrekking tot het gebruik van de valentiegrammatica, die in 2011 zo'n katalyserende rol hebben gespeeld, al bij een volgende leerplanvernieuwing zijn versoepeld. Verder onderscheidt het meest recente leerplan Latijn eerste graad van KathOndVla zich van het voorgaande door een nog sterkere focus op leesvaardigheid.⁷⁹ De indeling van de in totaal 27 leerplandoelen in de volgende categorieën spreekt boekdelen (onze cursivering): taalsysteem ter ondersteuning van *leesvaardigheid*, *leesstrategieën* en elementair *tekstbegrip*, *tekstreflectie*, cultuurreflectie als uitdieping van *lectuur* (KathOndVla, 2019).

Wat betreft de leesmethode tenslotte hebben de leerplannen van KathOndVla zich nooit expliciet uitgesproken voor deze of gene methode. Terwijl leraren in het GO! hun leerlingen al ettelijke decennia het Latijn enkel van links naar rechts (mogen!) leren lezen, hebben de meeste latinisten in het vrije onderwijsnet lang vastgehouden aan de traditionele analytische methode (in elke zin eerst de persoonsvorm en het onderwerp bepalen). Maar vandaag stellen we vast dat ook in het vrij onderwijs de vakwerkgroepen die een zuiver analytische leesmethode hanteren hoe dan ook in de minderheid zijn. Dit komt omdat Vlaamse lerarenopleidingen, nascholingen en publicaties al enige tijd een leesdidactiek promoten waarin een belangrijke rol is weggelegd voor het maken van inhoudelijke voorspellingen. Sinds de uitwerking door

⁷⁸ Na de leerplanvernieuwing in de eerste graad van KathOndVla volgden in 2013 en 2015 de tweede en derde graad. Van de vier nieuwe of grondig herziene lesboeken die vanaf 2011 zijn gelanceerd, wordt anno 2024 op Vlaamse katholieke scholen in de eerste plaats *Pegasus Novus* gebruikt, gevolgd door *Ars Legendi*.

⁷⁹ Tegelijk zijn enkele grammaticale leerinhouden doorgeschoven naar de tweede graad. Eenzelfde verschuiving is overigens ook waar te nemen in het GO! als we de leerplannen eerste graad van 2004 en 2020 met elkaar vergelijken.

Kristien Hulstaert (2016) van de positionele methode (Verbaal, 2015) voor het middelbaar onderwijs is in het werkveld ook de interesse in het kolometrisch aanbieden van (authentieke) Latijnse tekstpassages sterk toegenomen.⁸⁰

7.3 Onderzoeksmethode

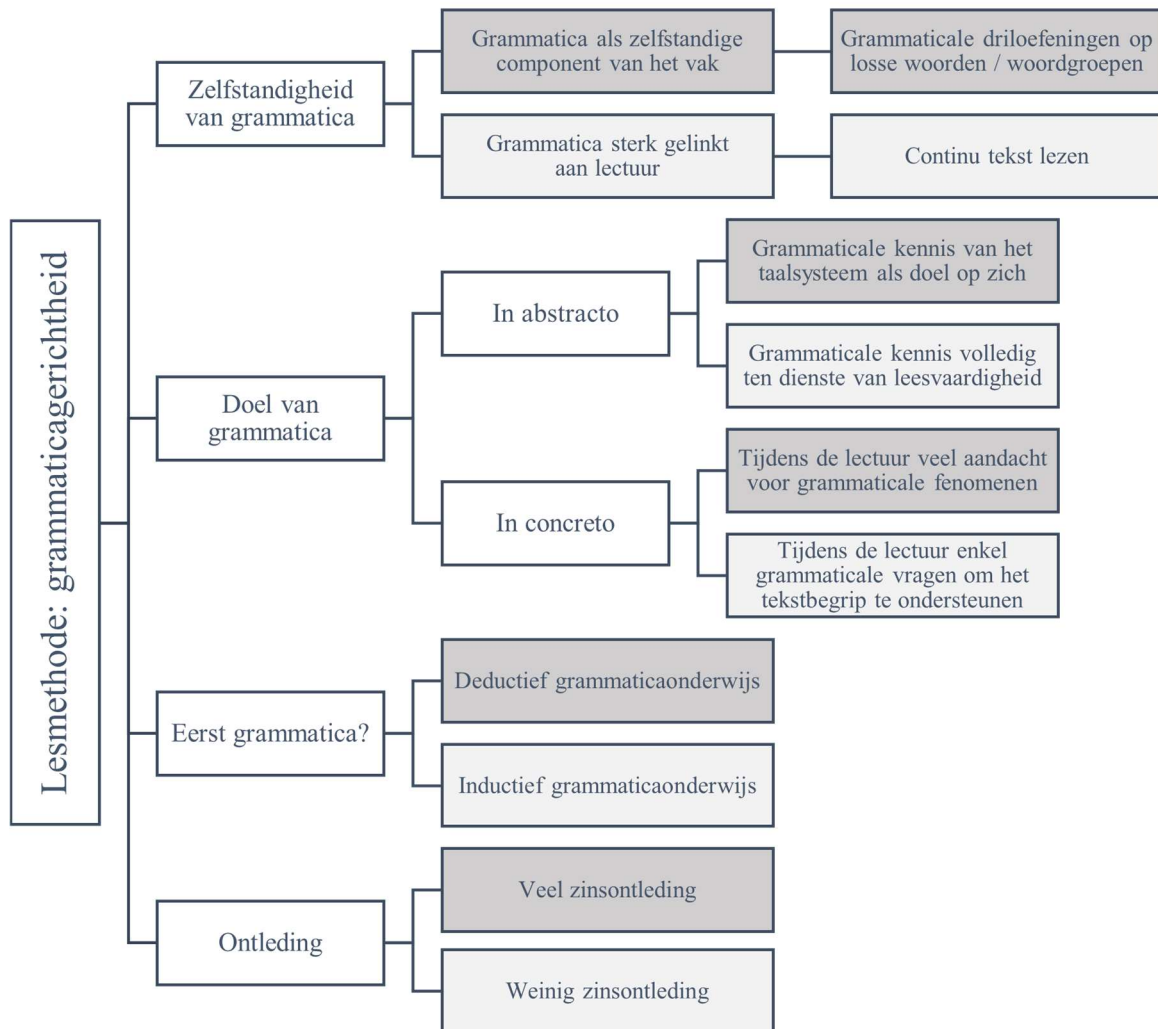
7.3.1 Instrumenten en dataverzameling

Voor dit onderzoek hebben we via verschillende wegen data verzameld, van zowel kwantitatieve als kwalitatieve aard. Het hoofdinstrument is een door ons ontworpen vragenlijst met een twintigtal stellingen die betrekking hebben op de handelingen en attitudes van leraren Latijn. Per stelling moeten respondenten op een schaal van zeven punten aangeven in welke mate ze het ermee eens zijn. Er wordt een onderscheid gemaakt naargelang de graad; leraren die zowel in de eerste als in de tweede graad lesgeven, duiden bij iedere stelling dus twee keer een antwoord aan. Vooraf wordt gevraagd om uit te gaan van klassikaal onderwijs onder normale omstandigheden (met andere woorden, om het uitzonderlijke afstandsonderwijs tijdens de coronacrisis buiten beschouwing te laten). De stellingen zijn ingedeeld in twee reeksen. Halverwege en aan het einde van elke reeks zijn er ook vrijblijvende commentaarvelden voorzien waar respondenten extra uitleg en bedenkingen kunnen schrijven.

De eerste reeks bestaat uit tien stellingen die samen één schaal vormen om de “grammaticagerichtheid” van een onderwijspraktijk te meten. Voor de ontwikkeling van deze schaal voerden we eerst een verkennende literatuurstudie uit en organiseerden we een focusgroep over onderwijspraktijken voor Latijn. Op basis hiervan definieerden we het concept grammaticagerichtheid als een continuüm van aan de ene kant zeer lektuurgericht en beïnvloed door de *reading method* tot aan de andere kant zeer grammaticagericht en beïnvloed door de grammatica-vertaalmethode. Na talloze rondes van validering en verfijning kwamen we

⁸⁰ De positionele methode is recent weer onder de aandacht van de *Lampas*-lezers gebracht door Regtuit (2024) en Van der Keur (2024).

uiteindelijk tot vijf indicatoren (zie Figuur 5), die hun weerslag vonden in tien stellingen (zie zowel Figuur 5 als Tabel 3). De schaal vertoont een goede interne consistentie en samenhang.⁸¹



Figuur 5 Indicatoren voor grammaticagerichtheid: zelfstandigheid van grammatica, doel van grammatica in abstracto, doel van grammatica in concreto, eerst grammatica? en ontleding

⁸¹ Cronbach's α , een veelgebruikte betrouwbaarheidsmaat voor vragenlijsten, is hier gelijk aan 0,8 voor de eerste graad en 0,7 voor de tweede graad, wat wijst op een goede tot zeer goede interne consistentie. Factoranalyse leverde een één-factoroplossing op, wat erop wijst dat alle stellingen daadwerkelijk eenzelfde onderliggend concept bevragen.

Tabel 3 Stellingen reeks 1

1.	Voor mij staat het aanleren van grammatica volledig ten dienste van leesvaardigheid.
2.	Haast iedere les ben ik met de leerlingen een Latijnse tekst aan het lezen.
3.	Bijna iedere Latijnse zin ga ik met de leerlingen volledig syntactisch ontleden.
4.	Ik onderwijs nieuwe grammaticale items voordat we een tekst beginnen te lezen waarin deze items voorkomen.
5.	Ik behandel grammaticale kennis van het Latijnse taalsysteem als een doel op zich van het vak.
6.	In ieder hoofdstuk geef ik driloefeningen op losse woord(groep)en aan alle leerlingen.
7.	Bij de lectuur van een tekst besteed ik veel aandacht aan grammaticale fenomenen, ook als de betekenis van de zin al duidelijk is.
8.	Tijdens klassikale lectuur besteed ik een groot deel van de tijd aan de grammaticale functies van de zinsdelen.
9.	Tijdens klassikale lectuur stel ik grammaticale vragen enkel met de bedoeling het tekstbegrip te ondersteunen.
10.	Ik leid nieuwe grammaticale items inductief af uit een reeds gelezen tekst.

Onze schaal laat toe om grammaticagerichtheid te kwantificeren en een grammaticagerichtheidsscore (vanaf nu GG-score) te berekenen. Dit wordt gedaan door de antwoorden (gaande van “sterk oneens” tot “sterk eens”) numeriek te coderen,⁸² de scores van alle vragen bij elkaar op te tellen, en het totaal op 100 te zetten. Lage scores (< 50) leunen dan aan bij de *reading method* en hoge scores (> 50) bij sterk grammaticagericht onderwijs. Bij weinig uitgesproken scores (in de buurt van 50) kunnen we gewagen van een eclectische methode. Grammatica en lectuur sluiten elkaar natuurlijk niet uit: de vraag is niet óf leerlingen nog grammatica krijgen aangeleerd, maar wel hoe het aanreiken van grammaticale kennis zich in het onderwijsproces verhoudt tot het verwerven van leesvaardigheid.

De stellingen van de tweede reeks vormen geen samengestelde schaal zoals de eerste reeks, maar bevragen daarentegen elk één concept. We kozen voor een handvol saillante thema's in het hedendaagse klassieketalenonderwijs, waarvan we de meeste echter onvermeld moeten laten met het oog op de *brevitas* en *perspicuitas* van dit artikel.⁸³ We zullen ons verderop daarom beperken tot twee thema's en hun bijbehorende stellingen die een mooie aanvulling vormen op het hoofdthema van grammaticagerichtheid, namelijk actieve kennis van basismodellen en

⁸² De antwoorden op stellingen 3, 4, 5, 6, 7 en 8 werden positief gecodeerd van 0 tot en met 6. De antwoorden op stellingen 1, 2, 9 en 10 werden omgekeerd gecodeerd van 6 tot en met 0.

⁸³ De stellingen van reeks 2 zijn thematisch gegroepeerd als volgt: stellingen 1 tot 2 houden verband met de verwachting van actieve taalkennis, stelling 3 met geïntegreerde evaluatie, stellingen 4 tot 6 met de communicatieve methode, stelling 7 met cultuuronderricht, en stelling 8 tot 12 met gevarieerde aspecten van lectuur (lay-out, woordenboekgebruik, vertaling, woordhulp en de toetsing van niet-bestudeerde tekst).

cultuuronderricht. Ook voor reeks 2 hebben we de stellingen in meerdere rondes gevalideerd en verfijnd alvorens de finale versie van deze reeks vast te leggen.

De vragenlijst biedt een rijke bron van data: enerzijds is er de kwantitatieve respons op de stellingen, anderzijds is er de kwalitatieve toelichting in de vier commentaarvelden. In het kader van een ander onderzoek werd bovendien een kortere versie van de vragenlijst, bestaande uit enkel de eerste reeks stellingen, in een beperkt aantal scholen afgenomen bij de volledige vakgroep Latijn van het tweede jaar. Ter voorbereiding op de vragenlijst hebben we in de loop van 2021 ook zes kwalitatieve een-op-een- interviews uitgevoerd met ervaren leraren. We stelden hiervoor zelf een semigestructureerde interviewleidraad op.

7.3.2 Steekproefname en deelnemers aan de enquête

De vragenlijst werd online afgenomen bij een gelegenheidssteekproef van Vlaamse leraren Latijn die in jaren 1 tot 4 doceren. We hebben de link naar de vragenlijst verspreid via vaktijdschriften, flyers, nieuwsbrieven en sociale media. De enquête liep van januari tot april 2022 en kende een hoge responsgraad. Er namen 229 leraren deel, wat in ieder geval neerkomt op meer dan een kwart van de doelpopulatie.⁸⁴ Hiervan gaven er destijds 64 enkel les in de eerste graad, 75 enkel in de tweede graad, en 90 in de beide graden. Met andere woorden, we beschikken over deels overlappende steekproeven van 154 respondenten voor de eerste graad en 165 respondenten voor de tweede graad.

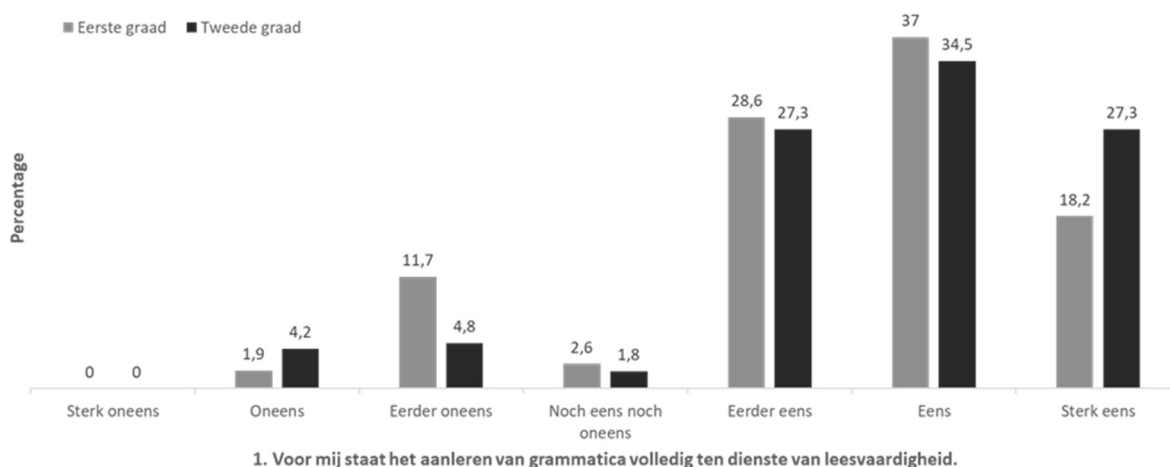
Niet alleen bereikten we een representatief aantal leraren, diverse sub- groepen in de populatie zijn ook goed vertegenwoordigd. De leeftijd van de respondenten varieert van 21 tot 67 jaar, de twee uitersten voor Vlaamse leraren, met 41 jaar als gemiddelde. Geografisch is er een mooie spreiding over alle regio's van Vlaanderen. Dat de provincie Oost-Vlaanderen maar liefst 71 deelnemers levert en dus enigszins oververtegenwoordigd is, mag niet verwonderen aangezien de enquête uitging van de Universiteit Gent. Qua onderwijsnet weerspiegelt de verhouding in de steekproef die in de populatie: 80% van de deelnemers is werkzaam in het vrije onderwijsnet, 20% in het officiële net.

⁸⁴ Exacte cijfers over het aantal leraren Latijn zijn niet beschikbaar. Volgens informatie naar het Ministerie van Onderwijs en Vorming zijn er in Vlaanderen in totaal “900 à 1000” leraren Latijn (VLOT, 2023); de groep die enkel in de eerste vier jaren lesgeeft zal dus nog wat kleiner zijn.

7.4 Bevindingen⁸⁵

7.4.1 Lectuur centraal in beide onderwijsnetten

Algemeen kunnen we stellen dat de Vlaamse lespraktijk Latijn in de eerste vier jaren (veel) meer leatuurgericht dan grammaticagericht is. Dat leatuur het uiteindelijke doel is van het vak, daarover bestaat een noemenswaardige eensgezindheid: respectievelijk is 83,8% (eerste graad) en 89,1% (tweede graad) van de respondenten het in meer of mindere mate eens met de eerste stelling, die zegt dat het aanleren van grammatica ten dienste moet staan van de leesvaardigheid van de leerlingen (zie Figuur 6).



Figuur 6 Verdeling respons op stelling 1

Hoe zit het nu met de grammaticagerichtheidsscore, die we hebben berekend op basis van de eerste tien stellingen samen? De gemiddelde GG-scores bedragen 39,13 in de eerste graad en 33,47 in de tweede graad.⁸⁶ De *reading method* heeft met andere woorden de overhand. De GG-score wordt onmiskenbaar beïnvloed door het onderwijsnet (zie Tabel 4): gemiddeld scoorden de leraren uit het vrije net zo'n 14 punten hoger – dus meer richting

⁸⁵ In de lopende tekst zullen we de resultaten bespreken en interpreteren; de nodige statistische onderbouwing vindt de lezer in voetnoten.

⁸⁶ Dit verschil tussen de graden is niet groot, maar wel statistisch significant (t-toets voor deels overlappende samples; $t_{184} = 4,53$; $p < 0,001$).

grammaticagericht onderwijs – in de eerste graad en zo’n 7 punten in de tweede graad.⁸⁷ Wanneer we opsplitsen volgens net, wordt snel duidelijk dat er alleen in het vrije net een significant verschil bestaat tussen de graden⁸⁸ en dat in het officiële net de GG-score quasi constant blijft over de jaren heen.

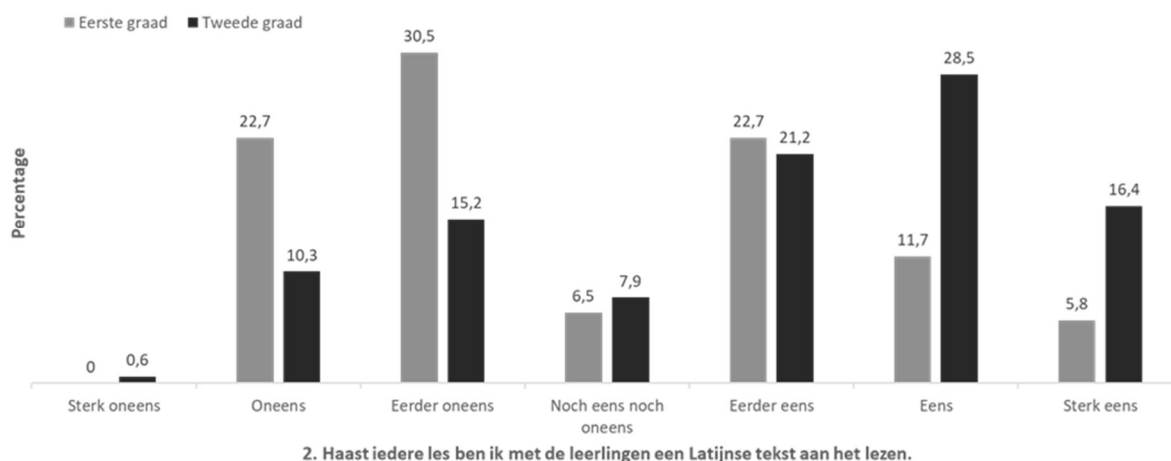
Tabel 4 GG-score volgens onderwijsnet en graad; gemiddelde, bereik, en aantal respondenten

Onderwijsnet		GG-score	
		Eerste graad	Tweede graad
Officiële net	Gemiddelde	28,29	28,10
	Minimum	5,00	6,67
	Maximum	60,00	58,33
	N	35	36
Vrije net	Gemiddelde	42,32	34,97
	Minimum	0,00	0,00
	Maximum	75,00	68,33
	N	119	129

Dat de gemiddelde GG-score lager is in het officiële onderwijsnet en dat dit verschil zich sterker manifesteert in de eerste graad dan in de tweede graad ligt in de lijn van de verwachting. De leerplannen van het GO! schrijven immers al zes decennia een inductieve taaldidactiek voor, die is geënt op (de voornaamste principes van) de *reading method*. Dit uit zich onder meer in de lestijdbesteding: van de GO!-leraren in de eerste graad is de helft het (eerder tot sterk) eens met de tweede stelling, “Haast iedere les ben ik met de leerlingen een Latijnse tekst aan het lezen” (zie Figuur 7 voor de respons over beide netten heen).

⁸⁷ Deze resultaten bieden sterk bewijs, in het bijzonder voor de eerste graad (ANOVA; $F_{(1,152)} = 27,42$; $p < 0,001$) maar ook voor de tweede graad (ANOVA; $F_{(1,163)} = 7,50$; $p = 0,007$), dat op populatieniveau de grammaticagerichtheid verschilt naargelang het onderwijsnet.

⁸⁸ Dit wordt bevestigd door statistische inferentie (t-toets voor deels overlappende samples; $t_{148} = 5,31$; $p < 0,001$).



Figuur 7 Verdeling respons op stelling 2

In het vrije onderwijsnet daarentegen is de integratie van taal, lectuur en cultuur op operationeel niveau, in de concrete doelen, een veel recenter fenomeen in de eerste graad (zie Sectie 7.2.2.2). Onze steekproef wijst evenwel op een snelle inhaalbeweging. Amper elf jaar na de fundamentele leerplanwijziging in 2011 vinden we ook de gemiddelde Vlaamse latinist die werkzaam is in de eerste graad van het vrije net terug aan de lectuurgerichte zijde van het spectrum, met een GG-score van 42. De KathOndVla-leerplanvernieuwers schijnen dus in hun missie te zijn geslaagd. Volgens ons hebben vooral de volgende factoren hierin een rol gespeeld:

- (1) In de eerste graad moeten zo goed als alle leerplandoelen met betrekking tot taal, lectuur en cultuur worden bereikt bij bestudeerde en bij analoge niet vooraf bestudeerde Latijnse teksten (KathOndVla, 2019, p. 10). Wil een leraar hieraan voldoen, dan kan hij niet anders dan aan lectuur een belangrijke plaats toe te kennen in zijn lespraktijk.
- (2) In de vormgeving van hun lespraktijk laten leraren zich gewoonlijk sterk leiden door de gekozen leergang, zeker wat betreft Latijnse tekst- fragmenten en oefeningen in de aanvangsjaren.⁸⁹ Zowel *Pegasus Novus*, de lesboekenreeks die in de eerste graad door 85% van de leerlingen in het vrij onderwijs wordt gebruikt, als de minder gebruikte reeks *Ars Legendi* zijn zeer goed afgestemd op de lectuurgerichte visie van de nieuwe leerplannen. Met name de “contextuele oefeningen” in deze lesboeken hebben er

⁸⁹ Zo stellen bijvoorbeeld De Ferrante en Van Oeveren (2022, p. 6) over Nederlandse docenten klassieke talen. Wat de Vlaamse docenten betreft, is uit een recente enquête onder een duizendtal taalleraren (Arteel et al. 2022, p. 47) gebleken dat ook zij hun lessen voornamelijk ontwikkelen op basis van leermiddelen zoals handboeken (80,7%). Twee op de drie (67,2%) baseert zich (ook) op de leerplannen.

volgens ons zeker toe bijgedragen dat ook in de eerste graad van het vrije net 37% van de respondenten beaamt dat lectuur bijna elke les aan bod komt (zie opnieuw Figuur 7). Zoals een respondent getuigt: “In de methode die wij gebruiken (*Pegasus Novus*) zijn de meeste oefeningen als een tekst opgevat. Dat maakt dat er veel tekst(jes) gelezen worden.”

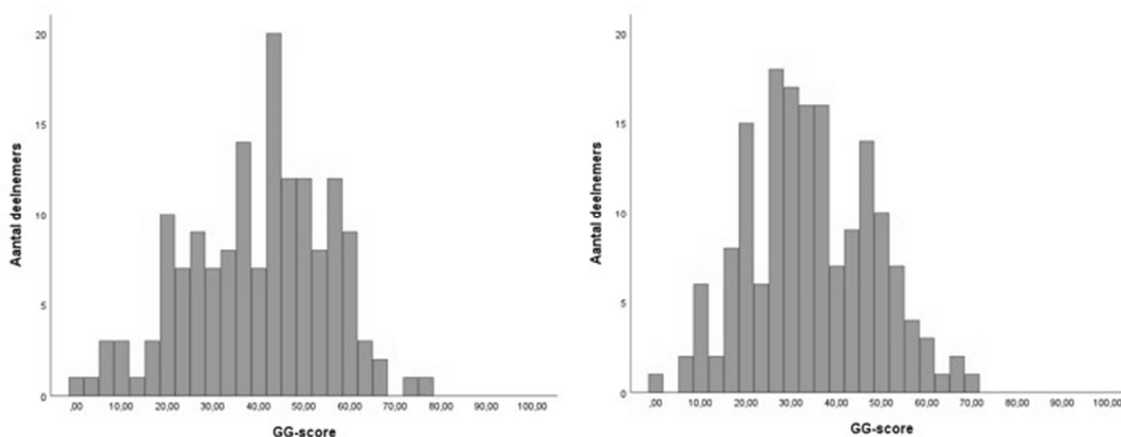
Dat de leerplannormen globaal genomen trouw worden uitgevoerd, betekent uiteraard niet per definitie dat uitvoerders het intrinsiek allemaal met de recente visie eens zijn. Af en toe weerklinkt in de commentaren en interviews – zelfs al sluit de GG-score van de leraar in kwestie nauw aan bij het gemiddelde – heimwee naar het verleden. Tegelijk is het plichtsgevoel groot. Een van de deelnemers aan de interviews stelt het als volgt: “dat je conform het leerplan moet werken, dat is nu eenmaal de wet. *Dura lex, sed lex.*”

7.4.2 Grote verschillen tussen individuele leraren

Groepsgemiddelden zeggen ook lang niet alles. Laten we daarom even kijken naar de spreiding van de GG-scores. De afstand tussen de uiterste scores, die overigens aan beide kanten bereikt worden in het vrije net, bedraagt 75 punten (zie Tabel 4 hierboven voor het bereik). Leraar A met een GG-score van 75 zweert bij een traditionele deductieve aanpak en besteedt behoorlijk wat les- en leestijd aan grammaticale analyse; leraar B met een GG-score van 0 beschouwt lezen tegelijk als einddoel en startpunt van het onderwijs en stemt zijn lespraktijk volledig af op leesvaardigheid. Op beide profielen komen we verder terug.

Evengoed treffen we alle mogelijke tussenposities en bijhorende GG-scores aan (zie Figuur 8 voor de verdeling over de netten heen). Scores van minder dan 20 of meer dan 60 komen beduidend weinig voor. In de eerste graad observeren we dat het gros van de scores zich binnen deze twee grenzen bevindt en dit ruwweg gelijk verspreid, zij het met een duidelijk zwaartepunt begin 40.⁹⁰ In de grafiek van de tweede graad zien we dat zich twee pieken aftekenen, eentje rond de 30 en eentje rond de 50, die zich respectievelijk laten identificeren als *reading method* en eclectische methode. Kortom, hoewel de gemiddelden wijzen op “lectuurgerichtheid” (zie Sectie 7.4.1), zijn er ook heel wat leraren van wie de respons getuigt van een eerder eclectische benadering.

⁹⁰ Van 20 tot 30: 18,7% van de scores; van 30 tot 40: 18,7%; van 40 tot 50: 28,3%; van 50 tot 60: 17,4%.



Figuur 8 Verdeling GG-scores in de eerste graad (links) en in de tweede graad (rechts)

Hoe kunnen we de grote interindividuele variatie in grammaticagerichtheid duiden? Leeftijd en regio blijken alvast geen rol te spelen. Het onderwijsnet verklaart een deel.⁹¹ Voorts stelden we vast dat er in de tweede graad van het vrije net een – weliswaar zeer lage – negatieve correlatie⁹² bestaat met het aantal gevolgde nascholingen: binnen deze categorie leraren is dus een zwakke tendens merkbaar dat wie vaker nascholingen bijwoont minder grammaticagericht is in zijn/haar lespraktijk.

Ongetwijfeld wordt de keuze voor deze of gene aanpak in hoofdzaak bepaald door tal van persoonsgebonden factoren die niet in deze eerste, exploratieve enquête bevraagd zijn. In het bijzonder denken we dan aan afspraken binnen de vakwerkgroep en de school. Het lijkt aannemelijk dat leraren hun lespraktijk afstemmen op het pedagogische project en het evaluatiebeleid van hun school en op de specifieke leerlingenpopulatie, bijvoorbeeld het aantal leerlingen dat Nederlands niet als thuistaal heeft. Ook onderstrepen de leerplannen van beide onderwijsnetten vanouds het belang van duidelijke afspraken binnen de vakwerkgroep, in de eerste plaats met betrekking tot de leesmethode die leerlingen krijgen aangeleerd (zie Sectie 7.2.2). Of de leesmethode verplicht lineair is (GO!) dan wel in onderling overleg vrij te kiezen door de vakwerkgroep (KathOndVla), ze moet over de jaren heen consequent worden toegepast. Nochtans blijkt uit de data van de scholen waar de kortere versie van de vragenlijst werd afgenomen, dat ook binnen een vakwerkgroep van eenzelfde school zeer uiteenlopende

⁹¹ Regressieanalyse wees uit dat onderwijsnet als enige variabele een significante predictor is voor GG-score op het 0,01 significantieniveau. Per graad verklaart het net respectievelijk rond de 15% en 5% van de totale variabiliteit in GG-scores.

⁹² $r = -0,18$; $p = 0,039$

visies en praktijken leven wat betreft de verhouding tussen grammaticale kennis en leesvaardigheid: individuele verschillen van 20 à 30 punten op onze GG-schaal zijn geen uitzondering.⁹³

7.4.3 Extreme lectuurgerichtheid en de Ørbergmethode

Iets anders dat opvalt aan Figuur 8, is dat de uiterste rechterzijdes van de grafieken leeg blijven. Extreem hoge scores van 80 of meer komen in het geheel niet voor in onze steekproef, extreem lage van 20 of minder daarentegen wel.⁹⁴ Het zal niet verbazen dat we bij deze lage scores verhoudingsgewijs meer leraren van het officieel onderwijs aantreffen, terwijl zij bij de hoogste scores van boven de 60 afwezig zijn. De voorrang van lectuur op grammatica zit immers al decennialang in het DNA van de GO!-leerplannen. Of zoals een respondent het uitdrukt: “De visie van het GO! is net dat lectuur (van authentieke!) teksten voorop staat en dat je tijdens een lectuurles enkel grammaticale vragen stelt als dat nodig is voor het tekstbegrip. Daarna worden grammaticale regels geïnduceerd en ingeoefend! Ik kan niet begrijpen dat je dat anders zou aanpakken...”

Maar wanneer we de zeven allerlaagste GG-scores (viermaal 5, eenmaal 3 en tweemaal 0!) onder de loep nemen, doen we een opmerkelijke vaststelling met betrekking tot het gebruikte leermateriaal en onderwijsnet. Dit aspect vormde geen expliciet onderdeel van de bevraging, onder meer omdat, zoals eerder aangehaald, de variatie in Vlaamse leergangen momenteel heel klein is en het gebruik ervan sterk samenvalt met het onderwijsnet. In de commentaarvelden wordt echter meer dan eens naar leerboeken verwezen. Wat blijkt er uit de commentaren? Leraren die volgens onze schaal extreem lectuurgericht werken, gebruiken in minstens vier van de zeven gevallen de lesboekenreeks *Lingua latina per se illustrata* van Hans Ørberg (*LLPSI*) en van deze *LLPSI*-leraren werkt drie op de vier in het vrije net.

Wereldwijd is *LLPSI* de enige leergang Latijn die bijna integraal is gestoeld op het principe van contextuele inductie. *LLPSI* was na een bescheiden lokale populariteit in de jaren 1980 en 1990 (Pijls, 1995) verdwenen uit Vlaanderen, maar is sinds vijf jaar weer terug, en wordt nu

⁹³ Een verschil van 30 punten komt overeen met twee standaarddeviaties... een behoorlijk groot verschil dus.

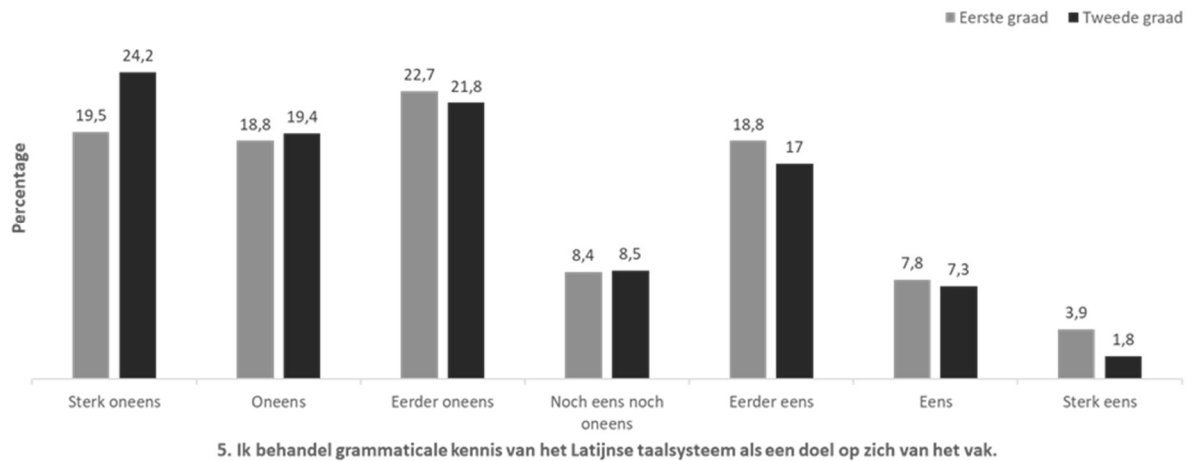
⁹⁴ In 8,4% (eerste graad) en 13,3% (tweede graad) van de gevallen bedroeg de GG-score minder dan 20, en in 4,5% (eerste graad) en 4,2% (tweede graad) van de gevallen zelfs minder dan 10.

voor zover wij weten op een zevental scholen gebruikt. De methode verschilt radicaal van de in het vrije net zo populaire leergang *Pegasus Novus*: zowel de Latijnse woordenschat als grammatica komen met veel meer geleidelijkheid en herhaling aan bod in teksten, die ook veel ruimer in aantal zijn. De opbouw van de teksten in *LLPSI* is zo doordacht dat leerlingen op natuurlijke wijze Latijn leren lezen, van gefingeerd bij aanvang tot originele passages op het einde van leerjaar 3. Woordbetekenissen en grammaticale fenomenen kunnen integraal worden begrepen uit de teksten, zonder dat Nederlandstalige woordhulp is voorzien, zoals gebruikelijk in andere methodes of syllabi.

Dit verklaart meteen ook waarom *LLPSI* minder gebruikt wordt in het officiële net en we, anders dan verwacht, de allerlaagste scores niet daar aantreffen. Zoals we in Sectie 7.2.2.1. hebben gezien, schrijft het GO!-leerplan naast het gebruik van inductie ook het lezen van “primaire tekstbronnen” voor – van teksten dus die niet zijn geschreven met geleidelijke taalverwerving als doel. Hoe moeilijker een tekst, des te meer zijn Nederlandstalige woordhulp of externe hulpmiddelen zoals een woordenlijst of grammaticaal naslagwerk vereist, en des te minder lijkt Latijn lezen bijgevolg op wat we gewoonlijk in andere talen onder “lezen” verstaan. In dat opzicht laat *LLPSI* met zijn niet-authentieke teksten zich minder goed rijmen met het leerplan van GO! dan met dat van KathOndVla. Niettemin geven ook twee respondenten uit het officiële net te kennen dat ze recent zijn gestart met de Ørbergmethode.

7.4.4 Welke plaats heeft zuivere grammatica nog?

De overwegende lectuurgerichtheid die het Vlaamse klassieketalenonderwijs typeert en waarmee we onze bevindingen begonnen zijn, moet enigszins worden genuanceerd. Zoals we in Sectie 7.4.2 reeds kort aanstipten, is naast de *reading method* in de meer strikte zin op de tweede plaats ook de eclectische methode prominent aanwezig in Vlaanderen. Het is zeker niet zo dat zuivere grammaticalesen uit de lespraktijk verdwenen zijn. Voor de beide graden geldt dat, hoewel slechts ongeveer een tiende van de leraren het uitgangspunt van de *reading method* dat grammatica ten dienste moet staan van leesvaardigheid durft te betwisten (stelling 1; zie ook Figuur 6), en helemaal niemand het er “sterk oneens” mee is, toch meer dan een kwart van de leraren aangeeft dat ze grammaticale kennis als een doel op zich behandelen van het vak Latijn (stelling 5; zie ook Figuur 9).



Figuur 9 Verdeling respons op stelling 5

Ook actieve beheersing van morfologische basiskennis wordt hoog in het vaandel gedragen. Een van de stellingen van de tweede reeks (zie Sectie 7.3.1) vroeg de respondenten om te kennen te geven of hun leerlingen in staat moeten zijn om een verbogen/vervoegde vorm van een basismodel te geven. Ongeacht de graad is de overgrote meerderheid, meer dan 85%, het erover eens dat paradigma's actief gekend moeten zijn. De mate waarin men het hiermee eens is, hangt tot op zekere hoogte samen met de mate van grammaticagerichtheid.⁹⁵

De overtuigend sterkste voorspeller voor de totale GG-score⁹⁶ is in beide graden de mate waarin men het eens is met de vierde stelling van reeks 1, “Ik onderwijs nieuwe grammaticale items voordat we een tekst beginnen te lezen waarin deze items voorkomen”. Anders gezegd, stel dat we van een individuele leraar weten hoe vaak hij/zij deductie toepast in de les, dan zouden we een beredeneerde gok kunnen wagen omtrent het belang dat diegene hecht aan grammatica: wie een vrij tot zeer sterke voorstander is van deductief taalonderwijs, gaat waarschijnlijk over de hele lijn vrij tot zeer grammaticagericht te werk, en omgekeerd. Of nog: leraren die hoog scoren op onze schaal vertonen doorgaans een voorliefde voor zelfstandig grammaticaonderwijs.

⁹⁵ Dit in de eerste graad meer ($r = 0,32$; $p < 0,001$) dan in de tweede graad ($r = 0,16$; $p = 0,045$).

⁹⁶ Dit zowel in de eerste ($r = 0,72$; $p < 0,001$) als de tweede graad ($r = 0,74$; $p < 0,001$).

7.4.5 Voorstanders van een meer grammaticagerichte aanpak

Laten we deze kant van het GG-spectrum verder exploreren. Wat zijn de praktijken en de visies van leraren die grammaticagerichter werken dan gemiddeld? Waarom vinden zij grammaticaonderwijs waardevol *an sich*? Een eerste veel gebruikt argument is dat inzicht in de Latijnse grammatica bijdraagt tot een groter taalbewustzijn en dat het de studie van andere vreemde talen vergemakkelijkt.⁹⁷ Sommige respondenten verwijzen daarenboven expliciet naar de troef die systematische taalstudie vormt als het erop aankomt het vak Latijn te promoten of te rechtvaardigen:

“Grammatica expliciet uitleggen is nodig en werkt voor leerlingen naar mijn aanvoelen nog steeds het best. Verder helpt het hen om taalbewust- zijn op te bouwen, ook met het oog op begrijpen en verwerven van andere talen, iets waar leerkrachten vaak mee pronken op opendeurdagen.”

Sommigen verruimen het transferargument van taal naar cognitie in het algemeen:⁹⁸ het grondig bestuderen van de structuur van de Latijnse taal zou de geesten van de leerlingen vormen en “kneden”. Een van de geïnterviewde leraren, die enorm betreurt dat de studie van het taalsysteem geen zelfstandige component meer uitmaakt in het eerstegraadsleerplan van KathOndVla, gaat zelfs zover om het bestaansrecht van het vak hiertoe te herleiden (licht geredigeerd citaat):

“Latijn heb je voor niets echt nodig, maar als een leerling dan toch de studie aangaat, laat ons die dan de waarde geven die het kan hebben. Waarom halen we dan de groenten uit de soep? Waarom laten we hen dan bouillon drinken?”

Een tweede argument waarmee leraren met een hoge GG-score hun stellingname dikwijls motiveren, is dat de “soepgroente” in het citaat hierboven, expliciet en systematisch

⁹⁷ Het bijkomende taalvoordeel dat latinisten zouden ondervinden van hun studie is onder leraren klassieke talen inderdaad een veelvoorkomend en breed gedragen argument, waarvoor vooralsnog echter weinig tot geen wetenschappelijke onderbouwing bestaat. Deze hypothese wordt momenteel in de Lage Landen voor het eerst grondig empirisch onderzocht. Zie Vereeck (2023) voor een voorstelling van dit lopende onderzoek.

⁹⁸ Hier geldt hetzelfde; zie voorgaande voetnoot.

grammaticaonderwijs dus, een *condicio sine qua non* vormt om Latijnse teksten te kunnen lezen.⁹⁹

Een van de respondenten die het oneens is met stelling 1 verwoordt het als volgt:

“Als grammatica louter ten dienste staat van het vertalen, wordt er geselecteerd op basis van frequentie van grammaticale fenomenen, wat volledig ten koste gaat van de consistentie, logica en helderheid van het systeem, waardoor leerlingen hun grammatica moeizamer in plaats van vlotter onder de knie krijgen. Het creëert een gevoel van willekeur en geeft aanleiding tot “radend” vertalen. Bovendien heb ik telkens weer ervaren dat een grammaticaal fenomeen dat los van een tekst is aangeleerd en ingeoefend, later veel makkelijker herkend en vlot vertaald wordt in tekst.”

Opvallend is dat deze leraar – geenszins als enige – “lezen” gelijkstelt aan “vertalen”. Hiermee wordt niet enkel ingegaan tegen het woordgebruik in de vragenlijst, waar we consequent voor de term “lezen” hebben geopteerd, maar ook tegen de terminologie en de visie van de actuele leerplannen. Zij gebruiken in hun doelstellingen enkel de werkwoorden “lezen” (GO!) of “begrijpen” (KathOndVla), en stellen dat vertalen slechts één manier is om tekstbegrip te demonstreren.

Een derde argument voor een traditionele, sterk grammaticagerichte aanpak, is dat het doelpubliek ertoe noopt of erom vraagt. Respondenten en deelnemers aan de interviews verwijzen naar “zwakke leerlingen” die grammaticaal houvast nodig hebben, naar het toenemende aantal anderstalige leerlingen in hun klas, of naar het feit dat de huidige generatie leerlingen minder voorkennis heeft op vlak van taalbeschouwing.¹⁰⁰ Aan de andere kant stellen verscheidene leraren in onze dataset dat leerlingen Latijn sowieso geboeid zijn door grammatica: “een latinist vindt dat leuk”. Uit de enquête die Lehenbre (2019) afnam onder ca. 200 Vlaamse tweedeklassers Latijn uit verschillende scholen, spreekt nochtans een ander beeld:

⁹⁹ Er bestaat echter geen empirisch bewijs dat een grondig inzicht in het taalsysteem leerlingen leesvaardig(er) maakt. Hunt (2022, pp. 12-13) formuleert hierover interessante bedenkingen. Hij suggereert dat sommige leraren geneigd zijn hun eigen leesvaardigheid rechtstreeks toe te schrijven aan het feit dat ze zelf op de middelbare school veel abstract grammaticaonderwijs hebben genoten, waarbij ze dan uit het oog verliezen dat ze tijdens hun universitaire studie en later als leraar simpelweg heel veel Latijn hebben gelezen.

¹⁰⁰ Dat een grammaticagerichte aanpak daadwerkelijk inclusiever zou zijn, is niet bewezen. Hetzelfde geldt evengoed voor de methodes aan de andere zijde van het spectrum, zoals de Ørbergmethode. In de internationale literatuur gebruiken aanhangers van beide uitersten inclusie als argument om hun voorkeursmethode te verdedigen (zie bij voorbeeld Piantaggini, 2020, die zijn pijlen richt op de grammatica-vertaalmethode).

op de vraag welk onderdeel van het vak ze het minst graag deden, antwoordde de meerderheid “grammatica”. Menig respondent met een hoge GG-score lijkt volgens ons dan ook indirect te verwijzen naar zijn persoonlijke liefde voor taalstudie, met uitspraken als: “De schoonheid van grammatica als systeem mag benadrukt worden, ook zonder link naar de functionaliteit”.

7.4.6 *Multum legendum est an multa?*¹⁰¹ De grenzen van de geïntegreerde benadering

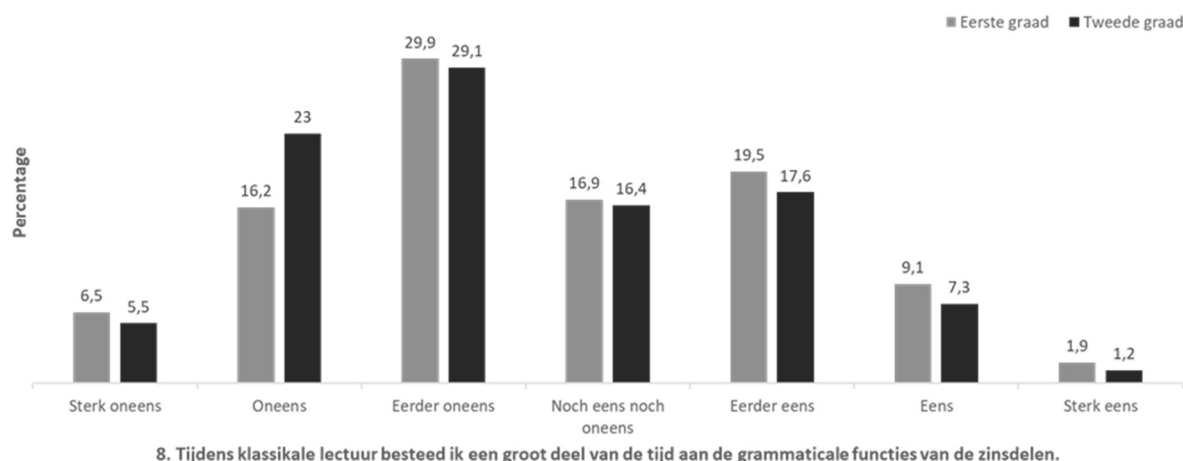
Uit de enquête komt naar voren dat Latijn lezen een activiteit is die heel divers wordt ingevuld, zelfs los van de vraag of lezen al dan niet neerkomt op vertalen. De geïntegreerde aanpak van taal en lectuur die de beide netten vandaag voorschrijven, is een uitnodiging om veel te lezen. De data vertellen ons dat zeker in de tweede graad het merendeel van de leraren dit effectief doet. “Veel” kan in de praktijk echter op diverse wijzen worden begrepen: als “*multa*” (veel teksten / snel en extensief lezen / met tekstbegrip als eerste en voornaamste focus) maar ook als “*multum*” (weinig teksten / traag en intensief lezen / met veel aandacht voor grammatica).

Het is dus misschien geen toeval dat de respons op de tweede stelling (“haast iedere les tekst lezen”; zie Figuur 7) statistisch wat minder blijkt samen te hangen met die op de andere stellingen in de grammaticagerichtheidsschaal. Zo is er bijvoorbeeld zo goed als geen verband tussen hoe vaak leraren zeggen aan het lezen te zijn en hoeveel aandacht ze – in hun antwoord op stelling 8 – zeggen te besteden aan zinsontleding.¹⁰² Hoeveel tijd er tijdens de lectuur naar zinsontleding gaat, is dan weer wel een tamelijk sterke indicator voor hoeveel of weinig grammaticagericht iemand in het algemeen is.¹⁰³ De grafiek in Figuur 10 visualiseert de respons op stelling 8.

¹⁰¹ Naar Plinius, *Epistulae* 7.9.15.

¹⁰² Er bestaat geen correlatie tussen de antwoorden op stelling 2 en stelling 3, noch in de eerste ($r = 0,07$; $p = 0,402$), noch in de tweede graad ($r = -0,08$; $p = 0,308$). Tussen de antwoorden op stelling 2 en stelling 8 bestaat een zwakke correlatie in de eerste graad ($r = 0,25$; $p = 0,002$) en geen correlatie in de tweede graad ($r = 0,08$; $p = 0,290$). Dit neemt voor alle duidelijkheid niet weg dat stelling 2 haar waarde heeft binnen de GG-schaal, zoals blijkt uit de correlaties met de andere stellingen, alsook uit het feit dat Cronbach's α zou dalen mocht dit item worden weggelaten. Bij stelling 2 lijkt vooral het doelpubliek van belang: in de eerste graad kunnen veel minder leraren zich in deze stelling vinden dan in de tweede graad.

¹⁰³ Eerste graad: $r = 0,66$; $p < 0,001$. Tweede graad: $r = 0,51$; $p < 0,001$.



Figuur 10 Verdeling respons op stelling 8

De hoeveelheid antwoorden aan de eens-zijde in deze grafiek tonen aan dat intensief of “grammaticaal lezen” zeker niet volledig uit de Vlaamse onderwijspraktijk is verdwenen. Voorts valt op dat relatief veel leraren hier “noch eens noch oneens” hebben aangeduid. Deze schijnbare onbeslistheid laat zich verklaren door de intra-individuele verschillen die in de commentaren tot uiting komen. Het ligt voor de hand dat leraren hun focus aanpassen naargelang de grammaticale moeilijkheidsgraad van de Latijnse tekst in kwestie. De focus kan ook verschillen naargelang het leerjaar waarin iemand aan het lesgeven is. Zo lichten sommige docenten van de tweede graad toe dat hun les- en leespraktijk in klas 3 meer grammaticagericht is dan in klas 4.

Volgens ons is er nog een fundamentele verklaring voor het gebrek aan uitgesproken antwoorden op deze stelling, die te vinden is in de actuele leerplannen van beide onderwijsnetten. Zij roepen immers op om taal zoveel mogelijk in een lectuurcontext aan te bieden, maar onderstrepen tegelijk het belang van “leesplezier”.¹⁰⁴ Het is maar de vraag of die twee premissen überhaupt te combineren zijn: én al lezend grammatica inoefenen, én doorlezen met zo weinig mogelijk grammaticaal oponthoud, zodat er vaart blijft in het verhaal en extensief lezen mogelijk wordt. De moeilijkheid om beide praktijken met elkaar te versmelten is ook een reden voor het relatief hoge aandeel “noch eens noch oneens”.

¹⁰⁴ Bijvoorbeeld: “De hoeveelheid grammaticale analysevragen tijdens de lectuur wordt het best beperkt tot de plaatsen waar dat noodzakelijk is om tot tekstbegrip te komen. Leesplezier staat centraal.” (KathOndVla, 2019, p. 10).

Zowel in de interviews als in de commentaren getuigen leraren dat ze heel bewust afwisselen tussen enerzijds “grammaticaal lezen”, bijvoorbeeld om nieuwe taalfenomenen te introduceren of om gekende grammatica te herhalen, en anderzijds wat ze “genietend lezen” noemen. Anderen zijn dan weer de mening toegedaan dat een geïntegreerde aanpak van taal en tekst niet te verenigen is met leesplezier, en blijven onder meer daarom vasthouden aan de traditionele scheiding van de componenten. Ter illustratie een (licht geredigeerd) citaat van een leraar met een betrekkelijk hoge GG-score van 56,7:

“Ook al staat grammatica ten dienste van lectuur, toch moet die apart aangeleerd en ingeoeffend worden in de eerste graad: omdat je anders te traag leest, omdat ongeziene tekst een stevige grammaticale basis veronderstelt, en omdat het jammer is om een leuke tekst op te offeren aan het aanleren van de grammatica. Grammatica op zich kan trouwens ook boeiend zijn.”

Hoe zit het trouwens met de derde component, naast grammatica en lectuur, namelijk cultuur? In de geest van de *reading method* is de integratie van teksten met cultuurinhouden sinds jaar en dag de norm in het GO!. In het vrij onderwijs was de koppeling tussen lectuur en cultuur veel minder aanwezig, zeker in de eerste graad. Ook op dit vlak lijkt de leerplanvernieuwing vanaf 2011 enig effect te hebben gehad: 59% van de leraren in de eerste graad en 68,5% in de tweede graad gaat *niet* akkoord met de stelling “Ik behandel sommige cultuurinhouden zonder dat we hierover een Latijnse tekst lezen”.

Wie wel akkoord gaat met de stelling (dit gaat in de eerste graad over toch een derde van de respondenten en in de tweede graad nog een kwart, ongeacht het onderwijsnet), wijkt dus enigszins af van de leerplannorm. Hiervoor kunnen verschillende redenen worden aangehaald. Leerlingen motiveren zal zeker een factor zijn: het leerlingenonderzoek van Lehembre (2019) bevestigt het wellicht algemene aanvoelen dat leerlingen cultuur het fijnste vakonderdeel vinden. Uit ons onderzoek blijkt ook een zwakke positieve correlatie met grammaticagerichtheid:¹⁰⁵ leraren met hoge GG-scores hebben vaker de neiging om af te wijken van de verplichte integratie van cultuur en lectuur. We veronderstellen dat de manier waarop zij lestijd besteden aan lectuur – *multum, non multa* – hier zeker mee te maken heeft. Een van de keerzijdes van “grammaticaal” en dus traag lezen is immers dat heel wat Latijnse teksten

¹⁰⁵ Eerste graad: $r = 0,19$; $p = 0,021$. Tweede graad: $r = 0,26$; $p = 0,001$.

uit het lesboek ongelezen blijven, waardoor de noodzaak ontstaat om de cultuuronderwerpen die aan deze overgeslagen teksten zijn gelinkt apart te behandelen.

7.5 Slot

7.5.1 Beperkingen van onze onderzoeksopzet en wenken voor verder onderzoek

Iedere onderzoeksopzet brengt bepaalde beperkingen met zich mee. Ons voornaamste meetinstrument was een zelfrapporteringsmaat over de onder- wijspraktijk in de vorm van een vragenlijst. We beschouwen de antwoorden op de stellingen als *proxy* voor wat er zich daadwerkelijk in de klas afspeelt. De validiteit van onze bevindingen is dus gebaseerd op de aanname dat respondenten de vragenlijst ten eerste naar eer en geweten invullen, en ten tweede ook adequaat, wat wil zeggen dat ze erin slagen om door reflectie een representatieve weerslag te bieden van hun feitelijke praktijken. In dat opzicht is de validiteit van een zelfrapporteringsmaat beperkter dan van observationeel onderzoek. Ideaal zou zijn om deze studie uit te breiden met lesobservaties. Daar staat natuurlijk tegenover dat die werkwijze vele malen arbeidsintensiever en tijdrovender is.

Een tweede interessante aanvulling zou kunnen zijn om leerlingenkenmerken te bevragen, en na te gaan in welke mate zij al dan niet bepalen hoe veel of weinig grammaticagericht docenten Latijn te werk gaan. Een andere bedenking die we na afloop van de enquête maakten met betrekking tot onze vragenlijst, is dat het misschien niet ideaal is om de beide jaren van de tweede graad samen te nemen, vermits de aanpak in de derde en de vierde klas soms erg uiteen kan lopen. In het vrije net begint men bijvoorbeeld doorgaans in het vierde jaar met de lectuur van volledig authentieke teksten.

Wat de steekproef betreft, gaan we ervan uit dat deze representatief is vanwege de hoge responsgraad en de evenwichtige vertegenwoordiging van subgroepen die we *a priori* in rekening hebben gebracht. We weten echter niet met zekerheid of de leraren die de vragenlijst hebben ingevuld al dan niet verschillen van de leraren die de vragenlijst niet hebben ingevuld. Acht op tien van onze respondenten volgen één à twee nascholingen per jaar, en bijna al de rest volgt er drie of meer. We zouden ons bijvoorbeeld de vraag kunnen stellen of we leraren

die geen nascholingen volgen wel bereikt hebben, en of zij er geen andere percepties en praktijken op na houden. Uiteraard is dit gegeven een elementaire beperking van ieder onderzoek dat door steekproefneming beoogt om uitspraken te doen op populatieniveau.¹⁰⁶ Dat alles gezegd zijnde, menen we de volgende conclusies te mogen trekken.

7.5.2 Conclusies

De Vlaamse onderwijspraktijk voor Latijn in de aanvangsjaren is op relatief korte tijd fundamenteel veranderd. De resultaten van onze enquête wijzen erop dat de grammatica-vertaalmethode, die in het numeriek belangrijkste vrije onderwijsnet lange tijd dominant was, ook daar plaats heeft geruimd voor een overwegend leatuurgerichte aanpak. Vandaag zijn de meeste leraren uit beide netten het ten gronde eens met het uitgangspunt dat taalstudie in functie staat van leatuur. De relatief lage gemiddelde grammaticagerichtheidsscores tonen aan dat de *reading method* het overheersende paradigma is geworden, gevolgd door de eclectische methode. In de grote evolutie in het vrije net speelden leerboekmakers volgens ons een cruciale rol: ze ver- ruimden het tekstaanbod, integreerden de inhouden van de teksten met de lessen cultuur, maakten een inductieve aanpak van de grammatica mogelijk, en ontwierpen contextuele oefeningen. Op die manier stimuleerden, zo niet dwongen ze de leraar om de nieuwe leerplanvoorschriften (vanaf 2011) daadwerkelijk toe te passen.

Tegelijkertijd bevestigen onze data dat de verschillen in lespraktijk tussen de onderwijsnetten groot blijven. In het GO! zweert men nog steeds bij het lezen van authentieke teksten vanaf het prille begin en werken de meeste leraren met eigen syllabi. In het vrije net mag er dan wel vrij grote eensgezindheid heersen rond het doel (tekstbegrip), het instrument (de lesboekenreeksen *Pegasus Novus* en *Ars Legendi*) en de geleidelijke overgang van aangepaste naar authentieke teksten, toch is de manier waarop de *reading method* gestalte krijgt, heel divers.

Ook tussen individuele leraren kan de grammatica- dan wel leatuurgerichtheid namelijk zeer sterk verschillen, zelfs tussen collega's uit dezelfde vakwerkgroep op een school. Een kleine groep leraren van alle leeftijden betreurt de eenzijdige leatuurgerichtheid, beschouwt de geïntegreerde aanpak van taal, leatuur en cultuur niet als meerwaarde, en houdt vast aan

¹⁰⁶ Daar komt bij dat in deze studie geen post-hoc correcties werden uitgevoerd voor het aantal analyses. Er is bijgevolg een verhoogde kans op valse significanties.

deductief grammaticaonderwijs. De premissen van de actuele leerplannen combineren – al lezend taal leren én leesplezier bereiken – is sowieso geen makkelijke opgave. Aan de andere kant van het spectrum vinden we onder meer aanhangers van het natuurlijke taalleren via de Ørbergmethode, met maximale taalinput vanuit de teksten. Bij leraren die trouw de leerplannen volgen en die zich kunnen vinden in de didactiek van het leerboek bestaat er evenzeer (grote) variatie in tijdsbesteding, leestempo en al dan niet grammaticale focus tijdens het lezen.

Kortom, voor Vlaamse leerlingen leiden nog altijd vele wegen naar Rome.

Werken geciteerd in dit hoofdstuk

- Arteel, I., Bernaerts, L., Buysse, L., Corteville, I., Lippens, J., & Mortelmans, T. (2022). *Naar een talenonderwijs in topvorm: Een bevraging van de Vlaamse taalleerkracht*. Vlaams Talenplatform.
- Bosteels, R. (1997). *Perge quo coepisti*. Over het aanvangsonderricht Latijn. *Kleio. Tijdschrift voor oude talen en antieke cultuur*, 26(3), 97–110.
- Classica Vlaanderen. (2023). Cijfermateriaal. *Classica Vlaanderen vzw*.
<http://www.classicavlaanderen.be/informatie/cijfermateriaal/index.html>.
- Claus, F. (2021). Zeventig jaar leerplannen klassieke talen van het Katholiek Onderwijs in Vlaanderen. *Kleio. Tijdschrift voor oude talen en antieke cultuur*. Online publicatie.
www.kleiotijdschrift.net.
- Coghe, R., Panhuis, D., & Vermeersch, R. (2006). *Tolle, lege 1*. De Boeck.
- de Ferrante, M., & van Oeveren, K. (2022). *Startnotitie klassieke talen: Actualisatie examenprogramma's*. Stichting Leerplanontwikkeling.
- de Snoo, F. (2011). Ovidius in de onderbouw. Een evaluatie van beginnersteksten in de Latijnse lesmethoden Disco en Tolle Lege. *Lampas*, 44(2), 121–137.
- Gay, B. (2003). The theoretical underpinning of the main Latin courses. In J. Morwood (Ed.), *The teaching of Classics* (pp. 73–84). Cambridge University Press.
- GO!. (2020). *Didactische fiche Klassieke Studïen 1ste graad*.
<https://pro.g-o.be/themas/leerplannen/go-navigator/>.
- Hulstaert, K. (2016). *Latijn: Lezen zien begrijpen. De positionele methode in de klas*. Skribis.
- Hunt, S. (2016). *Starting to teach Latin*. Bloomsbury Academic.
- Hunt, S. (2022). *Teaching Latin: Contexts, theories, practices*. Bloomsbury Academic.
- KathOndVla. (2019). *Leerplan Basisoptie klassieke talen, 1e graad, A-stroom*.
- Kroon, C., & Sluiter, I. (2010). *Het geheim van de blauwe broer: Eindrapport van de Verkenningcommissie Klassieke Talen*. Stichting Leerplanontwikkeling.
- Laes, C. (2019). Quo vadimus? Latin and Greek in Flemish secondary schools. *Cursor Latein4EU. Zeitschrift für Freunde der lateinischen Sprache und europäischen Kultur*, 15, 58–60.
- Lehembre, D. (2019). Metire ut scias. Welke leerlingentypes stoppen met Latijn en hoe kunnen we dat vermijden? *Kleio. Tijdschrift voor oude talen en antieke cultuur*, 48(3), 98–109.
- NVKSO (Nationaal Verbond van het Katholiek Secundair Onderwijs) (1989). *Leerplan SO, Klassieke Studies 1ste leerjaar, Grieks-Latijn 2de leerjaar*.
- Piantaggini, L. (2020). Grammar-translation: What is it—really—for students? *Journal of Classics Teaching*, 21(42), 92–94.

- Pijls, D. (1995). De natuurmethode van Hans H. Ørberg in de eenheidsstructuur. *Kleio. Tijdschrift voor oude talen en antieke cultuur*, 24(3), 121–128.
- Regtuit, R. (2015). Latijn is een taal, geen puzzel. *Lampas*, 48(3), 286–297.
- Regtuit, R. (2024). In hapklare brokken? Of hoe een Griek zijn zinnen construeert. *Lampas*, 57(1), 47–71.
- van der Keur, M. (2024). Livius als film. Lezen volgens de positionele methode. *Lampas*, 57(1), 72–95.
- Velle, T., & Maes, Y. (2021). *Latijn anders? Waarom er nood is aan ander Latijn in de klas*. Skribis.
- Verbaal, W. (2015). Leren lezen als Romeinen. Niet enkel een kwestie van syntaxis! *Didactica Classica Gandensia*, 51, 7–38.
- Vereeck, A. (2018). Valentiegrammatica in Pegasus. Een kritische analyse. *Kleio. Tijdschrift voor oude talen en antieke cultuur*, 47(3), 99–115.
- Vereeck, A. (2023). *Non scholae sed vitae?* Klassieke talen en cognitieve transfer. *Kleio. Tijdschrift voor oude talen en antieke cultuur*, 52(4), 146–156.
- VLOT. (2023). Onderwijs. *VLOT vzw*.
<https://www.vlot-vzw.be/faq/onderwijs/index.html>.
- VVKSO (Vlaams Verbond van het Katholiek Secundair Onderwijs) (2011). *Leerplan secundair onderwijs: Klassieke Studiën – Latijn – Grieks, 1ste graad*.

Corrigendum bij Hoofdstuk 7

De term “net” is in dit artikel oneigenlijk gebruikt. Anders dan gepostuleerd in Sectie 7.2.1., telt Vlaanderen in totaal geen twee, maar drie onderwijsnetten. Binnen het officieel onderwijs bestaan er namelijk twee aparte netten, het gemeenschapsonderwijs enerzijds (GO!, rechtstreeks ingericht door de Vlaamse Overheid) en het gesubsidieerd officieel onderwijs anderzijds (dat OVSG en POV omvat, respectievelijk ingericht door de stads- en gemeentebesturen en door de provinciebesturen). Binnen het vrij onderwijs is er één net, het gesubsidieerd vrij onderwijs (gvo): naast katholieke scholen, die zoals gezegd veruit de grootste groep vormen, vallen hieronder ook kleinere onderwijsverstrekkers zoals scholen met andere geloofsovertuigingen en vrije niet-confessionele scholen.

Afgezien van deze terminologische onnauwkeurigheid, houden de tweedeling die de enquête maakt tussen officieel en vrij onderwijs en de focus die het artikel legt op de leerplannen van het GO! en van Katholiek Onderwijs Vlaanderen wel degelijk steek in de context van klassieketalenonderwijs. Het GO! is immers altijd richtinggevend geweest voor de leerplannen klassieke talen in de beide officiële onderwijsnetten; zo kan ook worden aangestipt dat ten tijde van de enquête het OVSG en het POV geen eigen didactische fiches voor Latijn aanboden. Voor het vrije onderwijsnet geldt dat het aantal niet-katholieke scholen dat Latijn aanbiedt op een hand te tellen valt, en dat er in dit net nooit enige andere leerplannen Latijn dan die van de katholieke koepel hebben bestaan.

Chapter 8: Methods and Results of the Longitudinal Study

Together with Chapter 9, this chapter will be reworked into an article reflecting the entire doctoral project.

Chapter 8

Methods and Results of the Longitudinal Study

8.1 Methods

8.1.1 From design to data

The design chosen to investigate the cognitive transfer effects of studying classical languages is a longitudinal observational study with repeated measures. Given that a randomised experiment is not an available option in this educational context, a longitudinal quasi-experiment is the most reliable way to determine causal relations.

A sample of general secondary education pupils, consisting of a program group of latinists and a comparison group of non-latinists, was followed for three school years and measured once a year with the same battery of tests. The first measurement took place in 2021-2022 when the pupils were in their second year, the second measurement in 2022-2023 when the pupils were in their third year, and the third measurement in 2023-2024 when the pupils were in their fourth year. It was a deliberate choice to let the study commence in the second rather than the first year of secondary education, so that we could avoid the large drop-out that happens after year one, and follow up the pupils until the more advanced stage of year four. As a consequence, the test of transfer is slightly more conservative, since hypothetically one year of classical language education might have had cognitive benefits for both the latinists and likely some of the non-latinists (i.e., those who dropped out after one year) that now go undetected.

Another critical aspect of the design is that participants (and evidently their parents or guardians) as well as the majority of the school staff were left in the dark about the true motive behind the study. Its purpose was rephrased in more general terms as “a study on the cognitive

development of Flemish secondary pupils”, without any reference to study option being a key variable. In this way, we sought to safeguard the pupils’ performance against the influence of any stereotypes surrounding classical language education.

8.1.2 From participants to cases

8.1.2.1 Recruitment of schools and pupils

With the intention of making comparisons between schools with different types of classical language teaching methods (objective 2, cf. General Introduction), the schools with the highest number of second-year classical language pupils in Flanders were approached. Information about enrolment numbers was obtained from the pedagogical support services. The enrolment data revealed that the schools with the most classical language pupils were all part of the Catholic network (the largest educational provider in Flanders, see also Section 7.2.1), which in turn explains the lack of teaching method variation we ultimately found (cf. Chapter 7).

An example of an information letter for schools, distributed after initial interest had been established, is given in Appendix 10. Eventually, four schools committed to cooperate in the longitudinal study. The schools are situated in three different provinces. Participants were recruited via the typical school-to-home-to-school process (see also Section 6.6.1), by means of a package of documents comprising a cover letter, an information letter with consent form, and a background data form (described in more detail in Section 8.1.3). Examples of these documents are given in Appendices 11 to 13. Together, the four schools supplied a starting sample of a little under a thousand pupils.

The school staff was rewarded with a large quantity of chocolate bonbons, displayed in the staff room around the time of data collection with flyers about the research that was going on. The flyer for the third measurement is given in Appendix 14. Schools were also offered the opportunity to receive their data, but none of the schools responded to this offer. Pupils were encouraged to participate via two types of incentives. Firstly, either their parents or guardians or the pupils themselves could request to receive a personal report of their performance after the third measurement. An anonymised example of such a feedback report is given in Appendix 15. Secondly, vouchers were raffled after each data collection round among the pupils who had

completed all tests of that round. It would seem that the vouchers did have some appeal: Only a few pupils deliberately abandoned their research participation in the course of the study.¹⁰⁷

8.1.2.2 Composition of the program group and the comparison group

The program group is made up of pupils who study Latin, either in isolation or in combination with Ancient Greek. None of the cooperating schools offers the option to take Ancient Greek without Latin in the first four years. In this research study, therefore, latinists include hellenists, and all hellenists are latinists. In advance, we expected a drop-out ratio of ca. 40%, seeing that in the school years prior to the longitudinal study this had been the average Flemish ratio of pupils dropping out of classical languages between their second and fourth year (Vereeck, 2024).

The comparison group is made up of pupils who follow a general (i.e., not technical or vocational) study option without a classical language. No further contrasts are made between study options, as this would make the design much too complex and require an unfeasible number of participants because of the many options and frequent switches between them. The “purity” of the comparison group has been maintained in the sense that pupils who changed from a Latin to a non-Latin study option in between measuring moments were either not invited for the following measurement, or they were invited but their data were immediately removed, depending on the organisational preferences of the school.¹⁰⁸ In other words, the comparison group is not contaminated by Latin drop-outs, but only consists of pupils who never studied a classical language, disregarding the first year of secondary education.

8.1.2.3 Sample size and sample variants

The sample size evolved as follows. At the start of the study, 948 pupils partook in the first measurement, of which 398 were part of the program group and 550 of the comparison group.

¹⁰⁷ Exact numbers cannot be supplied because causes of attrition were not consistently recorded, but the main cause of loss of participants (apart from drop-out from Latin) was pupils changing schools.

¹⁰⁸ In the occasions where it was preferred to re-invite everyone, data from a few dozen Latin drop-outs were collected during the second and third measuring moments. As it was not part of the design to look at the performance of pupils who did not continue studying classical languages until the end of the research study, these data were not used for any analyses, and these cases were not counted in the participant numbers reviewed in Section 8.1.2.3.

Afterwards, 123 participants were lost due to drop-out from Latin (i.e., these pupils stayed in the same school but switched to a different study option), and another 156 due to attrition (for a variety of reasons, e.g., change of school, scheduling issues, withdrawal of consent, or absence at the day of testing). The next year, 669 pupils from before and 2 new pupils partook in the second measurement, of which 264 were part of the program group and 407 of the comparison group. Afterwards, 34 participants were lost due to drop-out from Latin,¹⁰⁹ and another 110 due to attrition. The final year, 527 pupils from before and 2 new pupils¹¹⁰ partook in the third measuring moment, of which 219 were part of the program group and 310 of the comparison group.

These data have given rise to five sample variants.

- (1) **Full samples per measuring moment** (N = 948 for the first, N = 671 for the second, and N = 529 for the third). These full samples were used for the revision of the Dutch language test (cf. Section 4.3).
- (2) **The longitudinal sample** of pupils who sat the Dutch language test and/or the general-cognitive test at each of the three measurements. There are exactly 500 participants that meet this description. Of these *Pentakosioi*, (as I nicknamed them, after the Ancient Greek word for five hundred) 206 are latinists and 294 are non-latinists. It must be noted that these data are not entirely complete; there are some missing values. Depending on the outcome variable, N varies between 480 (for Orthography) and 485 (for Lexicon). Missingness is caused by pupil absence and technical difficulties with devices and testing platforms. Background data are complete for this sample, however, thanks to the repeated distribution of background data forms (cf. the “follow-up measures” mentioned in Section 6.6.1).

¹⁰⁹ The drop-out ratio here, though low in absolute terms, is actually even a little above the Flemish average from previous school years of ca. 10% (Vereeck, 2024). Changing study options between the third and fourth year is definitely possible but not highly common, since in Flemish secondary education these two years form one cycle or *graad* in Dutch (see also Section 7.2.1).

¹¹⁰ It happened that pupils joined later measurements without having partaken in the first, for example because they were absent the first year or because they transferred to the testing school. For those pupils, a background check was performed during data cleaning with regard to their previous study option, in order to avoid Latin drop-outs. At the second measurement, two pupils were allowed into the sample who came from a non-Latin study option. At the third measurement, two latinists were allowed into the sample. None of these new pupils were included in the longitudinal sample as defined further in this section.

- (3) **The filtered longitudinal sample**, from which on the one hand untrustworthy and on the other hand dyslectic participants were filtered out, with the respective goals of mitigating so-called “careless and insufficient effort responding” (Ulitzsch et al., 2024; see also Section 6.8) and of achieving better balance between the program and comparison group. A participant was deemed “untrustworthy” if on one or more outcome variables across all tests and measuring moments they had scored more than 3 standard deviations below the mean score for that measuring moment. Applying this criterion, 22 cases were deleted. Additionally, 18 cases of dyslectic participants were deleted as well, seeing they were all non-latinists; filtering them out thus remedied the need to control for language disorder in the statistical models. The filtered longitudinal sample is used for the analyses presented in Section 8.2 and throughout Chapter 9.
- (4) **A “Greek” subset of the filtered longitudinal sample**, where the program group is subdivided into two groups: latinists *stricto sensu* who do not study Greek, and hellenists who do study Greek alongside Latin. Similar to the procedure described in Section 8.1.2.2, contamination by drop-out was avoided by excluding the 33 latinists who had studied Greek in their second and/or third year but then dropped out before their fourth year. This subset is used for the analyses presented in Section 8.3.
- (5) **A matched subset of the filtered longitudinal sample**. Statistical matching is introduced in Section 8.1.4.2. The matching procedure is described in Section 8.4.1, and the matched subset is used for the analyses presented in the remainder of Section 8.4.

8.1.3 From materials to variables

This section merges information about which tests and surveys were administered to the participants, what other data were gathered, and how all data components are reflected in the dataset as variables. For more details about test administration, see Appendix 16, which contains the latest version of the testing manual that was handed to university students and school teachers who assisted with data collection (in accordance with the guidelines on data collection formulated in Section 6.7).

8.1.3.1 Time-fixed data components

The following components are time-fixed, which means that their values do not vary across the three measurements in the longitudinal design.

- **School**, i.e., which of the four schools the participant in question went to. The schools are numbered 1, 3, 4, 5 in the dataset, since the collaboration with the school originally given number 2 was discontinued not long after the launch of the longitudinal study.
- **Study option**. It was closely monitored which pupils followed which study option, and who dropped out of their original study option in the course of the longitudinal study, as explained above. Accordingly, the grouping variable **Latinist** remains stable over time, with 0 = non-latinist (n = 257 in sample variant 3) and 1 = latinist (n = 203 in sample variant 3), as does the grouping variable **Hellenist**, with 0 = non-latinist (n = 257 in sample variant 4), 1 = latinist (n = 127 in sample variant 4), and 2 = hellenist (n = 43 in sample variant 4).
- **Background data**. Prospective participants were asked to fill out the abovementioned background data form together with their parents or guardians. The majority of participants in the longitudinal sample complied at the start of the study, and some returned the form at a later point. Via this form, the following subject data were collected.
 - **Sex**. The form offered the answer options *man* (male), *vrouw* (female) and X. No X's ended up in the longitudinal sample. The variable Sex is coded 0 = boy and 1 = girl.
 - **Socio-economic status (SES)**, operationalised as the highest parental educational level. In the form there were two separate questions, one in regard to the mother figure and one in regard to the father figure. Six answer options were offered: none, primary education, secondary education, bachelor (or equivalent), master (or equivalent), and PhD. In addition, there was a seventh option "does not apply", which was processed as a missing value. The other six answers were coded from 0 to 5. Subsequently, an overarching SES variable was created that expressed the highest educational level obtained by either parental figure, preserving the six categories. For example, if the mother had a master's degree (= 4) and the father a bachelor's degree (= 3), the value would be 4.

Since in the longitudinal sample there was just 1 case in category 0 (= no education), just 1 case in category 1 (= primary education as the highest level) as well, and merely 26 cases (reduced to 22 after filtering) in category 2 (= secondary education as the highest level), the SES variable was restructured from 6 to 4 categories. The values of this adjusted SES variable, which was used in all analyses

presented hereafter, represent the following levels: 0 = no higher education, 1 = bachelor, 2 = master, 3 = PhD.

- **Native language.** The form asked to indicate whether the child's native language was Dutch, a language other than Dutch, or a combination of the two. This component is reflected in the variable **NonNative**, which is coded 0 = Dutch as a native language, and 1 = Dutch as a non-native language. This coding scheme was chosen so as to make Dutch natives the reference group.
- **Language disorder.** The form also included a yes/no question about whether the child had a language disorder such as dyslexia. As mentioned above in Section 8.1.2.3, all participants for which this question had been answered affirmatively were eventually filtered out of the longitudinal sample.
- **Dutch language capacity**, which is present a priori before the start of the study and the three years of differentiated education (classical language(s) vs. no classical languages). In practice, this is equal to total performance on the language test at the first measurement. This baseline score is named **Dutch_T1** and equals the percentage of correct answers on the entire language test, including the orthographical scale alongside the lexical and verbal scales (cf. Section 4.3).
- **Cognitive capacity**, which is present a priori before the start of the study and the three years of differentiated education (classical language(s) vs. no classical languages). In practice, this is equal to total performance on the general-cognitive test at the first measurement. This baseline score is named **G_T1**.
- **Empathy.** In the course of the study, it was felt that an affective or emotional component was missing. Therefore, at the third measurement the affective empathy (Cronbach's $\alpha = .70$ in the full sample) and cognitive empathy (Cronbach's $\alpha = .67$ in the full sample) scales of the Empathy Questionnaire for Children and Adolescents developed at Leiden University (Overgaauw et al., 2017) were interspersed among the Dutch language test. Some wording adjustments were made to the items in order to better suit an audience from Flanders rather than the Netherlands; this Flemish version

is given in Appendix 17.¹¹¹ Since these two scales were administered just once in the final year of the study, the results are strictly exploratory and not part of the longitudinal design (more in Section 9.2.4).

- **Reading motivation.** In the course of the study, I happened to come across a government report on inequality in Flemish education (Boone, 2011). A finding from this report that particularly interested me, is that in last-year primary pupils, reading motivation is a significant predictor of preferring to study Latin in secondary education over another general study option without Latin. Therefore, at the third measurement a brief scale measuring reading motivation was added (Cronbach's $\alpha = .65$ in the full sample), given in Appendix 18. Again, since this scale was administered just once in the final year of the study, the results are strictly exploratory and not part of the longitudinal design (more in Section 9.2.4).

8.1.3.2 Time-varying data components

The values of the following components do vary across the three measurements in the longitudinal design.

- **Time**, i.e., an index of which measurement the data belong to. In the longitudinal data, the variable Time is coded as $T1 = 0$, $T2 = 1$, and $T3 = 2$, so that the first measuring moment is the point of reference.
- **Non-cognitive factors.** A questionnaire assessing the non-cognitive factors grit, self-control and motivation was programmed in LimeSurvey. With a view to minimising the total on-site administration time, initially participants were directed to the questionnaire via an url written in the background data form, so that they could fill it in at home. However, because the block of two class hours that was allotted to the Dutch language test turned out to be more than ample, the questionnaire was later coupled to the Dutch language test.

¹¹¹ Initially, I only had access to the English version of the questionnaire. A fellow linguist and I, both of us Flemings, first independently translated the items from English into Dutch, and then we compared our translations and decided on the best translation. When upon my request I later received the Dutch version of the questionnaire from the authors, I put their items and the items translated by my colleague and me side to side. It soon became apparent that the original items contained a lot of (Northern-) Dutch vocabulary that was hard to grasp for Flemish speakers of Dutch, so I settled on the Flemish version.

The non-cognitive factors figure as (potential) predictors in the statistical models of the following sections. Mean scores for latinists and non-latinists per measurement are given in Appendix 19. The general trend is that scores remain constant or decline over time, one exception being that for non-latinists controlled motivation for school (cf. *infra*) appears to become more important as they grow older. As with any self-report measure, responses to this questionnaire are liable to inaccuracies and social desirability bias (e.g., Sjöström & Holst, 2002).

- **Grit** was measured with the Short Grit Scale (Duckworth et al., 2007; Duckworth & Quinn, 2009) in a Dutch translation adopted from the SIMON database (Fonteyne et al., 2017). In the full samples per measuring moment, Cronbach's α respectively equals .69, .71, and .78. Grit has two facets: consistency of interest and perseverance of effort. The latter especially overlaps substantially with the personality trait conscientiousness (Ponnock et al., 2020). Because of the conceptual difference between the two facets, it was eventually decided that in the analyses the variables **Consistency** and **Perseverance** would be entered separately.
- **Self-control** was measured with the Brief Self-control Scale (Tangney et al., 2004) in a Dutch translation adopted from the SIMON database (Fonteyne et al., 2017). In the full samples per measuring moment, Cronbach's α respectively equals .81, .82, and .85. Self-control is an aspect of conscientiousness as well (Roberts et al., 2005).
- **Motivation for school** was measured with the Academic Self-regulation Scale (Vansteenkiste et al., 2009) in a Dutch translation adopted from the SIMON database (Fonteyne et al., 2017). Within self-determination theory (e.g., Deci & Ryan, 2000), motivation encompasses two distinct constructs: controlled motivation and autonomous motivation. The former is defined as motivation stemming from external regulation, such as meeting expectations or avoiding punishment, whereas the latter is defined as intrinsic and internalised motivation. In the full samples per measuring moment, Cronbach's α for the controlled motivation scale respectively equals .74, .77, and .77, and Cronbach's α for the autonomous motivation scale equals .89 each time. The variable names are abbreviated to **ControlledMot** and **AutoMot**.
- **Motivation for Latin**. Latinists were also presented with an adapted version of the Academic Self-regulation Scale, where references to studying for school in general

had been replaced by references to studying for their Latin course. Owing to a programming error in the second year of the longitudinal study, data from all schools are only available for the first and third measuring moment. For controlled motivation, Cronbach's α equals .82 and .81 respectively; for autonomous motivation, Cronbach's α equals .93 and .90 respectively. These variables are not included in the longitudinal analyses, since they only apply to half of the sample.

- **Dutch language ability.** The self-designed Dutch language test, compiling existing test items from several sources, is described in detail in Chapter 4 of Part II. Ideally, there were supposed to be three outcome variables, namely Lexicon, Verbs and Orthography, but the third was left out of the analyses because of serious ceiling effects.
 - **Lexicon** is a mixed variable, encompassing both word knowledge and reading ability. The majority of vocabulary items is of non-Latin origin. Among the items that are derived from a Latin word, none were borrowed straight from Latin (i.e., no items along the lines of *veto*, *fiat* or such) and their meanings generally do not follow from their etymology in a straightforward way (e.g., *paviljoen*, 'pavilion', from *papilio*, 'butterfly', or *bagatelliseren*, 'to downplay', from *baca*, 'berry').
 - **Verbs** refers to conjugating verbs in accordance with applicable spelling rules.
- **General cognitive ability.** As announced and explained in the introductory remarks to Part II, general cognitive ability was measured with the COVAT-3 (Magez et al., forthcoming), the digital successor of the CoVaT-CHC Basic Version (Magez et al., 2015). However, administering the COVAT-3 in its entirety takes well over three hours, which would have been too time consuming. The authors of the test therefore suggested that I use a selection of four subtests, which could be administered within the span of one 50-minute class hour and which related to four broad cognitive abilities, both verbal (Gc and Gsm) and non-verbal (Gf and Gv). The number of items and scoring rules differ per subtest. Here, all total scores were adjusted to a 100-point scale.

Despite the many advantages of the COVAT-3 over WISC and ICAR, it is still a test primarily designed for a diagnostic setting of one-on-one administration, and not for a research setting of group administration. The difference between the two is that in the first setting the test leader can keep a close eye on everything and intervene whenever necessary, but in second setting the test leader's attention is inevitably divided, and so administration should proceed as automatically as possible. The clinical background of the COVAT-3 especially manifested itself in the fact that user-friendliness and ease of

use left something to be desired in some areas, which would not really raise a problem when testing one person, but does create difficulties when testing a large group. Occasionally, data of one or more subtests were lost because participants did not interact with the testing platform correctly as instructed. For the cases where data of only one subtest went missing, the value was imputed as the mean score on the other three subtests; this procedure was performed on the full samples.

- **Crystallised intelligence or Gc** was measured with the subtest *Tegenstellingen* (Antonyms), in which participants are presented with a word and have to select the best antonym from a list of four other words. In the full samples per measuring moment, Cronbach's α respectively equals .81, .69, and .65. Values were imputed in 0.1% of cases for measurement 1, 0.6% of cases for measurement 2, and 0.4% of cases for measurement 3.
- **Short-term memory or Gsm** was measured with the subtest *Verbaal Geheugen* (Verbal memory), in which participants have to learn by heart a list of words and later retrieve the words from memory. In the full samples per measuring moment, Cronbach's α respectively equals .89, .83, and .86. Values were imputed in 1.1% of cases for measurement 1, 1.1% of cases for measurement 2, and 0.6% of cases for measurement 3.
- **Fluid intelligence or Gf** was measured with the subtest *Puntreeksen* (Dot sequences), in which participants have to complete logical sequences of dice or dominoes. In the full samples per measuring moment, Cronbach's α respectively equals .88, .77, and .71. Values were imputed in 0.9% of cases for measurement 1, 1.4% of cases for measurement 2, and 0.2% of cases for measurement 3.
- **Visuo-spatial processing or Gv** was measured with the subtest *Gedraaide Figuren* (Rotated figures), in which participants have to indicate all figures that are a rotated version but not a mirror image of the first given figure. In the full samples per measuring moment, Cronbach's α respectively equals .95, .94, and .94. Values were imputed in 2.9% of cases for measurement 1, 0.3% of cases for measurement 2, and 0.2% of cases for measurement 3.
- **Overall general cognitive ability or G** is estimated with the following formula: $2 \times Gf + 2 \times Gc + Gv + Gsm$. Fluid and crystallised intelligence have double weight, because these broad cognitive abilities load most strongly onto the G-factor (Schneider & McGrew, 2012).

- **School achievement.** As the final part of every measurement, grades from the corresponding term were requested from the school administration. Besides total achievement (**AchTotal**), also known as grade point average, the following courses are included as outcome variables: Dutch (**AchDutch**), English (**AchEnglish**), French (**AchFrench**), and mathematics (**AchMaths**). Latinists' Latin grades (**AchLatin**) were registered as well. All grades are expressed as percentages. The analyses for school achievement are based on only three schools, because the fourth school did not have traditional report cards with total scores per course and per term. These achievement data have to be approached with caution, since they do not stem from standardised tests and are therefore a lot less reliable. School, study option and teacher are determining factors for what subject matter is tested, how it is tested, and how the tests are graded, making comparability across different schools, study options and class groups altogether limited.

8.1.4 From hypotheses to analyses

8.1.4.1 Operationalising cognitive transfer

The main research hypothesis is that the study of classical languages has tangible transfer effects on Dutch language ability, general cognitive ability, and school achievement. Translated to the variables in the longitudinal study, this would mean that latinists make greater progress over time than non-latinists on the outcome variables Lexicon, Verbs, Gc, Gsm, Gf, Gv, G, AchDutch, AchEnglish, AchFrench, AchMaths, and AchTotal. Additionally, it is hypothesised that studying Greek has a complementary effect, so that hellenists make even more progress over time than latinists who do not study Greek. The focus of analysis will lie on the data from the standardised linguistic and general-cognitive test; the school achievement data will be interpreted less in-depth, given their lower reliability.

What exactly does “progress” look like in the data? For neither the Dutch language test nor the COVAT-3 age norms have been established, so all scores used here are raw scores. Because of cognitive development in general due to aging and schooling, combined with learning effects of filling out the same tests multiple times, it is expected that all pupils improve their test scores from measurement to measurement. Data exploration tells us that this is indeed the case at

sample level, even though individual cognitive reports like the one given in Appendix 15 sometimes show different evolutions. If latinists experience a cognitive benefit from learning Latin and/or Greek, their average test scores should rise faster than those of non-latinists. For school achievement, the story is slightly different. It could be said that school grades do have a certain age correction built in to them, as increasingly difficult subject matter needs to be mastered in order to obtain the same numerical grade. The general trend is that grades decline throughout secondary school; hence, “progress” is to be interpreted as a less steep decline.

8.1.4.2 Dealing with confounding influences, notably selection bias

An effect of study option is only meaningful if it can be proven to exist independently from other confounding influences. In particular, we want to compensate for the fact that pupils were not assigned randomly to their study options and that there may be a selection bias. Moreover, pupil characteristics that can affect performance, such as their demographic and non-cognitive profile, need to be taken into account as well. There are two different approaches to dealing with confounding influences, which will both be applied in the results sections. The first approach (applied in Section 8.2 to Latin and in Section 8.3 to Greek) is to rely exclusively on statistical modelling and add in all known confounders as control variables. For every outcome variable, I will control for school, sex, SES, native language, cognitive capacity, consistency of interest, perseverance of effort, self control, controlled motivation for school, and autonomous motivation for school. For language-related outcome variables (Lexicon, Verbs, Gc and Gsm), I will also control for Dutch language capacity.

The second approach (applied in Section 8.4) is to preprocess the data prior to analysis by matching the groups that are being compared. Iacus, King and Porro (2012, p. 1) give the following definition of matching.

“Matching is a nonparametric method of controlling for the confounding influence of pretreatment control variables in observational data. The key goal of matching is to prune observations from the data so that the remaining data have better *balance* between the treated and control groups, meaning that the empirical distributions of the covariates (X) in the groups are more similar. Exactly balanced data mean that controlling further for X is unnecessary (since it is unrelated to the treatment variable), and so a simple difference in means on the matched data can estimate the causal effect; approximately balanced data require controlling for X with a model (such as the same model that would have been used without matching) (...).”

By trimming the sample through matching, an artificial situation can be created in which there are no a priori differences between latinists and non-latinists; cognitive transfer can then be examined regardless of whether there is preselectivity at play in reality. The concept of matching aligns well with what the actual quantity of interest is here, namely the effect of studying classical languages on the people who study them. Since comparing real-world latinists to alternate-reality versions of themselves who did not opt for Latin is impossible, the best approximation of the quantity of interest is attained by comparing latinists to other pupils who are completely similar except for the fact that they did not opt for Latin. Another benefit of matching is that it makes results and inferences less model-dependent (Ho et al., 2007). A methodological downside of matching is that sample size is diminished, thus reducing the statistical power to detect small effects. Furthermore, there is a risk that the matched subgroups are less representative of their populations in some unforeseen way.

Different matching methods are currently in use. A highly popular method is propensity score matching (King & Nielsen, 2019). Simply put, a propensity score is a kind of composite score that quantifies the likelihood of a subject being in the program group given the values of the control variables. While this score has its uses, matching is not one of them: It has been convincingly shown that matching on the basis of propensity scores often leads to more imbalance, model dependence and bias – the exact things that matching supposedly seeks to reduce (Ho et al., 2007; Iacus et al., 2012; King & Nielsen, 2019). Instead, it is preferable to look for matches directly on the basis of the values of the control variables. This can be achieved with a method called coarsened exact matching (CEM), which Iacus, King and Porro (2012, p. 8) define as follows.

“The basic idea of CEM is to coarsen each variable by recoding so that substantively indistinguishable values are grouped and assigned the same numerical value (...). Then, the ‘exact matching’ algorithm is applied to the coarsened data to determine the matches and to prune unmatched units. Finally, the coarsened data are discarded and the original (uncoarsened) values of the matched data are retained.”

If there are sufficient matches to be found, exact matching can be executed for all control variables; if not, matching can also be executed for a selection of variables. The software written by Iacus, King and Porro (2009) that is used in Section 8.4 has a coarsening algorithm built in that automatically pursues the balance – sample size frontier for a given set of matching variables (King et al., 2017), simultaneously optimising both the degree of similarity and the

size of the matched groups. Note that matching is a feasible approach to test the main hypothesis, but not the additional hypothesis regarding Greek. Not only is matching three groups (non-latinists, latinists *stricto sensu* and hellenists) statistically a lot more complicated than matching two groups (non-latinists and latinists), the small number of hellenists (43 in the longitudinal samples) and pronounced preselectivity effects (cf. Section 8.3) also thwart a matching solution.

8.1.4.3 Longitudinal linear mixed modelling

Finally, a note about the structure of the data and the type of analysis. Pupil data were collected at three points in time, or in statistical terms, per outcome variable there are three observations “nested within” each participant. To account for the hierarchical nature of longitudinal data, linear mixed models are needed. Each model has two levels: The observed values on the outcome variable in question make up the lower level, and the pupils form the higher level that structures the observed values in clusters of three. Observations are identified by the time index (with 0 being the first measurement, 1 the second, and 2 the third), and pupils are identified by a unique ID.

8.2 Results: Latin, without matching

8.2.1 Models and interpretation

All of the analyses in this section are conducted on the filtered longitudinal sample (cf. Section 8.1.2.3), using the GAMLj package in Jamovi. The method of estimation is Maximum Likelihood. The displayed graphs are effect plots, which take into account the effects of the other variables in the model. No noteworthy violations were detected against the assumption of normality of variables and residuals.

First, a base model is defined, containing a fixed intercept, fixed effects of Time, Latinist, School and of the interaction between Latinist and Time, a random intercept, and a random effect of Time. The interaction effect represents the transfer hypothesis (cf. Section 8.1.4.1) that latinists make more progress than non-latinists. The other fixed elements reflect the basic anatomy of the data: longitudinal (Time), with two participant groups (Latinist) originating

from four different schools (School). The random elements allow for individual variation between participants in terms of both their baseline performance on the outcome variable (Intercept) and their evolution over time (Time), which ensures a more accurate estimation of the fixed effects. The general formula of the base model is: $\text{Outcome} \sim 1 + \text{Time} + \text{Latinist} + \text{School} + \text{Latinist}:\text{Time} + (1 + \text{Time} | \text{ID})$.

Then, a full model is defined, by adding the control variables as fixed effects. The general formula of the full model is: $\text{Outcome} \sim 1 + \text{Time} + \text{Latinist} + \text{School} + \text{Sex} + \text{SES} + \text{NonNative} + \text{Dutch_T1} + \text{G_T1} + \text{Consistency} + \text{Perseverance} + \text{SelfControl} + \text{ControlledMot} + \text{AutoMot} + \text{Latinist}:\text{Time} + (1 + \text{Time} | \text{ID})$. As stated above in Section 8.1.4.2, Dutch_T1 is only included in the models for Dutch language ability and verbal intelligence.

In the original full model, Dutch language capacity and cognitive capacity are controlled for by including baseline performance on Dutch language ability (Dutch_T1) and on general cognitive ability (G_T1). For test variables where some evidence emerges for a transfer effect (namely Lexicon, Verbs and Gsm), further refinement of the model is sought by testing the hypothesis that a greater Dutch language capacity and cognitive capacity could lead to faster progress on the outcome variable. This is done by extending the full model with interaction effects of Dutch_T1 and G_T1 with Time. The general formula of the extended model is: $\text{Outcome} \sim 1 + \text{Time} + \text{Latinist} + \text{School} + \text{Sex} + \text{SES} + \text{NonNative} + \text{Dutch_T1} + \text{G_T1} + \text{Consistency} + \text{Perseverance} + \text{SelfControl} + \text{ControlledMot} + \text{AutoMot} + \text{Latinist}:\text{Time} + \text{Dutch_T1}:\text{Time} + \text{G_T1}:\text{Time} + (1 + \text{Time} | \text{ID})$.

In all models for all outcome variables, the effect of Time is highly significant. This observation will therefore be left out in the interpretation of the tables below. At this stage, effects with a p-value of .05 or lower are interpreted as significant, and effects with a p-value above .05 as non-significant. Although I acknowledge that p-values are continuous and that this cut-off point is arbitrary, the traditional threshold of 5% still seems a good starting point to think about the strength of the evidence. The multiple comparisons problem will be addressed in Section 9.3.2.

For each test variable, the evolution per school year is quantified for the average latinist and non-latinist. “Average” in this context means the following: averaged across schools, averaged across sexes, with parents of average education level (for this sample), native speaker of Dutch, and with average scores on all covariates.

Before moving to the longitudinal linear mixed models, the descriptive statistics of the outcome variables are reported and interpreted, comparing test variables' group means from the first measurement through t-tests.

8.2.2 Descriptive statistics

	Latinist	Time	N	Mean	SD	Minimum	Maximum
Lexicon	0	0	249	44.4	12.60	7.69	72.2
		1	249	50.3	13.11	16.67	80.6
		2	249	55.7	14.06	22.22	94.4
	1	0	200	53.1	14.35	25.00	88.9
		1	197	61.3	15.22	22.22	91.7
		2	199	67.3	15.27	27.78	97.2
Verbs	0	0	248	54.6	26.99	0.00	100.0
		1	249	59.7	29.13	0.00	100.0
		2	249	57.8	26.57	0.00	100.0
	1	0	200	71.5	25.98	0.00	100.0
		1	197	78.3	24.01	0.00	100.0
		2	199	80.2	21.45	10.00	100.0
Gf	0	0	244	62.5	9.51	39.58	83.3
		1	255	68.0	9.54	41.67	87.5
		2	257	72.5	7.56	50.00	89.6
	1	0	202	63.6	9.00	41.67	85.4
		1	200	66.8	8.48	43.75	85.4
		2	203	71.9	7.34	52.08	87.5
Gc	0	0	244	57.4	8.57	31.11	77.8
		1	255	61.0	8.18	42.22	88.9
		2	257	64.2	7.72	46.67	82.2
	1	0	202	63.1	8.50	40.00	84.4
		1	200	66.3	9.25	42.22	86.7
		2	203	70.1	8.73	44.44	91.1
Gv	0	0	244	43.9	14.95	0.00	82.9
		1	255	50.2	14.97	5.71	78.6
		2	257	57.2	15.25	15.71	92.9
	1	0	202	44.6	14.44	1.43	80.0
		1	200	52.4	14.32	22.86	87.1
		2	203	58.1	14.56	21.43	94.3
Gsm	0	0	244	38.6	16.55	0.00	83.3
		1	255	47.9	16.41	0.00	100.0
		2	257	51.5	18.35	0.00	95.8

Descriptives

	Latinist	Time	N	Mean	SD	Minimum	Maximum
G	1	0	202	44.3	14.85	0.00	72.9
		1	200	55.0	16.62	0.00	100.0
		2	203	61.0	15.84	0.00	100.0
	0	0	244	322.3	40.23	218.31	422.9
		1	255	356.1	40.70	243.53	486.3
		2	257	382.1	38.59	274.60	489.5
	1	0	202	342.1	41.78	232.54	437.7
		1	200	373.7	40.81	246.53	469.7
		2	203	403.1	39.29	283.41	501.4
AchDutch	0	0	257	76.5	7.64	53.30	95.7
		1	171	72.4	8.07	49.90	90.0
		2	167	68.4	10.52	33.12	94.0
	1	0	203	80.6	7.49	56.70	96.4
		1	132	78.6	6.61	52.90	93.0
		2	131	74.8	8.62	51.88	93.0
AchEnglish	0	0	257	79.5	8.63	58.00	97.0
		1	169	74.6	9.13	48.00	95.3
		2	167	73.1	8.94	50.70	100.0
	1	0	203	84.6	8.02	53.00	98.0
		1	132	78.9	9.67	42.00	95.2
		2	132	80.5	9.06	52.50	96.7
AchFrench	0	0	257	71.9	10.38	35.00	90.8
		1	171	64.8	12.66	34.00	91.7
		2	167	63.9	12.76	33.00	92.3
	1	0	203	80.6	8.29	49.00	97.9
		1	132	70.7	11.92	41.00	97.3
		2	131	72.7	12.20	40.00	99.0
AchMaths	0	0	257	70.9	11.86	34.90	96.0
		1	171	67.6	12.65	35.00	98.0
		2	167	64.2	14.60	25.00	97.7
	1	0	203	77.5	10.27	25.00	98.1
		1	132	72.4	11.17	44.20	93.0
		2	131	68.3	13.32	37.00	94.5
AchTotal	0	0	171	75.8	6.79	57.50	92.0
		1	171	71.8	7.88	52.00	91.0
		2	168	69.1	8.88	48.93	89.0
	1	0	132	80.8	5.85	67.00	92.9
		1	132	76.7	6.94	61.00	93.4
		2	132	74.4	8.32	53.84	93.0

From the descriptives it can be gathered that the mean test scores show upward trends over time and that the mean achievement scores show downward trends over time, as expected (cf. Section 8.1.4.1).

Furthermore, latinists obtain notably higher mean scores on all variables except Gf and Gv. At the outset, latinists already score 8.7 percentage points (henceforth in this chapter: p.p.) higher on Lexicon (Cohen's $d = .649$), 16.9 p.p. on Verbs (Cohen's $d = .633$), 5.7 p.p. on Gc (Cohen's $d = .665$), 5.7 p.p. on Gsm (Cohen's $d = .362$), and 19.8 points out of 600 on G (Cohen's $d = .485$); all these effects are highly significant ($p < .001$). The initial difference in school achievement ranges between 4.1 p.p. (for Dutch) and 8.7 p.p. (for French).

8.2.3 Dutch language ability

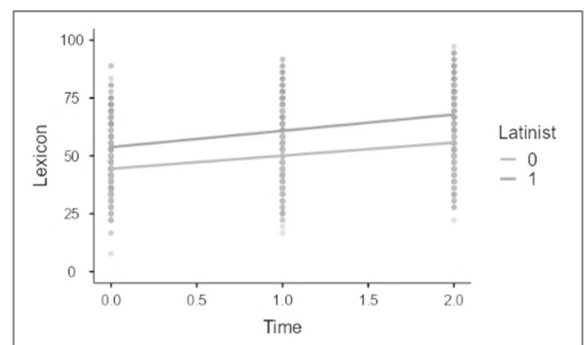
8.2.3.1 Lexicon

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	647.31	1	449	< .001
Latinist	56.88	1	449	< .001
School	6.92	3	449	< .001
Time * Latinist	8.00	1	449	0.005

Note. Satterthwaite method for degrees of freedom

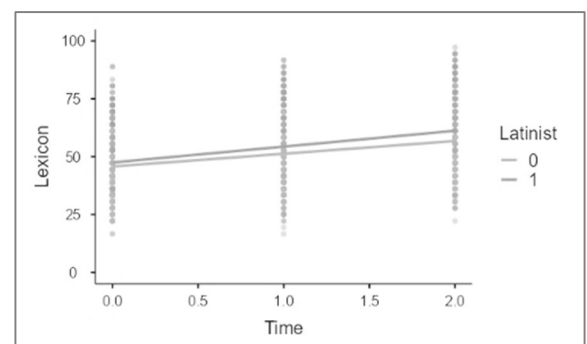


The main effect of Latinist is highly significant, indicating that latinists score higher on Lexicon across time. The interaction effect with Time is significant as well, indicating that latinists also make more progress over time.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	578.958	1	465	< .001
Latinist	2.591	1	465	0.108
School	2.407	3	441	0.067
Sex	5.272	1	443	0.022
SES	1.941	3	441	0.122
NonNative	12.414	1	441	< .001



Fixed Effect Omnibus tests

	F	Num df	Den df	p
Dutch_T1	148.569	1	445	< .001
G_T1	56.978	1	445	< .001
Consistency	2.548	1	1141	0.111
Perseverance	1.130	1	1195	0.288
SelfControl	1.580	1	1263	0.209
ControlledMot	0.186	1	1216	0.666
AutoMot	1.815	1	1216	0.178
Time * Latinist	7.014	1	450	0.008

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	42.950	1.563	39.887	46.014	454	27.480	< .001
Time	Time	5.509	0.344	4.835	6.183	455	16.021	< .001
Latinist1	1 - 0	1.724	1.071	-0.375	3.824	465	1.610	0.108
School1	3 - 1, 3, 4, 5	1.448	0.865	-0.246	3.143	441	1.675	0.095
School2	4 - 1, 3, 4, 5	-1.678	0.715	-3.080	-0.276	442	-2.345	0.019
School3	5 - 1, 3, 4, 5	0.699	0.799	-0.867	2.264	439	0.875	0.382
Sex1	0 - 1	2.085	0.908	0.305	3.865	443	2.296	0.022
SES1	1 - 0	-0.271	2.248	-4.677	4.135	439	-0.120	0.904
SES2	2 - 0	1.629	2.150	-2.584	5.842	445	0.758	0.449
SES3	3 - 0	3.378	2.406	-1.337	8.093	443	1.404	0.161
NonNative1	0 - 1	5.595	1.588	2.483	8.707	441	3.523	< .001
Dutch_T1	Dutch_T1	6.599	0.541	5.538	7.660	445	12.189	< .001
G_T1	G_T1	3.715	0.492	2.750	4.679	445	7.548	< .001
Consistency	Consistency	-0.511	0.320	-1.137	0.116	1141	-1.596	0.111
Perseverance	Perseverance	-0.369	0.347	-1.049	0.311	1195	-1.063	0.288
SelfControl	SelfControl	0.503	0.400	-0.281	1.288	1263	1.257	0.209
ControlledMot	ControlledMot	-0.120	0.278	-0.665	0.425	1216	-0.432	0.666
AutoMot	AutoMot	-0.422	0.313	-1.036	0.192	1216	-1.347	0.178
Time * Latinist1	Time * 1 - 0	1.344	0.508	0.349	2.339	450	2.648	0.008

The main effect of Latinist is no longer significant, suggesting that the control variables adequately control for the a priori difference between the two groups. The interaction effect with Time remains significant, meaning that even when taking into account the control variables latinists make additional progress on Lexicon. The average non-latinist improves their

Lexicon score 5.5 p.p. per school year, and the average latinist 6.9 p.p. The size of this effect is very small ($\Delta R^2 = .006$; Cohen's $f^2 = .012$).

The effect of School, which was highly significant in the base model, no longer reaches significance, suggesting that the control variables capture the differences between schools. Control variables that make a significant contribution to the model are Sex, NonNative, Dutch_T1 and G_T1. Keeping all other variables constant, boys score 2 p.p. higher than girls. The effects of NonNative, Dutch_T1 and G_T1 have the expected direction: Higher scores on Lexicon tend to be obtained by native speakers of Dutch, subjects with a greater Dutch language capacity and subjects with a greater cognitive capacity. Neither SES nor any of the non-cognitive variables seem to explain differences in Lexicon scores on top of the other variables.

Extended model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	579.6595	1	465	< .001
Latinist	2.2837	1	451	0.131
School	2.4024	3	441	0.067
Sex	5.2861	1	443	0.022
SES	1.9422	3	441	0.122
NonNative	12.4287	1	441	< .001
Dutch_T1	138.2290	1	449	< .001
G_T1	50.1296	1	443	< .001
Consistency	2.5552	1	1141	0.110
Perseverance	1.1302	1	1196	0.288
SelfControl	1.5483	1	1263	0.214
ControlledMot	0.1656	1	1216	0.684
AutoMot	1.7073	1	1216	0.192
Time * Latinist	6.9549	1	450	0.009
Time * Dutch_T1	0.4248	1	439	0.515
Time * G_T1	0.0342	1	436	0.853

Note. Satterthwaite method for degrees of freedom

The interaction effects of Dutch_T1 and G_T1 with Time are non-significant. For Lexicon, there is no evidence to support the hypothesis that a greater Dutch language capacity or cognitive capacity leads to faster progress. The full model indeed has a slightly better fit ($AIC = 9518.71$; $BIC = 9637.74$) than the extended model ($AIC = 9522.28$; $BIC = 9651.67$).

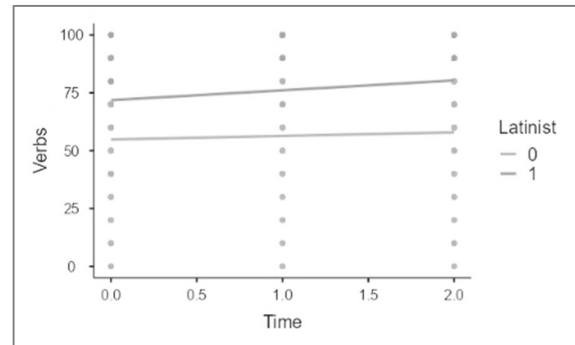
8.2.3.2 Verbs

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	17.99	1	449	<.001
Latinist	53.40	1	448	<.001
School	18.69	3	449	<.001
Time * Latinist	3.92	1	449	0.048

Note. Satterthwaite method for degrees of freedom



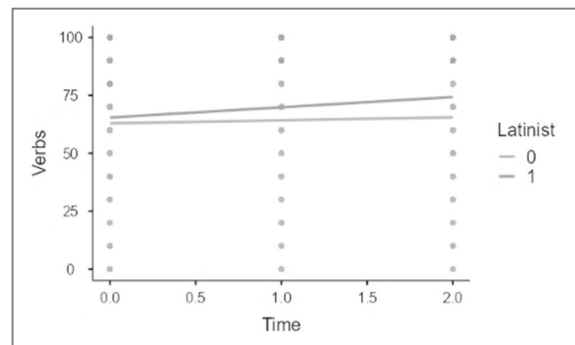
The main effect of Latinist is highly significant, indicating that latinists score higher on Verbs across time. The interaction effect with Time is just about significant as well, indicating that latinists also make more progress over time.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	14.290	1	956	<.001
Latinist	1.986	1	957	0.159
School	10.858	3	1207	<.001
Sex	0.758	1	1221	0.384
SES	0.525	3	1208	0.665
NonNative	2.184	1	1205	0.140
Dutch_T1	548.025	1	1209	<.001
G_T1	11.889	1	1214	<.001
Consistency	0.242	1	1256	0.623
Perseverance	5.013	1	1286	0.025
SelfControl	1.239	1	1296	0.266
ControlledMot	0.270	1	1278	0.603
AutoMot	0.847	1	1276	0.358
Time * Latinist	4.270	1	951	0.039

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	64.55477	2.270	60.106	69.003	1264	28.44239	< .001
Time	Time	1.30723	1.022	-0.696	3.311	952	1.27893	0.201
Latinist1	1 - 0	2.49426	1.770	-0.975	5.964	957	1.40910	0.159
School1	3 - 1, 3, 4, 5	0.00903	1.208	-2.360	2.378	1204	0.00747	0.994
School2	4 - 1, 3, 4, 5	4.46015	1.001	2.499	6.422	1210	4.45688	< .001
School3	5 - 1, 3, 4, 5	1.00544	1.114	-1.179	3.190	1206	0.90217	0.367
Sex1	0 - 1	-1.10902	1.274	-3.606	1.388	1221	-0.87053	0.384
SES1	1 - 0	0.37233	3.138	-5.778	6.522	1208	0.11866	0.906
SES2	2 - 0	1.31395	3.020	-4.605	7.233	1211	0.43509	0.664
SES3	3 - 0	3.01989	3.372	-3.589	9.628	1206	0.89566	0.371
NonNative1	0 - 1	-3.27976	2.219	-7.629	1.070	1205	-1.47793	0.140
Dutch_T1	Dutch_T1	17.76848	0.759	16.281	19.256	1209	23.40994	< .001
G_T1	G_T1	-2.38053	0.690	-3.734	-1.027	1214	-3.44798	< .001
Consistency	Consistency	0.36352	0.739	-1.085	1.812	1256	0.49171	0.623
Perseverance	Perseverance	1.73839	0.776	0.217	3.260	1286	2.23892	0.025
SelfControl	SelfControl	-0.97277	0.874	-2.686	0.740	1296	-1.11313	0.266
ControlledMot	ControlledMot	0.31588	0.608	-0.875	1.507	1278	0.51987	0.603
AutoMot	AutoMot	-0.61499	0.668	-1.925	0.695	1276	-0.92028	0.358
Time * Latinist1	Time * 1 - 0	3.12567	1.513	0.161	6.091	951	2.06628	0.039

The main effect of Latinist is no longer significant, suggesting that the control variables adequately control for the a priori difference between the two groups. The interaction effect with Time remains significant, meaning that even when taking into account the control variables latinists make additional progress on Verbs. The average non-latinist improves their Verbs score 1.3 p.p. per school year, and the average latinist 4.4 p.p. The size of this effect is very small ($\Delta R^2 = .008$; Cohen's $f^2 = .015$).

The effect of School, which was highly significant in the base model, remains highly significant. Pupils of school no. 4 in particular score significantly higher on Verbs. Variables that also make a significant contribution to the model are Dutch_T1, G_T1 and Perseverance of effort. The effects of Dutch_T1 and Perseverance of effort have the expected direction. The effect of G_T1, however, does not, as G_T1 is inversely related with Verbs here. Sex, SES and the remaining non-cognitive variables do not seem to explain differences in Verbs scores on top of the other variables.

Extended model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	18.229	1	1023	< .001
Latinist	0.405	1	903	0.525
School	11.577	3	1214	< .001
Sex	0.680	1	1229	0.410
SES	0.557	3	1215	0.643
NonNative	2.266	1	1211	0.133
Dutch_T1	571.742	1	913	< .001
G_T1	15.380	1	887	< .001
Consistency	0.218	1	1273	0.641
Perseverance	5.385	1	1295	0.020
SelfControl	1.503	1	1299	0.220
ControlledMot	0.558	1	1290	0.455
AutoMot	0.310	1	1288	0.578
Time * Latinist	26.963	1	1019	< .001
Time * Dutch_T1	72.461	1	1013	< .001
Time * G_T1	3.824	1	1009	0.051

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	66.1225	2.221	61.77009	70.475	1278	29.7760	< .001
Time	Time	-1.0611	1.035	-3.08958	0.967	1018	-1.0252	0.305
Latinist1	1 - 0	-1.1495	1.807	-4.69171	2.393	903	-0.6360	0.525
School1	3 - 1, 3, 4, 5	-0.0974	1.173	-2.39645	2.202	1211	-0.0830	0.934
School2	4 - 1, 3, 4, 5	4.4797	0.972	2.57553	6.384	1217	4.6109	< .001
School3	5 - 1, 3, 4, 5	1.0839	1.082	-1.03638	3.204	1213	1.0020	0.317
Sex1	0 - 1	-1.0202	1.237	-3.44536	1.405	1229	-0.8245	0.410
SES1	1 - 0	0.5010	3.046	-5.46926	6.471	1215	0.1645	0.869
SES2	2 - 0	1.4282	2.932	-4.31807	7.175	1218	0.4871	0.626
SES3	3 - 0	3.1159	3.273	-3.29878	9.531	1213	0.9520	0.341
NonNative1	0 - 1	-3.2422	2.154	-7.46404	0.980	1211	-1.5052	0.133
Dutch_T1	Dutch_T1	22.6036	0.945	20.75078	24.456	913	23.9111	< .001
G_T1	G_T1	-3.4043	0.868	-5.10572	-1.703	887	-3.9217	< .001
Consistency	Consistency	0.3367	0.722	-1.07747	1.751	1273	0.4667	0.641
Perseverance	Perseverance	1.7560	0.757	0.27286	3.239	1295	2.3205	0.020
SelfControl	SelfControl	-1.0432	0.851	-2.71129	0.625	1299	-1.2258	0.220
ControlledMot	ControlledMot	0.4426	0.593	-0.71884	1.604	1290	0.7469	0.455

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
AutoMot	AutoMot	-0.3632	0.652	-1.64140	0.915	1288	-0.5570	0.578
Time * Latinist1	Time * 1 - 0	8.3876	1.615	5.22170	11.554	1019	5.1926	< .001
Time * Dutch_T1	Time * Dutch_T1	-7.1291	0.837	-8.77051	-5.488	1013	-8.5124	< .001
Time * G_T1	Time * G_T1	1.5342	0.785	-0.00349	3.072	1009	1.9555	0.051

The interaction effect between Dutch_T1 and Time is highly significant, and the interaction effect between G_T1 is just about significant. The hypothesis that a greater Dutch language capacity and cognitive capacity lead to faster progress on Verbs is not supported, however. The parameter estimates indicate that a greater Dutch language capacity restrains progress, which suggests a ceiling effect in the higher ranges of performance. When removing the interaction between Dutch_T1 and Time, the interaction between G_T1 and Time becomes non-significant ($p = .474$).

8.2.4 General cognitive ability

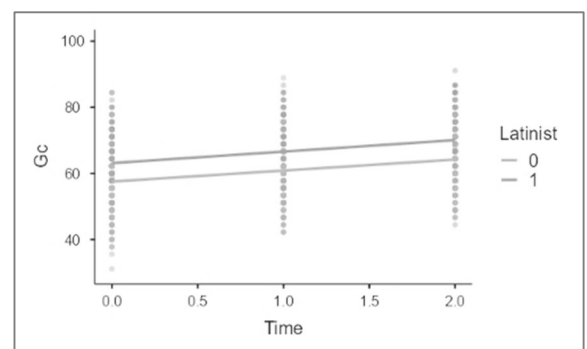
8.2.4.1 Verbal intelligence: crystallised intelligence (Gc)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	384.771	1	453	< .001
Latinist	49.988	1	452	< .001
School	3.162	3	459	0.024
Time * Latinist	0.188	1	453	0.665

Note. Satterthwaite method for degrees of freedom



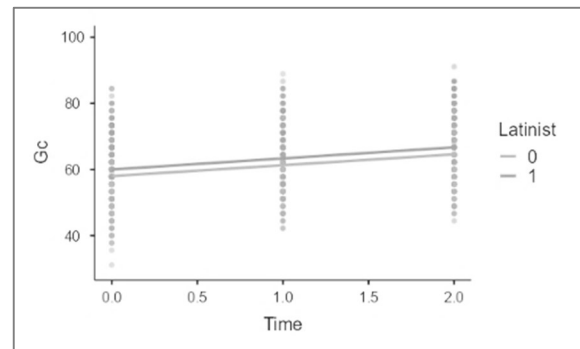
The main effect of Latinist is highly significant, indicating that latinists score higher on Gc across time. The interaction effect with Time is non-significant, indicating that latinists do not make more progress over time.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	348.1016	1	462	< .001
Latinist	8.6647	1	477	0.003
School	0.6372	3	439	0.591
Sex	8.6391	1	445	0.003
SES	1.8705	3	441	0.134
NonNative	5.9812	1	440	0.015
Dutch_T1	48.4193	1	443	< .001
G_T1	162.7464	1	445	< .001
Consistency	6.9940	1	1258	0.008
Perseverance	4.54e-4	1	1294	0.983
SelfControl	6.3183	1	1286	0.012
ControlledMot	0.5492	1	1294	0.459
AutoMot	4.4138	1	1262	0.036
Time * Latinist	0.0302	1	449	0.862

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	56.85665	0.926	55.042	58.6716	480	61.3997	< .001
Time	Time	3.29514	0.239	2.827	3.7633	453	13.7946	< .001
Latinist1	1 - 0	1.99073	0.676	0.665	3.3162	477	2.9436	0.003
School1	3 - 1, 3, 4, 5	0.69468	0.503	-0.291	1.6803	437	1.3814	0.168
School2	4 - 1, 3, 4, 5	-0.19078	0.417	-1.009	0.6270	441	-0.4572	0.648
School3	5 - 1, 3, 4, 5	-0.21725	0.464	-1.128	0.6930	436	-0.4678	0.640
Sex1	0 - 1	1.56045	0.531	0.520	2.6010	445	2.9392	0.003
SES1	1 - 0	0.21077	1.312	-2.361	2.7828	441	0.1606	0.872
SES2	2 - 0	1.46850	1.258	-0.997	3.9338	449	1.1675	0.244
SES3	3 - 0	1.99989	1.404	-0.753	4.7524	444	1.4240	0.155
NonNative1	0 - 1	2.26359	0.926	0.450	4.0776	440	2.4457	0.015
Dutch_T1	Dutch_T1	2.19652	0.316	1.578	2.8152	443	6.9584	< .001
G_T1	G_T1	3.66739	0.287	3.104	4.2308	445	12.7572	< .001
Consistency	Consistency	-0.60220	0.228	-1.048	-0.1559	1258	-2.6446	0.008
Perseverance	Perseverance	0.00519	0.244	-0.472	0.4827	1294	0.0213	0.983
SelfControl	SelfControl	0.69485	0.276	0.153	1.2366	1286	2.5136	0.012
ControlledMot	ControlledMot	0.14399	0.194	-0.237	0.5248	1294	0.7411	0.459
AutoMot	AutoMot	-0.45704	0.218	-0.883	-0.0307	1262	-2.1009	0.036

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
Time * Latinist1	Time * 1 - 0	0.06131	0.353	-0.630	0.7528	449	0.1738	0.862

The main effect of Latinist remains significant, suggesting that the control variables do not fully cover the a priori difference between the two groups. The interaction effect with Time remains non-significant. The average non-latinist improves their Gc score 3.3 p.p. per school year, and the average latinist 3.4 p.p.

The effect of School, which was significant in the base model, no longer reaches significance, suggesting that the control variables capture the differences between schools. Variables that make a significant contribution to the model are Sex, NonNative, Dutch_T1, G_T1, Consistency of interest, Self-control and Autonomous motivation for school. Keeping all other variables constant, boys score 1.6 p.p. higher than girls. The effects of NonNative, Dutch_T1, G_T1 and Self-control have the expected direction. The effects of Consistency of interest and Autonomous motivation, however, do not, as these variables are inversely related with Gc here. SES, Perseverance of effort and Controlled motivation for school do not seem to explain differences in Gc scores on top of the other variables.

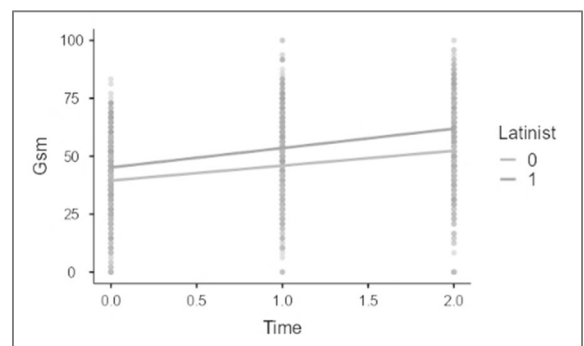
8.2.4.2 Verbal intelligence: short-term memory (Gsm)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	325.96	1	442	< .001
Latinist	15.30	1	453	< .001
School	1.35	3	457	0.256
Time * Latinist	5.56	1	442	0.019

Note. Satterthwaite method for degrees of freedom



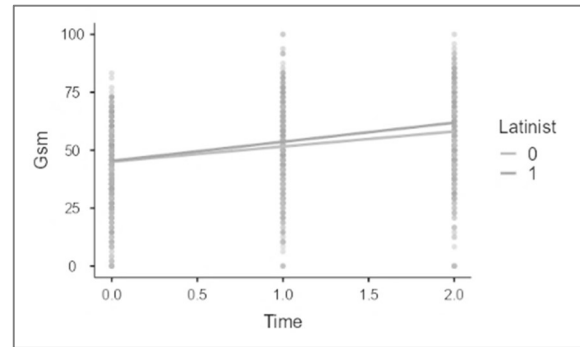
The main effect of Latinist is highly significant, indicating that latinists score higher on Gsm across time. The interaction effect with Time is significant as well, indicating that latinists also make more progress over time.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	310.2653	1	712	< .001
Latinist	0.0577	1	559	0.810
School	0.4260	3	448	0.734
Sex	0.2020	1	453	0.653
SES	3.5651	3	450	0.014
NonNative	5.6872	1	451	0.018
Dutch_T1	10.5828	1	452	0.001
G_T1	169.7856	1	453	< .001
Consistency	0.7974	1	1252	0.372
Perseverance	0.4720	1	1279	0.492
SelfControl	0.0663	1	1280	0.797
ControlledMot	0.8908	1	1248	0.345
AutoMot	0.6775	1	1217	0.411
Time * Latinist	4.3451	1	691	0.037

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	47.293	1.845	43.677	50.909	489	25.634	< .001
Time	Time	6.509	0.562	5.407	7.610	698	11.582	< .001
Latinist1	1 - 0	0.317	1.320	-2.269	2.903	559	0.240	0.810
School1	3 - 1, 3, 4, 5	-0.769	1.009	-2.746	1.208	447	-0.763	0.446
School2	4 - 1, 3, 4, 5	0.648	0.836	-0.991	2.287	450	0.775	0.439
School3	5 - 1, 3, 4, 5	0.565	0.931	-1.259	2.390	445	0.607	0.544
Sex1	0 - 1	-0.477	1.062	-2.559	1.604	453	-0.449	0.653
SES1	1 - 0	-2.793	2.631	-7.951	2.364	451	-1.062	0.289
SES2	2 - 0	-6.122	2.526	-11.073	-1.172	460	-2.424	0.016
SES3	3 - 0	-4.017	2.820	-9.544	1.510	455	-1.425	0.155
NonNative1	0 - 1	-4.429	1.857	-8.069	-0.789	451	-2.385	0.018
Dutch_T1	Dutch_T1	2.059	0.633	0.819	3.300	452	3.253	0.001
G_T1	G_T1	7.507	0.576	6.377	8.636	453	13.030	< .001
Consistency	Consistency	-0.448	0.502	-1.431	0.535	1252	-0.893	0.372
Perseverance	Perseverance	0.368	0.536	-0.683	1.420	1279	0.687	0.492
SelfControl	SelfControl	-0.157	0.608	-1.348	1.035	1280	-0.258	0.797
ControlledMot	ControlledMot	0.400	0.424	-0.431	1.231	1248	0.944	0.345
AutoMot	AutoMot	0.388	0.471	-0.536	1.311	1217	0.823	0.411

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
Time * Latinist1	Time * 1 - 0	1.732	0.831	0.103	3.360	691	2.085	0.037

The main effect of Latinist is no longer significant, suggesting that the control variables adequately control for the a priori difference between the two groups. The interaction effect with Time remains significant, meaning that even when taking into account the control variables latinists make additional progress on Gsm. The average non-latinist improves their Gsm score 6.5 p.p. per school year, and the average latinist 8.2 p.p. The size of this effect is very small ($\Delta R^2 = .005$; Cohen's $f^2 = .008$).

Variables that make a significant contribution to the model are SES, NonNative, Dutch_T1 and G_T1. The effect of SES is not in line with expectations: The highest mean score on Gsm is obtained by pupils whose parents do not have higher education degrees, and the lowest mean score by pupils of which at least one parent has a master's degree. The effects of Dutch_T1 and G_T1 have the expected direction. The effect of NonNative, however, does not, as non-native speakers of Dutch tend to score higher on Gsm. Neither Sex nor any of the non-cognitive variables seem to explain differences in Gsm scores on top of the other variables.

Extended model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	323.0916	1	736	<.001
Latinist	2.07e-4	1	548	0.989
School	0.4070	3	448	0.748
Sex	0.1937	1	453	0.660
SES	3.5697	3	450	0.014
NonNative	5.6571	1	451	0.018
Dutch_T1	5.2258	1	544	0.023
G_T1	193.5383	1	538	<.001
Consistency	0.7270	1	1256	0.394
Perseverance	0.4399	1	1285	0.507
SelfControl	0.0962	1	1286	0.757
ControlledMot	0.7436	1	1258	0.389
AutoMot	0.5217	1	1226	0.470
Time * Latinist	6.3336	1	714	0.012
Time * Dutch_T1	2.1241	1	702	0.145

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time * G_T1	26.9548	1	697	< .001

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	47.4549	1.851	43.828	51.082	492	25.6433	< .001
Time	Time	6.2602	0.578	5.127	7.393	720	10.8276	< .001
Latinist1	1 - 0	-0.0194	1.347	-2.660	2.621	548	-0.0144	0.989
School1	3 - 1, 3, 4, 5	-0.7482	1.010	-2.728	1.231	447	-0.7408	0.459
School2	4 - 1, 3, 4, 5	0.6227	0.837	-1.019	2.264	451	0.7436	0.458
School3	5 - 1, 3, 4, 5	0.5697	0.932	-1.257	2.397	446	0.6112	0.541
Sex1	0 - 1	-0.4681	1.064	-2.553	1.616	453	-0.4402	0.660
SES1	1 - 0	-2.8121	2.635	-7.976	2.352	451	-1.0674	0.286
SES2	2 - 0	-6.1396	2.529	-11.096	-1.184	460	-2.4281	0.016
SES3	3 - 0	-4.0157	2.823	-9.549	1.517	455	-1.4225	0.156
NonNative1	0 - 1	-4.4226	1.859	-8.067	-0.778	451	-2.3785	0.018
Dutch_T1	Dutch_T1	1.6184	0.708	0.231	3.006	544	2.2860	0.023
G_T1	G_T1	8.9976	0.647	7.730	10.265	538	13.9118	< .001
Consistency	Consistency	-0.4236	0.497	-1.397	0.550	1256	-0.8527	0.394
Perseverance	Perseverance	0.3525	0.531	-0.689	1.394	1285	0.6632	0.507
SelfControl	SelfControl	-0.1869	0.603	-1.368	0.994	1286	-0.3101	0.757
ControlledMot	ControlledMot	0.3628	0.421	-0.462	1.187	1258	0.8623	0.389
AutoMot	AutoMot	0.3382	0.468	-0.579	1.256	1226	0.7223	0.470
Time * Latinist1	Time * 1 - 0	2.2692	0.902	0.502	4.036	714	2.5167	0.012
Time * Dutch_T1	Time * Dutch_T1	0.6790	0.466	-0.234	1.592	702	1.4574	0.145
Time * G_T1	Time * G_T1	-2.2654	0.436	-3.121	-1.410	697	-5.1918	< .001

Whereas the interaction effect between Dutch_T1 and Time is non-significant, the interaction effect between G_T1 and Time is. The hypothesis that a greater cognitive capacity lead to faster progress on Gsm is not supported, however. The parameter estimates indicate that a greater cognitive capacity restrains progress, which suggests a ceiling effect in the higher ranges of performance.

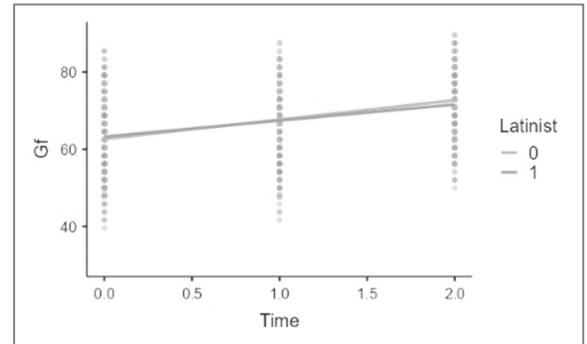
8.2.4.3 Non-verbal intelligence: fluid intelligence (Gf)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	373.284	1	790	< .001
Latinist	0.525	1	466	0.469
School	0.212	3	462	0.888
Time * Latinist	3.276	1	790	0.071

Note. Satterthwaite method for degrees of freedom



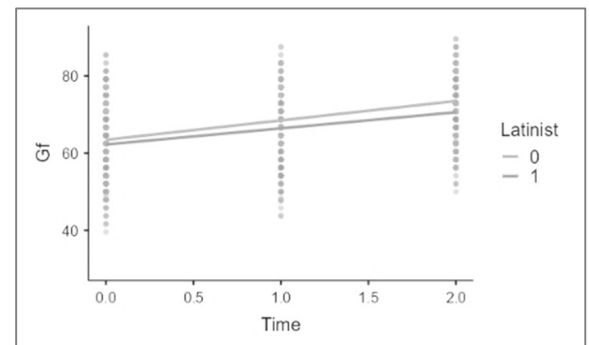
The main effect of Latinist is non-significant, indicating that as far as Gf is concerned there is no a priori difference between latinists and non-latinists. The interaction effect with Time is non-significant as well, indicating that latinists do not make more progress over time either.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	365.80451	1	905	< .001
Latinist	2.85372	1	528	0.092
School	0.32755	3	450	0.805
Sex	6.86786	1	460	0.009
SES	0.72417	3	453	0.538
NonNative	1.34849	1	452	0.246
G_T1	214.65583	1	454	< .001
Consistency	3.24e-6	1	1299	0.999
Perseverance	0.03286	1	1242	0.856
SelfControl	0.02466	1	1189	0.875
ControlledMot	0.00426	1	1211	0.948
AutoMot	0.26288	1	1155	0.608
Time * Latinist	3.19190	1	886	0.074

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	63.9853	0.926	62.170	65.8005	542	69.09035	< .001
Time	Time	5.0280	0.322	4.396	5.6597	893	15.60042	< .001
Latinist1	1 - 0	-1.2082	0.715	-2.610	0.1936	528	-1.68930	0.092
School1	3 - 1, 3, 4, 5	-0.4381	0.495	-1.409	0.5329	448	-0.88440	0.377
School2	4 - 1, 3, 4, 5	-0.0394	0.406	-0.835	0.7567	454	-0.09705	0.923
School3	5 - 1, 3, 4, 5	0.1635	0.443	-0.704	1.0308	448	0.36944	0.712
Sex1	0 - 1	-1.3574	0.518	-2.373	-0.3422	460	-2.62066	0.009
SES1	1 - 0	1.7174	1.296	-0.822	4.2571	455	1.32535	0.186
SES2	2 - 0	1.7230	1.247	-0.721	4.1669	464	1.38182	0.168
SES3	3 - 0	1.2689	1.390	-1.455	3.9929	457	0.91299	0.362
NonNative1	0 - 1	-1.0606	0.913	-2.851	0.7295	452	-1.16125	0.246
G_T1	G_T1	3.9005	0.266	3.379	4.4222	454	14.65114	< .001
Consistency	Consistency	5.01e-4	0.278	-0.545	0.5445	1299	-0.00180	0.999
Perseverance	Perseverance	0.0533	0.294	-0.523	0.6294	1242	0.18127	0.856
SelfControl	SelfControl	-0.0514	0.327	-0.693	0.5901	1189	-0.15705	0.875
ControlledMot	ControlledMot	-0.0152	0.233	-0.471	0.4410	1211	-0.06525	0.948
AutoMot	AutoMot	-0.1321	0.258	-0.637	0.3729	1155	-0.51272	0.608
Time * Latinist1	Time * 1 - 0	-0.8543	0.478	-1.792	0.0829	886	-1.78659	0.074

Both the main effect of Latinist and the interaction effect with Time remain non-significant. The average non-latinist improves their Gf score 5.0 p.p. per school year, and the average latinist 4.2 p.p.

Apart from Time, the only variables that make a significant contribution to the model are Sex and G_T1. Keeping all other variables constant, girls score 1.4 p.p. higher than boys. The effect of G_T1 has the expected direction: Higher scores on Gf tend to be obtained by subjects with a greater cognitive capacity.

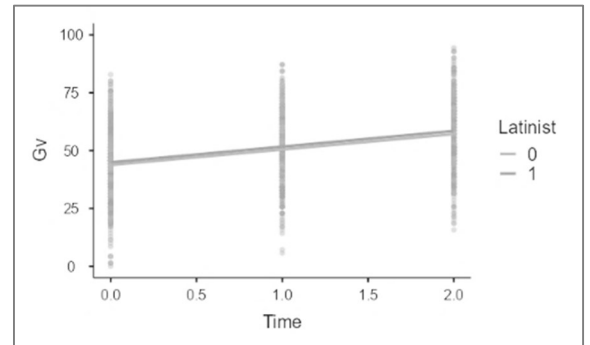
8.2.4.4 Non-verbal intelligence: visuo-spatial processing (Gv)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	617.7447	1	453	< .001
Latinist	0.6936	1	459	0.405
School	0.2588	3	460	0.855
Time * Latinist	0.0467	1	453	0.829

Note. Satterthwaite method for degrees of freedom



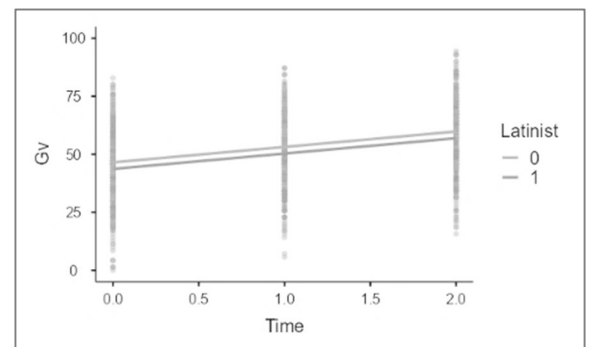
The main effect of Latinist is non-significant, indicating that as far as Gv is concerned there is no a priori difference between latinists and non-latinists. The interaction effect with Time is non-significant as well, indicating that latinists do not make more progress over time either.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	554.75713	1	472	< .001
Latinist	6.41742	1	466	0.012
School	1.16290	3	448	0.323
Sex	17.44847	1	451	< .001
SES	0.85896	3	449	0.462
NonNative	0.05060	1	449	0.822
G_T1	254.48041	1	451	< .001
Consistency	0.13322	1	1171	0.715
Perseverance	1.34732	1	1229	0.246
SelfControl	1.16205	1	1284	0.281
ControlledMot	0.38697	1	1236	0.534
AutoMot	4.25428	1	1236	0.039
Time * Latinist	0.00279	1	457	0.958

Note. Satterthwaite method for degrees of freedom



Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	46.7026	1.712	43.347	50.0581	462	27.2795	< .001
Time	Time	6.6463	0.376	5.910	7.3827	462	17.6903	< .001
Latinist1	1 - 0	-2.8158	1.112	-4.994	-0.6372	466	-2.5333	0.012
School1	3 - 1, 3, 4, 5	-1.0161	0.961	-2.900	0.8674	447	-1.0574	0.291
School2	4 - 1, 3, 4, 5	-0.3631	0.785	-1.901	1.1747	450	-0.4628	0.644
School3	5 - 1, 3, 4, 5	-0.2826	0.858	-1.964	1.3990	447	-0.3294	0.742
Sex1	0 - 1	4.1551	0.995	2.205	6.1047	451	4.1771	< .001
SES1	1 - 0	-1.2739	2.503	-6.180	3.6321	449	-0.5089	0.611
SES2	2 - 0	-1.3792	2.397	-6.077	3.3190	456	-0.5754	0.565
SES3	3 - 0	0.9769	2.682	-4.281	6.2345	452	0.3642	0.716
NonNative1	0 - 1	-0.3976	1.768	-3.862	3.0667	449	-0.2249	0.822
G_T1	G_T1	8.1987	0.514	7.191	9.2060	451	15.9524	< .001
Consistency	Consistency	-0.1307	0.358	-0.833	0.5712	1171	-0.3650	0.715
Perseverance	Perseverance	-0.4529	0.390	-1.218	0.3119	1229	-1.1607	0.246
SelfControl	SelfControl	0.4826	0.448	-0.395	1.3600	1284	1.0780	0.281
ControlledMot	ControlledMot	-0.1941	0.312	-0.806	0.4175	1236	-0.6221	0.534
AutoMot	AutoMot	-0.7271	0.353	-1.418	-0.0362	1236	-2.0626	0.039
Time * Latinist1	Time * 1 - 0	-0.0294	0.556	-1.120	1.0611	457	-0.0529	0.958

The main effect of Latinist is now significant. When taking into account the control variables, the non-latinists display a small advantage in terms of Gv. The interaction effect with Time remains non-significant. Both the average non-latinist and the average latinist improve their Gv score 6.6 p.p. per school year.

Apart from Time, the only variables that make a significant contribution to the model are Sex and G_T1. Keeping all other variables constant, boys score 4.2 p.p. higher than girls. The effect of G_T1 has the expected direction: Higher scores on Gv tend to be obtained by subjects with a greater cognitive capacity.

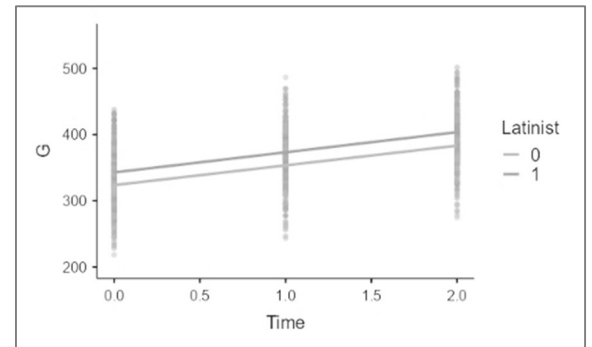
8.2.4.5 Overall general cognitive ability (G)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	1441.124	1	893	< .001
Latinist	25.924	1	460	< .001
School	1.872	3	459	0.133
Time * Latinist	0.150	1	893	0.699

Note. Satterthwaite method for degrees of freedom



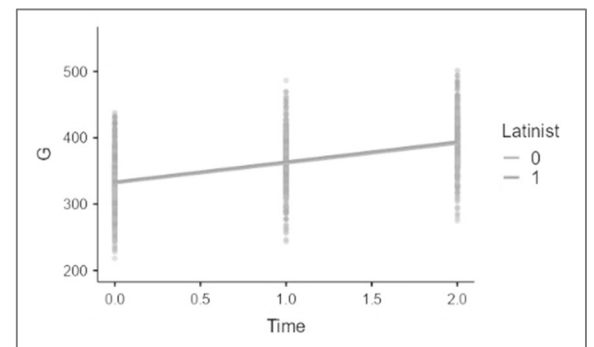
The main effect of Latinist is highly significant, indicating that latinists score higher on G across time. The interaction effect with Time is non-significant, indicating that latinists do not make more progress over time.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	1130.2715	1	525	< .001
Latinist	1.1390	1	877	0.286
School	0.3178	3	641	0.813
Sex	1.7544	1	649	0.186
SES	0.5233	3	643	0.666
NonNative	0.0168	1	642	0.897
G_T1	2314.6098	1	645	< .001
Consistency	4.7752	1	1112	0.029
Perseverance	0.2217	1	1122	0.638
SelfControl	3.2591	1	1135	0.071
ControlledMot	0.2278	1	1099	0.633
AutoMot	3.7355	1	1089	0.054
Time * Latinist	0.0495	1	518	0.824

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	331.901	2.510	326.982	336.8208	737	132.2374	< .001
Time	Time	29.822	1.198	27.473	32.1712	520	24.8836	< .001
Latinist1	1 - 0	1.842	1.726	-1.541	5.2249	877	1.0673	0.286
School1	3 - 1, 3, 4, 5	-0.927	1.393	-3.658	1.8032	639	-0.6657	0.506
School2	4 - 1, 3, 4, 5	0.170	1.139	-2.062	2.4019	644	0.1494	0.881
School3	5 - 1, 3, 4, 5	1.095	1.244	-1.343	3.5321	639	0.8803	0.379
Sex1	0 - 1	1.910	1.442	-0.916	4.7355	649	1.3246	0.186
SES1	1 - 0	0.101	3.632	-7.017	7.2194	644	0.0279	0.978
SES2	2 - 0	-0.112	3.499	-6.969	6.7456	654	-0.0319	0.975
SES3	3 - 0	2.648	3.906	-5.009	10.3040	648	0.6778	0.498
NonNative1	0 - 1	-0.333	2.566	-5.362	4.6964	642	-0.1297	0.897
G_T1	G_T1	35.894	0.746	34.432	37.3566	645	48.1104	< .001
Consistency	Consistency	-1.714	0.785	-3.252	-0.1767	1112	-2.1852	0.029
Perseverance	Perseverance	0.396	0.842	-1.254	2.0463	1122	0.4708	0.638
SelfControl	SelfControl	1.716	0.951	-0.147	3.5799	1135	1.8053	0.071
ControlledMot	ControlledMot	0.314	0.659	-0.977	1.6055	1099	0.4773	0.633
AutoMot	AutoMot	-1.407	0.728	-2.833	0.0198	1089	-1.9327	0.054
Time * Latinist1	Time * 1 - 0	0.396	1.781	-3.094	3.8860	518	0.2225	0.824

The main effect of Latinist is no longer significant, suggesting that the control variables adequately control for the a priori difference between the two groups. The interaction effect with Time remains non-significant. The average non-latinist improves their G score with 30 points (on a scale of 600) per school year, as does the average latinist.

The only other variables that make a significant contribution to the model are G_T1 and Consistency of interest. Unsurprisingly, baseline performance on G is by far the best predictor for G over time. The effect of Consistency of interest does not have the expected direction, as Consistency of interest is inversely related with G here.

8.2.5 School achievement

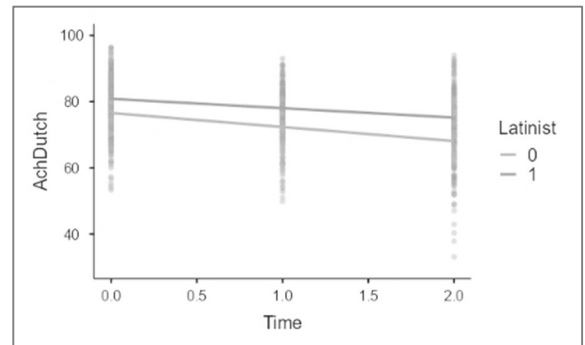
8.2.5.1 Achievement for Dutch

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	39.03	1	459	< .001
School	1.68	3	457	0.171
Time	171.74	1	310	< .001
Latinist * Time	7.15	1	329	0.008

Note. Satterthwaite method for degrees of freedom



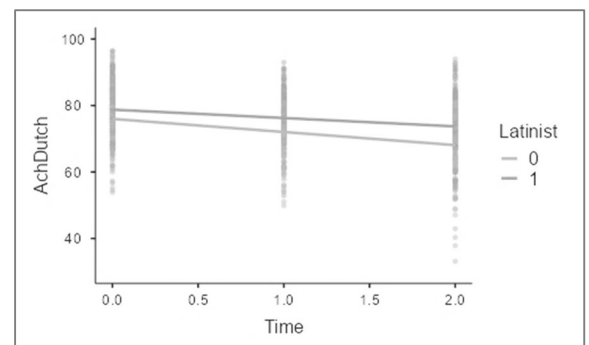
The main effect of Latinist is highly significant, indicating that latinists achieve higher grades for Dutch across time. The interaction effect with Time is significant as well, indicating that latinists' grades drop less over time.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	17.207	1	463	< .001
School	2.064	3	450	0.104
Time	136.637	1	318	< .001
Sex	45.574	1	445	< .001
SES	1.310	3	433	0.271
NonNative	1.422	1	414	0.234
G_T1	38.506	1	444	< .001
Consistency	2.014	1	953	0.156
Perseverance	7.767	1	975	0.005
SelfControl	18.486	1	994	< .001
ControlledMot	3.693	1	990	0.055
AutonMot	0.635	1	977	0.426
Latinist * Time	7.237	1	345	0.007

Note. Satterthwaite method for degrees of freedom



After adding the control variables, both the main effect of Latinist and the interaction effect with Time remain significant.

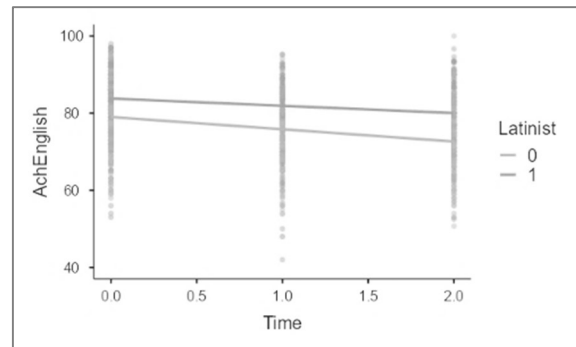
8.2.5.2 Achievement for English

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	36.24	1	443	< .001
School	5.81	3	454	< .001
Time	84.55	1	305	< .001
Latinist * Time	6.02	1	331	0.015

Note. Satterthwaite method for degrees of freedom



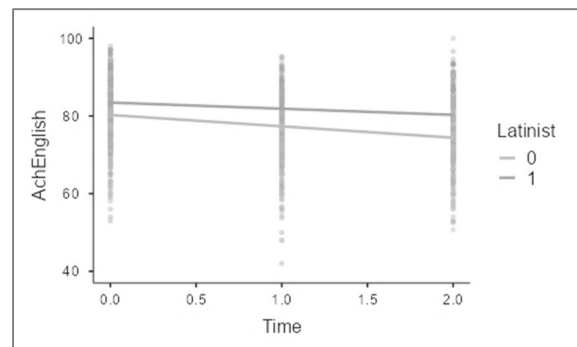
The main effect of Latinist is highly significant, indicating that latinists achieve higher grades for English across time. The interaction effect with Time is significant as well, indicating that latinists' grades drop less over time.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	15.493	1	450	< .001
School	4.819	3	440	0.003
Time	61.599	1	306	< .001
Sex	7.435	1	436	0.007
SES	1.290	3	422	0.277
NonNative	4.895	1	402	0.027
G_T1	31.671	1	434	< .001
Consistency	0.775	1	965	0.379
Perseverance	1.907	1	992	0.168
SelfControl	2.522	1	1014	0.113
ControlledMot	2.216	1	1014	0.137
AutoMot	3.981	1	1014	0.046
Latinist * Time	6.012	1	333	0.015

Note. Satterthwaite method for degrees of freedom



After adding the control variables, both the main effect of Latinist and the interaction effect with Time remain significant.

The effect of School, which was highly significant in the base model, remains highly significant, suggesting that schools do not use comparable grading systems.

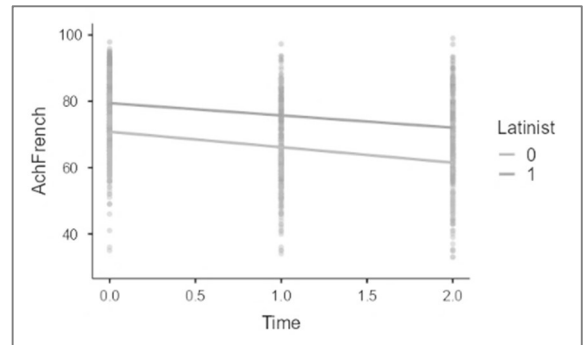
8.2.5.3 Achievement for French

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	110.00	1	487	< .001
School	34.23	3	465	< .001
Time	209.75	1	435	< .001
Latinist * Time	2.84	1	421	0.093

Note. Satterthwaite method for degrees of freedom



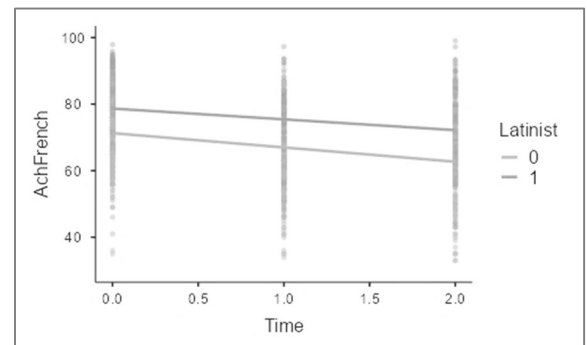
The main effect of Latinist is highly significant, indicating that latinists achieve higher grades for French across time. The interaction effect with Time is non-significant, indicating that the drop in grades over time is about equally steep for latinists and non-latinists.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	77.06	1	495	< .001
School	38.57	3	448	< .001
Time	159.10	1	447	< .001
Sex	19.99	1	459	< .001
SES	2.25	3	453	0.082
NonNative	5.21	1	444	0.023
G_T1	9.87	1	461	0.002
Consistency	3.69e-4	1	919	0.985
Perseverance	8.35	1	956	0.004
SelfControl	6.51	1	984	0.011
ControlledMot	1.08	1	953	0.299
AutoMot	3.31	1	950	0.069
Latinist * Time	3.16	1	423	0.076

Note. Satterthwaite method for degrees of freedom



After adding the control variables, the main effect of Latinist remains significant and the interaction effect with Time remains non-significant.

The effect of School, which was highly significant in the base model, remains highly significant, suggesting that schools do not use comparable grading systems.

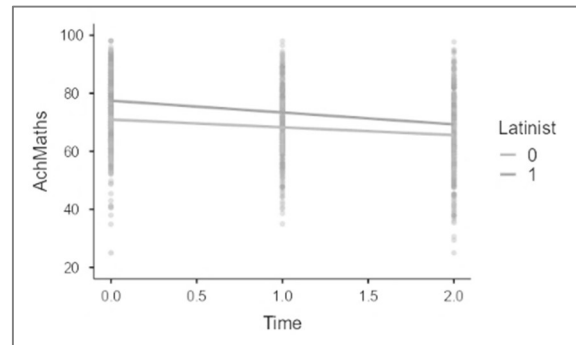
8.2.5.4 Achievement for mathematics

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	41.90	1	462	< .001
School	11.79	3	452	< .001
Time	75.24	1	308	< .001
Latinist * Time	3.43	1	325	0.065

Note. Satterthwaite method for degrees of freedom



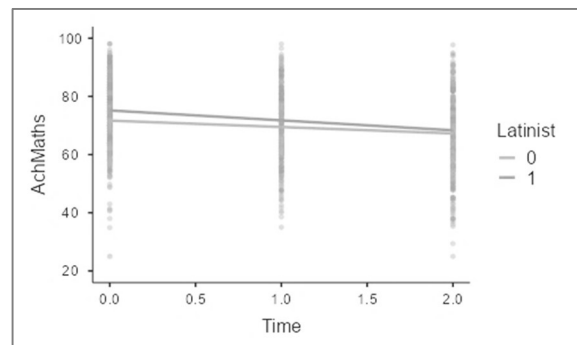
The main effect of Latinist is highly significant, indicating that latinists achieve higher grades for mathematics across time. The interaction effect with Time is non-significant, indicating that the drop in grades over time is about equally steep for latinists and non-latinists.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	12.5470	1	463	< .001
School	11.5409	3	438	< .001
Time	51.7028	1	313	< .001
Sex	0.0374	1	448	0.847
SES	3.8308	3	439	0.010
NonNative	0.8551	1	426	0.356
G_T1	34.9399	1	448	< .001
Consistency	0.4239	1	935	0.515
Perseverance	15.6261	1	966	< .001
SelfControl	5.2361	1	995	0.022
ControlledMot	3.6556	1	980	0.056
AutoMot	3.9454	1	968	0.047
Latinist * Time	2.5402	1	325	0.112

Note. Satterthwaite method for degrees of freedom



After adding the control variables, the main effect of Latinist remains significant and the interaction effect with Time remains non-significant.

The effect of School, which was highly significant in the base model, remains highly significant, suggesting that schools do not use comparable grading systems.

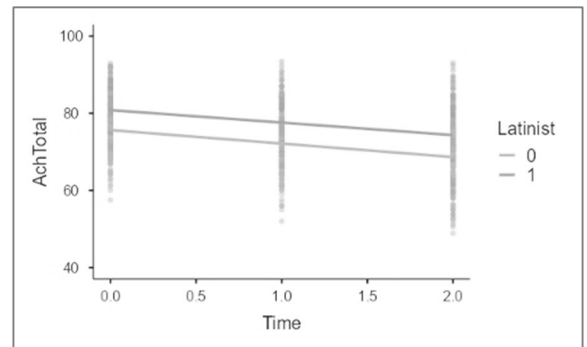
8.2.5.5 Total achievement

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	51.971	1	303	< .001
School	12.259	2	303	< .001
Time	378.133	1	298	< .001
Latinist * Time	0.549	1	298	0.459

Note. Satterthwaite method for degrees of freedom



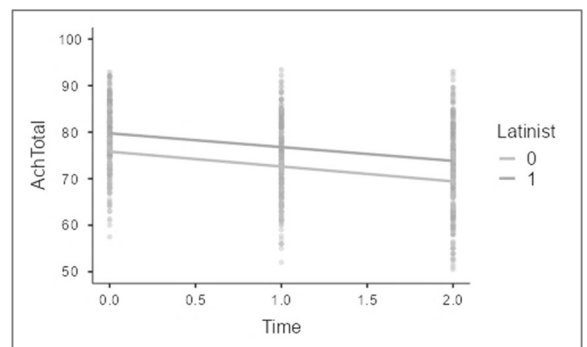
The main effect of Latinist is highly significant, indicating that latinists achieve higher total grades across time. The interaction effect with Time is non-significant, indicating that the drop in grades over time is about equally steep for latinists and non-latinists.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Latinist	34.024	1	298	< .001
School	10.394	2	290	< .001
Time	299.889	1	302	< .001
Sex	20.951	1	289	< .001
SES	0.810	3	290	0.489
NonNative	0.123	1	292	0.726
G_T1	24.734	1	292	< .001
Consistency	1.031	1	676	0.310
Perseverance	7.449	1	716	0.007
SelfControl	13.905	1	782	< .001
ControlledMot	3.590	1	736	0.059
AutoMot	3.276	1	739	0.071
Latinist * Time	0.443	1	298	0.506

Note. Satterthwaite method for degrees of freedom



After adding the control variables, the main effect of Latinist remains significant and the interaction effect with Time remains non-significant.

The effect of School, which was highly significant in the base model, remains highly significant, suggesting that schools do not use comparable grading systems.

8.3 Results: Greek

8.3.1 Models and interpretation

All of the analyses in this section are conducted on the filtered longitudinal sample, with further exceptions of latinists who dropped out of Greek (cf. Section 8.1.2.3), using the GAMLj package in Jamovi. The method of estimation is Maximum Likelihood. The displayed graphs are effect plots, which take into account the effects of the other variables in the model. No noteworthy violations were detected against the assumption of normality of variables and residuals.

The base model and full model are defined in exactly the same way as in the previous results section, but replacing the grouping variable Latinist by the grouping variable Hellenist. Since there are now three groups (non-latinists, latinists *stricto sensu* and hellenists) as opposed to two, contrast analyses are added whenever the main effect of Hellenist is significant. Extended models are omitted, as the hypothesis that greater capacity equals faster progress (cf. Section 8.2.1) was previously not confirmed.

In all models for all outcome variables, the effect of Time is highly significant. This observation will therefore be left out in the interpretation of the tables below. The definition of significance as having a p-value equal to or below .05 is maintained.

“Average” is defined in the same way as well: averaged across schools, averaged across sexes, with parents of average education level (for this sample), native speaker of Dutch, and with average scores on all covariates. To check the evolution per school year for the average non-latinist the variable Hellenist was temporarily dummy-coded (with non-latinists being the reference group), whereas in the tables a repeated coding scheme is reported.

Interpretation will be limited to descriptions of the main effects of Hellenist and its interaction effects with Time. The contributions of control variables to the full models will not be made explicit here for the following reason: In terms of the significance of control variables, the structures of all Hellenist models are identical to the structures of their Latinist counterparts, the only exception being that Autonomous motivation reaches significance in the Hellenist model for Gv while it did not in the Latinist model.

In line with the hypothesis that Greek has a complementary effect on top of Latin (cf. Section 8.1.4.1), focus lies on comparing non-latinists and latinists on the one hand and latinists

and hellenists on the other hand. It is expected that hellenists perform equally well or better than latinists, and as we will see, this expectation is in fact largely fulfilled. Therefore, if for a given outcome variable the difference between non-latinists and latinists – either in absolute scores or evolution over time – is significant, then it logically follows and is implied that the difference between non-latinists and hellenists is significant as well; this holds true for all outcome variables except Gsm, which is also explicitly mentioned in the interpretation of the results for Gsm below.

Before moving to the longitudinal linear mixed models, the descriptive statistics of the outcome variables are reported and interpreted, comparing test variables' group means from the first measurement through t-tests.

8.3.2 Descriptive statistics

Descriptives							
	Hellenist	Time	N	Mean	SD	Minimum	Maximum
Lexicon	0	0	249	44.4	12.60	7.69	72.2
		1	249	50.3	13.11	16.67	80.6
		2	249	55.7	14.06	22.22	94.4
	1	0	126	50.5	14.05	25.00	83.3
		1	123	58.6	14.82	25.81	91.7
		2	126	64.3	15.19	27.78	94.4
	2	0	42	60.0	13.28	33.33	88.9
		1	42	68.4	16.00	22.22	88.9
		2	42	75.7	14.75	33.33	97.2
Verbs	0	0	248	54.6	26.99	0.00	100.0
		1	249	59.7	29.13	0.00	100.0
		2	249	57.8	26.57	0.00	100.0
	1	0	126	69.5	27.42	0.00	100.0
		1	123	78.0	24.18	0.00	100.0
		2	126	78.3	22.86	10.00	100.0
	2	0	42	76.7	22.60	10.00	100.0
		1	42	84.0	18.62	10.00	100.0
		2	42	86.9	16.45	30.00	100.0
Gf	0	0	244	62.5	9.51	39.58	83.3
		1	255	68.0	9.54	41.67	87.5
		2	257	72.5	7.56	50.00	89.6
	1	0	127	63.4	8.91	41.67	85.4
		1	125	67.0	8.27	43.75	83.3
		2	127	71.4	7.35	52.08	87.5

Descriptives

	Hellenist	Time	N	Mean	SD	Minimum	Maximum
Gc	2	0	43	64.4	9.36	47.92	79.2
		1	42	66.3	8.99	50.00	83.3
		2	43	74.0	7.08	56.25	87.5
	0	0	244	57.4	8.57	31.11	77.8
		1	255	61.0	8.18	42.22	88.9
		2	257	64.2	7.72	46.67	82.2
	1	0	127	61.6	8.21	40.00	82.2
		1	125	64.5	8.95	42.22	86.7
		2	127	68.8	8.36	44.44	86.7
	2	0	43	66.2	9.21	46.67	84.4
		1	42	69.6	9.78	46.67	82.2
		2	43	73.2	9.92	46.67	91.1
Gv	0	0	244	43.9	14.95	0.00	82.9
		1	255	50.2	14.97	5.71	78.6
		2	257	57.2	15.25	15.71	92.9
	1	0	127	42.9	12.97	1.43	80.0
		1	125	51.7	13.29	25.71	87.1
		2	127	56.9	12.83	25.71	92.9
	2	0	43	50.0	16.25	17.14	78.6
		1	42	55.6	16.69	22.86	87.1
		2	43	62.7	16.57	31.43	94.3
Gsm	0	0	244	38.6	16.55	0.00	83.3
		1	255	47.9	16.41	0.00	100.0
		2	257	51.5	18.35	0.00	95.8
	1	0	127	43.3	15.34	0.00	72.9
		1	125	56.9	17.32	10.42	100.0
		2	127	61.6	16.93	0.00	100.0
	2	0	43	48.1	13.17	20.83	68.8
		1	42	55.0	11.48	33.33	81.3
		2	43	60.5	15.31	25.00	85.4
G	0	0	244	322.3	40.23	218.31	422.9
		1	255	356.1	40.70	243.53	486.3
		2	257	382.1	38.59	274.60	489.5
	1	0	127	336.0	38.75	238.04	431.8
		1	125	371.7	41.75	246.53	469.7
		2	127	398.8	38.74	283.41	494.0
	2	0	43	359.2	44.74	261.96	432.8
		1	42	382.4	39.98	262.86	458.4
		2	43	417.5	39.91	355.56	501.4
AchDutch	0	0	257	76.5	7.64	53.30	95.7
		1	171	72.4	8.07	49.90	90.0
		2	167	68.4	10.52	33.12	94.0

Descriptives

	Hellenist	Time	N	Mean	SD	Minimum	Maximum
AchEnglish	1	0	127	81.2	6.94	66.00	96.4
		1	71	78.0	6.85	52.90	91.3
		2	71	73.4	8.54	51.88	93.0
	2	0	43	80.8	7.56	68.00	95.0
		1	33	80.0	7.38	62.00	93.0
		2	33	78.5	9.09	57.21	92.5
	0	0	257	79.5	8.63	58.00	97.0
		1	169	74.6	9.13	48.00	95.3
		2	167	73.1	8.94	50.70	100.0
	1	0	127	84.2	8.97	53.00	97.0
		1	71	76.2	9.87	42.00	92.0
		2	71	79.1	8.94	52.50	93.3
	2	0	43	85.5	6.59	72.40	98.0
		1	33	83.8	7.29	66.00	93.0
		2	33	83.7	8.21	62.50	96.7
AchFrench	0	0	257	71.9	10.38	35.00	90.8
		1	171	64.8	12.66	34.00	91.7
		2	167	63.9	12.76	33.00	92.3
	1	0	127	80.5	8.55	49.00	95.0
		1	71	70.0	10.84	48.00	92.4
		2	71	71.8	11.59	40.00	93.2
	2	0	43	83.2	6.63	68.30	97.9
		1	33	73.5	12.32	41.00	97.3
		2	33	75.7	11.24	43.00	99.0
AchMaths	0	0	257	70.9	11.86	34.90	96.0
		1	171	67.6	12.65	35.00	98.0
		2	167	64.2	14.60	25.00	97.7
	1	0	127	77.6	10.93	25.00	98.0
		1	71	73.1	11.62	44.20	93.0
		2	71	68.5	13.90	37.00	93.9
	2	0	43	78.3	8.66	60.40	94.0
		1	33	74.1	10.19	47.00	92.0
		2	33	69.7	13.17	38.02	94.5
AchTotal	0	0	171	75.8	6.79	57.50	92.0
		1	171	71.8	7.88	52.00	91.0
		2	168	69.1	8.88	48.93	89.0
	1	0	71	80.3	5.87	67.00	91.5
		1	71	76.3	6.79	63.00	88.6
		2	71	73.7	8.33	55.00	93.0
	2	0	33	82.4	5.55	72.00	92.2
		1	33	79.1	6.84	61.00	93.4
		2	33	76.6	8.32	61.73	92.3

Even when separating the hellenists from the latinists *stricto sensu*, latinists still obtain notably higher mean scores than non-latinists on all variables except Gf and Gv. Thus it appears likely that the results regarding Latin will remain unchanged.

Hellenists, in turn, obtain notably higher mean scores than latinists on all variables, except for Gf and the follow-up measurements of Gsm. At the outset, hellenists score 9.5 p.p. higher than latinists on Lexicon ($p_{holm} < .001$; Cohen's $d = .724$), 7.1 p.p. on Verbs (non-significant, $p_{holm} = .134$; Cohen's $d = .267$), 4.6 p.p. on Gc ($p_{holm} = .002$; Cohen's $d = .544$), 4.8 p.p. on Gsm (non-significant, $p_{holm} = .088$; Cohen's $d = .302$), 7.1 p.p. on Gv ($p_{holm} = .018$; Cohen's $d = .489$), and 23.2 points out of 600 on G ($p_{holm} = .002$; Cohen's $d = .576$). The initial difference in school achievement is often negligible and no larger than 2.7 p.p. (for French).

8.3.3 Dutch language ability

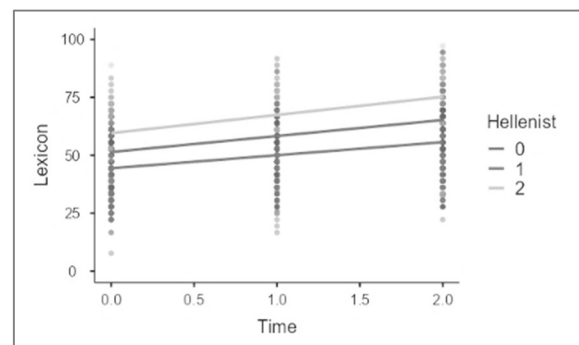
8.3.3.1 Lexicon

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	417.42	1	417	<.001
School	4.85	3	417	0.002
Hellenist	31.33	2	418	<.001
Time * Hellenist	4.54	2	417	0.011

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison									
Hellenist	Hellenist	Difference	SE	t	df	p	pbonferroni	pholm	
0	- 1	-2.12	1.15	-1.85	430	0.065	0.194	0.065	
0	- 2	-7.05	1.80	-3.91	439	<.001	<.001	<.001	
1	- 2	-4.92	1.78	-2.76	430	0.006	0.018	0.012	

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	51.783	0.821	50.17	53.393	419	63.055	< .001
Time	Time	6.793	0.332	6.14	7.444	417	20.430	< .001
School1	3 - 1, 3, 4, 5	0.841	1.185	-1.48	3.163	417	0.710	0.478
School2	4 - 1, 3, 4, 5	-0.586	0.956	-2.46	1.288	417	-0.613	0.540
School3	5 - 1, 3, 4, 5	3.173	1.054	1.11	5.239	417	3.009	0.003
Hellenist1	0 - 1	-7.011	1.412	-9.78	-4.243	418	-4.965	< .001
Hellenist2	1 - 2	-8.175	2.306	-12.69	-3.656	418	-3.546	< .001
Time * Hellenist1	Time * 0 - 1	-1.273	0.577	-2.40	-0.142	417	-2.207	0.028
Time * Hellenist2	Time * 1 - 2	-0.917	0.940	-2.76	0.925	417	-0.976	0.330

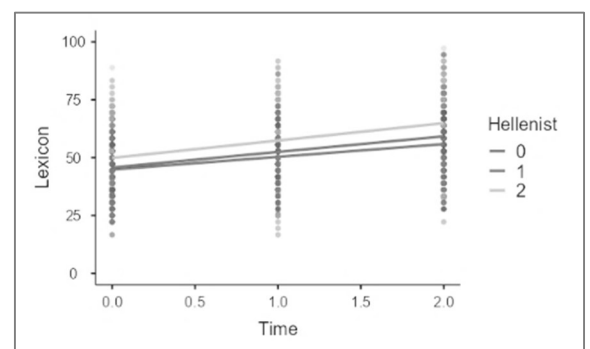
The main effect of Hellenist is highly significant, indicating that there are group differences in absolute scores on Lexicon. Pairwise comparisons show that all group differences are significant. As already shown by the descriptive statistics, hellenists score higher than latinists, and latinists score higher than non-latinists.

The interaction effect between Hellenist and Time is significant as well, indicating that there are also group differences in progress over time on Lexicon. The difference in progress between latinists and non-latinists is significant, confirming previous results. The difference in progress between hellenists and latinists is not, however.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	386.578	1	418	< .001
School	1.511	3	408	0.211
Sex	3.796	1	411	0.052
SES	1.658	3	410	0.176
NonNative	12.500	1	409	< .001
Dutch_T1	136.692	1	412	< .001
G_T1	43.390	1	414	< .001
Consistency	3.701	1	1075	0.055
Perseverance	0.649	1	1117	0.421
SelfControl	1.764	1	1172	0.184
ControlledMot	0.411	1	1134	0.522
AutoMot	2.880	1	1131	0.090
Hellenist	3.850	2	423	0.022



Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time * Hellenist	3.834	2	413	0.022

Note. Satterthwaite method for degrees of freedom

Post Hoc Comparisons - Hellenist

Comparison									
Hellenist	Hellenist	Difference	SE	t	df	p	pbonferroni	pholm	
0	- 1	-2.12	1.15	-1.85	430	0.065	0.194	0.065	
0	- 2	-7.05	1.80	-3.91	439	<.001	<.001	<.001	
1	- 2	-4.92	1.78	-2.76	430	0.006	0.018	0.012	

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	43.847	1.553	40.8034	46.8909	426	28.235	<.001
Time	Time	6.584	0.335	5.9280	7.2407	418	19.661	<.001
School1	3 - 1, 3, 4, 5	1.374	0.916	-0.4203	3.1691	408	1.501	0.134
School2	4 - 1, 3, 4, 5	-1.344	0.736	-2.7855	0.0981	409	-1.827	0.068
School3	5 - 1, 3, 4, 5	0.377	0.831	-1.2512	2.0061	407	0.454	0.650
Sex1	0 - 1	1.834	0.941	-0.0111	3.6792	411	1.948	0.052
SES1	1 - 0	0.384	2.387	-4.2938	5.0627	408	0.161	0.872
SES2	2 - 0	2.521	2.320	-2.0267	7.0688	414	1.087	0.278
SES3	3 - 0	3.283	2.599	-1.8105	8.3758	411	1.263	0.207
NonNative1	0 - 1	5.909	1.671	2.6334	9.1852	409	3.536	<.001
Dutch_T1	Dutch_T1	6.543	0.560	5.4458	7.6393	412	11.692	<.001
G_T1	G_T1	3.374	0.512	2.3699	4.3776	414	6.587	<.001
Consistency	Consistency	-0.648	0.337	-1.3073	0.0121	1075	-1.924	0.055
Perseverance	Perseverance	-0.293	0.363	-1.0046	0.4192	1117	-0.806	0.421
SelfControl	SelfControl	0.554	0.417	-0.2636	1.3716	1172	1.328	0.184
ControlledMot	ControlledMot	-0.187	0.291	-0.7577	0.3843	1134	-0.641	0.522
AutoMot	AutoMot	-0.554	0.327	-1.1946	0.0859	1131	-1.697	0.090
Hellenist1	0 - 1	-0.904	1.163	-3.1844	1.3759	424	-0.777	0.437
Hellenist2	1 - 2	-4.141	1.811	-7.6917	-0.5912	417	-2.286	0.023
Time * Hellenist1	Time * 0 - 1	-1.218	0.583	-2.3595	-0.0755	416	-2.090	0.037
Time * Hellenist2	Time * 1 - 2	-0.780	0.936	-2.6138	1.0540	408	-0.833	0.405

The main effect of Hellenist remains significant, meaning that there is a still a group difference in absolute scores on Lexicon. Pairwise comparisons suggest that the control variables capture

the a priori difference between latinists and non-latinists, but not between hellenists and the other groups.

The interaction effect between Hellenist and Time remains significant as well. Taking into account the control variables, there is still a significant difference in progress between latinists and non-latinists, and not between hellenists and latinists. As we saw previously, the average non-latinist improves their Lexicon score 5.5 p.p. per school year. The average latinist now improves their Lexicon score 6.7 p.p. per school year, and the average hellenist 7.5 p.p.

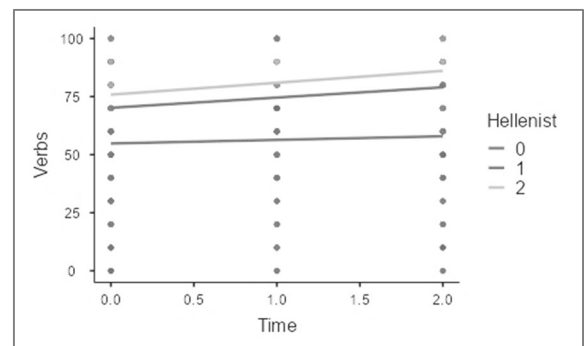
8.3.3.2 Verbs

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	16.27	1	417	<.001
School	15.20	3	417	<.001
Hellenist	23.66	2	419	<.001
Time * Hellenist	2.22	2	417	0.110

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison									
Hellenist	Hellenist	Difference	SE	t	df	p	pbonferroni	pholm	
0	- 1	-18.22	2.16	-8.44	424	<.001	<.001	<.001	
0	- 2	-24.59	3.28	-7.50	422	<.001	<.001	<.001	
1	- 2	-6.37	3.52	-1.81	423	0.071	0.214	0.071	

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	66.936	1.581	63.84	70.035	422	42.326	<.001
Time	Time	3.692	0.915	1.90	5.486	417	4.034	<.001
School1	3 - 1, 3, 4, 5	-1.896	1.870	-5.56	1.768	418	-1.014	0.311
School2	4 - 1, 3, 4, 5	4.954	1.509	2.00	7.912	418	3.284	0.001
School3	5 - 1, 3, 4, 5	6.794	1.663	3.53	10.054	417	4.084	<.001
Hellenist1	0 - 1	-15.369	2.729	-20.72	-10.021	420	-5.632	<.001
Hellenist2	1 - 2	-5.656	4.449	-14.38	3.064	419	-1.271	0.204

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
Time * Hellenist1	Time * 0 - 1	-2.852	1.588	-5.96	0.260	417	-1.796	0.073
Time * Hellenist2	Time * 1 - 2	-0.714	2.587	-5.78	4.355	416	-0.276	0.783

The main effect of Hellenist is highly significant, indicating that there are group differences in absolute scores on Verbs. Pairwise comparisons show that the differences with non-latinists are significant, but the difference between latinists and hellenists is not.

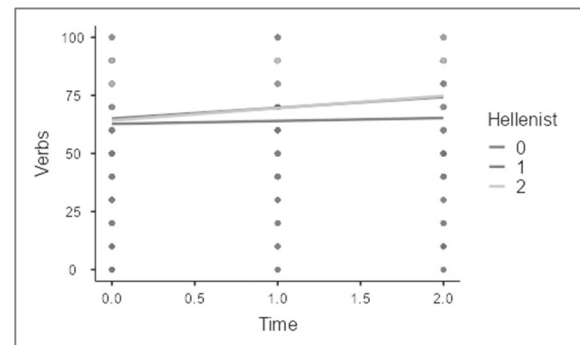
The interaction effect between Hellenist and Time is non-significant, indicating that there are no group differences in progress over time on Verbs. The difference in progress between latinists and non-latinists is not significant at the 5%-level, contradicting previous results. The difference in progress between hellenists and latinists is not at all significant.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	15.454	1	496	< .001
School	7.303	3	479	< .001
Sex	0.369	1	488	0.544
SES	0.707	3	483	0.548
NonNative	0.871	1	480	0.351
Dutch_T1	509.584	1	483	< .001
G_T1	14.921	1	487	< .001
Consistency	0.634	1	1110	0.426
Perseverance	4.558	1	1091	0.033
SelfControl	2.187	1	1085	0.139
ControlledMot	0.297	1	1040	0.586
AutoMot	0.511	1	1032	0.475
Hellenist	0.687	2	779	0.503
Time * Hellenist	2.739	2	493	0.066

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	65.135	2.358	60.513	69.757	594	27.619	< .001
Time	Time	3.737	0.951	1.874	5.601	496	3.931	< .001
School1	3 - 1, 3, 4, 5	0.159	1.347	-2.481	2.799	477	0.118	0.906
School2	4 - 1, 3, 4, 5	4.219	1.084	2.094	6.344	481	3.891	< .001
School3	5 - 1, 3, 4, 5	0.201	1.222	-2.193	2.596	477	0.165	0.869
Sex1	0 - 1	-0.845	1.390	-3.570	1.880	488	-0.608	0.544
SES1	1 - 0	-1.992	3.515	-8.880	4.897	481	-0.567	0.571
SES2	2 - 0	-0.821	3.437	-7.557	5.915	490	-0.239	0.811
SES3	3 - 0	1.567	3.837	-5.954	9.088	484	0.408	0.683
NonNative1	0 - 1	-2.299	2.463	-7.125	2.528	480	-0.933	0.351
Dutch_T1	Dutch_T1	18.655	0.826	17.035	20.275	483	22.574	< .001
G_T1	G_T1	-2.927	0.758	-4.413	-1.442	487	-3.863	< .001
Consistency	Consistency	0.615	0.773	-0.899	2.130	1110	0.796	0.426
Perseverance	Perseverance	1.745	0.817	0.143	3.347	1091	2.135	0.033
SelfControl	SelfControl	-1.359	0.919	-3.160	0.442	1085	-1.479	0.139
ControlledMot	ControlledMot	0.348	0.639	-0.905	1.601	1040	0.545	0.586
AutoMot	AutoMot	-0.506	0.707	-1.891	0.880	1032	-0.715	0.475
Hellenist1	0 - 1	-2.286	1.963	-6.132	1.561	784	-1.165	0.244
Hellenist2	1 - 2	0.804	3.086	-5.246	6.853	764	0.260	0.795
Time * Hellenist1	Time * 0 - 1	-3.371	1.658	-6.620	-0.122	495	-2.034	0.043
Time * Hellenist2	Time * 1 - 2	-0.656	2.669	-5.888	4.576	489	-0.246	0.806

The main effect of Hellenist is no longer significant, suggesting that the control variables adequately control for the a priori group differences.

While the overall interaction effect between Hellenist and Time remains non-significant, the difference in progress between non-latinists and latinists does reach significance at the 5%-level. As we saw previously, the average non-latinist improves their Verbs score 1.3 p.p. per school year. The average latinist now improves their Verbs score 4.6 p.p. per school year, and the average hellenist 5.3 p.p.

8.3.4 General cognitive ability

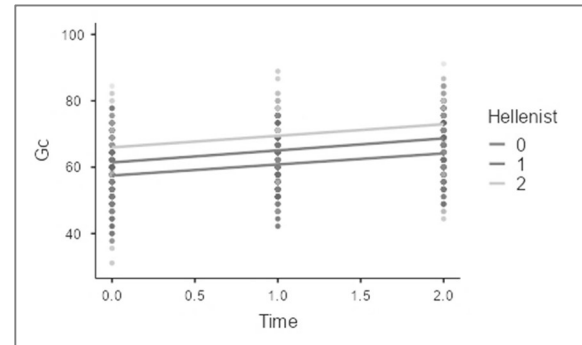
8.3.4.1 Verbal intelligence: crystallised intelligence (Gc)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	228.211	1	416	< .001
School	2.604	3	426	0.051
Hellenist	23.501	2	420	< .001
Time * Hellenist	0.275	2	417	0.760

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison									
Hellenist	Hellenist	Difference	SE	t	df	p	Pbonferroni	Pholm	
0	- 1	-4.24	0.801	-5.29	432	< .001	< .001	< .001	
0	- 2	-8.64	1.211	-7.14	432	< .001	< .001	< .001	
1	- 2	-4.40	1.304	-3.38	432	< .001	0.002	< .001	

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	61.6244	0.530	60.5860	62.663	423	116.318	< .001
Time	Time	3.4790	0.230	3.0276	3.930	416	15.106	< .001
School1	3 - 1, 3, 4, 5	0.3910	0.697	-0.9749	1.757	426	0.561	0.575
School2	4 - 1, 3, 4, 5	0.0905	0.556	-0.9987	1.180	428	0.163	0.871
School3	5 - 1, 3, 4, 5	1.1747	0.616	-0.0334	2.383	426	1.906	0.057
Hellenist1	0 - 1	-3.9392	0.915	-5.7316	-2.147	421	-4.307	< .001
Hellenist2	1 - 2	-4.5388	1.485	-7.4490	-1.629	419	-3.057	0.002
Time * Hellenist1	Time * 0 - 1	-0.2954	0.402	-1.0834	0.493	418	-0.735	0.463
Time * Hellenist2	Time * 1 - 2	0.1337	0.650	-1.1404	1.408	415	0.206	0.837

The main effect of Hellenist is highly significant, indicating that there are group differences in absolute scores on Gc. Pairwise comparisons show that all group differences are highly

significant. As already shown by the descriptive statistics, hellenists score higher than latinists, and latinists score higher than non-latinists.

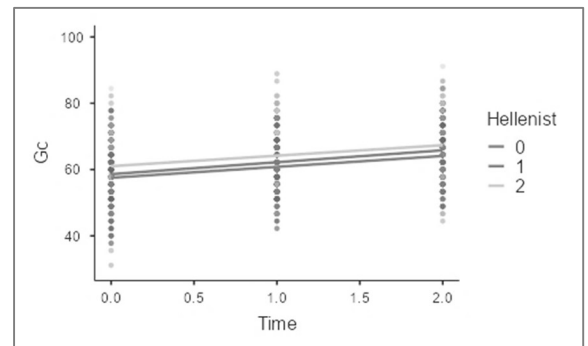
The interaction effect between Hellenist and Time is non-significant, indicating that there are no group differences in progress over time on Gc. The difference in progress between latinists and non-latinists is indeed non-significant, confirming previous results, as is the difference in progress between latinists and hellenists.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	205.1759	1	418	< .001
School	0.3828	3	408	0.765
Sex	6.0811	1	415	0.014
SES	1.3007	3	412	0.274
NonNative	6.8605	1	411	0.009
Dutch_T1	43.5940	1	412	< .001
G_T1	135.6502	1	415	< .001
Consistency	7.0355	1	1173	0.008
Perseverance	0.0149	1	1199	0.903
SelfControl	6.0692	1	1198	0.014
ControlledMot	0.7347	1	1205	0.392
AutoMot	4.5902	1	1182	0.032
Hellenist	4.6258	2	429	0.010
Time * Hellenist	0.3622	2	414	0.696

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison									
Hellenist	Hellenist	Difference	SE	t	df	p	pbonferroni	pholm	
0	- 1	-1.40	0.666	-2.09	427	0.037	0.110	0.074	
0	- 2	-3.37	1.052	-3.20	438	0.001	0.004	0.004	
1	- 2	-1.97	1.037	-1.90	431	0.058	0.173	0.074	

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	57.7483	0.934	55.918	59.5784	452	61.845	< .001
Time	Time	3.3494	0.234	2.891	3.8077	418	14.324	< .001
School1	3 - 1, 3, 4, 5	0.5099	0.540	-0.548	1.5683	406	0.944	0.346
School2	4 - 1, 3, 4, 5	0.0603	0.435	-0.792	0.9129	411	0.139	0.890
School3	5 - 1, 3, 4, 5	-0.2369	0.490	-1.197	0.7236	406	-0.483	0.629
Sex1	0 - 1	1.3761	0.558	0.282	2.4699	415	2.466	0.014
SES1	1 - 0	0.4298	1.414	-2.342	3.2013	413	0.304	0.761
SES2	2 - 0	1.6090	1.378	-1.092	4.3097	421	1.168	0.244
SES3	3 - 0	1.5680	1.539	-1.449	4.5853	415	1.019	0.309
NonNative1	0 - 1	2.5884	0.988	0.652	4.5252	411	2.619	0.009
Dutch_T1	Dutch_T1	2.1844	0.331	1.536	2.8328	412	6.603	< .001
G_T1	G_T1	3.5325	0.303	2.938	4.1270	415	11.647	< .001
Consistency	Consistency	-0.6325	0.238	-1.100	-0.1651	1173	-2.652	0.008
Perseverance	Perseverance	0.0311	0.255	-0.468	0.5301	1199	0.122	0.903
SelfControl	SelfControl	0.7102	0.288	0.145	1.2752	1198	2.464	0.014
ControlledMot	ControlledMot	0.1743	0.203	-0.224	0.5730	1205	0.857	0.392
AutoMot	AutoMot	-0.4870	0.227	-0.933	-0.0415	1182	-2.142	0.032
Hellenist1	0 - 1	-1.0881	0.748	-2.555	0.3788	429	-1.454	0.147
Hellenist2	1 - 2	-2.4077	1.174	-4.709	-0.1064	417	-2.051	0.041
Time * Hellenist1	Time * 0 - 1	-0.3071	0.407	-1.105	0.4908	417	-0.754	0.451
Time * Hellenist2	Time * 1 - 2	0.4330	0.654	-0.849	1.7147	408	0.662	0.508

The main effect of Hellenist remains significant, meaning that there is a still a group difference in absolute scores on Gc. Pairwise comparisons show that even when taking into account the control variables especially hellenists' scores still stand out.

The interaction effect between Hellenist and Time remains non-significant. As we saw previously, the average non-latinist improves their Gc score 3.3 p.p. per school year. The average latinist now improves their Gc score 3.6 p.p. per school year, and the average hellenist 3.2 p.p.

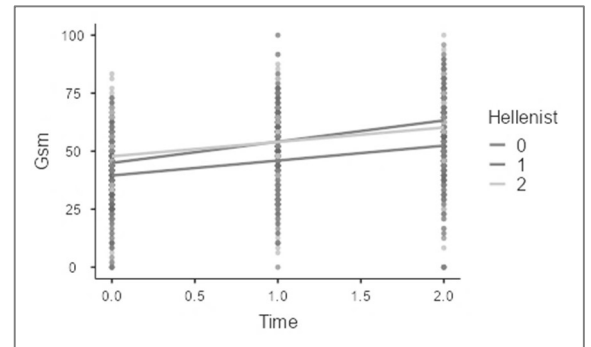
8.3.4.2 Verbal intelligence: short-term memory (Gsm)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	179.32	1	406	< .001
School	1.26	3	424	0.288
Hellenist	8.80	2	419	< .001
Time * Hellenist	4.45	2	407	0.012

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison								
Hellenist	Hellenist	Difference	SE	t	df	p	pbonferroni	pholm
0	- 1	-8.176	1.48	-5.5195	432	< .001	< .001	< .001
0	- 2	-8.041	2.24	-3.5896	432	< .001	0.001	< .001
1	- 2	0.134	2.41	0.0557	431	0.956	1.000	0.956

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	44.091	0.965	42.1993	45.983	422	45.679	< .001
Time	Time	7.257	0.542	6.1947	8.319	406	13.391	< .001
School1	3 - 1, 3, 4, 5	-0.486	1.285	-3.0040	2.032	423	-0.378	0.706
School2	4 - 1, 3, 4, 5	-0.974	1.026	-2.9844	1.036	426	-0.950	0.343
School3	5 - 1, 3, 4, 5	2.124	1.136	-0.1030	4.352	423	1.869	0.062
Hellenist1	0 - 1	-5.435	1.668	-8.7035	-2.167	421	-3.259	0.001
Hellenist2	1 - 2	-2.835	2.704	-8.1349	2.464	418	-1.049	0.295
Time * Hellenist1	Time * 0 - 1	-2.713	0.946	-4.5661	-0.859	409	-2.869	0.004
Time * Hellenist2	Time * 1 - 2	2.939	1.530	-0.0595	5.938	405	1.921	0.055

The main effect of Hellenist is highly significant, indicating that there are group differences in absolute scores on Gsm. Pairwise comparisons show that the differences with non-latinists are significant, but the difference between latinists and hellenists is not.

The interaction effect between Hellenist and Time is significant as well, indicating that there are also group differences in progress over time on Gsm. The difference in progress between latinists and non-latinists is significant, confirming previous results. The difference in progress

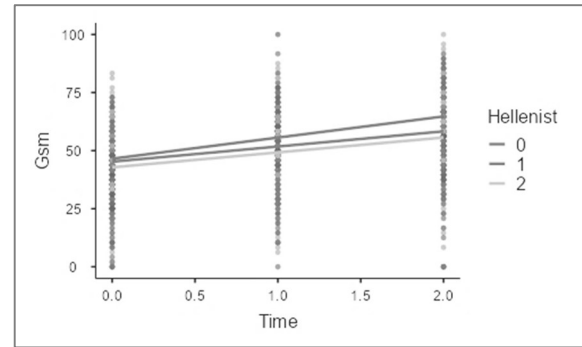
between hellenists and latinists is marginally non-significant at the 5%-level. However, it is estimated that latinists make more progress than hellenists, and indeed the amount of progress made by hellenists does not significantly differ from that made by non-latinists (not reported in the parameter estimate table).

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	176.36201	1	630	< .001
School	0.75557	3	415	0.520
Sex	0.00385	1	421	0.951
SES	4.36468	3	420	0.005
NonNative	4.54102	1	420	0.034
Dutch_T1	12.32209	1	419	< .001
G_T1	168.65860	1	422	< .001
Consistency	0.92528	1	1166	0.336
Perseverance	0.74394	1	1183	0.389
SelfControl	0.50063	1	1183	0.479
ControlledMot	0.78427	1	1145	0.376
AutoMot	2.73285	1	1127	0.099
Hellenist	1.43302	2	526	0.240
Time * Hellenist	3.79490	2	623	0.023

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	46.9148	1.825	43.337	50.492	462	25.7031	< .001
Time	Time	7.3521	0.554	6.267	8.437	630	13.2802	< .001
School1	3 - 1, 3, 4, 5	-1.4736	1.058	-3.547	0.600	413	-1.3928	0.164
School2	4 - 1, 3, 4, 5	0.2462	0.852	-1.424	1.917	418	0.2889	0.773
School3	5 - 1, 3, 4, 5	1.0110	0.960	-0.870	2.892	413	1.0533	0.293
Sex1	0 - 1	-0.0678	1.092	-2.208	2.072	421	-0.0621	0.951
SES1	1 - 0	-3.6922	2.773	-9.127	1.742	421	-1.3317	0.184
SES2	2 - 0	-7.2591	2.707	-12.565	-1.953	430	-2.6814	0.008
SES3	3 - 0	-4.3206	3.022	-10.243	1.602	424	-1.4298	0.154
NonNative1	0 - 1	-4.1316	1.939	-7.932	-0.332	420	-2.1310	0.034
Dutch_T1	Dutch_T1	2.2770	0.649	1.006	3.548	419	3.5103	< .001
G_T1	G_T1	7.7241	0.595	6.558	8.890	422	12.9869	< .001
Consistency	Consistency	-0.5049	0.525	-1.534	0.524	1166	-0.9619	0.336

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
Perseverance	Perseverance	0.4829	0.560	-0.614	1.580	1183	0.8625	0.389
SelfControl	SelfControl	-0.4476	0.633	-1.688	0.792	1183	-0.7075	0.479
ControlledMot	ControlledMot	0.3919	0.443	-0.476	1.259	1145	0.8856	0.376
AutoMot	AutoMot	0.8125	0.491	-0.151	1.776	1127	1.6531	0.099
Hellenist1	0 - 1	-1.2831	1.436	-4.098	1.532	526	-0.8933	0.372
Hellenist2	1 - 2	3.7106	2.250	-0.700	8.121	521	1.6489	0.100
Time * Hellenist1	Time * 0 - 1	-2.5643	0.965	-4.455	-0.673	625	-2.6579	0.008
Time * Hellenist2	Time * 1 - 2	2.6775	1.551	-0.362	5.717	618	1.7264	0.085

The main effect of Hellenist is no longer significant, suggesting that the control variables adequately control for the a priori group differences.

The interaction effect between Hellenist and Time remains significant. Taking into account the control variables, there is still a significant difference in progress between latinists and non-latinists, and not between hellenists and latinists. Notably, the difference in progress between hellenists and non-latinists remains non-significant as well (not reported in the parameter estimate table). As we saw previously, the average non-latinist improves their Gsm score 6.5 p.p. per school year. The average latinist now improves their Gsm score 9.1 p.p. per school year, and the average hellenist 6.4 p.p.

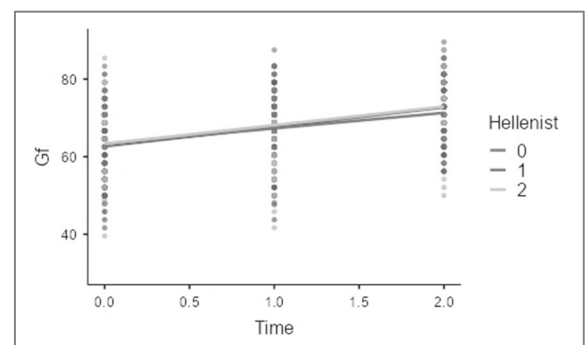
8.3.4.3 Non-verbal intelligence: fluid intelligence (Gf)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	214.298	1	731	<.001
School	0.292	3	429	0.831
Hellenist	0.277	2	434	0.758
Time * Hellenist	1.716	2	731	0.181

Note. Satterthwaite method for degrees of freedom



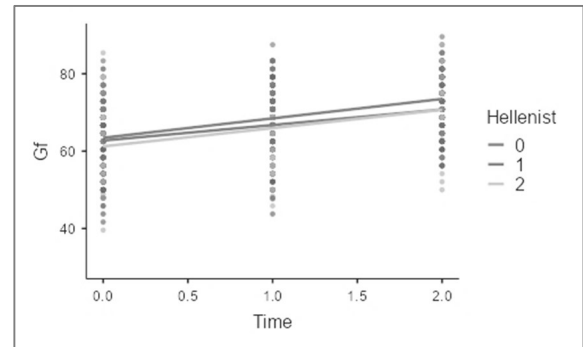
Both the main effect of Hellenist and the interaction effect with Time are non-significant, indicating that as far as Gf is concerned there are no group differences in absolute scores or progress over time, which confirms previous results.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	212.63619	1	823	< .001
School	0.11872	3	416	0.949
Sex	7.93537	1	426	0.005
SES	1.04469	3	422	0.373
NonNative	2.19822	1	420	0.139
G_T1	204.18482	1	421	< .001
Consistency	0.00418	1	1196	0.948
Perseverance	2.83e-4	1	1151	0.987
SelfControl	0.03739	1	1104	0.847
ControlledMot	0.00169	1	1112	0.967
AutoMot	0.17337	1	1076	0.677
Hellenist	1.68643	2	487	0.186
Time * Hellenist	1.67436	2	817	0.188

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	63.18720	0.940	61.3453	65.029	506	67.2371	< .001
Time	Time	4.58650	0.315	3.9700	5.203	823	14.5813	< .001
School1	3 - 1, 3, 4, 5	-0.17453	0.525	-1.2033	0.854	413	-0.3325	0.740
School2	4 - 1, 3, 4, 5	-0.06073	0.418	-0.8794	0.758	420	-0.1454	0.884
School3	5 - 1, 3, 4, 5	0.26558	0.460	-0.6361	1.167	414	0.5773	0.564
Sex1	0 - 1	-1.51706	0.539	-2.5726	-0.462	426	-2.8170	0.005
SES1	1 - 0	2.35538	1.380	-0.3498	5.061	424	1.7065	0.089
SES2	2 - 0	2.33728	1.350	-0.3091	4.984	433	1.7311	0.084
SES3	3 - 0	2.19165	1.507	-0.7621	5.145	425	1.4543	0.147
NonNative1	0 - 1	-1.43022	0.965	-3.3209	0.460	420	-1.4826	0.139
G_T1	G_T1	3.97755	0.278	3.4320	4.523	421	14.2893	< .001
Consistency	Consistency	-0.01868	0.289	-0.5853	0.548	1196	-0.0646	0.948
Perseverance	Perseverance	-0.00515	0.306	-0.6051	0.595	1151	-0.0168	0.987
SelfControl	SelfControl	-0.06575	0.340	-0.7322	0.601	1104	-0.1934	0.847
ControlledMot	ControlledMot	0.00996	0.242	-0.4649	0.485	1112	0.0411	0.967
AutoMot	AutoMot	-0.11176	0.268	-0.6379	0.414	1076	-0.4164	0.677
Hellenist1	0 - 1	0.71743	0.805	-0.8608	2.296	483	0.8909	0.373
Hellenist2	1 - 2	1.49859	1.299	-1.0477	4.045	481	1.1535	0.249
Time * Hellenist1	Time * 0 - 1	1.00517	0.550	-0.0723	2.083	819	1.8284	0.068

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
Time * Hellenist2	Time * 1 - 2	-0.70952	0.882	-2.4389	1.020	811	-0.8041	0.422

Both the main effect of Hellenist and the interaction effect with Time remain non-significant. As we saw previously, the average non-latinist improves their Gf score 5.0 p.p. per school year. The average latinist now improves their Gf score 4.0 p.p. per school year, and the average hellenist 4.7 p.p.

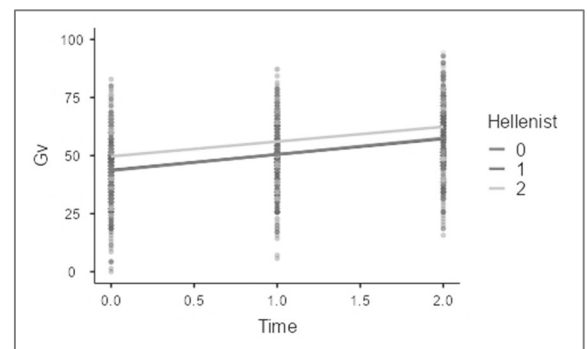
8.3.4.4 Non-verbal intelligence: visuo-spatial processing (Gv)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	349.133	1	416	< .001
School	0.118	3	427	0.950
Hellenist	3.456	2	425	0.032
Time * Hellenist	0.253	2	417	0.777

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison		Difference	SE	t	df	p	p _{bonferroni}	p _{holm}
Hellenist	Hellenist							
0	- 1	-0.0544	1.45	-0.0375	433	0.970	1.000	0.970
0	- 2	-5.5581	2.19	-2.5330	432	0.012	0.035	0.035
1	- 2	-5.5037	2.36	-2.3286	432	0.020	0.061	0.041

The main effect of Hellenist is significant, indicating that there are group differences in absolute scores on Gv. Pairwise comparisons show that hellenists excel, while the difference between latinists and non-latinists is non-significant.

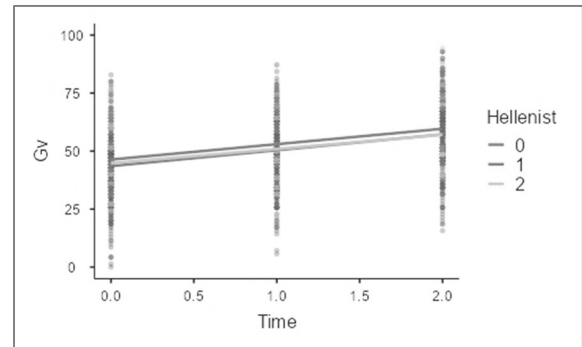
The interaction effect between Hellenist and Time is non-significant, indicating that there are no group differences in progress over time on Gv, which confirms previous results.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	321.60148	1	425	< .001
School	0.79309	3	415	0.498
Sex	17.03466	1	420	< .001
SES	0.51264	3	418	0.674
NonNative	0.07431	1	417	0.785
G_T1	216.47483	1	420	< .001
Consistency	0.00518	1	1101	0.943
Perseverance	1.21605	1	1144	0.270
SelfControl	0.62248	1	1194	0.430
ControlledMot	0.41100	1	1160	0.522
AutoMot	4.34962	1	1160	0.037
Hellenist	2.51026	2	426	0.082
Time * Hellenist	0.23042	2	420	0.794

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	45.1564	1.745	41.737	48.5756	436	25.8848	< .001
Time	Time	6.5323	0.364	5.818	7.2462	425	17.9327	< .001
School1	3 - 1, 3, 4, 5	-0.5325	1.024	-2.539	1.4737	414	-0.5202	0.603
School2	4 - 1, 3, 4, 5	-0.3833	0.811	-1.972	1.2059	417	-0.4727	0.637
School3	5 - 1, 3, 4, 5	-0.5488	0.897	-2.306	1.2086	414	-0.6120	0.541
Sex1	0 - 1	4.2893	1.039	2.252	6.3261	420	4.1273	< .001
SES1	1 - 0	-2.0146	2.675	-7.257	3.2278	418	-0.7532	0.452
SES2	2 - 0	-1.9141	2.603	-7.015	3.1872	425	-0.7354	0.463
SES3	3 - 0	-0.2550	2.918	-5.975	5.4652	420	-0.0874	0.930
NonNative1	0 - 1	-0.5109	1.874	-4.184	3.1624	417	-0.2726	0.785
G_T1	G_T1	7.9466	0.540	6.888	9.0052	420	14.7131	< .001
Consistency	Consistency	-0.0270	0.375	-0.763	0.7085	1101	-0.0720	0.943
Perseverance	Perseverance	-0.4487	0.407	-1.246	0.3488	1144	-1.1027	0.270
SelfControl	SelfControl	0.3669	0.465	-0.545	1.2783	1194	0.7890	0.430
ControlledMot	ControlledMot	-0.2093	0.326	-0.849	0.4306	1160	-0.6411	0.522
AutoMot	AutoMot	-0.7662	0.367	-1.486	-0.0461	1160	-2.0856	0.037
Hellenist1	0 - 1	2.7926	1.253	0.338	5.2477	424	2.2294	0.026
Hellenist2	1 - 2	-1.2293	2.019	-5.187	2.7287	422	-0.6087	0.543
Time * Hellenist1	Time * 0 - 1	-0.1968	0.635	-1.442	1.0488	422	-0.3097	0.757

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
Time * Hellenist2	Time * 1 - 2	0.6894	1.017	-1.305	2.6837	414	0.6776	0.498

The main effect of Hellenist is no longer significant, suggesting that the control variables adequately control for the a priori advantage of the hellenists.

The interaction effect between Hellenist and Time remains non-significant. As we saw previously, the average non-latinist improves their Gv score 6.6 p.p. per school year. The average latinist now improves their Gv score 6.8 p.p. per school year, and the average hellenist 6.1 p.p.

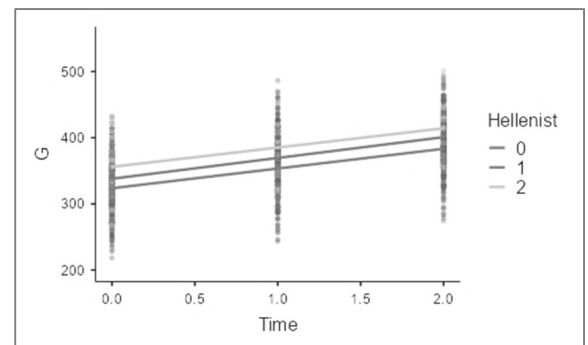
8.3.4.5 Overall general cognitive ability (G)

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	823.493	1	830	<.001
School	1.556	3	426	0.199
Hellenist	14.920	2	426	<.001
Time * Hellenist	0.483	2	830	0.617

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison		Hellenist	Hellenist	Difference	SE	t	df	p	pbonferroni	pholm
Hellenist										
0	-	1		-16.0	3.85	-4.16	432	<.001	<.001	<.001
0	-	2		-31.7	5.82	-5.44	432	<.001	<.001	<.001
1	-	2		-15.7	6.27	-2.50	432	0.013	0.038	0.013

The main effect of Hellenist is highly significant, indicating that there are group differences in absolute scores on G. Pairwise comparisons show that all group differences are significant. As already shown by the descriptive statistics, hellenists score higher than latinists, and latinists score higher than non-latinists.

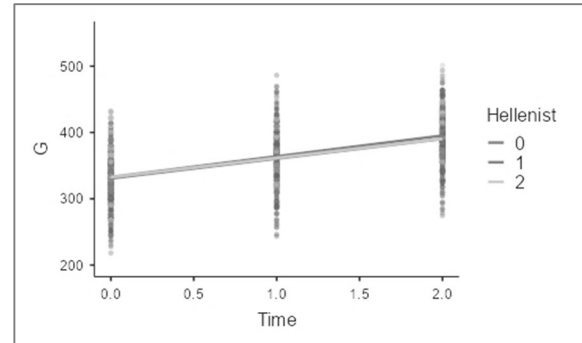
The interaction effect between Hellenist and Time is non-significant, indicating that there are no group differences in progress over time on G, which confirms previous results.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	658.3766	1	483	< .001
School	0.4851	3	592	0.693
Sex	1.4902	1	601	0.223
SES	0.2540	3	597	0.858
NonNative	0.0207	1	595	0.886
G_T1	2031.7274	1	598	< .001
Consistency	4.4757	1	1039	0.035
Perseverance	0.1682	1	1049	0.682
SelfControl	2.2675	1	1059	0.132
ControlledMot	0.4478	1	1017	0.504
AutoMot	2.3707	1	1019	0.124
Hellenist	0.6034	2	805	0.547
Time * Hellenist	0.3467	2	481	0.707

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	332.244	2.593	327.162	337.325	681	128.1424	< .001
Time	Time	30.012	1.170	27.719	32.304	483	25.6581	< .001
School1	3 - 1, 3, 4, 5	-0.964	1.496	-3.896	1.968	589	-0.6442	0.520
School2	4 - 1, 3, 4, 5	0.247	1.189	-2.083	2.576	595	0.2076	0.836
School3	5 - 1, 3, 4, 5	1.459	1.312	-1.113	4.030	590	1.1118	0.267
Sex1	0 - 1	1.858	1.522	-1.125	4.840	601	1.2207	0.223
SES1	1 - 0	0.616	3.924	-7.075	8.306	598	0.1569	0.875
SES2	2 - 0	0.283	3.843	-7.250	7.815	608	0.0735	0.941
SES3	3 - 0	2.352	4.293	-6.062	10.766	600	0.5479	0.584
NonNative1	0 - 1	-0.395	2.750	-5.786	4.995	595	-0.1438	0.886
G_T1	G_T1	35.767	0.794	34.212	37.322	598	45.0747	< .001
Consistency	Consistency	-1.743	0.824	-3.358	-0.128	1039	-2.1156	0.035
Perseverance	Perseverance	0.363	0.885	-1.372	2.098	1049	0.4101	0.682
SelfControl	SelfControl	1.501	0.997	-0.453	3.456	1059	1.5058	0.132
ControlledMot	ControlledMot	0.463	0.692	-0.893	1.818	1017	0.6692	0.504
AutoMot	AutoMot	-1.182	0.768	-2.686	0.323	1019	-1.5397	0.124
Hellenist1	0 - 1	-2.052	1.952	-5.879	1.775	803	-1.0511	0.294

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
Hellenist2	1 - 2	0.268	3.145	-5.896	6.431	799	0.0851	0.932
Time * Hellenist1	Time * 0 - 1	-1.454	2.046	-5.464	2.557	482	-0.7104	0.478
Time * Hellenist2	Time * 1 - 2	2.252	3.288	-4.193	8.696	477	0.6848	0.494

The main effect of Hellenist is no longer significant, suggesting that the control variables adequately control for the a priori group differences.

The interaction effect between Hellenist and Time remains non-significant. As we saw previously, the average non-latinist improves their G score with 30 points (on a scale of 600) per school year. The average latinist now improves their G score with 31 points per school year, and the average hellenist with 29 points per school year.

8.3.5 School achievement

The results regarding the significance of the main effect and interaction effect with Time of the grouping variable are identical to the previous results. Moreover, in the cases of French, mathematics and total achievement, no significant differences between latinists and hellenists are found in either the base model or the full model. Reported below are therefore only achievement for Dutch and for English.

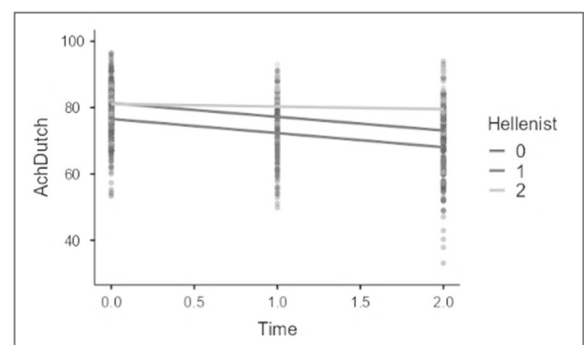
8.3.5.1 Achievement for Dutch

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Hellenist	21.54	2	424	<.001
School	2.33	3	415	0.073
Time	80.45	1	283	<.001
Hellenist * Time	8.43	2	293	<.001

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison									
Hellenist	Hellenist	Difference	SE	t	df	p	pbonferroni	pholm	
0	- 1	-4.88	0.805	-6.06	410	<.001	<.001	<.001	
0	- 2	-7.45	1.161	-6.42	374	<.001	<.001	<.001	
1	- 2	-2.57	1.264	-2.03	382	0.043	0.128	0.043	

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	79.674	0.455	78.78	80.566	419	174.984	<.001
Hellenist1	0 - 1	-4.748	0.789	-6.29	-3.202	429	-6.021	<.001
Hellenist2	1 - 2	0.221	1.279	-2.29	2.728	421	0.173	0.863
School1	3 - 1, 3, 4, 5	-0.267	0.633	-1.51	0.973	384	-0.423	0.673
School2	4 - 1, 3, 4, 5	-0.662	0.557	-1.75	0.429	445	-1.189	0.235
School3	5 - 1, 3, 4, 5	-0.578	0.561	-1.68	0.522	388	-1.030	0.304
Time	Time	-3.060	0.341	-3.73	-2.391	283	-8.969	<.001
Hellenist1 * Time	0 - 1 * Time	-0.155	0.630	-1.39	1.080	295	-0.246	0.806
Hellenist2 * Time	1 - 2 * Time	-3.324	0.947	-5.18	-1.468	291	-3.510	<.001

The main effect of Hellenist is highly significant, indicating that there are group differences in absolute grades for Dutch. Pairwise comparisons show that all group differences are significant. As already shown by the descriptive statistics, hellenists score higher than latinists, and latinists score higher than non-latinists.

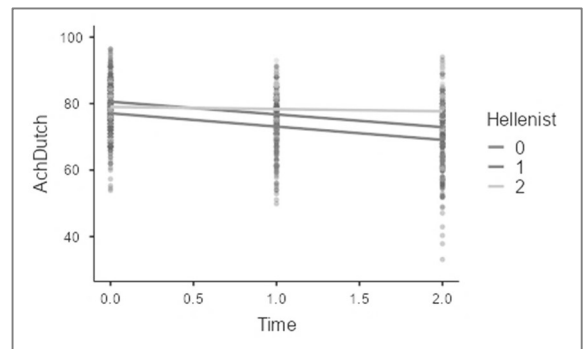
The interaction effect between Hellenist and Time is highly significant as well, indicating that there are also group differences in how the grades evolve over time. The difference in evolution between latinists and non-latinists is non-significant, contradicting previous results. The difference in evolution between latinists and hellenists is, however. It is worth noting – as nicely illustrated by the graph as well as by the descriptive statistics – that on average hellenists' scores remain almost constant across the three years.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Hellenist	10.82	2	419	<.001
School	2.71	3	408	0.045
Time	68.49	1	282	<.001
G_T1	32.45	1	423	<.001
Consistency	1.77	1	872	0.183
Perseverance	3.74	1	893	0.053
SelfControl	17.88	1	905	<.001
ControlledMot	1.63	1	894	0.203
AutoMot	1.40	1	892	0.236
Hellenist * Time	7.72	2	293	<.001

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison									
Hellenist	Hellenist	Difference	SE	t	df	p	pbonferroni	pholm	
0	- 1	-3.63	0.774	-4.688	396	<.001	<.001	<.001	
0	- 2	-4.75	1.142	-4.155	375	<.001	<.001	<.001	
1	- 2	-1.12	1.201	-0.930	373	0.353	1.000	0.353	

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	78.888	0.434	78.03799	79.738	415	181.878	<.001
Hellenist1	0 - 1	-3.485	0.753	-4.96075	-2.009	423	-4.629	<.001
Hellenist2	1 - 2	1.558	1.201	-0.79675	3.913	412	1.297	0.195
School1	3 - 1, 3, 4, 5	-0.572	0.590	-1.72790	0.584	373	-0.970	0.333
School2	4 - 1, 3, 4, 5	-0.424	0.525	-1.45371	0.606	433	-0.806	0.420
School3	5 - 1, 3, 4, 5	-0.562	0.523	-1.58695	0.463	379	-1.075	0.283
Time	Time	-2.828	0.342	-3.49719	-2.158	282	-8.276	<.001
G_T1	G_T1	1.834	0.322	1.20292	2.465	423	5.697	<.001
Consistency	Consistency	-0.376	0.283	-0.93048	0.178	872	-1.331	0.183
Perseverance	Perseverance	0.594	0.307	-0.00810	1.196	893	1.934	0.053
SelfControl	SelfControl	1.460	0.345	0.78300	2.136	905	4.228	<.001
ControlledMot	ControlledMot	-0.318	0.250	-0.80733	0.171	894	-1.275	0.203
AutoMot	AutoMot	0.326	0.275	-0.21322	0.866	892	1.185	0.236
Hellenist1 * Time	0 - 1 * Time	-0.171	0.633	-1.41213	1.070	296	-0.270	0.787

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
Hellenist2 * Time	1 - 2 * Time	-3.166	0.942	-5.01205	-1.320	289	-3.361	<.001

After adding the control variables, both the main effect of Hellenist and the interaction effect with Time remain significant. The absolute difference between latinists and hellenists vanishes, but their difference in evolution over time remains significant.

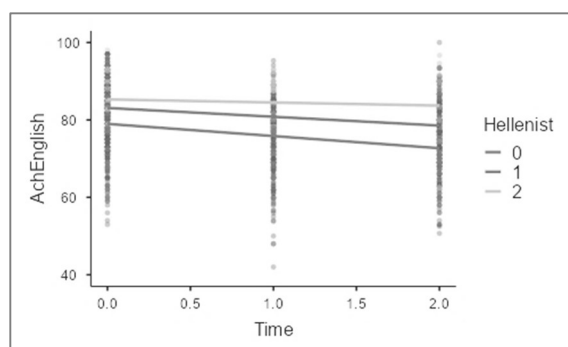
8.3.5.2 Achievement for English

Base model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Hellenist	16.02	2	411	<.001
School	4.25	3	420	0.006
Time	33.82	1	287	<.001
Hellenist * Time	3.82	2	304	0.023

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison										
Hellenist			Hellenist	Difference	SE	t	df	p	$p_{\text{bonferroni}}$	p_{holm}
0	-	1		-4.84	0.858	-5.64	420	< .001	< .001	< .001
0	-	2		-8.29	1.246	-6.65	384	< .001	< .001	< .001
1	-	2		-3.45	1.355	-2.54	392	0.011	0.034	0.011

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	82.4498	0.545	81.3815	83.518	408	151.270	<.001
Hellenist1	0 - 1	-4.1164	0.945	-5.9679	-2.265	416	-4.357	<.001
Hellenist2	1 - 2	-2.2036	1.533	-5.2076	0.800	409	-1.438	0.151
School1	3 - 1, 3, 4, 5	0.7406	0.698	-0.6284	2.110	361	1.060	0.290
School2	4 - 1, 3, 4, 5	1.3795	0.667	0.0717	2.687	430	2.067	0.039

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
School3	5 - 1, 3, 4, 5	0.0873	0.620	-1.1287	1.303	369	0.141	0.888
Time	Time	-2.0691	0.356	-2.7666	-1.372	287	-5.815	<.001
Hellenist1 * Time	0 - 1 * Time	-0.8645	0.642	-2.1227	0.394	308	-1.347	0.179
Hellenist2 * Time	1 - 2 * Time	-1.4817	0.973	-3.3896	0.426	302	-1.522	0.129

The main effect of Hellenist is highly significant, indicating that there are group differences in absolute grades for English. Pairwise comparisons show that all group differences are significant. As already shown by the descriptive statistics, hellenists score higher than latinists, and latinists score higher than non-latinists.

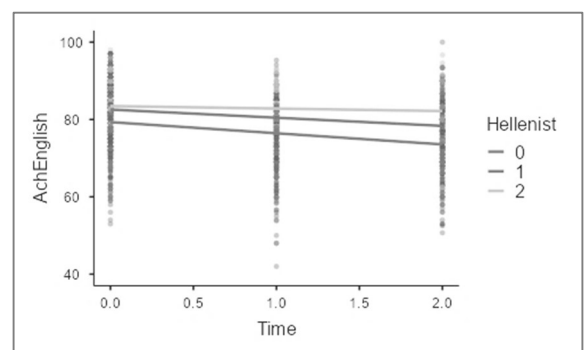
The interaction effect between Hellenist and Time is significant as well, indicating that there are also group differences in how the grades evolve over time. The difference in evolution between latinists and non-latinists is non-significant, contradicting previous results. Only the difference in evolution between non-latinists and hellenists is significant. It is worth noting – as nicely illustrated by the graph as well as by the descriptive statistics – that on average hellenists' scores remain almost constant across the three years.

Full model

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Hellenist	8.05	2	407	<.001
School	4.32	3	403	0.005
Time	26.41	1	281	<.001
G_T1	21.76	1	403	<.001
Consistency	2.27	1	902	0.133
Perseverance	1.43	1	913	0.233
SelfControl	3.28	1	927	0.071
ControlledMot	1.46	1	922	0.226
AutoMot	2.13	1	936	0.144
Hellenist * Time	3.22	2	300	0.041

Note. Satterthwaite method for degrees of freedom



Post Hoc Comparisons - Hellenist

Comparison									
Hellenist	Hellenist	Difference	SE	t	df	p	pbonferroni	pholm	
0	- 1	-3.90	0.854	-4.56	408	< .001	< .001	< .001	
0	- 2	-6.03	1.276	-4.73	385	< .001	< .001	< .001	
1	- 2	-2.14	1.334	-1.60	386	0.110	0.330	0.110	

Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	81.7807	0.542	80.7191	82.842	404	150.9923	< .001
Hellenist1	0 - 1	-3.2408	0.941	-5.0846	-1.397	410	-3.4451	< .001
Hellenist2	1 - 2	-0.8979	1.504	-3.8459	2.050	398	-0.5969	0.551
School1	3 - 1, 3, 4, 5	0.5451	0.681	-0.7888	1.879	345	0.8010	0.424
School2	4 - 1, 3, 4, 5	1.5331	0.658	0.2438	2.822	413	2.3306	0.020
School3	5 - 1, 3, 4, 5	0.0365	0.605	-1.1488	1.222	355	0.0604	0.952
Time	Time	-1.8792	0.366	-2.5960	-1.162	281	-5.1386	< .001
G_T1	G_T1	1.7612	0.378	1.0212	2.501	403	4.6643	< .001
Consistency	Consistency	-0.4961	0.329	-1.1419	0.150	902	-1.5055	0.133
Perseverance	Perseverance	0.4249	0.356	-0.2722	1.122	913	1.1947	0.233
SelfControl	SelfControl	0.7229	0.399	-0.0599	1.506	927	1.8099	0.071
ControlledMot	ControlledMot	-0.3510	0.290	-0.9194	0.217	922	-1.2102	0.226
AutoMot	AutoMot	0.4682	0.321	-0.1601	1.096	936	1.4605	0.144
Hellenist1 * Time	0 - 1 * Time	-0.7766	0.662	-2.0737	0.520	305	-1.1735	0.242
Hellenist2 * Time	1 - 2 * Time	-1.4677	0.993	-3.4130	0.478	294	-1.4787	0.140

After adding the control variables, both the main effect of Hellenist and the interaction effect with Time remain significant. The absolute difference between latinists and hellenists vanishes. The difference in evolution between non-latinists and hellenists over time remains significant.

8.4 Results: Latin, with matching

8.4.1 Matching procedure, models and interpretation

In this section, the same analyses are repeated on a matched subset of the filtered longitudinal sample. Matching was carried out using the *cem* package in R (Iacus et al., 2009). Having experimented with various selections of matching variables, I resolved to match latinists and non-latinists on Dutch_T1, Gc_T1, and Gsm_T1. The resulting matching solution also eliminated the mean difference on overall G_T1 and to some extent on the non-cognitive variables, while preserving a reasonable sample size. 80 matches were found which displayed similar distributions of the three matching variables at group level. Multivariate imbalance was thus lowered from $L1 = .802$ to $L1 = .613$. Since exact balance could not be achieved, statistical modelling remains a necessity. Below, the characteristics of the chosen matching solution are discussed in more detail.

As for the linear mixed models, for efficiency only the base models are reported here. Adding in the control variables never altered the outcome of the significance test, as it had a negligible impact on the p-value of the interaction effect between Time and Latinist. For the test variables, the absolute value of the change in the parameter estimate of the interaction effect between Time and Latinist when going from the base to the full model ranged between 0.002 (for Gf) and 0.33 p.p. (for Verbs). Time, Dutch_T1 and G_T1 were the only variables that consistently made a significant contribution to the full models, with the one exception of the effect of G_T1 being marginally non-significant in the model of English achievement. Also worth noting is that the previously observed effects of School on Verbs and on French and mathematics achievement persisted, as did the effects of Sex on Gc and Gv.

Interpretation will be limited to the interaction effect between Time and Latinist. The main effect of Time is always significant, as before. The main effect of Latinist is always non-

significant, due to the matching procedure.¹¹² The effect of School is only significant in the case of Verbs, as just noted.

8.4.2 Matching solution

8.4.2.1 Balance in terms of sex

Before matching

Contingency Tables

		Sex		Total
		1	0	
0	Observed	131	126	257
	Expected	130	127	257
1	Observed	101	102	203
	Expected	102	101	203

χ^2 Tests

	Value	df	p
χ^2	0.0674	1	0.795
N	460		

After matching

Contingency Tables

		Sex		Total
		1	0	
0	Observed	46	34	80
	Expected	43	37	80
1	Observed	40	40	80
	Expected	43	37	80

χ^2 Tests

	Value	df	p
χ^2	0.905	1	0.341
N	160		

There is good gender balance in both the unmatched and the matched samples.

¹¹² Jamovi does not allow to include an interaction effect without including the main effect as well.

8.4.2.2 Balance in terms of SES

Before matching

Contingency Tables

		SES				Total
		0	1	2	3	
0	Observed	13	71	151	22	257
	Expected	13.4	57	154	33	257
1	Observed	11	31	124	37	203
	Expected	10.6	45	121	26	203

χ^2 Tests

	Value	df	p
χ^2	16.2	3	0.001
N	460		

After matching

Contingency Tables

		SES				Total
		0	1	2	3	
0	Observed	2	22	50	6	80
	Expected	3.5	16	49.5	11	80.0
1	Observed	5	10	49	16	80
	Expected	3.5	16	49.5	11	80.0

χ^2 Tests

	Value	df	p
χ^2	10.3	3	0.016
N	160		

Before matching, the variables Latinist and SES are not independent. When we take a closer look at the distribution, we see that there are relatively more non-latinists whose parents have a bachelor degree, and relatively more latinists whose parents have a doctoral degree. However, in the previous longitudinal analyses the SES variable consistently failed to explain differences in performance on the tests, and for this reason it was not included in the matching solution.

After matching, the same imbalance persists, i.e. more bachelor degrees among the non-latinists' parents and more doctoral degrees among the latinists' parents.

8.4.2.3 Balance in terms of native language

Before matching

Contingency Tables

		NonNative		Total
		0	1	
0	Observed	233	24	257
	Expected	234	23.5	257
1	Observed	185	18	203
	Expected	184	18.5	203

χ^2 Tests

	Value	df	p
χ^2	0.0304	1	0.862
N	460		

After matching

Contingency Tables

		NonNative		Total
		0	1	
0	Observed	75	5	80
	Expected	73	7	80
1	Observed	71	9	80
	Expected	73	7	80

χ^2 Tests

	Value	df	p
χ^2	1.25	1	0.263
N	160		

The proportion of non-native speakers of Dutch in each group is comparable, and remains more or less constant after matching.

8.4.2.4 Balance in terms of non-cognitive variables

Before matching

One-Way ANOVA (Welch's)

	F	df1	df2	p
Consistency_T1	15.759	1	400	<.001
Perseverance_T1	17.227	1	410	<.001
SelfControl_T1	12.716	1	429	<.001
ControlledMot_T1	0.331	1	431	0.565
AutoMot_T1	18.246	1	416	<.001

After matching

One-Way ANOVA (Welch's)

	F	df1	df2	p
Consistency_T1	2.7829	1	150	0.097
Perseverance_T1	3.7245	1	151	0.055
SelfControl_T1	2.3578	1	158	0.127
ControlledMot_T1	0.0151	1	158	0.902
AutoMot_T1	6.0648	1	156	0.015

Before matching, the latinists and non-latinists differ very significantly on all non-cognitive variables except controlled motivation (the differences being to the advantage of the latinists; see Descriptives above).

After matching, the differences in Consistency of interest and Self control become non-significant. The difference in Perseverance of effort flirts with the significance threshold. The difference in Autonomous motivation remains significant, but is more than ten times less pronounced.

8.4.2.5 Balance in terms of Dutch language ability

Before matching

One-Way ANOVA (Welch's)

	F	df1	df2	p
Dutch_T1	92.6	1	422	<.001
Lexicon_T1	45.4	1	399	<.001
Verbs_T1	44.8	1	432	<.001

After matching

One-Way ANOVA (Welch's)

	F	df1	df2	p
Dutch_T1	0.0691	1	158	0.793
Lexicon_T1	0.0173	1	158	0.896
Verbs_T1	0.0962	1	158	0.757

The highly significant differences between latinists and non-latinists on all Dutch language variables that exist prior to matching (the differences being to the advantage of the latinists; see Descriptives above) are completely erased by the matching procedure.

8.4.2.6 Balance in terms of general cognitive ability

Before matching

One-Way ANOVA (Welch's)

	F	df1	df2	p
G_T1	25.771	1	422	< .001
Gc_T1	48.945	1	430	< .001
Gsm_T1	14.776	1	441	< .001
Gf_T1	1.457	1	436	0.228
Gv_T1	0.212	1	434	0.646

After matching

One-Way ANOVA (Welch's)

	F	df1	df2	p
G_T1	0.00258	1	157	0.960
Gc_T1	0.01927	1	158	0.890
Gsm_T1	0.08038	1	158	0.777
Gf_T1	0.27185	1	157	0.603
Gv_T1	1.59489	1	157	0.209

The highly significant differences between latinists and non-latinists on verbal intelligence and overall general cognitive ability that exist prior to matching (the differences being to the advantage of the latinists; see Descriptives above) are completely erased by the matching procedure.

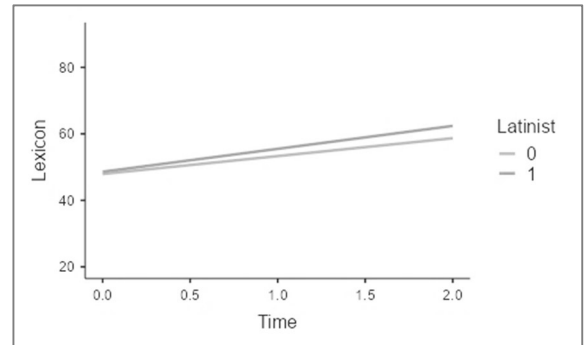
8.4.3 Dutch language ability

8.4.3.1 Lexicon

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	204.891	1	160	< .001
School	1.864	3	160	0.138
Latinist	0.116	1	161	0.734
Time * Latinist	3.145	1	160	0.078

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	47.936	1.379	45.2338	50.64	161	34.769	< .001
Time	Time	5.377	0.606	4.1887	6.57	160	8.867	< .001
School1	3 - 1, 3, 4, 5	0.610	1.866	-3.0467	4.27	160	0.327	0.744
School2	4 - 1, 3, 4, 5	-1.361	1.420	-4.1449	1.42	160	-0.958	0.339
School3	5 - 1, 3, 4, 5	3.079	1.549	0.0428	6.12	160	1.988	0.049
Latinist1	1 - 0	0.660	1.941	-3.1433	4.46	161	0.340	0.734
Time * Latinist1	Time * 1 - 0	1.521	0.858	-0.1599	3.20	160	1.774	0.078

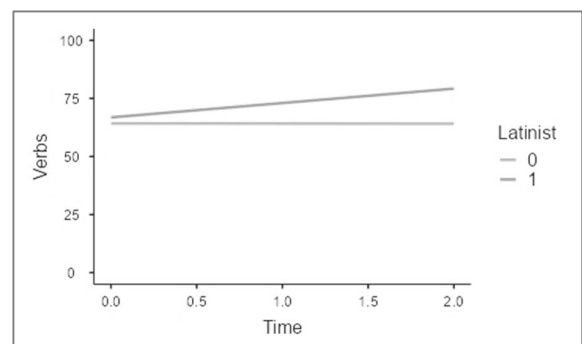
In the matched sample, the interaction effect between Time and Latinist fails to reach significance at the 5%-level. Nonetheless, latinists make 1.5 p.p. more progress on Lexicon per school year than non-latinists.

8.4.3.2 Verbs

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	7.613	1	160	0.006
School	6.557	3	160	< .001
Latinist	0.488	1	159	0.486
Time * Latinist	7.927	1	160	0.005

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	64.2232	2.64	59.04	69.40	160	24.2962	< .001
Time	Time	-0.0625	1.57	-3.14	3.01	160	-0.0398	0.968
School1	3 - 1, 3, 4, 5	1.8748	3.07	-4.15	7.90	160	0.6102	0.543
School2	4 - 1, 3, 4, 5	5.6968	2.34	1.11	10.28	160	2.4347	0.016
School3	5 - 1, 3, 4, 5	2.9909	2.55	-2.01	7.99	160	1.1725	0.243
Latinist1	1 - 0	2.6036	3.73	-4.70	9.91	159	0.6988	0.486
Time * Latinist1	Time * 1 - 0	6.2500	2.22	1.90	10.60	160	2.8155	0.005

The interaction effect between Time and Latinist is significant. Latinists make 6.3 p.p. more progress on Verbs per school year than non-latinists, whose performance remains about constant or even slightly declines across time.

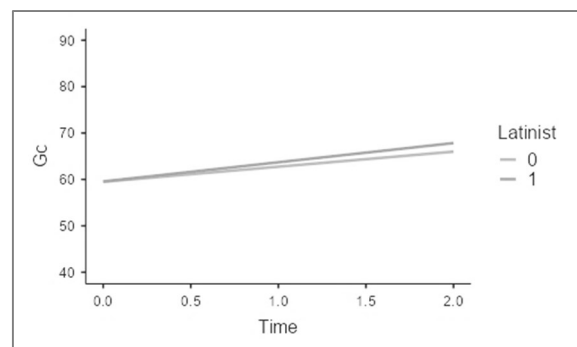
8.4.4 General cognitive ability

8.4.4.1 Verbal intelligence: crystallised intelligence (Gc)

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	180.76148	1	305	< .001
School	0.14306	3	160	0.934
Latinist	0.00282	1	167	0.958
Time * Latinist	2.66062	1	305	0.104

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	59.5004	0.787	57.957	61.04	167	75.5627	< .001
Time	Time	3.2433	0.388	2.482	4.00	305	8.3535	< .001
School1	3 - 1, 3, 4, 5	-0.0704	1.067	-2.161	2.02	160	-0.0660	0.947
School2	4 - 1, 3, 4, 5	0.5055	0.812	-1.086	2.10	160	0.6224	0.535
School3	5 - 1, 3, 4, 5	-0.1713	0.886	-1.907	1.56	160	-0.1935	0.847
Latinist1	1 - 0	0.0589	1.108	-2.114	2.23	167	0.0531	0.958
Time * Latinist1	Time * 1 - 0	0.8956	0.549	-0.181	1.97	305	1.6311	0.104

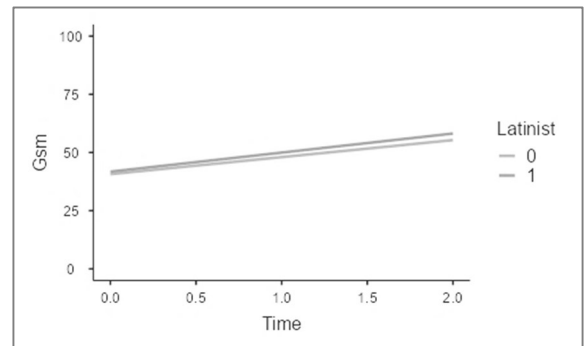
The interaction effect between Time and Latinist is non-significant.

8.4.4.2 Verbal intelligence: short-term memory (Gsm)

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	151.155	1	301	< .001
School	0.672	3	161	0.570
Latinist	0.222	1	176	0.638
Time * Latinist	0.487	1	301	0.486

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	40.698	1.596	37.57	43.83	176	25.498	< .001
Time	Time	7.316	0.892	5.57	9.06	301	8.200	< .001
School1	3 - 1, 3, 4, 5	0.625	2.103	-3.50	4.75	161	0.297	0.767
School2	4 - 1, 3, 4, 5	-0.281	1.602	-3.42	2.86	161	-0.175	0.861
School3	5 - 1, 3, 4, 5	1.785	1.746	-1.64	5.21	161	1.022	0.308
Latinist1	1 - 0	1.058	2.247	-3.35	5.46	176	0.471	0.638
Time * Latinist1	Time * 1 - 0	0.881	1.262	-1.59	3.35	301	0.698	0.486

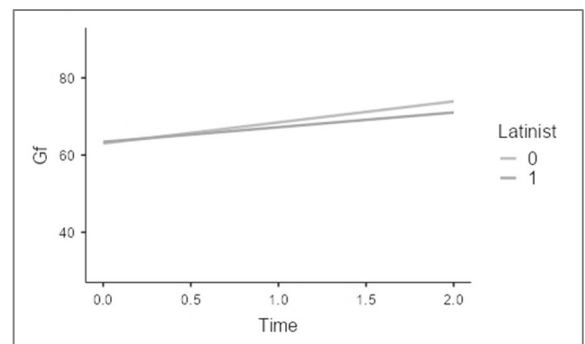
In the matched sample, the interaction effect between Time and Latinist is non-significant.

8.4.4.3 Non-verbal intelligence: fluid intelligence (Gf)

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	139.7472	1	160	< .001
School	0.0438	3	160	0.988
Latinist	0.0754	1	161	0.784
Time * Latinist	4.3490	1	160	0.039

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	63.01508	1.018	61.02	65.0106	162	61.89101	<.001
Time	Time	5.44947	0.554	4.36	6.5356	160	9.83366	<.001
School1	3 - 1, 3, 4, 5	0.17484	1.145	-2.07	2.4195	160	0.15267	0.879
School2	4 - 1, 3, 4, 5	0.00280	0.872	-1.71	1.7123	160	0.00321	0.997
School3	5 - 1, 3, 4, 5	0.15876	0.951	-1.70	2.0224	160	0.16696	0.868
Latinist1	1 - 0	0.39425	1.435	-2.42	3.2075	161	0.27466	0.784
Time * Latinist1	Time * 1 - 0	-1.63436	0.784	-3.17	-0.0983	160	-2.08543	0.039

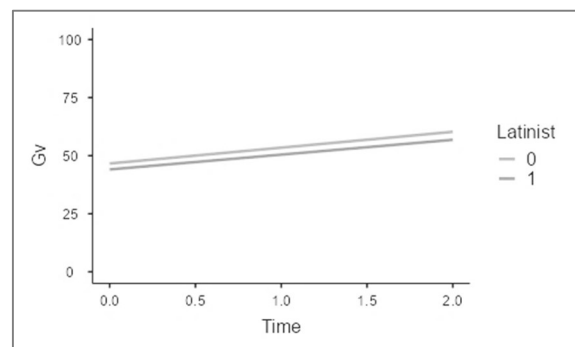
In the matched sample, the interaction effect between Time and Latinist is significant, to the advantage of the non-latinists.

8.4.4.4 Non-verbal intelligence: visuo-spatial processing (Gv)

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	199.237	1	160	<.001
School	0.705	3	160	0.550
Latinist	1.388	1	159	0.240
Time * Latinist	0.224	1	160	0.637

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	46.610	1.540	43.59	49.63	159	30.260	<.001
Time	Time	6.827	0.662	5.53	8.12	160	10.315	<.001
School1	3 - 1, 3, 4, 5	1.101	2.134	-3.08	5.28	160	0.516	0.607
School2	4 - 1, 3, 4, 5	-0.450	1.625	-3.63	2.73	160	-0.277	0.782
School3	5 - 1, 3, 4, 5	-2.256	1.772	-5.73	1.22	160	-1.274	0.205
Latinist1	1 - 0	-2.554	2.168	-6.80	1.69	159	-1.178	0.240
Time * Latinist1	Time * 1 - 0	-0.443	0.936	-2.28	1.39	160	-0.473	0.637

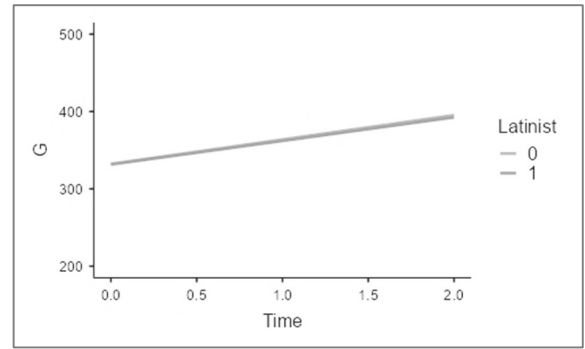
The interaction effect between Time and Latinist is non-significant.

8.4.4.5 Overall general cognitive ability (G)

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	579.9062	1	318	<.001
School	0.0104	3	160	0.999
Latinist	0.0129	1	163	0.910
Time * Latinist	0.1630	1	318	0.687

Note. Satterthwaite method for degrees of freedom



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	(Intercept)	332.238	4.38	323.66	340.81	163	75.9299	<.001
Time	Time	31.528	1.82	27.96	35.10	318	17.3135	<.001
School1	3 - 1, 3, 4, 5	0.400	5.72	-10.82	11.62	160	0.0698	0.944
School2	4 - 1, 3, 4, 5	0.161	4.36	-8.38	8.70	160	0.0369	0.971
School3	5 - 1, 3, 4, 5	0.282	4.75	-9.03	9.60	160	0.0593	0.953
Latinist1	1 - 0	-0.700	6.16	-12.77	11.38	163	-0.1136	0.910
Time * Latinist1	Time * 1 - 0	-1.040	2.58	-6.09	4.01	318	-0.4037	0.687

The interaction effect between Time and Latinist is non-significant.

8.4.5 School achievement

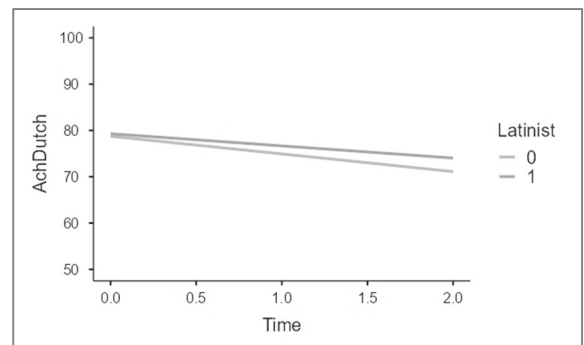
In the matched sample, the interaction effects between Time and Latinist are non-significant for all school achievement variables.

8.4.5.1 Achievement for Dutch

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	49.550	1	107	<.001
School	1.674	3	156	0.175
Latinist	0.237	1	159	0.627
Time * Latinist	1.789	1	115	0.184

Note. Satterthwaite method for degrees of freedom

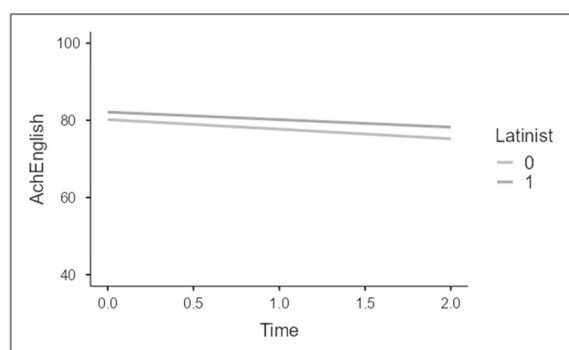


8.4.5.2 Achievement for English

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	20.719	1	118	< .001
School	2.005	3	168	0.115
Latinist	1.969	1	155	0.163
Time * Latinist	0.353	1	130	0.553

Note. Satterthwaite method for degrees of freedom

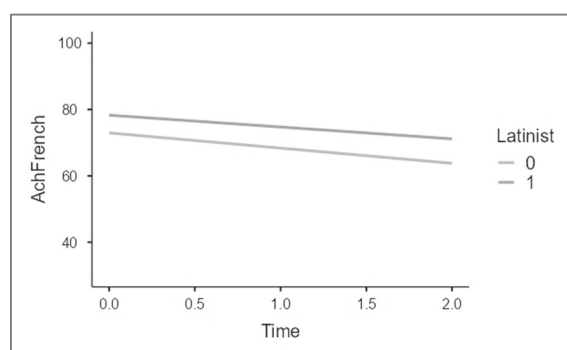


8.4.5.3 Achievement for French

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	65.55	1	158	< .001
School	5.67	3	160	0.001
Latinist	14.32	1	171	< .001
Time * Latinist	1.03	1	151	0.311

Note. Satterthwaite method for degrees of freedom

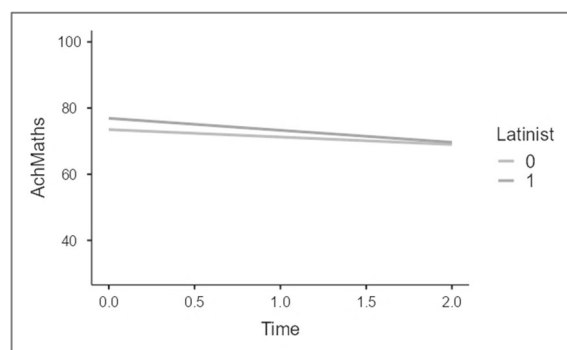


8.4.5.4 Achievement for mathematics

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	22.40	1	107	< .001
School	2.84	3	156	0.040
Latinist	4.09	1	161	0.045
Time * Latinist	1.27	1	112	0.262

Note. Satterthwaite method for degrees of freedom

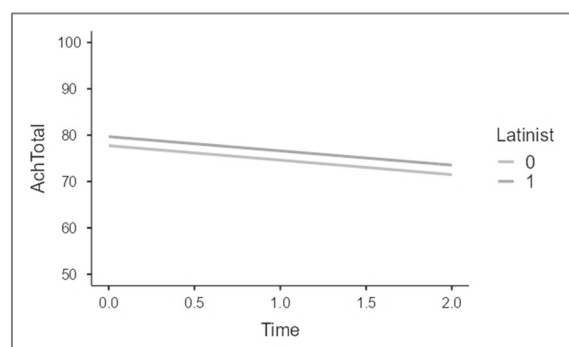


8.4.5.5 Total achievement

Fixed Effect Omnibus tests

	F	Num df	Den df	p
Time	109.04985	1	106	< .001
School	1.95578	2	106	0.147
Latinist	3.03585	1	106	0.084
Time * Latinist	0.00822	1	106	0.928

Note. Satterthwaite method for degrees of freedom



Works cited in this chapter

- Boone, S. (2011). *Sociale ongelijkheid bij de overgang van basis- naar secundair onderwijs: Een onderzoek naar de oriënteringspraktijk*. Vlaams Ministerie van Onderwijs en Vorming.
- Deci, E., & Ryan, R. (2000). The ‘what’ and ‘why’ of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268.
- Duckworth, A., Peterson, C., Matthews, M., & Kelly, D. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087–1101.
- Duckworth, A., & Quinn, P. (2009). Development and validation of the Short Grit Scale (grit-s). *Journal of Personality Assessment*, 91(2), 166–174.
- Fonteyne, L., Duyck, W., & De Fruyt, F. (2017). Program-specific prediction of academic achievement on the basis of cognitive and non-cognitive factors. *Learning and Individual Differences*, 56, 34–48.
- Ho, D., Imai, K., King, G., & Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3), 199–236.
- Iacus, S., King, G., & Porro, G. (2009). cem: Software for coarsened exact matching. *Journal of Statistical Software*, 30(9), 1–27.
- Iacus, S., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1), 1–24.
- King, G., Lucas, C., & Nielsen, R. (2017). The balance-sample size frontier in matching methods for causal inference. *American Journal of Political Science*, 61(2), 473–489.
- King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27(4), 435–454.
- Magez, W., Tierens, M., Van Huynegem, J., Van Parijs, K., Decaluwé, V., & Bos, A. (2015). *CoVaT-CHC Basisversie*. Psychodiagnostisch Centrum Thomas More.
- Magez, W., Van Parijs, K., Joris, S., & Tierens, M. (Forthcoming). *COVAT-3*. Psychodiagnostisch Centrum Thomas More.
- Overgaauw, S., Rieffe, C., Broekhof, E., Crone, E., & Güroğlu, B. (2017). Assessing empathy across childhood and adolescence: Validation of the Empathy Questionnaire for Children and Adolescents (EmQue-CA). *Frontiers in Psychology*, 8, no. 870.
- Ponnock, A., Muenks, K., Morell, M., Seung Yang, J., Gladstone, J., & Wigfield, A. (2020). Grit and conscientiousness: Another jangle fallacy. *Journal of Research in Personality*, 89, no. 104021.
- Roberts, B., Chernyshenko, O., Stark, S., & Goldberg, L. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, 58(1), 103–139.
- Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). The Guilford Press.
- Sjöström, O., & Holst, D. (2002). Validity of a questionnaire survey: Response patterns in different subgroups and the effect of social desirability. *Acta Odontologica Scandinavica*, 60(3), 136–140.
- Tangney, J., Baumeister, R., & Luzio Boone, A. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2), 271–324.
- Ulitzsch, E., Shin, H. J., & Lüdtke, O. (2024). Accounting for careless and insufficient effort responding in large-scale survey data—Development, evaluation, and application of a screen-time-based weighting procedure. *Behavior Research Methods*, 56(2), 804–825.
- Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology*, 101(3), 671–688.
- Vereeck, A. (2024). Een nieuwe blik op de Vlaamse leerlingenaantallen klassieke talen. *Prora*, 29(1), 7–12.

Chapter 9: Discussion of Results and Some Additional Analyses

Together with Chapter 8, this chapter will be reworked into an article reflecting the entire doctoral project.

Chapter 9

Discussion of Results and Some Additional Analyses

9.1 Introduction

This final chapter provides a discussion of what the results laid out in the previous chapter reveal about studying classical languages in Flemish secondary education, complemented with a couple of additional analyses (see Sections 9.2.3, 9.2.4 and 9.3.3).

In Section 9.2, I will go over the empirical findings. To start, group differences in absolute performance (i.e., a significant main effect of the grouping variable Latinist / Hellenist) will be passed in review, with the goal of sketching the profiles of pupils who do and do not study classical languages. Extra attention will be paid to the first measuring moment, which took place in the second year of secondary school, in order to answer the question of whether there is preselectivity. Then, transfer effects will be looked at, operationalised above as a difference in progress on test variables or a difference in decline on school achievement variables (i.e., a significant interaction effect of the grouping variable Latinist / Hellenist with the variable Time), which cannot be explained by the control variables. Although cognitive transfer and preselectivity are often presented as two opposite explanations of why classical language pupils perform well, they are not two different sides of the same coin. Preselectivity does not immediately exclude cognitive transfer; rather, the two can co-occur. Subsequently, some other observations from the longitudinal analyses will be shared. To conclude, the empathy and reading motivation variables will be analysed, which were not included until the last measurement and were therefore absent from the longitudinal analyses in Chapter 8.

In Section 9.3, I will evaluate to what extent the empirical findings ultimately permit to speak about cognitive transfer effects of studying classical languages. There are five subsections. The

first two subsections discuss which outcome variables qualify as potential candidates for cognitive transfer and how strong the supporting evidence is. The following two subsections add some nuances regarding individual differences and the importance of context. In the American research to which the article in Chapter 2 is dedicated, contextuality was generally not stressed enough, so I wish to make it abundantly clear that all results and findings discussed here only apply to general secondary education in Flanders, and in particular to the (type of) schools and pupils in this sample. The fifth subsection synthesises some key considerations. Finally, Section 9.4 contains the limitations of this research and its methodology.

9.2 Overview of empirical findings

9.2.1 Group differences in absolute performance across time: pupil profiles

A first finding of the longitudinal study is that, as a rule, the latinist group does better than the non-latinist group, confirming what we assumed to know from the large quantities of previous scientific and anecdotic evidence (cf. Chapters 1 and 2). Looking at the main effect of Latinist in the base models of the unmatched sample, we see that latinists score higher on all language-related test variables (i.e., Lexicon, Verbs, Gc and Gsm), overall intelligence (i.e., G), and all school achievement variables (i.e., achievement for Dutch, English, French and mathematics, as well as total achievement). In terms of non-verbal intelligence (i.e., Gf and Gv), however, there are no noticeable differences between the two groups. This first finding has two further implications.

One implication is that there is a preselection mechanism operating at the beginning of general secondary education in Flanders, the classical language study options having an influx of pupils with higher cognitive abilities, mainly verbal and linguistic. In the second year of secondary school, the pupils in this sample who opt for classical languages already score significantly higher than their peers on verbal intelligence (and hence overall intelligence) as well as Dutch language ability; these differences are statistically small to medium. In addition to the cognitive aspect, the classical language pupils arrive at the start with a little more grit, self-control and autonomous motivation; these differences are statistically small. Those results

are in line with the results of a recent study on a different sample of Flemish schools, which used a cross-sectional design (Hauspie et al., 2024). From the current study and the one by Hauspie and colleagues, preselectivity can now be considered proven with regard to classical language education in Flanders.

Within the education field, there have long been suspicions to that effect, and in the general public as well there is the idea that Latin is something only for the cleverest children. Nevertheless, the constataion of preselectivity is remarkable given both the wide availability of classical languages and their diminishing popularity. Secondary schools offering classical languages are numerous across the whole Flemish region (unlike the situation of classics poverty in the UK for example, cf. Section 1.4.2) and anyone with a primary school diploma can enrol for Latin (unlike the gymnasia in the Netherlands for example). In se, there does not seem to be ground to presuppose that access or a lack thereof contributes to preselectivity in Flanders, although it has been shown that social stereotypes play into perceived accessibility (Boone, 2011). From an utilitarianist perspective, however, classical languages are increasingly looked down upon (cf. Section 1.4.3), which is reflected by steadily declining enrolment numbers (Vereeck, 2024). Alternative study options are available that also enjoy a good reputation and appeal to top pupils, in the last decade notoriously STEM (cf. Section 1.4.3). Despite these facts – or perhaps precisely because of the waning prestige of classical languages – study options with Latin attract a cognitively stronger audience, especially those with a talent for language, the current study suggests.

Another implication that must not be overlooked is that latinists manage to retain their lead in the time span from the second to the fourth year of secondary school. The average cognitive development of the latinists proceeds with at least the same pace as that of the non-latinists (if not faster, cf. *infra*). To put it differently, the fact that these *a priori* smarter pupils invest a considerable amount of their school time and learning efforts in Latin and/or Ancient Greek certainly does not cause them to later fall behind in linguistic and cognitive abilities in any way. At the very minimum, one could therefore say that there seem to be no negative effects associated with studying classical languages for the pupils who have chosen to study them. Regardless of cognitive transfer, there is nothing to lose and potentially much to gain from the study of classical languages.

When we have a look at the main effect of Latinist in the full models of the unmatched sample, some other interesting features of the Latinist group come to light. For the outcome variables Lexicon, Verbs, Gsm and G, controlling for Dutch language capacity and cognitive

capacity (in combination with the other control variables) renders the main effect of Latinist non-significant. This is what logic predicts: Given that Dutch language capacity is nothing else than a composite score of baseline performance on Lexicon, Verbs and Orthography, and cognitive capacity is nothing else than a composite score of baseline performance on Gf, Gc, Gv and Gsm, one would expect that adding these variables to the model offsets the existing group differences. For Gc, by contrast, the main effect of Latinist remains significant in the full model. This means that the latinists in this sample particularly excel in crystallised intelligence, more so than would be expected based on their Dutch language capacity, cognitive capacity, and other characteristics (i.e., grit, self-control, motivation, sex, SES, and native language).¹¹³

For school achievement, the same pattern is observed: The main effect of Latinist is not only highly significant in the base model but in the full model as well. Analogously to Gc, we can interpret this as latinists achieving higher grades than would be expected of them based on their cognitive, non-cognitive and demographic data alone. There is unfortunately no way of knowing the driving force behind this disparity between latinists and non-latinists. A possible explanation could be that there is a difference in mentality, not captured by the non-cognitive variables, which causes latinists to be more successful in school relative to their intelligence levels. Another plausible explanation is that, for whatever reason, teachers are more inclined to award latinists higher grades. There might well be a Pygmalion effect at play, whereby teachers' high expectancies of Latin pupils positively affect their school achievement.

What about the profile of the hellenists? If we separate the latinists who also study Greek from the latinists who only study Latin, the most notable finding is supplementary cognitive preselectivity in the Greek study option. In the Greek results, there always two things to look out for: whether or not the hellenists outperform the other latinists (i.e., whether or not the difference between hellenists and latinists *stricto sensu* is significant), and whether or not the hellenists were responsible for the group differences found earlier in the Latin results (i.e., whether or not the difference between non-latinists and latinists *stricto sensu* is significant). Starting with the latter, the differences between non-latinists and latinists in performance on Lexicon, Verbs, Gc, Gsm, G and school achievement all persist when excluding the hellenists,

¹¹³ The opposite is true of Gv, where the effect of Latin is non-significant in the base model but significant in the full model: For visuo-spatial processing, it is the non-latinists who excel in respect to their cognitive capacity and other characteristics.

which indicates that there is indeed a preselection effect of Latin itself. The only deviation compared to the Latin results from before is that the main effect of the difference between non-latinists and latinists is no longer significant in the full model of Gc.

The hellenists in this sample distinguish themselves across time from the other latinists for Lexicon, Gc, Gv, and G. These differences are already significant, and even medium in size, in the second year of secondary school. For the non-cognitive variables, there are no significant differences between second-year hellenists and other latinists. In the full models, the main effect of the difference between hellenists and latinists *stricto sensu* remains significant for Lexicon and Gc. In other words, it turns out that it is primarily the hellenists who particularly excel in crystallised intelligence, and not so much the other latinists; furthermore, the hellenists also particularly excel in lexical ability in respect to their Dutch language capacity, cognitive capacity, and other characteristics. As far as school achievement is concerned, the absolute performance of all latinists is similar, hellenist or not.

9.2.2 Group differences in evolution over time: to transfer or not to transfer?

Next we turn to the central hypothesis of this research, namely that the study of classical languages produces transfer effects to other cognitive domains. At this point, we have unequivocally established (cf. Section 9.2.1) that there are a priori differences between the latinist and non-latinist groups. In the analyses, these a priori differences have been cancelled out by matching and/or statistical modelling, so that a significant interaction between study option and time can be ascribed to the study option. The outcome variables will be reviewed one by one, focusing on the explanatory value of the interaction effect on top of the control variables.

For Lexicon, the results from the unmatched sample point towards a modest transfer effect of Latin, as latinists make a little bit more progress over time and this interaction between Latinist and Time is significant. Remember that the variable Lexicon does not only entail Dutch vocabulary, but also text structure and reading comprehension (cf. Sections 4.3 and 8.1.3.2). The results from the matched sample provide weaker support for transfer, the interaction effect being significant merely at the 10%-level. The difference in progress on Lexicon between latinists and non-latinists is nevertheless the same with or without matching, lying around 1.5 percentage points. This discrepancy raises the suspicion of a lack of statistical power in the

matched analysis, due to the fact that the matched sample ($N = 160$) is only about one-third the size of the complete unmatched sample ($N = 449$). The effect of Latin still exists when leaving the hellenists out of consideration, and there is no evidence for an added effect of Greek.

For Verbs, the results from both the unmatched and the matched sample point towards a modest transfer effect of Latin. More conspicuous than the progress of the latinists, arguably, is the lack of progress displayed by the non-latinists in the area of Dutch verb conjugation. The effect of Latin still exists when leaving the hellenists out of consideration, and there is no evidence for an added effect of Greek.

The results concerning verbal intelligence (Gc and Gsm) are mixed. For Gc, there is no evidence to suggest a transfer effect from either Latin or Greek. This is rather surprising in view of the Lexicon results, for the Antonyms subtest of the COVAT-3 measuring Gc essentially taps into vocabulary. Moreover, Lexicon and Gc correlate strongly at each of the three measurements, with an average correlation of $r = .644$. However, across all test variables, Antonyms is the only subtest that has Cronbach's alphas of below .70, signalling that the reliability of the Gc results may be questionable.

For Gsm, the results from the unmatched sample point towards a modest transfer effect of Latin. After splitting up the program group into hellenists and latinists *stricto sensu*, it becomes apparent that only the latinists make more progress over time compared to the non-latinists, whereas the hellenists do not. This finding is highly irregular, and it seems untenable that only latinists who do not study Greek would enjoy transfer. At first sight, the results from the matched sample do not support the idea of a transfer effect to Gsm. Nonetheless, there is still a difference in progress of 0.72 percentage points (= parameter estimate in the full model) to the advantage of the latinists, which falls well within the 95% confidence interval postulated by the full model of the unmatched sample.

As for non-verbal (Gf and Gv) and general intelligence (G), there is no evidence to suggest a transfer effect from either Latin or Greek. In fact, the results from the matched sample indicate that the non-latinists make slightly more progress on Gf over time.

Regarding school achievement, latinists are better at upholding their modern language grades, in particular Dutch and English, than non-latinists, whose modern language grades decline more strongly as years go by. The Greek results reveal that the hellenists are actually behind these effects, since their average Dutch and English grades stay roughly the same from the second until fourth year of secondary school. In the results from the matched sample, the

interaction effects for Dutch and English achievement no longer reach significance, even though the parameter estimates for these effects go up, which again casts doubt on the amount of statistical power.

9.2.3 Other observations from the longitudinal analyses

This section assembles some observations about the behaviour of the other independent variables in the models for Latin without matching (besides Latinist and Time). These observations are not directly connected to a research question, but are worth noting anyway.

- Overall, the chosen set of control variables appears to control for a priori group differences rather well, except in the case of school achievement, where an unexplained performance gap remains (as discussed earlier, cf. Section 9.2.1).
- Between schools, there are significant differences in performance on Lexicon, Verbs and Gc, as well as in school achievement (though not for Dutch). The control variables eliminate the school differences for Lexicon and Gc. For Verbs, the effect of school is significant in every model. Upon closer inspection, this effect is primarily dependent on school no. 4, which stands out in the positive sense when it comes to Dutch verb conjugation. For school achievement, the obvious conclusion – as noted multiple times in the first result section – is that the schools in this study do not use comparable grading systems.
- On average, boys perform better than girls on Lexicon, Gc and Gv, and girls perform better than boys on Gf, although in the matched sample only the sex differences for Gc and Gv are significant.
- SES, or parental educational level, cannot explain differences on the outcome variables, apart from achievement for mathematics: Math scores go up with every additional degree obtained by the parents.
- Native language is a significant predictor for most language-related variables, including Lexicon, Gc, Gsm, English achievement and French achievement. The effect is reversed for Gsm, which means that non-native speakers of Dutch score higher. If we split the mean scores on Gsm by Time, Latinist and NonNative, as in Table 5 below, we see that it is only among non-latinists that non-natives surpass natives.
- The effects of Dutch_T1 and G_T1 are always significant and nearly always have the expected direction. Cognitive capacity is inversely related to Verbs and to total school

achievement, although weakly. There is no sensible explanation for these inverse relations, so they are likely just statistical artefacts.

- The non-cognitive variables contribute remarkably little to the models of the test variables, rarely reaching significance and sometimes having opposite-direction effects. In the models of the school achievement variables, they seem to have more explanatory value.
- Taken together, the effects of Latinist, Time, School and the control variables explain a decent amount of the total sample variance. For most test variables, conditional R^2 of the full model lies between .605 (Gsm) and .871 (G); only the model of Gf lags behind, with a conditional R^2 of just .455. For the school achievement variables, conditional R^2 lies between .658 (AchEnglish) and .856 (AchTotal).

Table 5 Mean Gsm scores for Time x Latinist x NonNative

Time	Latinist	NonNative	N	Mean	SD
0	0	1	24	42.7	18.4
		0	220	38.1	16.3
	1	1	18	44.3	17.5
		0	184	44.3	14.6
1	0	1	24	52.8	15.2
		0	231	47.3	16.5
	1	1	17	55.0	20.7
		0	183	55.0	16.3
2	0	1	24	54.9	17.1
		0	233	51.1	18.5
	1	1	18	55.9	14.4
		0	185	61.5	15.9

9.2.4 Group differences in empathy and reading motivation

On the cognitive empathy scale, the fourth-year latinists and non-latinists both obtained an average score of 7.4 (minimum = 3; maximum = 9), so no group difference there. On the affective empathy scale, latinists obtained an average score of 12.6 and non-latinists of 12.1 (minimum = 6; maximum = 18). The difference is small (Cohen's $d = .202$), but significant at the 5%-level. Unfortunately, because this scale was administered only during the third measuring moment, it is not possible to attribute this effect to preselectivity or transfer; that will be for future research to decide.

For reading motivation, the scores diverged somewhat more: Latinists obtained an average score of 7.9 and non-latinists of 6.6 (minimum = 3; maximum = 15). This difference is highly significant ($p < .001$) and medium in size (Cohen's $d = .480$). The government report by Boone (2011) shows that pupils opting for Latin in the first year of secondary school are more eager to read than their peers, so preselectivity likely plays a part. Apparently, this difference is still present in the fourth year, whether it is further reinforced by classical language study or not.

To make sure that the results of the language-related outcome variables were not confounded by this fact, the full models of the unmatched sample were re-run with Reading motivation added as a covariate. Although its effect was significant for Lexicon, Verbs and Gc, the addition of Reading motivation did not alter the findings and had little impact on the model overall, as detailed in Table 6. The observed effects of Latin therefore do not seem to be caused by the pupils' inclination to reading, and the original models still stand. Nevertheless, in future research it may be advisable to measure reading motivation longitudinally and include it as a time-varying component.

Table 6 Explanatory value of Reading motivation

Outcome variable	<i>P</i> -value of Reading motivation	Impact on <i>p</i> -value of Time * Latinist	ΔR^2
Lexicon	< .001	+ 0.002	- 0.001
Verbs	0.027	- 0.012	+ 0.011
Gc	< .001	+ 0.052	- 0.001
Gsm	0.404	+ 0.003	- 0.002

9.3 Evaluation of the cognitive transfer effects of studying classical languages in Flemish secondary education

9.3.1 Potential candidates for cognitive transfer

From the results of the longitudinal study, far transfer from classical language learning to non-language-related domains seems highly improbable. There are tentative indications for near transfer to Dutch language ability and verbal short-term memory, albeit evidence is limited. There is an inkling of near transfer to modern foreign language ability as well, more particularly from Greek to English, but this is really to be taken with a grain of salt given that school achievement is not a reliable measure (cf. Section 8.1.3.2.). If we consider previous psychological research on cognitive transfer in general, this divergence between near and far transfer is not a novel outcome: The possibility of near transfer has been established and accepted, but the existence of far transfer is heavily contested till this day (cf. Sections 3.3 and 3.5.2; see also the addenda to Chapter 3). This study does not alter the status quo in that regard. The results are also in line with those of previous studies that assessed transfer from classical language learning to both language ability and general cognitive ability (Goik, 2021 and Haag & Stern, 2000, both summarised in the addenda to Chapter 2).

The skills quantified by Lexicon, Verbs and Gsm can easily be connected with common practices within classical language education. It makes sense that classical language pupils would expand their Dutch vocabulary more rapidly (cf. Lexicon), gain more insight into how texts are structured (cf. Lexicon), become better readers (cf. Lexicon), and cultivate the capacity more strongly to hold a string of words in working memory (cf. Gsm). After all, on a regular basis they spend time memorising Latin / Greek vocabulary, analysing Latin / Greek texts, and translating or paraphrasing these texts into Dutch. The variable Verbs is a somewhat curious case. One would expect all pupils to improve their mastery of the spelling rules for Dutch verbs over time, but instead the ordinary course of events seems to be that this competence stagnates or even deteriorates.¹¹⁴ Classical language pupils do make progress, presumably because they

¹¹⁴ This also holds true for school no. 4 in isolation. The average Verbs score may be higher in this school for all pupils, but a similar pattern is found of non-latinists more or less stagnating and latinists progressing.

engage more with the Dutch language and develop a deeper understanding of its verb system as it is placed next to the Latin and /or Ancient Greek verb system(s).

This research is agnostic as to the locus of any observed effects that are reconcilable with transfer. An observational study of the kind performed here does not allow to determine the true mechanism causing such effects, but merely permits to speculate about their possible underlying cause. It therefore remains completely open to debate whether actual cognitive transfer as envisioned in Chapter 3 occurs in the brains of classical language pupils, or if the effects they exhibit are side effects of classical language education in some other way. For example, it could be that classical language pupils show more improvement in Dutch proficiency purely because of the amount of attention devoted to the Dutch language in a Latin / Greek class, rather than that knowledge about Latin / Greek automatically transfers to Dutch.

A different hypothesis altogether – which many teachers and other dialogic partners of mine have hinted at – is that, rather than the subject matter itself bringing about transfer, it is the favourable academic atmosphere characteristic of classical language class groups that leads these pupils to experience more cognitive growth. While differences in atmosphere or mentality may very well be there, *an sich* this hypothesis has some difficulty explaining why language-related effects would be more likely than non-language-related effects. One could imagine a link with social desirability and high expectancies, especially with something like verb conjugation. The ability to spell verb forms correctly, or the lack thereof, is a very visible and sensitive aspect of native language proficiency – the infamous Dutch “dt-errors” (see e.g., Sandra et al., 2001). Latin pupils may feel more pressure to maintain a higher standard of spelling in their writing and avoid dt-errors.

9.3.2 Strength of the evidence

Despite the obscurity of the locus of effects, I will continue to refer to the significant interaction effects that were observed between study option and time as “transfer effects”. Now, how statistically significant are these effects really? To clarify the matter of strength of the evidence, it may be salutary to return to the basics for a moment and dwell on how inferential statistics for group comparisons work. The ideal situation would be to study all subjects in the population of interest, i.e. all pupils in Flemish general secondary education. Very often, as is the case here, this ideal is practically unattainable, so a sample is drawn and inferential statistics is applied to the sample data. Typically, two groups within the sample will not have the exact

same scores on the variable of interest. The p-value tells the probability of finding group differences equal to or more extreme than those found in the sample, assuming that there is no difference at population level. If this probability is small enough, it is reasonable to reject the null hypothesis and infer that there is indeed a difference at population level.

The conventional definition in the social and behavioural sciences of a “small enough” probability is 5%, which is also the threshold that has been used here. This brings us to the multiple comparisons problem: If multiple group comparisons are being made using the raw p-values, on average 5% of results will be false positives. Since common correction methods like Bonferroni are very restrictive and render significance tests overly conservative, I have deliberately chosen not to correct the p-values within the analyses but to address the multiple comparisons problem afterwards, here in the discussion. The present study counts 12 outcome variables (Lexicon, Verbs, Gc, Gsm, Gf, Gv, G, AchDutch, AchEnglish, AchFrench, AchMaths, and AchTotal), so one can expect the null hypothesis to be unduly rejected in approximately 0 to 1 out of 12 cases, which is an acceptable risk in this context. It must simply be kept in mind that among the variables that qualify as potential candidates for cognitive transfer (Lexicon, Verbs, Gsm, AchDutch, AchEnglish) there might perhaps be a false positive.

Another issue is constituted by the conflicting results between the full filtered longitudinal sample and the matched subset. Verbs is the only variable for which there is unambiguous evidence for a transfer effect in both the unmatched and the matched sample. For Lexicon, Gsm, AchDutch and AchEnglish, the significant interaction between study option and time vanishes in the analysis of the matched sample. Yet, as already proposed above, it would appear that the matched analyses suffer from a lack of power to detect small differences, seeing that the estimated size of effects continues to fall within the same range. The parameter estimate of the interaction effect stays the same for Lexicon, goes up for AchDutch and AchEnglish, but down for Gsm.

Aside from the reduction in sample size as a consequence of matching, one may ask the question whether the matched subset is sufficiently representative of the whole sample: Because the baseline linguistic and general-cognitive performance of the non-latinist group is lower, the worst-performing non-latinists as well as the best-performing latinists had to be pruned in order to achieve a balanced matching solution. Moreover, it cannot be ruled out that the latinists that were selected for matching and those that were not also differ from each other in other ways. In sum, while matching on baseline performance is the conceptually more

elegant method to test the development of group differences over time, some reserve about the results is in order due to the just mentioned disadvantages of the method.

That being said, the group effects are admittedly quite modest, even for behavioural research. Based on the full models of the unmatched sample, it can be claimed with 95% confidence that the additional progress per school year made by the latinists lies between 0.35 and 2.34 percentage points for Lexicon, between 0.16 and 6.09 percentage points for Verbs, and between 0.10 and 3.36 percentage points for Gsm. There is reason to suspect, however, that these values underestimate the true effects, for the extended models uncovered more ceiling effects (cf. Section 8.2) in addition to the one previously established for Orthography (cf. Section 4.3). In other words, the subtests are not difficult enough for the best performing participants, curtailing the room available to them to improve their scores.

This study was unable to prove any added transfer effects of Greek. In view of the above, there is no doubt that this study is too underpowered (with just 43 cases making up the Hellenist group) and the measuring instruments' sensitivity is too limited to detect any such effects, which, if they do exist, are likely to be small. What did become clear, is that Greek is not a requirement for transfer: Leaving out the hellenists did not alter the results of Latin, so the observed effects are not dependent on the presence of hellenists in the sample.

9.3.3 Group versus individual

A crucial assumption underlying this empirical research that has not been made explicit so far, was the following: If it is the case that classical language learning brings about cognitive transfer, these effects should be obviously demonstrable at the level of the group, aggregated across all classical language pupils. In hindsight, this is a rather bold assumption to make. Different pupils probably get different things from their classical language classes, and it is a well-known adage that what is taught does not equal what is learnt. It is therefore implausible that all classical language pupils would benefit from cognitive transfer to a similar extent: Some may experience a lot of transfer, and others none at all. This awareness of within-group variation casts new light on the modest size of the group effects. In the end, it is not the impact at the abstract group level that is of interest, but the impact that studying classical language can have on individual pupils.

This methodological contradiction is not unique to this particular research: The tension between the group and the individual is nothing new within psychology. In 1957, in his

presidential speech at the American Psychological Association's convention, Lee Cronbach (1916–2001) famously spoke of “the two disciplines of scientific psychology”: experimental psychology, which studies variance among treatments, and correlational psychology, which studies individual differences. Since then, entire books have been written about the matter of these two disciplines and how they might be united. One way of looking at it, already proposed by Cronbach (1957) himself, is in terms of “aptitude x treatment interactions”. The average treatment effect or group effect only occurs *in abstracto* at the aggregated group level. For an individual person from this group, the *in concreto* treatment effect may be larger, smaller, absent or negative, depending on their so-called “aptitude” for the treatment. One obstacle in the way of modelling individual differences is that many more measurements are needed: Just like conclusions at the group level require a large sample of participants, conclusions at the individual level require a large number of trials per participant (Boogert et al., 2018; Rouder & Haaf, 2019).

The longitudinal observational study may have been conceived in the tradition of experimental psychology, but there is some information at our disposal to reflect about individual characteristics that could influence treatment aptitude. Logically, a necessary condition for transfer from classical languages is that the subject matter is mastered reasonably well. It is unthinkable that transfer to Dutch language ability and verbal memory would happen fully automatically as a result of coming into contact with Latin, even if the pupil in question is not at all invested in their classical language study. Could motivation for and/or proficiency in Latin be the determining influences? An exploratory correlation analysis can (dis)credit these suggestions. Motivation appears to be of secondary importance, as there are no significant correlations between either controlled or autonomous motivation for Latin and any of the outcome variables at the third measurement.¹¹⁵ Achievement for Latin, on the other hand, does correlate significantly ($p < .05$ unless given) with Lexicon ($r = .214$), Verbs ($r = .337$) and Gsm ($r = .207$) at the second measurement, and with Lexicon ($r = .263$), Verbs ($r = .182$), Gc ($r = .217$) and Gsm ($r = .169$; $p = .054$) at the third measurement. Therefore, it seems possible that pupils who are more adept at Latin are also more susceptible to transfer from Latin to other domains.

¹¹⁵ As stated in Section 8.1.3.2, the second-measurement data of motivation for Latin are incomplete.

9.3.4 Context is everything

Having examined the cognitive transfer effects of studying classical languages, it remains to devote some thought to the last part of the title of both Section 9.3 and the entire dissertation itself: “in Flemish secondary education”. The generalisability of this and all research is restricted by its context. The findings presented here cannot just be transposed to other educational systems outside of Flanders. What is more, it is a question to what extent these findings even generalise to Flemish secondary education as a whole. It is worth repeating that all four schools are part of the Catholic network, an inevitable implication of prioritising schools with the highest number of classical language pupils in the second year. Especially striking is the educational level of the parents: Over 94% of participants in the longitudinal sample have at least one parent or direct caretaker who has a higher education degree. Highly educated families are thus overrepresented.

Moreover, this study cannot pronounce upon the cognitive transfer potential of classical languages on their own, but merely upon the occurrence of transfer effects as a consequence of studying classical languages within Flemish secondary education, more specifically for children who go to large Catholic schools and whose parents are highly educated. As mentioned throughout Chapters 1 and 2, Latin and Ancient Greek have also been deployed in completely different contexts, as an intervention aimed particularly at underprivileged children. Such an intervention is further removed from the everyday realities of children whose home environment provides few opportunities for intellectual stimulation, and classical languages are more likely to have a stronger impact in these contexts.

The pupils in the current sample, on the other hand, attend Flemish general secondary education, which is academically oriented, of high quality (in spite of rapidly declining PISA scores), and already rich in modern foreign language instruction (with French and English being mandatory for all, German being mandatory in some study options, and Spanish and sometimes other languages being optional), and they come from privileged backgrounds as well (going on parental educational level). In this context, it likely makes less of a difference whether someone studies classical languages or not; this is yet another reason, alongside ceiling effects and individual differences, why it must not amaze that group effects are only small.

9.3.5 Synthesis

In recapitulation, the present study found support for near transfer effects of studying Latin in Flemish secondary education. The evidence is not beyond irrefutable, but the results justify that near transfer be given the benefit of the doubt. Clearly, these transfer effects are not so large as to manifest as highly significant regardless of any design or analysis choice. Instead, the obtained results are more opaque and require more meticulous consideration, looking beyond rigid cutoffs for statistical significance every now and then. In view of the imperfections of the measuring instruments, the fact that abstraction was made from individual differences, and the specifics of the context, the evidence seems reasonably strong. Enrolling in classical languages is probably not a guarantee for transfer, but there are indications that classical language pupils tend to experience transfer to Dutch language ability and to verbal short-term memory, although it cannot be denied that the evidence for Dutch language ability is more convincing than that for verbal short-term memory.

9.4 Limitations and implications for further research

A discussion is not complete without stating research limitations, which I have divided into six groups. The limitations concern the study as it was designed; I will not repeat the larger conceptual considerations of studying groups versus individuals, touched upon in Section 9.3.3. Each of these limitations has implications for further research.

- (1) There is a risk of unknown factors affecting the (latent) outcome variables. It is typical of observational studies that they offer strong ecological validity but minimal control. While the amount of explained variance in this study is high (cf. Section 9.2.3), there is nevertheless a myriad of other, unmodelled factors that can affect cognitive performance (examples include diet, exercise, lifestyle et cetera). It is not impossible that the effects found here and interpreted as cognitive transfer are not attributable to classical language education at all: Theoretically, the classical language pupils could have some other life circumstance or experience in common that causes faster progress in Dutch language ability and verbal short-term memory. We have no information about

the pupils' lives outside of school (e.g., reading habits or other hobbies).¹¹⁶ Neither do we know about anything before the second year of secondary school; the influence of the first year of Latin vs. no Latin is a black box.

- (2) There is a risk of test-taking motivation clouding the (observed) results. Performance on a cognitive test is always a composite score, as it were, of cognition and of the way the test-taker relates to the test (because of personality, interests, mood et cetera). As treated in more detail in Section 6.8, test-taking motivation is a very important variable. However, it is conspicuously absent from the research designs of previous classical language impact studies (cf. Chapter 2), as well as from previous guides on school-based research (cf. Chapter 6). When designing this study, we too failed to take into account test-taking motivation. It is questionable whether all pupils took the tests as seriously. If not, careless and insufficient effort responding may have obscured the “true” scores they would have obtained had they given their best effort. The individual cognitive reports do not augur well in this regard, as some show rather peculiar fluctuations in test scores from year to year. Participants with extremely low scores were filtered out prior to analysis, but future research would do well to at least register response times so that more fine-grained filtering can be carried out, possibly in combination with strategies to increase test-taking motivation.
- (3) There is a risk of cheating clouding the (observed) results. Just like some pupils may have given minimal effort, it cannot be ruled out that some pupils may have cheated one way or another to artificially raise their scores. Besides illicit collaboration with peers, all language-related subtests could be made easier by looking up words or, in the case of Gsm, by copying the words that had to be remembered. We did not have a secure digital testing environment at our disposal to prevent such cheating behaviours. The only precaution that could be taken was having the participants disable their devices' spellcheck. The participants were of course under constant surveillance, but there were never enough invigilators to keep an eye on every screen at any given time.
- (4) The measuring instruments are suboptimal. The Dutch language test in particular was a make-shift solution. Time was lacking to go through a fully fledged development

¹¹⁶ Speaking of hobbies, in light of recent findings about the interaction between music ability and language ability (e.g., Smit et al., 2023) and the far transfer effects of music training (e.g., Jamey et al., 2024) it might be worthwhile to include participation in a part-time music programme as a background variable.

process, as outlined for instance by Brysbaert (2024), or to conduct a thorough validation, comparable to the one for Ch-ICAR (cf. Chapter 5). Hence ceiling effects stayed under the radar until the instrument was used in the main study. Furthermore, for neither the Dutch language test nor the COVAT-3 there were sufficient items to set up a counterbalanced design in order to avoid retest effects. Retesting is therefore another factor that may have affected the outcome variables. It might be problematic to assume that effects of retesting are equal for all participants and can thus be ignored. As noted earlier (cf. Section 9.2.2), subtest reliability was generally good to very good, but not for Gc.

- (5) The scope of this study is limited. Mainly because of constraints of time and available measuring instruments, several potential outcome variables were not measured. A lot of terrain has yet to be cultivated, so to speak. Further refinement with regard to the aspects of Dutch language ability is desirable. Of extraordinary interest would be to study transfer to writing ability.¹¹⁷ Furthermore, there are less prototypical cognitive traits like historical insight, non-cognitive traits like (affective) empathy, metacognitive traits like self-monitoring, and so on, which classical language education could possibly have an impact on as well.
- (6) The matter of classical language teaching method has not been not resolved. For reasons explained in the introductory remarks to Chapter 7, in the end I was unable to investigate whether teaching method has a moderating influence on the cognitive transfer effects of studying classical languages. Several research desiderata remain in this area. There is of course the starting hypothesis of this research project, which is in line with both contemporary (cf. Section 7.4.5) and earlier (cf. Introductory remarks to Part I) comments to that effect and which is supported by the theory of declarative memory

¹¹⁷ Not only is being able to produce coherent and polished texts a valuable skill in life, in my experience teachers often mention writing ability as a prime example of the difference between latinists and non-latinists, and personally I am also inclined to believe that studying classical languages has benefitted my own writing ability. Furthermore, Vanheule (2015) already demonstrated the likelihood of transfer to syntactic competence, which is indispensable for writing complex sentences. The difficulty of researching writing ability lies in the grading of essays: Either this has to be done manually by multiple correctors, which is very labour-intensive, or it is done automatically, which requires a strong affinity with Natural Language Processing on the part of the researcher.

(cf. Section 3.4.1), that a more heavily grammar-oriented method (insofar that still exists in Flanders, cf. Chapter 7) is conducive to transfer.

It would also be very enlightening to compare the typical way of teaching classical languages, which does not aim for communication skills but for reading knowledge – be it with varying degrees of focus on grammar – with the communicative method, which tries to mimic natural language acquisition and restricts the use of the native language to a minimum. The communicative method has been mentioned multiple times throughout this dissertation, and yet it was omitted from the theoretical framework and the research design for two reasons: because it highlights the similarity between classical learning and modern foreign language learning rather than the difference, and because in Flanders it is practiced by only a handful of teachers. In an Anglo-Saxon context, the communicative method is more prevalent and so a comparison with this method would be more feasible. Including the communicative method in the picture could advance our understanding of near transfer from classical language learning to native language ability, by separating the otherwise intertwined entities of the Latin / Greek languages themselves and the typical way of teaching them.

Works cited in this chapter

- Boogert, N., Madden, J., Morand-Ferron, J., & Thornton, A. (2018). Measuring and understanding individual differences in cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756), no. 20170280.
- Boone, S. (2011). *Sociale ongelijkheid bij de overgang van basis- naar secundair onderwijs: Een onderzoek naar de oriënteringspraktijk*. Vlaams Ministerie van Onderwijs en Vorming.
- Brysbaert, M. (2024). Designing and evaluating tasks to measure individual differences in experimental psychology: A tutorial. *Cognitive Research: Principles and Implications*, 9, no. 11.
- Cronbach, L. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12(11), 671–684.
- Goik, L.-S. (2021). *Zur Wirkung des Lateinunterrichts. Ergebnisse einer Längsschnittstudie* (U. Röhr-Sendlmeier, Ed.). Logos Verlag.
- Haag, L., & Stern, E. (2000). Non scholae sed vitae discimus? Auf der Suche nach globalen und spezifischen Transfereffekten des Lateinunterrichts. *Zeitschrift für Pädagogische Psychologie*, 14(2/3), 146–157.
- Hauspie, C., Duyck, W., Schelfhout, S., Vereeck, A., Janse, M., & Szmalec, A. (2024). Is studying Latin associated with (non-)linguistic cognitive transfer? A large-scale cross-sectional study. *Contemporary Educational Psychology*, 79, no. 102325.
- Jamey, K., Foster, N., Hyde, K., & Dalla Bella, S. (2024). Does music training improve inhibition control in children? A systematic review and meta-analysis. *Cognition*, 252, no. 105913.

- Rouder, J., & Haaf, J. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, 26(2), 452–467.
- Sandra, D., Brysbaert, M., Frisson, S., & Daems, F. (2001). Paradoxen van de Nederlandse werkwoordspelling. De logica van dt-fouten. *De Psycholoog*, 36(6), 282–287.
- Smit, E., Rathcke, T., & Keller, P. (2023). Tuning the musical mind: Next steps in solving the puzzle of the cognitive transfer of musical training to language and back. *Music & Science*, 6.
- Vanheule, J. (2015). Latijn en grammaticale competentie in het Nederlands. Verslag van een nieuw onderzoek. *Didactica Classica Gandensia*, 51, 89–107.
- Vereeck, A. (2024). Een nieuwe blik op de Vlaamse leerlingenaantallen klassieke talen. *Prora*, 29(1), 7–12.

General Conclusion

General Conclusion in Layman's Terms

Main study on cognitive transfer

Do smart pupils choose to study classical languages, or does studying classical languages make you smart? That is the central issue of this doctoral research. Earlier American research from the previous century suggested that pupils taking Latin progressed more quickly than their peers in both language proficiency and mathematical insight, thanks to the study of Latin. However, this research is outdated and not generalisable to our contemporary context. In Flanders, studies have shown that pupils who used to take Latin and/or Ancient Greek in secondary school tend to have higher success rates in higher education, and that they tend to score higher on tests of Dutch spelling and syntax. The question here is to what extent the fact that Latin pupils perform better can be explained by the enrolment of cognitively stronger pupils in classical language programs, or by a so-called “(cognitive) transfer” effect that is the result of studying classical languages, or by both. With my research, I aimed to provide a scientifically substantiated answer to this question.

To this end, I conducted a large-scale, multiyear observational study on the cognitive effects of studying classical languages. Such a study had never been performed before. More specifically, some thousand Flemish pupils in general secondary education were followed over a three-year period, from the second to the fourth year of secondary school. Nearly half of them studied Latin, while the rest pursued other general, academically-oriented study programs without a classical language. Once a year the pupils were tested on Dutch language ability and general cognitive abilities (comparable to an IQ test) in order to chart their cognitive development. Their school grades were collected as well. By repeatedly measuring the same group of pupils over time and taking into account various other factors such as the pupils’

gender, socio-economic status, and motivation, it was possible to isolate the effect of classical language learning and establish causal relationships.

Generally speaking, the study found two things: (1) On average, pupils who enrol for classical languages at the beginning of secondary school are indeed cognitively stronger, and retain this cognitive lead throughout secondary school, and (2) on average, classical language pupils do indeed exhibit a transfer effect of studying Latin on Dutch language ability, and possibly on short-term memory. There were no indications of a transfer effect on other general-cognitive abilities, or of an added effect of Greek, although not enough pupils in the sample studied Greek to tell for sure. In sum, in addition to the many invaluable lessons about cultural history, language, philosophy et cetera that classical language study provides, Latin pupils are also likely to experience a beneficial impact on their Dutch language ability. Due to the limitations of the tests that were used, this study does not permit to state with definitive certainty which aspects of Dutch language ability are impacted, or exactly how large the size of the transfer effect is. Nevertheless, generic vocabulary, reading comprehension and spelling are all likely to benefit. It also seems probable that the difference with other general study programs is not very large at the group level, but that the amount of experienced transfer varies from individual to individual.

Complementary study on teaching method

While official educational guidelines shape teaching practices, Latin is not taught in the same way in every Flemish classroom. The aim of the complementary study was to examine recent developments in Latin instruction in Flanders, with particular focus on the role of grammar. Historically, the traditional method emphasising grammatical knowledge long dominated Latin instruction, especially in the larger, Catholic educational network. In the sixties, the smaller, governmental educational network pioneered the so-called “reading method”. This method emphasises text comprehension and champions a contextualised approach: Grammar and cultural content are integrated in the reading of texts, rather than taught separately. Currently, the educational guidelines also stimulate the use of the reading method.

For the first time in Flanders, a large-scale survey on classical language teaching method was distributed, to which over two hundred Latin teachers responded. The survey results revealed

a shift away from the traditional method towards the reading method. Across both educational networks, the reading method has come to prevail, in line with educational guidelines. However, differences between the two networks still exist. Teachers from the governmental network, which already introduced the reading method decades ago, are on average more reading-oriented. The Catholic network, by contrast, only recently transitioned to the reading method, a change that was largely driven by new textbooks such as *Pegasus Novus*. Accordingly, the teachers from this network are on average somewhat more grammar-oriented, devoting more time to grammatical instruction.

Despite the broad trend of teachers embracing the reading method and its contextualised approach, teaching practices remain highly diverse. Some teachers continue to favour a more traditional approach: They argue that adequate grammatical knowledge is a necessary condition for text comprehension, and that systematic study of Latin grammar also contributes to overall language ability (which the main study confirmed as a transfer domain) and general cognition (which the main study did not confirm as a transfer domain). Others adopt an eclectic approach, combining elements from the reading method and the traditional method. Extremely text-centred approaches also exist, in which pupils are immersed in Latin through continuous reading with minimal grammatical explanations. Within the regular reading method as well, there is considerable variation in how time is allocated, how much attention is paid to grammatical phenomena during reading, and how balance is achieved between reading enjoyment and text-based learning. Individual teachers sometimes also vary their approach. Variation is therefore a defining feature of the Latin teaching landscape in Flanders.

Appendices

Appendix 1

Instrument Selection Process

Op zoek naar een talig meetinstrument

Fase 1: een eerste triage

Uit mijn voorbereiding op het FWO-interview:

I have taken a wide range of Dutch language tests in consideration, which I eventually narrowed down by applying the following criteria. To begin with, the test needed to suit my age bracket, which already ruled out quite a number of tests. I was looking for a test that first of all would tell me something about a pupil's general ability in their native language – rather than just knowledge of complex spelling rules that are very idiosyncratic of Dutch – and second of all would result in a meaningful score. Several tests and exercise programs, however, resulted in Meijerink levels, which are used in the Netherlands but not in Flanders, and which merely give a very broad indication of someone's language ability. Furthermore, it had to be possible to re-test and to administer the test in written/digital form. Naturally I also paid attention to validity and reliability.

Finally, I arrived at the CELF as the only viable option. I will select subtests that cover morphological, semantic and syntactic competence, and the materials will be expanded based on the work of Vanheule and in consultation with the research unit of Dutch linguistics. To the CELF I added Diatekst and LexTALE, to include measures of reading comprehension and receptive vocabulary. LexTALE is so quick and user-friendly that I thought a shame not to incorporate it in my linguistic test battery.

Fase 2: verbreding en verdieping van de heuristiek

Samenvatting heuristische bevindingen:

TEST	TAALASPECT(EN)	COMMENTAAR
LexTALE	Receptieve woordenschat	Lexical decision task: aangeven of letterreeksen bestaande woorden zijn Veel woorden van Latijnse origine
Woordenschattest met meerkeuze antwoorden CRR (Brysbaert)	Receptieve woordenschat	Zinvoller geacht omdat naar de betekenis van de woorden wordt gepeild Moeilijkheidsgraad te hoog voor mijn doelgroep? (oorspronkelijk bedoeld voor universiteitsstudenten)
CELF-5-NL	Subtests: - Woordcategorieën - (Aanwijzingen volgen) - Zinnen formuleren - Zinnen herhalen - Tekstbegrip - Definities van woorden - Zinnen samenstellen - Semantische relaties	Arbeidsintensieve en dure individuele afnames vs. zelfgecreëerde digitale groepsafnames waarrond veel vraagtekens bestaan en die geen tijdswinst opleveren qua scoring Automatische verbetering enkel haalbaar voor Woordcategorieën Te zeer gericht op lln. met taalstoornis
Test bachelor-onderzoek Jolien Vanheule	Subtests: - Zintransformaties - Splitsen in zinsdelen - Grammaticale fouten aanduiden en verbeteren	Tijdrovende handmatige verbetering → ondersteuning?, interbeoordelaarsbetrouwbaarheid? Beperkt aantal items, ontwikkeling van extra items voor hertests wenselijk
Diataal	Begrijpend lezen	Blijkt bij nadere inspectie ook enkel Meijerink-niveaus op te leveren
Toolkit Onderwijs en Arbeidsmarkt (TOA)	Begrijpend lezen	Vragen omtrent doelgroep, score en kostprijs Niet te bereiken (nooit antwoord gekregen op herhaaldelijke mails)
TaalVaST / Lucit-test	Algehele Nederlandse taalvaardigheid Categorieën:	Niet alle categorieën geschikt, m.n. degene waar academisch aandoende teksten aan te pas komen of die

	- (Het juiste woord) - Eén woord voor drie contexten - De juiste vorm - Het juiste synoniem - (Lezen op betekenis) - (Lezen op tekststructuur) - Alineaopbouw - (Studievaardigheden)	betrekking hebben op studievaardigheden hoger onderwijs Opmerking Jordi Heeren: praktisch instrument om grote groepen studenten op relatief korte tijd te screenen, minder geschikt als onderzoeksinstrument want construct breed en aantal items klein (beperkte interne consistentie)
Taaltoets Nederlands Vrije Universiteit Amsterdam	Algehele Nederlandse taalvaardigheid Categorieën: - Spelling en interpunctie - Structuur - Grammatica - Woordenschat en woordkeus - Formuleren	Pool van 96 losse items gekregen, recent verwijderd vanwege toetsveiligheid en inhoudelijke veroudering Duidelijke indeling die kan worden verbonden met eerder onderzoek, goede variatie aan items, ogenschijnlijk doenbare moeilijkheidsgraad
New Group Reading Test (NGRT)	Sentence completion & begrijpend lezen	Opgaves begrijpend lezen potentieel vertaalbaar → vereist vertaalwerk + wat met plagiaat?
Spellingtest Secundair Onderwijs (SSO Desoete)	Subtests: - Woorden in zinnen aangeboden - (Pseudowoorden) - Werkwoordspelling - Zinnendictee	Bestaande woorden in een zin + Werkwoordspelling: veel te simpel, cf. normeringsonderzoek (resp. 28,55/30 en 23,89/30 in 3ASO) Zinnendictee: handmatig te verbeteren, maar misschien niet onoverkomelijk
Begrijpend Leestest Secundair Onderwijs (BLSO Desoete)	Begrijpend lezen van een verhalende tekst, twee paralleltests	Aan de makkelijke kant, cf. normeringsonderzoek (resp. 20,77/28 en 22,14/28 in 3ASO), maar bruikbaar

Bovenstaande tabel is de neerslag van een iteratief proces waarbij (1) actie werd ondernomen om materiaal in te kijken en te verwerven, (2) tests tot in detail werden bestudeerd, en (3) nieuwe tests in overweging werden genomen.

Fase 3: bundeling en overzicht van de in aanmerking komende items per taalaspect

TAALASPECT		ITEMS
Woordenschat		A. Makkelijkste items < woordenschat CRR (20) B. Items Amsterdam (6) C. Woordcategorieën < CELF-5-NL (29) D. Eén woord voor drie contexten + Het juiste synoniem < Luci (8)
Grammatica	Congruentie Hen/hun Sterke werkwoorden Verwijswoorden Zinsbouw Lidwoorden Voorzetsels Voegwoorden Woordtransformaties	Items Amsterdam (3) Items Amsterdam (3) Items Amsterdam (3) Items Amsterdam (10) Items Amsterdam (2) Items Amsterdam (3) Items Amsterdam (6) Items Amsterdam (4) De juiste vorm < Luci (5)
Tekststructuur	Alineagrenzen Alineaopbouw	Items Amsterdam (5) Items Amsterdam (2) + Alineaopbouw < Luci (1)
Spelling	Woordspelling passief Werkwoordspelling passief Werkwoordspelling actief Dictee	Items Amsterdam (11) Items Amsterdam (2) Items Amsterdam (10) Zinnendictee < SSO Desoete
Begrijpend lezen	Hoofdgedachte alinea Verhalende tekst	Items Amsterdam (3) BLSO Desoete

Weggelaten uit de items van Amsterdam zijn interpunctie, idiomatische uitdrukkingen (bv. “spijkers op laag water zoeken” enz.), taalfouten (bv. contaminatie, pleonasme) en registerkeuze.

Fase 4: selectie van items voor try-out en onderverdeling in paralleltests

TAALASPECT	SELECTIE VAN ITEMS	VERANTWOORDING
Woordenschat	1. Woordenschat zonder context = woordenschat CRR (20) 2. Woordenschat in context = items Amsterdam (6)	CELF: probleem copyright Luci: te moeilijk + duurt langer want meer lezen
Grammatica	3. Woordtransformaties met zinnen (3) 4. Woordtransformaties met titels (2) 5. Voorzetsels (6) 6. Voegwoorden (3) 7. Zinsbouw (2 + 1 item vanuit voegwoorden)	Congruentie, hen/hun, sterke werkwoorden: arbitraire taalnormen die bovendien aan het verschuiven zijn Verwijswoorden: veel bedenkingen en dubbelzinnigheden Lidwoorden: minder relevant voor moedertaalsprekers
Tekststructuur	8. Alineagrenzen (4) 9. Alineaopbouw (2)	Alineagrenzen: item over Nederlandse KNMI geschrapt + in andere items enkele aanpassingen voor Vlaams publiek Alineaopbouw: item Luci niet geselecteerd voor try-out om praktische redenen
Spelling	10. Werkwoordspelling < 15 moeilijkste items van SSO Werkwoordspelling, met gegeven tijd (10) 11. Werkwoordspelling < 15 moeilijkste items van SSO Werkwoordspelling + actieve opgaves Amsterdam, zonder gegeven tijd (5+10)	Passieve opgaves afgeraden SSO Woorddictee: te makkelijk SSO Zinnendictee niet geselecteerd voor try-out om praktische redenen
Begrijpend lezen	12. verkorte BLSO – Lina's eigen koers (12) 13. verkorte BLSO – Anticavia-actieplan! (12)	Per BLSO leestekst de 2 moeilijkste items van iedere categorie, i.e. verbaal begrip, causale inferentie, logische inferentie, anaforische inferentie, interpretatie op macroniveau, extrapolatie Hoofdgedachte Amsterdam weg

Items werden vervolgens gevarieerd verdeeld over 5 paralleltests.

Fase 5: reductie van items na try-out

Samenvatting data-analyse try-out taal:

Het aantal items werd gereduceerd op de volgende gronden:

- transparantie van de opgave;
- spreiding van de moeilijkheidsgraad in het tweede en vierde jaar;
- moeilijkheidsgraad van individuele items;
- aantal items per taalaspect;
- informatieve waarde in het geheel van de taalmeting;
- culturele overwegingen;
- (in beperkte mate) interne samenhang van de schalen.

Fase 6: voltooiing van het meetinstrument

Fase 6a: samenstelling van de testbatterij

= behouden try-out items + aangepaste versies van SSO Zinnendictiee (gekozen voor items met een zo hoog mogelijke moeilijkheidsgraad, cf. itemanalyse) en item alineaoopbouw < Luci (opgave vormelijk in overeenstemming gebracht met de andere items alineaoopbouw, en door de samenstelling van de antwoordopties getracht de moeilijkheid ietwat te verlagen)

Fase 6b: programmatie van de items voor digitale afname

Fase 6c: testrun, finetuning en laatste check

Op zoek naar een algemeen-cognitief meetinstrument

Drie pistes

1. Wechsler Intelligence Scale (WISC-V-NL)

- Meest gebruikte IQ-test met internationale bekendheid en positieve reviews in binnen- en buitenland
- Digitale groepsafname niet mogelijk, enkel individueel en 1:1 af te nemen
→ grote werklast, nood aan team van assistenten

- Kostenplaatje: iPads, licenties, ev. jobstudenten
 - Bezorgdheid om leereffecten en plafondeffecten
2. **International Cognitive Ability Resource (ICAR)**, naar de Nederlandse vertaling van Fontaine
- Internationaal open source project, in het publieke domein, al gebruikt in 80-tal studies
 - Ontwikkeld voor web-based afnames
 - Volledig zelf te programmeren in software naar keuze
 - Geen kant-en-klare test + gebruik bij jongeren onontgonnen terrein
→ pionierswerk: zelf samenstellen en valideren, vergt uitgebreide try-out
 - Weinig theoretisch onderbouwd, d.w.z. geen expliciete link naar CHC-model
3. **CoVaT-CHC**
- Van eigen bodem, gebaseerd op een decennium gedegen Vlaams onderzoek
 - Versie voor digitale groepsafname
 - Afgewerkt product
 - Afgestemd op een jongere populatie, nl. specifiek ontwikkeld voor jongeren
 - Sterk theoretisch onderbouwd
 - Mogelijkheid tot TIQ
 - Gratis maar data return gevraagd

Bewandeling van de CoVaT-piste

Na een grondige verkenning van de drie pistes werd uiteindelijk gekozen om de CoVaT-piste verder te bewandelen. Er werden onderhandelingen opgestart met Thomas More en TechTransfer om een gebruikersovereenkomst te sluiten waarin onder meer de voorwaarden aangaande de data transfer worden vastgelegd.

Toevoeging van niet-cognitieve schalen

Heuristiek

- (Brief) Self-Control Scale
- Academic Self-Regulation Scale
- (Short) Grit Scale
- College Academic Self-Efficacy Scale
- Metacognitive Awareness Inventory
- Big Five Inventory

Keuze

1. **Grit / Volharding:** SIMON-vertaling van de Short Grit Scale (8 items)
2. **Zelfcontrole:** SIMON-vertaling van de Brief Self-Control Scale (13 items)
3. **Motivatie:** SIMON-vertaling van de Academic Self-Regulation Scale (16 items)

Appendix 2

Try-out Dutch Language Test: Parallel Test no.1



INSTRUCTIES

Deze test maakt deel uit van een onderzoek van de Universiteit Gent. Bedankt om mee te helpen aan dit onderzoek! Jouw deelname is heel belangrijk voor ons.

De test bestaat uit vier oefeningen, die je vindt op pagina's 2 en 3 van deze bundel. Controleer zo dadelijk of je alle vier de oefeningen hebt gekregen.

Lees telkens aandachtig de opgave voor je aan een oefening begint.

Schrijf je antwoorden niet bij de vragen maar in de kaders van het antwoordformulier, dat je vindt op pagina 4 van deze bundel. Bij meerkeuzevragen volstaat het om de letter van je antwoord te noteren.

Vergeet ook niet om bovenaan het antwoordformulier de gevraagde gegevens in te vullen.

Je werkt individueel en maakt geen gebruik van kladpapier.

Doe je best om alle vragen zo goed mogelijk op te lossen. Je hebt hiervoor één lesuur de tijd. Zodra je leerkracht toestemming geeft, mag je beginnen.

Succes!

Oefening 1

Kies bij elk item het alternatief dat het best aansluit bij het woord aan het begin.

- | | | |
|---|---|--|
| 1) Tequila
A. Schelp
B. Pannenkoekje
C. Monster
D. Alcohol | 8) Gong
A. Slaginstrument
B. Sleepinstrument
C. Blaasinstrument
D. Houtinstrument | 15) Woelig
A. Tactvol
B. Turbulent
C. Delicaat
D. Ontroerd |
| 2) Rups
A. Hondjes
B. Larve
C. Taartjes
D. Aardig | 9) Matig
A. Krachtig blijvend
B. Voordelig blijvend
C. Efficiënt blijvend
D. Redelijk blijvend | 16) Schrokop
A. Domoor
B. Schroothoop
C. Vogelschrik
D. Gulzigaard |
| 3) Dressoir
A. Werktuig
B. Boom
C. Klimaat
D. Meubelstuk | 10) Rekwisieten
A. Beperkingen
B. Benodigdheden
C. Afbakeningen
D. Versnaperingen | 17) Kakofonie
A. Geheimschrift
B. Kabaal
C. Vuile praat
D. Signalisatie |
| 4) Publiekelijk
A. Bevallig
B. Aansprakelijk
C. Kostbaar
D. Openbaar | 11) Paviljoen
A. Bijgebouw
B. Bijbedoeling
C. Bijfiguur
D. Bijgerecht | 18) Knullig
A. Ontrouw
B. Flauw
C. Onhandig
D. Prullerig |
| 5) Acteur
A. Beheerder van goederen
B. Persoon verbonden aan het toneel
C. Ontwerper van auto's
D. Functionaris op treinen | 12) Onbekwaam
A. Aanstootgevend
B. Niet passend
C. Niet geschikt
D. Niet bezonnen | 19) Wrok
A. Bouwval
B. Keukengerei
C. Haat
D. Gierigaard |
| 6) Romig
A. Slaperig
B. Slordig
C. Dik en vloeibaar
D. Met lijm bedekt | 13) Feeks
A. Schroevendraaier
B. Boor
C. Tang
D. Hamer | 20) Teint
A. Specerij
B. Pesterij
C. Kleur
D. Gesp |
| 7) Rekrut
A. Soldaat
B. Reglement
C. Onmens
D. Hoedanigheid | 14) Bretel
A. Jas
B. Schoen
C. Broek
D. Pet | |

Oefening 2

Vul de juiste werkwoordsvormen in. Er is gegeven of het gaat om de tegenwoordige tijd (TT), de verleden tijd (VT) of het voltooid deelwoord (VD).

- 1) De piraat _____ [veroveren, TT] het schip.
- 2) _____ [Bieden, TT] jij je excuses aan?
- 3) Toen _____ [barsten, VT] hij in lachen uit.
- 4) De kok _____ [bereiden, VT] de maaltijd met veel zorg.
- 5) Mijn vrienden _____ [bezoeken, VT] mij vorige week in het ziekenhuis.
- 6) De meisjes _____ [brengen, VT] me toen een cadeautje mee.
- 7) We _____ [kruipen, VT] door smalle spleten in de rotsen.
- 8) Tijdens de wedstrijd _____ [springen, VT] de atleten zo hoog mogelijk.
- 9) Vroeger _____ [lusten, VT] ik dat niet graag, maar nu lust ik alles.
- 10) Ik _____ [laden, TT] mijn telefoon op regelmatige tijdstippen op.

Oefening 3

Lees de volgende zinnen. Gebruik het vetgedrukte woord erboven om een woord te vormen dat past.

Voorbeeld:

0) **Helpen**

Zonder _____ kan ik deze test niet maken.

→ Antwoord: hulp

1) **Docent**

Deze notities zijn een synthese van de stof die _____ is tijdens de lessen.

2) **Legitimeren**

Het is dus te simpel om te stellen dat regelingen _____ worden gevonden als zij de economische belangen van zoveel mogelijk burgers dienen.

3) **Einde**

In het kader van een _____ steekproef mogen de resultaten niet veralgemeend worden naar de gehele populatie.

Oefening 4

Lees de volgende titels en ondertitels. Gebruik het vetgedrukte woord erboven met het overeenkomstige symbool om een woord te vormen dat past.

1) **(*) Descriptie, (+) functie**

Titel: Extreem moeilijk gedrag bij meisjes.

Ondertitel: Een (*) _____ onderzoek naar achtergronden en kenmerken van meisjes die in een kinderbeschermingsinrichting werden opgenomen en naar het (+) _____ in de maatschappij jaren later.

2) **(*) Muteren, (+) impliceren**

Titel: De diversiteit van genetische (*) _____.

Ondertitel: Nieuwe inzichten en (+) _____ voor genetische diagnostiek.

ANTWOORDFORMULIER

School:

Leerjaar:

Studierichting:

Klas:

Klasnummer:

Welke studierichting(en) volgde je de voorbije jaren in het secundair onderwijs?

.....

.....

Oefening 1 (meerkeuze; gelieve enkel de letter van je antwoord te noteren)

1	8	15
2	9	16
3	10	17
4	11	18
5	12	19
6	13	20
7	14	

Oefening 2

1	6
2	7
3	8
4	9
5	10

Oefening 3

1
2
3

Oefening 4

1*	2*
1†	2†

Appendix 3

Full List of Items in the Dutch Language Test

Given below are the items with their corresponding instructions in italics. The correct answers are underlined. These items cannot be freely used, see Chapter 4 for more information.

<i>T1&2 (bovenaan de pagina): "Kies telkens het alternatief dat het best aansluit bij het woord aan het begin."</i> <i>T3 (bij ieder item): "Kies het alternatief dat het best aansluit bij dit woord:"</i>	
Q1	Rups A. Hondjes <u>B. Larve</u> C. Taartjes D. Aardig
Q2	Romig A. Slaperig B. Slordig <u>C. Dik en vloeibaar</u> D. Met lijm bedekt
Q3	Schrokop A. Domoor B. Schrootloop C. Vogelschrik <u>D. Gulzigaard</u>
Q4	Feeks A. Schroevendraaier B. Boor <u>C. Tang</u> D. Hamer
Q5	Rekrut <u>A. Soldaat</u> B. Reglement

	C. Onmens D. Hoedanigheid
Q6	Paviljoen <u>A. Bijgebouw</u> B. Bijbedoeling C. Bijfiguur D. Bijgerecht
Q7	Bretel A. Jas B. Schoen <u>C. Broek</u> D. Pet
Q8	Onbekwaam A. Aanstootgevend B. Niet passend <u>C. Niet geschikt</u> D. Niet bezonnen
Q9	Kakofonie A. Geheimschrift <u>B. Kabaal</u> C. Vuile praat D. Signalisatie
Q10	Rekwisieten A. Beperkingen <u>B. Benodigdheden</u> C. Afbakeningen D. Versnaperingen
Q11	Woelig A. Tactvol <u>B. Turbulent</u> C. Delicaat D. Ontroerd
Q12	Wrok A. Bouwval B. Keukengerei <u>C. Haat</u> D. Gierigaard
Q13	Teint A. Specerij B. Pesterij <u>C. Kleur</u> D. Gesp

Q14	Stramien A. Geheim B. Moeizaam <u>C. Patroon</u> D. Zeer hoog
Q15	Verbolgen A. Taboe B. Beduud C. Verbluft <u>D. Boos</u>

T1&2 (bovenaan de pagina): "Door welk woord kan je het vetgedrukte woord vervangen zonder dat de betekenis van de zin verandert?"

T3 (bij ieder item): "Door welk woord kan je het vetgedrukte woord vervangen zonder dat de betekenis van de zin verandert?"

Q16	Maak je niet druk over zo'n futiliteit . A. <u>kleinigheid</u> B. tegenslag C. uitbrander D. teleurstelling
Q17	Het toepassen van complementaire zorg is de afgelopen jaren sterk toegenomen. A. uiteenlopende B. <u>aanvullende</u> C. uitmuntende D. traditionele
Q18	De media speculeerden over de uitslag van de verkiezingen. A. zochten naar B. rapporteerden over C. gaven een verklaring van D. <u>gisten naar</u>
Q19	Deze verklaring voor de klimaatsverandering is heel plausibel . A. <u>aannemelijk</u> B. tegenstrijdig C. onlogisch D. merkwaardig
Q20	Het kabinet heeft de risico's die deelname aan militaire acties met zich meebrengt gebagatelliseerd . A. in kaart gebracht B. overdreven C. verzwegen D. <u>te klein voorgesteld</u>

T1 (bovenaan de pagina): “Vul het juiste woord in.”	
T2 (bovenaan de pagina): “Vul het juiste woord in. Noteer enkel het woord dat ontbreekt en NIET de volledige zin.”	
T3 (bij ieder item): “Vul het juiste woord in. (Noteer enkel het woord dat ontbreekt en NIET de volledige zin.)”	
Q21	In de 14e eeuw bezweek een derde van alle Europeanen <u>aan</u> de pest.
Q22	Deelname <u>aan</u> dit practicum betekent dat u de aanwijzingen nauwkeurig opvolgt.
Q23	Gabriel García Márquez’ uitzonderlijke schrijftalent komt het beste tot uitdrukking <u>in</u> zijn meesterwerk “Cien años de soledad”.
Q24	Geneeskunde stelt hoge eisen <u>aan</u> de studenten.
Q25	Toen de Olympische Spelen voor de deur stonden, werd China aangesproken <u>op</u> de mensenrechtensituatie.

T1&2 (bovenaan de pagina): “Lees de volgende zinnen. Gebruik het vetgedrukte woord tussen haakjes om een woord te vormen dat past. Voorbeeld: Helpen Zonder _____ kan ik deze test niet maken. → Antwoord: <i>hulp</i> ”	
T3 (bij ieder item): “Lees onderstaande zin, en vul de zin aan door een woord te vormen op basis van het vetgedrukte woord tussen haakjes.”	
Q26	Legitimeren Het is dus te simpel om te stellen dat regelingen <u>legitiem</u> worden gevonden als zij de economische belangen van zoveel mogelijk burgers dienen.
Q27	Docent Deze notities zijn een synthese van de stof die <u>gedoceerd</u> is tijdens de lessen.

T1&2 (bovenaan de pagina): “Lees de volgende titels en ondertitels. Gebruik het vetgedrukte woord erboven met het overeenkomstige symbool om een woord te vormen dat past. Voorbeeld: Helpen Titel: Zonder (*)_____ aan ouders schiet gerechtelijke jeugdhulp tekort. Ondertitel: Ex-medewerkers van de jeugdrechtsbank kaarten lacunes in familiale hulpverlening aan. → Antwoord: <i>hulp</i> ”	
T3 (bij ieder item): “Lees onderstaande titel en ondertitel, en vul ze aan door woorden te vormen op basis van de vetgedrukte woorden tussen haakjes.”	
Q28_SQ281	(*) Descriptie , (†) functie
Q28_SQ282	Titel: Extreem moeilijk gedrag bij meisjes. Ondertitel: Een (*) <u>descriptief</u> onderzoek naar achtergronden en kenmerken van meisjes die in een kinderschermingsinrichting werden opgenomen en naar het (†) <u>functioneren</u> in de maatschappij jaren later.

Q29_SQ291	(*) Muteren , (†) impliceren
Q29_SQ292	<i>Titel:</i> De diversiteit van genetische (*) <u>mutaties</u> . <i>Ondertitel:</i> Nieuwe inzichten en (†) <u>implicaties</u> voor genetische diagnostiek.

T1-3 (bovenaan de pagina): “Vul de juiste werkwoordsvormen in. Je moet zelf uit de zin afleiden hoe je het werkwoord moet vervoegen.”

Q30_SQ301 Q30_SQ302 Q30_SQ303	Volgend jaar <u>wordt</u> [worden] de Schoolstraat <u>verbreed</u> [verbreden] en <u>geasfalteerd</u> [asfalteren].
Q31_SQ311	Ik heb de glazen goed <u>gespoeld</u> [spoelen].
Q32_SQ321	Door dagelijkse hersengymnastiek kunnen de hersenen <u>gestimuleerd</u> [stimuleren] worden.
Q33_SQ331	De zaal is mooi versierd [versieren].
Q34_SQ341 Q34_SQ342	De havenregels zijn <u>veranderd</u> [veranderen]: schepen worden uitsluitend nog overdag <u>gelost</u> [lossen].
Q35_SQ351 Q35_SQ352	Nu onze omzet is <u>verdubbeld</u> [verdubbelen], wordt het tijd dat men ook mijn salaris <u>verveelvoudigt</u> [verveelvoudigen].
Q36_SQ361	Hij heeft zijn rekening al <u>betaald</u> [betalen].
Q37_SQ371	De onderzoeker verwacht dat zijn onderzoek veel nieuwe inzichten <u>oplevert</u> [opleveren].
Q38_SQ381	Het is niet duidelijk hoeveel de studenten <u>verdiend</u> [verdienen] hebben met de verkoop van hun spelletje.
Q39_SQ391	Het pas <u>gelande</u> [landen] vliegtuig had vertraging.
Q40_SQ401 Q40_SQ402	Naar <u>verluidt</u> [verluiden] heeft het kabinet nieuwe bezuinigingsvoorstellen <u>behandeld</u> [behandelen].
Q41_SQ411	Er wordt volgende week besloten of een van de meest <u>gehate</u> [haten] belastingen verdwijnt.

T1-3 (bovenaan de pagina): “Noteer de zinnen die je hoort. Vergeet geen hoofdletters te schrijven.”

Q42_SQ421 Q42_SQ422 Q42_SQ423 Q42_SQ424 Q42_SQ425 Q42_SQ426	<u>Gelukkig biedt Sams antigifmiddel</u> / de <u>oplossing</u> tegen die <u>slangenbeet</u> .
Q43_SQ431 Q43_SQ432 Q43_SQ433 Q43_SQ434 Q43_SQ435	De <u>leraar geschiedenis</u> <u>vertelt</u> / dat hij de <u>belangrijkste Parijse bezienswaardigheden</u> <u>bezoekt</u> heeft.

Q43_SQ436	
Q43_SQ437	
Q44_SQ441	Ik <u>belandde</u> in het ziekenhuis / met <u>geschaafde knieën</u> / en een <u>arm-</u> en <u>beenbreuk</u> .
Q44_SQ442	
Q44_SQ443	
Q44_SQ444	
Q44_SQ445	

T1&2 (bovenaan de pagina): "Lees onderstaande zinnen. Ze vormen telkens een alinea, maar ze staan niet in de goede volgorde. Geef aan wat de correcte volgorde is."

T3 (bij ieder item): "Onderstaande zinnen vormen een alinea, maar ze staan niet in de goede volgorde. Lees de zinnen, en geef aan wat de correcte volgorde is."

Q45	(1) Op grond van nieuwe wetenschappelijke inzichten zijn er echter sterke aanwijzingen dat in 2010 de kritische belasting voor verzuring nog steeds overschreden zal worden in bosgebieden. (2) Voorlopige berekeningen van de overschrijding van de kritische depositie op bosgebieden, gebaseerd op deze nieuwe inzichten, zijn weergegeven in onderstaande figuur. (3) In Europa wordt een gemiddelde reductie van het areaal waar de kritische depositie door verzurende stoffen wordt overschreden met 85% teruggebracht (ruim meer dan de doelstelling van 50%). A. 1 – 3 – 2 B. 2 – 1 – 3 C. 2 – 3 – 1 D. <u>3 – 1 – 2</u>
Q46	(1) Een groot deel van de wapenhandel naar conflictgebieden is in handen van particuliere tussenhandelaren. (2) Nog maar een klein aantal staten heeft wetgeving die de tussenhandel aan banden legt. (3) Vanuit gebieden met een zwakke controle op wapenexport organiseren die de verscheping van wapens en veiligheidsapparatuur naar eindgebruikers, onder wie criminelen, terroristen, gewapende oppositiegroepen en de overheden die zich aan ernstige mensenrechtenschendingen schuldig maken. A. <u>1 – 3 – 2</u> B. 2 – 1 – 3 C. 2 – 3 – 1 D. 3 – 1 – 2
Q47	(1) Eerstgeborenen zijn egoïstischer en minder behulpzaam dan mensen met een oudere broer of zus. (2) Enige kinderen daarentegen waren net als middelste kinderen en laatstgeborenen vrij genereus. (3) De wetenschappers besluiten dat kinderen behoedzamer worden voor de motieven van anderen wanneer ze er een broertje of een zusje bij krijgen.

	(4) Uiteindelijk bleek dat een vierde van de eerstgeborenen minder geld teruggaf dan dat ze in eerste instantie hadden ontvangen. (5) Bijgevolg zijn ze minder behulpzaam. (6) De mannen en vrouwen namen deel aan een spel waarin ze elkaar geld moesten geven. Elk bedrag werd verdrievoudigd wanneer het gedoneerd werd. (7) Tot die vaststelling zijn Franse onderzoekers gekomen na een experiment met meer dan 400 vrijwilligers. A. 1 – 7 – 4 – 6 – 2 – 5 – 3 B. 1 – 4 – 7 – 6 – 2 – 3 – 5 C. <u>1 – 7 – 6 – 4 – 2 – 3 – 5</u> D. 1 – 6 – 4 – 7 – 2 – 5 – 3
--	--

<i>T1&2 (bovenaan de pagina): "Lees onderstaande teksten. Geef aan waar de verschillende alinea's beginnen."</i> <i>T3: geen overkoepelende instructie</i>	
Q48	<i>T1&2: "Onderstaande tekst bestaat uit vier alinea's. De eerste alinea is gegeven, de tweede alinea begint bij zin (1). Geef aan waar de derde en vierde alinea beginnen."</i> <i>T3: "Onderstaande tekst bestaat uit vier alinea's. De eerste alinea is gegeven, de tweede alinea begint bij zin (1). Lees de tekst, en geef aan waar de derde en vierde alinea beginnen."</i> BOSSEN BETER DAN BIOBRANDSTOFFEN De ruimte die nodig zou zijn om gewassen te kweken voor biobrandstoffen, om daarmee de CO₂-uitstoot te beperken, kan beter gebruikt worden voor bossen. Bij gebrek aan beter moeten auto's voorlopig op fossiele brandstoffen blijven rijden. (1) Dat schrijven Renton Righelato van de World Land Trust en Dominick Spracklen van de Universiteit van Leeds in het wetenschappelijke tijdschrift Science. (2) Ze maakten een overzicht van alle soorten biobrandstoffen en rekenden uit hoeveel CO₂ ze netto besparen. (3) Het extraatje was dat ze dit keer rekening hielden met de hoeveelheid land die de gewassen innemen. (4) Dat is belangrijk, want op datzelfde land kunnen bossen groeien. (5) En die nemen CO₂ op. (6) De eindbalans was duidelijk: een hectare bos, en met name regenwoud, neemt twee tot negen keer meer CO₂ op dan een hectare van welk soort biobrandstof dan ook bespaart. (7) Zonde dus, die biobrandstoffen. (8) Bovendien, schrijven de auteurs, hebben bossen grote voordelen boven biobrandstofgewassen. (9) Ze voorkomen verwoestijning en helpen beschermde dieren en planten overleven. (10) Een akkertje suikerbieten kan niets daarvan. A. zin 5 en 8 B. zin 5 en 9 C. <u>zin 6 en 8</u> D. zin 6 en 9
Q49	<i>T1&2: "Onderstaande tekst bestaat uit drie alinea's. De eerste alinea is gegeven, de tweede alinea begint bij zin (1). Geef aan waar de derde alinea begint."</i>

	T3: “Onderstaande tekst bestaat uit drie alinea's. De eerste alinea is gegeven, de tweede alinea begint bij <u>zin</u> (1). Lees de tekst, en geef aan waar de derde alinea begint.” WAAROM IS DE ZEE ZOUT? Gemiddeld is het zoutgehalte in onze oceanen 3,5 procent. Dat betekent dat elke liter zeewater ongeveer 35 gram zouten bevat; voor het grootste deel (24 gram) natriumchloride. (1) Sir Edmond Halley (de Engelse astronoom, geofysicus, wiskundige en naamgever van de komeet) deed in 1715 een poging om de zoute zee te verklaren. (2) Volgens hem werden zouten en andere mineralen door neerslag uit de bodem gewassen en via rivieren naar de zee gebracht. (3) In werkelijkheid kwam het natrium lang geleden vrij bij de vorming van de oceaانبodem. (4) Gasvormige chlorides borrelden op uit het binnenste van de aarde. (5) Vervolgens belandden deze via onderzeese vulkanen en hydrothermale bronnen in het zeewater. A. zin 2 B. <u>zin 3</u> C. zin 4 D. zin 5
Q50	T1&2: “Onderstaande tekst bestaat uit drie alinea's. De eerste alinea begint bij <u>zin</u> (1), de derde alinea is gegeven. Geef aan waar de tweede alinea begint.” T3: “Onderstaande tekst bestaat uit drie alinea's. De eerste alinea begint bij <u>zin</u> (1), de derde alinea is gegeven. Lees de tekst, en geef aan waar de tweede alinea begint.” BEELDEN MAANLANDING KWIJT (1) Beetje slordig, zeg dat wel. (2) De ruimtevaartorganisatie zoekt al een jaar naar de originele opnames van de eerste landing op de maan op 20 juli 1969. (3) Het beeldmateriaal bevond zich lange tijd in het Amerikaanse Nationale Archief. (4) Eind jaren zeventig verhuisde het naar de Nasa. (5) En nu zijn de magneetbanden dus kwijt, bekend de Nasa met het schaamrood op de kaken. (6) Wel bestaan er gekopieerde versies van de beelden van de eerste maanlanding - "One small step for man, but a giant leap for mankind" - maar de kwaliteit daarvan is niet zo goed. (7) "Een kopie van een kopie", aldus de Nasa. (8) De beelden die miljoenen televisiekijkers in de zomer van 1969 aan het scherm gekluisterd hielden, waren namelijk met een televisiecamera opgenomen van de monitor waarop de originele, doorgestraalde beelden te zien waren. (9) De apparatuur van de Nasa sloot niet aan op de toenmalige televisietechnologie. In totaal zijn er 700 dozen kwijtgeraakt. In die dozen zitten niet alleen magneetbanden met het originele beeldmateriaal, maar ook gegevens over de gezondheidstoestand van de astronauten en de staat van hun ruimtecapsule. A. zin 4 B. <u>zin 6</u> C. zin 7 D. zin 8

T1-3 (bovenaan de pagina): "Lees onderstaande tekst en beantwoord de vragen.

LINA'S EIGEN KOERS – Barbara Axters

(...)"

Q51	Waarom zegt het meisje tegen haar vader "We zijn hier weg:"? A. Omdat het al laat is en ze naar huis wil. B. <u>Omdat ze zich schaamt.</u> C. Omdat ze boos is op haar papa omdat hij verder fietste. D. Omdat ze bang is voor de sleuteldrager.
Q52	De titel 'Lina 's eigen koers' verwijst naar... A. ...de fietstocht van een vader en zijn dochter. B. ...het meisje dat de fietstocht onderbreekt om naar de kerk te gaan. C. <u>...het meisje dat beslist om haar vader niet meer te volgen in zijn sportprestaties.</u> D. ...de fiets van het meisje.
Q53	'Dat er niemand meer in zijn wielsporen reed' (alinea 3) betekent... A. ...dat er geen voertuigen volgen. B. <u>...dat zijn dochter niet volgt.</u> C. ...dat er geen afdrukken van autobanden zichtbaar zijn. D. ...dat er geen sporen van de fiets te zien zijn.
Q54	De vader van het meisje wil een week later gaan klimmen. A. De vader en zijn dochter klimmen tot aan de top. B. <u>De vader en zijn zoon klimmen tot aan de top.</u> C. Broer en zus klimmen tot aan de top. D. Niemand klimt tot aan de top.
Q55	'Zijn vader had de Mount Everest nog beklommen' (alinea 1): 'zijn vader' verwijst naar... A. ...Lina's vader. B. <u>...Lina's grootvader.</u> C. ...Brieks vader. D. ...de vader van Lina's stiefbroer.
Q56	'Daarna zouden ze de hele weg langs zijn vinger volgen' (alinea 2): naar wie verwijst ze? A. Papa. B. Het meisje. C. Briek. D. <u>De luisteraars.</u>
Q57	Wat zal het meisje in de tekst doen als haar broer haar vraagt om mee te gaan klimmen? A. Zich zorgen maken omdat het haar toch niet zal lukken. B. Beginnen trainen om mee te kunnen gaan. C. Zeggen dat ze liever gaat fietsen met haar papa. D. <u>De klimtocht aanvangen en afwachten hoe ver ze geraakt.</u>

Q58	De zin ‘Toen ze buitenkwam, zag ze haar vader grinnikend tussen de menigte’ (laatste alinea) betekent dat haar vader... A. ...de enige is die blij is, in het midden van een groep mensen. B. ...de enige is die luid lacht. C. <u>...zacht spottend lacht.</u> D. ...giert van het lachen samen met alle aanwezigen.
Q59	Hoe oud is Briek op het moment van de fietstocht volgens jou? A. Jonger dan 13 jaar. B. Tussen 13 en 16 jaar. C. <u>Tussen 16 en 20 jaar.</u> D. Ouder dan 20 jaar.
Q60	‘Met pretlichtjes in zijn ogen vertelde hij aan de thuisblijvers’ (laatste alinea): waarom heeft hij pretlichtjes in de ogen? A. <u>Omdat hij het grappig vond wat zijn dochter overkwam.</u> B. Omdat hij een ondeugend iemand is. C. Omdat hij moe is van zo veel te fietsen. D. Omdat hij ontroerd is.
Q61	Deze fietstocht kan plaatsvinden... A. ...in december. B. ...in februari. C. <u>...in september.</u> D. ...in november.
Q62	Hoe zal het meisje zich later de fietstocht in Picardië herinneren? A. Een plaats waar oude mannen met een gevlochten koord rond hun middel rondlopen. B. Een streek met veel verlaten dorpen waar alles heel voorspelbaar verloopt. C. Een druk bezocht gebied in Frankrijk waar de kerken veel bezoekers lokken. D. <u>Een vakantiebestemming waar de relatie met haar vader erg veranderd is.</u>

Appendix 4

Classification of Items in the Dutch Language Test

LS = LimeSurvey.

Item	LS-vraag	Oefening	Subtest	Taalaspect	LS-pagina T1 & 2	LS-pagina T3					
Q1	Q1	Oefening 2 / Woorden- schat zonder context	Losse woorden- schat	Woorden- schat	Oefening 2	Woorden- schat					
Q2	Q2										
Q3	Q3										
Q4	Q4										
Q5	Q5										
Q6	Q6										
Q7	Q7										
Q8	Q8										
Q9	Q9										
Q10	Q10										
Q11	Q11										
Q12	Q12										
Q13	Q13										
Q14	Q14										
Q15	Q15										
Q16	Q16	Oefening 3 / Woorden- schat in context	Woorden- schat in context		Grammatica		Oefening 3	Grammatica			
Q17	Q17										
Q18	Q18										
Q19	Q19										
Q20	Q20										
Q21	Q21	Oefening 4 / Voorzetsel- gebruik	Voorzetsels	Grammatica		Oefening 4	Grammatica				
Q22	Q22										
Q23	Q23										
Q24	Q24										
Q25	Q25										
Q26	Q26	Oefening 5 / Woordtrans- formaties met zinnen	Woordtrans- formaties			Oefening 5					
Q27	Q27										
Q28_SQ281	Q28	Oefening 6 / Woordtrans- formaties met titels								Oefening 6	
Q28_SQ282											
Q29_SQ291	Q29										
Q29_SQ292											

Q30_SQ301	Q30	Oefening 7 / Werkwoord- spelling	Werkwoord- spelling	Spelling	Oefening 7	Spelling
Q30_SQ302						
Q30_SQ303						
Q31_SQ311	Q31					
Q32_SQ321	Q32					
Q33_SQ331	Q33					
Q34_SQ341	Q34					
Q34_SQ342						
Q35_SQ351	Q35					
Q35_SQ352						
Q36_SQ361	Q36					
Q37_SQ371	Q37					
Q38_SQ381	Q38					
Q39_SQ391	Q39					
Q40_SQ401	Q40					
Q40_SQ402						
Q41_SQ411						
Q42_SQ421	Q42	Oefening 1 / Dictee	Zinspelling	Spelling	Oefening 1	Dictee
Q42_SQ422						
Q42_SQ423						
Q42_SQ424						
Q42_SQ425						
Q42_SQ426						
Q43_SQ431	Q43					
Q43_SQ432						
Q43_SQ433						
Q43_SQ434						
Q43_SQ435						
Q43_SQ436						
Q43_SQ437	Q44					
Q44_SQ441						
Q44_SQ442						
Q44_SQ443						
Q44_SQ444						
Q44_SQ445						
Q45	Q45	Oefening 8 / Alinea- opbouw	Alinea- opbouw	Tekst- structuur	Oefening 8	Tekst- structuur
Q46	Q46					
Q47	Q47					
Q48	Q48	Oefening 9 / Alinea- grenzen	Alinea- grenzen		Oefening 9	
Q49	Q49					
Q50	Q50					
Q51	Q51	Oefening 10 / Lina's eigen koers	Lezen	Begrijpend lezen	Oefening 10	Begrijpend lezen
Q52	Q52					
Q53	Q53					
Q54	Q54					
Q55	Q55					
Q56	Q56					
Q57	Q57					
Q58	Q58					
Q59	Q59					
Q60	Q60					
Q61	Q61					
Q62	Q62					

Appendix 5

Scoring Protocol for the Dutch Language Test

Verbetering per vraagtype

Alle items worden binair gescoord: 1 = correct, en 0 = fout.

Dictee

Oogmerk: Een punt word toegekend voor ieder beoordeeld woord dat exact correct gespeld is, ongeacht of de volledige zin is ingevuld. Als er bij geen enkele zin iets is ingevuld, zijn alle items = missing.

Uitvoering: R-script Scoring.dictee toepassen op ruwe antwoorden in Excel, waarbij dicteevragen op zinsniveau worden gesplitst in subvragen op woordniveau; nadien handmatig 0-waarden verwijderen voor cases met drie maal “#N/B”

Voorzetsels

Oogmerk: Een punt wordt in principe toegekend als het juiste voorzetsel is ingevuld en verder niets; met hoofdletters wordt geen rekening gehouden. Bij meetmoment 1 wordt het antwoord ook goed gerekend als de volledige zin met het correcte voorzetsel is overgenomen (wegens onvoldoende expliciete instructie dat het niet de bedoeling was de zin over te schrijven). Een ontbrekend antwoord = 0 wanneer de rest van de LimeSurvey-pagina is gemaakt, en = missing wanneer de volledige LimeSurvey-pagina is overgeslagen.

Opmerking: Tijdens meetmomenten 1 en 2 stond er één oefening oftewel oorspronkelijke subtest per pagina. Tijdens meetmoment 3 werden per pagina alle items van een taalaspect gepresenteerd; enkel voor het aspect spelling werd de onderverdeling tussen werkwoordspelling enerzijds en zinspelling oftewel dictee anderzijds wel bewaard.

Uitvoering: voor meetmoment 1 scoring met de hand in SPSS, zelf rekening houdend met de voorwaarde dat minstens één item van die oefening moet zijn beantwoord; voor meetmomenten 2 en 3 hercodering in SPSS met specificering van een voorwaarde

Overige invulvragen

Woordtransformaties en Werkwoordspelling nemen eveneens de vorm aan van invulvragen.

Oogmerk: Een punt wordt toegekend als het juiste woord is ingevuld en verder niets; met hoofdletters wordt geen rekening gehouden. Een ontbrekend antwoord = 0 wanneer de rest van de LimeSurvey-pagina is gemaakt, en = missing wanneer de volledige LimeSurvey-pagina is overgeslagen.

Uitvoering: hercodering in SPSS met specificering van een voorwaarde

Meerkeuzevragen

Woordenschat, Tekststructuur en Begrijpend lezen nemen de vorm aan van meerkeuzevragen.

Oogmerk: Een punt wordt toegekend als de juiste antwoordoptie is aangeduid. Een ontbrekend antwoord = 0 wanneer de rest van de LimeSurvey-pagina is gemaakt, en = missing wanneer de volledige LimeSurvey-pagina is overgeslagen.

Uitvoering: hercodering in SPSS met specificering van een voorwaarde

Overzicht hercoderingen

Oefening	Item(s)	SPSS
2	Q5, Q6	'A' → 1 ELSE → 0
	Q1, Q9, Q10, Q11	'B' → 1 ELSE → 0
	Q2, Q4, Q7, Q8, Q12, Q13, Q14	'C' → 1 ELSE → 0
	Q3, Q15	'D' → 1 ELSE → 0
3	Q16, Q19	'A' → 1 ELSE → 0
	Q17	'B' → 1 ELSE → 0
	Q18, Q20	'D' → 1 ELSE → 0
4	Q21, Q22, Q24	'aan' → 1 'Aan' → 1 ELSE → 0
	Q23	'in' → 1 'In' → 1 ELSE → 0
	Q25	'op' → 1 'Op' → 1 ELSE → 0
5	Q26	'legitiem' → 1 'Legitiem' → 1 ELSE → 0
	Q27	'gedoceerd' → 1 'Gedoceerd' → 1 ELSE → 0

6	Q28_SQ281	‘descriptief’ → 1 ‘Descriptief’ → 1 ELSE → 0
	Q28_SQ282	‘functioneren’ → 1 ‘Functioneren’ → 1 ELSE → 0
	Q29_SQ291	‘mutaties’ → 1 ‘Mutaties’ → 1 ELSE → 0
	Q29_SQ292	‘implicaties’ → 1 ‘Implicaties’ → 1 ELSE → 0
7	Q30_SQ301	‘woord’ → 1 ‘Wordt’ → 1 ELSE → 0
	Q30_SQ302	‘verbreed’ → 1 ‘Verbreed’ → 1 ELSE → 0
	Q30_SQ303	‘geasfalteerd’ → 1 ‘Geasfalteerd’ → 1 ELSE → 0
	Q31_SQ311	‘gespoeld’ → 1 ‘Gespoeld’ → 1 ELSE → 0
	Q32_SQ321	‘gestimuleerd’ → 1 ‘Gestimuleerd’ → 1 ELSE → 0
	Q33_SQ331	‘versierd’ → 1 ‘Versierd’ → 1 ELSE → 0
	Q34_SQ341	‘veranderd’ → 1 ‘Veranderd’ → 1 ELSE → 0
	Q34_SQ342	‘gelost’ → 1 ‘Gelost’ → 1 ELSE → 0
	Q35_SQ351	‘verdubbeld’ → 1 ‘Verdubbeld’ → 1 ELSE → 0

	Q35_SQ352	‘verveelvoudigt’ $\rightarrow$ 1 ‘Verveelvoudigt’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q36_SQ361	‘betaald’ $\rightarrow$ 1 ‘Betaald’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q37_SQ371	‘oplevert’ $\rightarrow$ 1 ‘Oplevert’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q38_SQ381	‘verdiend’ $\rightarrow$ 1 ‘Verdiend’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q39_SQ391	‘gelande’ $\rightarrow$ 1 ‘Gelande’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q40_SQ401	‘verluidt’ $\rightarrow$ 1 ‘Verluidt’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q40_SQ402	‘behandeld’ $\rightarrow$ 1 ‘Behandeld’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q41_SQ411	‘gehate’ $\rightarrow$ 1 ‘Gehate’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
8	Q46	‘A’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q47	‘C’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q45	‘D’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
9	Q49, Q50	‘B’ $\rightarrow$ 1 ELSE $\rightarrow$ 0
	Q48	‘C’ $\rightarrow$ 1 ELSE $\rightarrow$ 0

10	Q60	'A' → 1 ELSE → 0
	Q51, Q53, Q54, Q55	'B' → 1 ELSE → 0
	Q52, Q58, Q59, Q61	'C' → 1 ELSE → 0
	Q56, Q57, Q62	'D' → 1 ELSE → 0

Overzicht voorwaarden

Oefening	T	SPSS
2 (Q1-15)	1&2	NOT (Q1='' & Q2='' & Q3='' & Q4='' & Q5='' & Q6='' & Q7='' & Q8='' & Q9='' & Q10='' & Q11='' & Q12='' & Q13='' & Q14='' & Q15='')
	3	NOT (Q1='' & Q2='' & Q3='' & Q4='' & Q5='' & Q6='' & Q7='' & Q8='' & Q9='' & Q10='' & Q11='' & Q12='' & Q13='' & Q14='' & Q15='' & Q16='' & Q17='' & Q18='' & Q19='' & Q20='')
3 (Q16-20)	1&2	NOT (Q16='' & Q17='' & Q18='' & Q19='' & Q20='')
	3	NOT (Q1='' & Q2='' & Q3='' & Q4='' & Q5='' & Q6='' & Q7='' & Q8='' & Q9='' & Q10='' & Q11='' & Q12='' & Q13='' & Q14='' & Q15='' & Q16='' & Q17='' & Q18='' & Q19='' & Q20='')
4 (Q21-25)	1	scoring met de hand indien Q21 t.e.m. Q25 niet allemaal leeg zijn
	2	NOT (Q21='' & Q22='' & Q23='' & Q24='' & Q25='')
	3	NOT (Q21='' & Q22='' & Q23='' & Q24='' & Q25='' & Q26='' & Q27='' & Q28_SQ281='' & Q28_SQ282='' & Q29_SQ291='' & Q29_SQ292='')
5 (Q26-27)	1&2	NOT (Q26='' & Q27='')
	3	NOT (Q21='' & Q22='' & Q23='' & Q24='' & Q25='' & Q26='' & Q27='' & Q28_SQ281='' & Q28_SQ282='' & Q29_SQ291='' & Q29_SQ292='')

6 (Q28-29)	1&2	NOT (Q28_SQ281='' & Q28_SQ282='' & Q29_SQ291='' & Q29_SQ292='')
	3	NOT (Q21='' & Q22='' & Q23='' & Q24='' & Q25='' & Q26='' & Q27='' & Q28_SQ281='' & Q28_SQ282='' & Q29_SQ291='' & Q29_SQ292='')
7 (Q30-41)	1, 2, 3	NOT (Q30_SQ301='' & Q30_SQ302='' & Q30_SQ303='' & Q31_SQ311='' & Q32_SQ321='' & Q33_SQ331='' & Q34_SQ341='' & Q34_SQ342='' & Q35_SQ351='' & Q35_SQ352='' & Q36_SQ361='' & Q37_SQ371='' & Q38_SQ381='' & Q39_SQ391='' & Q40_SQ401='' & Q40_SQ402='' & Q41_SQ411='')
8 (Q45-47)	1&2	NOT (Q45='' & Q46='' & Q47='')
	3	NOT (Q45='' & Q46='' & Q47='' & Q48='' & Q49='' & Q50='')
9 (Q48-50)	1&2	NOT (Q48='' & Q49='' & Q50='')
	3	NOT (Q45='' & Q46='' & Q47='' & Q48='' & Q49='' & Q50='')
10 (Q51-62)	1,2,3	NOT (Q51='' & Q52='' & Q53='' & Q54='' & Q55='' & Q56='' & Q57='' & Q58='' & Q59='' & Q60='' & Q61='' & Q62='')

Appendix 6

Factor Loadings of All Dutch Language Test Items

T1

	F1	F2	F3	h2
Q1value	0.15754	0.021947	0.26130	0.13453
Q2value	0.38225	-0.043347	0.22604	0.26513
Q3value	0.52268	-0.024360	0.08089	0.31238
Q4value	0.19387	-0.000251	0.06783	0.05390
Q5value	0.51454	-0.083827	-0.08157	0.22856
Q6value	0.40168	-0.011310	0.16204	0.24324
Q7value	0.28095	-0.071890	0.06583	0.09509
Q8value	-0.07149	0.013189	0.02334	0.00415
Q9value	0.35173	-0.027475	-0.00787	0.11861
Q10value	0.19473	0.041560	0.01976	0.04687
Q11value	0.41907	-0.031216	0.05243	0.19330
Q12value	0.22210	-0.012486	-0.01174	0.04635
Q13value	0.11553	-0.020517	0.08944	0.02918
Q14value	-0.01047	-0.046125	-0.01583	0.00319
Q15value	0.11749	0.232852	-0.16913	0.06890
Q16value	0.38994	0.009240	0.09680	0.19699
Q17value	0.23783	-0.039564	0.07015	0.07306
Q18value	0.58472	0.056626	-0.00208	0.35616
Q19value	0.25077	-0.055732	-0.00827	0.05929
Q20value	0.19276	0.046357	-0.07335	0.03362
Q21value	0.32230	-0.037214	0.04203	0.11391
Q22value	0.34818	0.069992	0.00280	0.13608
Q23value	0.43804	-0.011193	0.05948	0.21666
Q24value	-0.05581	-0.105474	0.19611	0.03455
Q25value	0.52907	-0.093602	-0.09978	0.23802
Q26value	0.42402	-0.100810	0.15960	0.25190
Q27value	0.47782	0.200168	0.21064	0.45932
Q28_sq281value	0.20497	0.082301	0.30407	0.21586
Q28_sq282value	0.45031	0.025574	0.05268	0.23231
Q29_sq291value	0.27980	0.111369	0.12641	0.15690
Q29_sq292value	0.36815	0.034368	0.12618	0.20098
Q30_sq301value	0.11409	-0.046271	0.27217	0.10860
Q30_sq302value	0.14632	0.323384	0.07567	0.17150
Q30_sq303value	0.27007	0.334500	0.15022	0.30250

Q31_SQ311value	0.08470	0.485208	0.13591	0.32010
Q32_SQ321value	0.13808	0.551058	0.18345	0.45854
Q33_SQ331value	0.02008	0.735099	-0.04127	0.53171
Q34_SQ341value	0.02005	0.720800	-0.09109	0.49849
Q34_SQ342value	0.21323	-0.044460	0.29461	0.18016
Q35_SQ351value	0.04018	0.742608	0.00754	0.56725
Q35_SQ352value	0.14204	-0.496781	0.16212	0.24677
Q36_SQ361value	-0.03631	0.807095	0.07641	0.67671
Q37_SQ371value	0.06428	-0.552589	0.21942	0.29535
Q38_SQ381value	0.01973	0.625102	0.08912	0.43358
Q39_SQ391value	0.30580	0.166421	-0.12314	0.11106
Q40_SQ401value	0.05579	-0.359531	0.10508	0.12201
Q40_SQ402value	-0.04862	0.781534	0.02438	0.60841
Q41_SQ411value	0.30519	0.111279	0.02520	0.12696
Q42_SQ421value	-0.14245	-0.033992	0.66685	0.37156
Q42_SQ422value	0.08886	-0.006318	0.50198	0.29792
Q42_SQ423value	-0.07492	-0.018344	0.18578	0.02681
Q42_SQ424value	0.14416	0.107622	0.33195	0.20920
Q42_SQ425value	-0.01804	0.106426	0.63257	0.43522
Q42_SQ426value	0.00661	0.018323	0.71141	0.51737
Q43_SQ431value	-0.10713	-0.010759	0.43645	0.15837
Q43_SQ432value	-0.07032	-0.050476	0.23900	0.04477
Q43_SQ433value	0.15168	-0.159432	0.31488	0.15573
Q43_SQ434value	-0.05789	-0.003080	0.44120	0.17459
Q43_SQ435value	0.20155	0.008650	0.43972	0.31575
Q43_SQ436value	0.31750	0.018087	0.28842	0.27084
Q43_SQ437value	0.08327	0.013259	0.47208	0.26867
Q44_SQ441value	-0.07997	0.075193	0.64738	0.40753
Q44_SQ442value	0.15981	-0.076344	0.47688	0.30378
Q44_SQ443value	0.02642	-0.029858	0.33735	0.11793
Q44_SQ444value	0.07868	-0.002475	0.68122	0.51718
Q44_SQ445value	0.13098	0.058046	0.49447	0.34027
Q45value	-0.00366	0.058454	-0.06598	0.00595
Q46value	0.12061	0.037574	-0.08608	0.01412
Q47value	0.19336	0.080223	0.16952	0.11447
Q48value	0.23288	0.051464	0.14557	0.11657
Q49value	0.28941	0.067987	-0.04838	0.08379
Q50value	-0.01717	0.099699	-0.09925	0.01593
Q51value	0.62840	-0.024333	-0.17266	0.32494
Q52value	0.40593	0.041579	0.08987	0.21527
Q53value	0.31379	0.067968	0.01201	0.11487
Q54value	0.08798	0.047325	0.09250	0.02957
Q55value	0.31434	0.009843	0.09260	0.13507
Q56value	0.35293	0.048519	0.14717	0.20489
Q57value	0.21029	0.061962	-0.02608	0.04782
Q58value	0.47749	0.016927	-0.08039	0.20276
Q59value	0.43172	0.055793	0.07879	0.23718
Q60value	0.53990	0.030993	-0.11675	0.25413
Q61value	0.23267	-0.009829	0.02304	0.05859
Q62value	0.39859	0.024959	0.05088	0.18450

T2

	F1	F2	F3	h2
q1value	0.6731	0.02655	-0.09918	0.41538
q2value	0.6106	0.13512	0.15807	0.54045
q3value	0.4764	-0.08009	0.11801	0.27199
q4value	0.3268	-0.02105	0.09965	0.13961
q5value	0.6697	0.01435	-0.26577	0.37662
q6value	0.5296	-0.00602	0.02873	0.29226
q7value	0.3599	-0.08825	0.13837	0.17713
q8value	0.0928	0.02674	-0.02782	0.00864
q9value	0.3983	0.02409	0.00779	0.16593
q10value	0.4604	-0.03641	-0.14343	0.17592
q11value	0.5187	0.00919	0.05615	0.29821
q12value	0.2968	-0.04244	0.00626	0.08602
q13value	0.2156	-0.01849	0.05159	0.05635
q14value	0.1773	-0.01520	-0.10029	0.02695
q15value	0.2919	0.08075	-0.05828	0.08871
q16value	0.5304	0.02976	0.08362	0.33325
q17value	0.3346	0.07435	-0.11130	0.10566
q18value	0.5530	0.03486	0.01701	0.32342
q19value	0.4651	0.02110	-0.15967	0.18419
q20value	0.2072	-0.02244	0.08190	0.06103
q21value	0.3500	-0.03586	-0.07480	0.10427
q22value	0.4001	-0.05968	0.20619	0.25666
q23value	0.5806	0.05013	-0.04751	0.33037
q24value	0.1160	0.05724	-0.18433	0.03052
q25value	0.4331	-0.17000	-0.01142	0.18272
q26value	0.3335	0.00213	0.22040	0.22010
q27value	0.4680	0.28088	0.21623	0.51432
q28_sq281value	0.3384	0.10312	0.05086	0.15914
q28_sq282value	0.3967	0.07617	0.14659	0.25055
q29_sq291value	0.4965	-0.05372	-0.02721	0.22869
q29_sq292value	0.3552	-0.01275	0.20974	0.22755
q30_sq301value	0.1940	-0.13442	0.35195	0.19902
q30_sq302value	0.1266	0.56592	0.02811	0.37849
q30_sq303value	0.3142	0.34913	0.10228	0.32220
q31_sq311value	0.1327	0.57611	-0.00313	0.38032
q32_sq321value	0.1337	0.58933	0.16285	0.49338
q33_sq331value	0.0278	0.79765	-0.00346	0.64481
q34_sq341value	-0.0990	0.66962	0.08829	0.46243
q34_sq342value	0.2312	0.03095	0.27813	0.19164
q35_sq351value	0.0249	0.74752	0.08784	0.61158
q35_sq352value	0.1596	-0.50249	0.19714	0.25617
q36_sq361value	0.0656	0.71616	0.04074	0.55623
q37_sq371value	0.0838	-0.59686	0.37874	0.39157
q38_sq381value	0.0200	0.65497	0.03658	0.44954
q39_sq391value	0.2307	0.34303	-0.08118	0.18069
q40_sq401value	0.2507	-0.31820	0.09495	0.14296
q40_sq402value	-0.0188	0.80975	0.06757	0.68222
q41_sq411value	0.3238	0.04091	0.00438	0.11334
q42_sq421value	-0.0394	-0.17038	0.44277	0.17523
q42_sq422value	0.1318	0.00819	0.47526	0.29672
q42_sq423value	0.0585	0.08594	0.01314	0.01432
q42_sq424value	0.0846	-0.03606	0.06864	0.01530
q42_sq425value	0.0106	0.00931	0.31001	0.10056
q42_sq426value	0.1866	-0.16544	0.54037	0.37580
q43_sq431value	0.0789	0.00962	0.40931	0.20252
q43_sq432value	0.0782	-0.11589	0.22819	0.06829
q43_sq433value	0.0645	-0.07627	0.52990	0.29502

Q43_SQ434value	-0.0419	0.03096	0.42986	0.17938
Q43_SQ435value	0.2142	-0.00781	0.28146	0.17229
Q43_SQ436value	0.3286	-0.04169	0.33973	0.30264
Q43_SQ437value	0.0596	0.06837	0.38103	0.18742
Q44_SQ441value	0.0310	0.11816	0.57159	0.39341
Q44_SQ442value	0.3043	0.13301	0.06409	0.15175
Q44_SQ443value	0.0236	0.20258	0.17212	0.09504
Q44_SQ444value	-0.0153	0.22570	0.69316	0.60456
Q44_SQ445value	-0.0520	0.18478	0.72020	0.59165
Q45value	0.0301	-0.04503	0.12941	0.01919
Q46value	0.0981	0.00562	0.08686	0.02461
Q47value	0.4056	0.08267	0.06559	0.21421
Q48value	0.0261	0.14575	0.22951	0.09881
Q49value	0.4355	0.09049	-0.07741	0.18924
Q50value	0.1705	-0.01191	-0.09673	0.02491
Q51value	0.4043	-0.09350	0.12937	0.20920
Q52value	0.4103	-0.08474	0.15845	0.23177
Q53value	0.2939	-0.06982	0.25912	0.20209
Q54value	0.1170	-0.02699	0.12643	0.03929
Q55value	0.2353	0.06386	0.13522	0.11448
Q56value	0.3645	-0.00156	0.14719	0.19779
Q57value	0.0789	-0.01182	0.11560	0.02603
Q58value	0.2955	-0.00980	0.06133	0.10438
Q59value	0.3889	0.07312	0.10810	0.21860
Q60value	0.4255	-0.05624	0.03905	0.18801
Q61value	0.2874	-0.05907	-0.00122	0.07870
Q62value	0.3386	0.08025	0.11462	0.18208

T3

	F1	F2	F3	h2
Q1value	0.46593	0.03822	0.33629	0.4752
Q2value	0.42982	0.16427	0.17962	0.3671
Q3value	0.66066	-0.11148	0.12833	0.4797
Q4value	0.22838	-0.08446	0.30764	0.1798
Q5value	0.68595	0.00472	-0.15245	0.4123
Q6value	0.63294	0.03832	-0.00989	0.4110
Q7value	0.49458	-0.11712	-0.03436	0.2155
Q8value	0.30840	0.03563	-0.09599	0.0860
Q9value	0.47991	0.02150	0.05411	0.2611
Q10value	0.29120	-0.04812	-0.01530	0.0763
Q11value	0.58359	0.03550	-0.14757	0.3036
Q12value	0.48458	-0.06345	0.04040	0.2364
Q13value	0.19166	0.05469	0.11438	0.0807
Q14value	0.17328	0.09201	-0.33149	0.0902
Q15value	0.38985	0.42043	-0.00355	0.4212
Q16value	0.51731	0.00393	0.05298	0.2935
Q17value	0.40538	0.08982	-0.11379	0.1625
Q18value	0.62569	0.03559	-0.02716	0.3922
Q19value	0.45483	0.14411	-0.07989	0.2348
Q20value	0.25911	-0.06040	0.15505	0.1109
Q21value	0.24030	0.05549	0.06019	0.0860
Q22value	0.39606	0.07539	0.10231	0.2279
Q23value	0.56562	0.04066	-0.01401	0.3284
Q24value	-0.06082	0.12029	-0.00395	0.0138
Q25value	0.18204	0.09355	-0.10265	0.0406

Q26value	0.59347	-0.05448	-0.03060	0.3242
Q27value	0.42415	0.20550	0.22849	0.4351
Q28_SQ281value	0.34017	-0.02579	0.18417	0.1914
Q28_SQ282value	0.61810	0.01433	0.00508	0.3899
Q29_SQ291value	0.54565	0.05776	-0.09842	0.2823
Q29_SQ292value	0.35046	0.09137	0.17304	0.2390
Q30_SQ301value	-0.07945	-0.01069	0.59300	0.3167
Q30_SQ302value	0.09575	0.44044	0.06448	0.2569
Q30_SQ303value	0.22752	0.30396	0.01149	0.1888
Q31_SQ311value	0.07698	0.62327	0.14993	0.5208
Q32_SQ321value	-0.02128	0.62148	0.14317	0.4610
Q33_SQ331value	0.08045	0.70506	-0.03090	0.5197
Q34_SQ341value	-0.06228	0.80659	-0.04509	0.6036
Q34_SQ342value	0.18550	0.05407	0.32029	0.2052
Q35_SQ351value	0.08148	0.77221	-0.00389	0.6369
Q35_SQ352value	0.14067	-0.35238	0.37357	0.2021
Q36_SQ361value	-0.01667	0.64429	0.00274	0.4104
Q37_SQ371value	0.02153	-0.33132	0.33594	0.1448
Q38_SQ381value	0.04592	0.59434	0.19007	0.4952
Q39_SQ391value	0.23983	0.26828	-0.10596	0.1374
Q40_SQ401value	0.12173	-0.17745	0.13005	0.0467
Q40_SQ402value	-0.08060	0.73586	0.05785	0.5439
Q41_SQ411value	0.26680	0.14670	0.05700	0.1366
Q42_SQ421value	0.05702	-0.01316	0.39840	0.1759
Q42_SQ422value	-0.05974	0.11950	0.53681	0.3225
Q42_SQ423value	-0.02833	0.17823	0.23351	0.1088
Q42_SQ424value	-0.05860	-0.04050	0.15839	0.0196
Q42_SQ425value	0.04181	0.13364	0.67177	0.5608
Q42_SQ426value	-0.08365	0.11404	0.61001	0.3961
Q43_SQ431value	0.05090	0.00324	0.38195	0.1649
Q43_SQ432value	0.08476	0.02059	0.33778	0.1504
Q43_SQ433value	0.09960	-0.12012	0.61852	0.3955
Q43_SQ434value	0.00997	-0.06161	0.49471	0.2303
Q43_SQ435value	0.13172	0.15107	0.27665	0.1870
Q43_SQ436value	0.41108	-0.04434	0.31026	0.3478
Q43_SQ437value	0.17759	-0.07867	0.44577	0.2659
Q44_SQ441value	-0.04991	0.11779	0.65172	0.4670
Q44_SQ442value	0.22350	-0.03093	0.28315	0.1709
Q44_SQ443value	0.04158	0.10956	0.16618	0.0625
Q44_SQ444value	0.05680	0.24841	0.45082	0.3770
Q44_SQ445value	0.08109	0.20064	0.60437	0.5473
Q45value	0.15191	-0.06713	0.14677	0.0538
Q46value	0.25308	-0.07150	0.02264	0.0626
Q47value	0.29927	0.01146	0.16619	0.1600
Q48value	0.33228	0.07134	0.14072	0.1932
Q49value	0.33304	0.14948	-0.08372	0.1380
Q50value	0.09679	-0.02488	0.07865	0.0194
Q51value	0.37010	-0.11534	0.21165	0.2149
Q52value	0.55806	0.02280	-0.00313	0.3179
Q53value	0.33829	0.03333	-0.18722	0.1025
Q54value	0.23789	0.04177	0.00784	0.0658
Q55value	0.30790	0.05464	0.12302	0.1575
Q56value	0.46807	-0.03853	0.23028	0.3422
Q57value	0.11314	-0.09490	0.12111	0.0329
Q58value	0.39577	-0.03871	0.01204	0.1529
Q59value	0.33250	0.04811	0.05575	0.1418
Q60value	0.46863	-0.01777	-0.05330	0.1989
Q61value	0.12623	0.09851	0.05455	0.0451
Q62value	0.38757	0.07155	0.13913	0.2405

Appendix 7

Factor Structure of the Revised Dutch Language Test

T1

Rotation: oblimin

Rotated factor loadings:

	F1	F2	F3	h2
Q2value1	0.391		0.237	0.2779
Q3value1	0.528			0.3151
Q4value1	0.192			0.0537
Q5value1	0.500			0.2192
Q6value1	0.403			0.2426
Q7value1	0.285			0.1001
Q9value1	0.369			0.1263
Q10value1				0.0451
Q11value1	0.430			0.1955
Q12value1	0.207			0.0427
Q15value1		0.240		0.0683
Q16value1	0.391			0.1947
Q17value1	0.221			0.0674
Q18value1	0.585			0.3598
Q19value1	0.254			0.0579
Q21value1	0.323			0.1143
Q22value1	0.324			0.1258
Q23value1	0.466			0.2338
Q25value1	0.543			0.2453
Q26value1	0.443			0.2495
Q27value1	0.462	0.252		0.4707
Q28_SQ281value1	0.193		0.302	0.2129
Q28_SQ282value1	0.461			0.2475
Q29_SQ291value1	0.278			0.1613
Q29_SQ292value1	0.378			0.2033
Q30_SQ302value1		0.318		0.1660
Q30_SQ303value1	0.269	0.344		0.3062
Q31_SQ311value1		0.512		0.3369
Q32_SQ321value1		0.563		0.4600
Q33_SQ331value1		0.752		0.5401

Q34_SQ341value1	0.738	0.5076
Q35_SQ351value1	0.744	0.5606
Q36_SQ361value1	0.836	0.7004
Q38_SQ381value1	0.640	0.4390
Q40_SQ402value1	0.789	0.6071
Q42_SQ421value1		0.669 0.3763
Q42_SQ422value1		0.477 0.2751
Q42_SQ425value1		0.623 0.4272
Q42_SQ426value1		0.727 0.5274
Q43_SQ431value1		0.456 0.1709
Q43_SQ433value1		0.302 0.1412
Q43_SQ434value1		0.475 0.1953
Q43_SQ435value1	0.207	0.445 0.3213
Q43_SQ436value1	0.325	0.280 0.2701
Q43_SQ437value1		0.481 0.2773
Q44_SQ441value1		0.590 0.3701
Q44_SQ442value1		0.454 0.2909
Q44_SQ444value1		0.703 0.5315
Q44_SQ445value1		0.508 0.3420
Q47value1		0.1149
Q49value1	0.288	0.0834
Q51value1	0.626	0.3213
Q52value1	0.426	0.2228
Q53value1	0.331	0.1198
Q55value1	0.307	0.1305
Q56value1	0.347	0.1957
Q58value1	0.481	0.2034
Q59value1	0.426	0.2319
Q60value1	0.503	0.2411
Q62value1	0.409	0.1715

Rotated SS loadings: 5.812 4.47 4.472

Factor correlations:

	F1	F2	F3
F1	1.000		
F2	0.211	1.000	
F3	0.450	0.301	1

T2

Rotation: oblimin

Rotated factor loadings:

	F1	F2	F3	h2
Q2value2	0.683			0.5602
Q3value2	0.550			0.3053
Q4value2	0.380			0.1582
Q5value2	0.636	-0.249		0.3513
Q6value2	0.548			0.2988
Q7value2	0.391			0.1712
Q9value2	0.406			0.1765
Q10value2	0.489	-0.211		0.2069
Q11value2	0.554			0.3118
Q12value2	0.298			0.0906
Q15value2	0.268			0.0866

Q16value2	0.596		0.3636
Q17value2	0.300		0.0935
Q18value2	0.530		0.3022
Q19value2	0.460		0.1933
Q21value2	0.312		0.0869
Q22value2	0.424		0.2454
Q23value2	0.574		0.3298
Q25value2	0.463		0.1875
Q26value2	0.385	0.216	0.2474
Q27value2	0.483	0.303	0.5205
Q28_SQ281value2	0.332		0.1595
Q28_SQ282value2	0.430		0.2618
Q29_SQ291value2	0.492		0.2274
Q29_SQ292value2	0.386		0.2260
Q30_SQ302value2		0.563	0.3712
Q30_SQ303value2	0.325	0.330	0.3269
Q31_SQ311value2		0.590	0.3844
Q32_SQ321value2		0.600	0.5176
Q33_SQ331value2		0.808	0.6491
Q34_SQ341value2		0.714	0.4852
Q35_SQ351value2		0.794	0.6425
Q36_SQ361value2		0.733	0.5625
Q38_SQ381value2		0.724	0.4967
Q40_SQ402value2		0.795	0.6521
Q42_SQ421value2		0.286	0.0872
Q42_SQ422value2	0.252	0.357	0.2562
Q42_SQ425value2		0.258	0.0811
Q42_SQ426value2	0.295	0.419	0.3004
Q43_SQ431value2		0.326	0.1643
Q43_SQ433value2		0.465	0.2541
Q43_SQ434value2		0.325	0.1267
Q43_SQ435value2	0.257	0.219	0.1450
Q43_SQ436value2	0.397	0.256	0.2776
Q43_SQ437value2		0.294	0.1466
Q44_SQ441value2		0.442	0.3346
Q44_SQ442value2	0.352		0.1659
Q44_SQ444value2		0.842	0.7875
Q44_SQ445value2		0.894	0.8015
Q47value2	0.416		0.2137
Q49value2	0.413		0.1810
Q51value2	0.428		0.2013
Q52value2	0.446		0.2349
Q53value2	0.331	0.244	0.2026
Q55value2	0.263		0.1068
Q56value2	0.419		0.1991
Q58value2	0.325		0.1124
Q59value2	0.385		0.2127
Q60value2	0.435		0.1846
Q62value2	0.321		0.1861

Rotated SS loadings: 7.846 3.305 5.031

Factor correlations:

	F1	F2	F3
F1	1.000		
F2	0.374	1.000	
F3	0.267	0.385	1

T3

Rotation: oblimin

Rotated factor loadings:

	F1	F2	F3	h2
Q2value3	0.330		0.277	0.3384
Q3value3	0.645			0.4785
Q4value3	0.246		0.272	0.1691
Q5value3	0.679			0.4013
Q6value3	0.650			0.4246
Q7value3	0.484			0.2282
Q9value3	0.449			0.2456
Q10value3	0.251			0.0644
Q11value3	0.554			0.2818
Q12value3	0.471			0.2366
Q15value3	0.359	0.340		0.3629
Q16value3	0.502			0.2724
Q17value3	0.409			0.1754
Q18value3	0.666			0.4485
Q19value3	0.491			0.2692
Q21value3	0.246			0.0842
Q22value3	0.415			0.2383
Q23value3	0.542			0.3170
Q25value3				0.0255
Q26value3	0.604			0.3349
Q27value3	0.405	0.199	0.280	0.4584
Q28_SQ281value3	0.327			0.1819
Q28_SQ282value3	0.607			0.4185
Q29_SQ291value3	0.540			0.2913
Q29_SQ292value3	0.346		0.219	0.2626
Q30_SQ302value3		0.407		0.2367
Q30_SQ303value3	0.212	0.258		0.1801
Q31_SQ311value3		0.580	0.205	0.5001
Q32_SQ321value3		0.580	0.263	0.4815
Q33_SQ331value3		0.654		0.4774
Q34_SQ341value3		0.895		0.7184
Q35_SQ351value3		0.768		0.6489
Q36_SQ361value3		0.601		0.3832
Q38_SQ381value3		0.609		0.4987
Q40_SQ402value3		0.761		0.5913
Q42_SQ421value3			0.392	0.1686
Q42_SQ422value3			0.596	0.3540
Q42_SQ425value3			0.601	0.5054
Q42_SQ426value3			0.588	0.3631
Q43_SQ431value3			0.413	0.1732
Q43_SQ433value3			0.631	0.3773
Q43_SQ434value3			0.516	0.2481
Q43_SQ435value3			0.319	0.2065
Q43_SQ436value3	0.421		0.402	0.4196
Q43_SQ437value3	0.193		0.444	0.2788
Q44_SQ441value3			0.677	0.4836
Q44_SQ442value3	0.213		0.338	0.1998
Q44_SQ444value3		0.231	0.470	0.3806
Q44_SQ445value3		0.208	0.582	0.5145
Q47value3	0.297			0.1333
Q49value3	0.310			0.1310
Q51value3	0.383			0.1981
Q52value3	0.537			0.3158
Q53value3	0.271			0.0602

Q55value3	0.305		0.1624
Q56value3	0.442	0.239	0.3237
Q58value3	0.384		0.1448
Q59value3	0.330		0.1409
Q60value3	0.495		0.2158
Q62value3	0.451		0.2705

Rotated SS loadings: 7.788 4.631 4.449

Factor correlations:

	F1	F2	F3
F1	1.000		
F2	0.278	1.000	
F3	0.402	0.358	1

Appendix 8

Difficulty Levels of the Revised Dutch Language Test

Lexical scale

Item	Mean score		
	T1	T2	T3
Q2value	0.86	0.94	0.95
Q3value	0.59	0.70	0.71
Q4value	0.45	0.53	0.59
Q5value	0.78	0.80	0.85
Q6value	0.59	0.67	0.74
Q7value	0.60	0.70	0.82
Q9value	0.55	0.65	0.71
Q10value	0.34	0.42	0.44
Q11value	0.37	0.51	0.62
Q12value	0.43	0.51	0.52
Q15value	0.07	0.08	0.12
Q16value	0.74	0.82	0.78
Q17value	0.46	0.57	0.73
Q18value	0.18	0.30	0.40
Q19value	0.24	0.27	0.46
Q21value	0.59	0.67	0.75
Q22value	0.62	0.74	0.84
Q23value	0.22	0.31	0.47
Q25value	0.04	0.07	0.11
Q26value	0.05	0.08	0.12
Q27value	0.09	0.20	0.23

Q28_SQ281value	0.16	0.41	0.47
Q28_SQ282value	0.11	0.18	0.18
Q29_SQ291value	0.09	0.14	0.22
Q29_SQ292value	0.29	0.40	0.51
Q47value	0.55	0.71	0.75
Q49value	0.53	0.61	0.68
Q51value	0.52	0.62	0.72
Q52value	0.69	0.76	0.83
Q53value	0.82	0.87	0.95
Q55value	0.44	0.55	0.65
Q56value	0.47	0.63	0.66
Q58value	0.60	0.67	0.75
Q59value	0.62	0.63	0.77
Q60value	0.50	0.58	0.63
Q62value	0.54	0.65	0.75

Verbal scale

Item	Mean score		
	T1	T2	T3
Q30_SQ302value	0.51	0.57	0.59
Q30_SQ303value	0.47	0.58	0.67
Q31_SQ311value	0.75	0.80	0.81
Q32_SQ321value	0.72	0.81	0.82
Q33_SQ331value	0.60	0.67	0.65
Q34_SQ341value	0.61	0.63	0.61
Q35_SQ351value	0.56	0.62	0.62
Q36_SQ361value	0.61	0.70	0.73
Q38_SQ381value	0.44	0.57	0.54
Q40_SQ402value	0.61	0.66	0.64

Orthographical scale

Item	Mean score		
	T1	T2	T3
Q42_SQ421value	0.85	0.90	0.93
Q42_SQ422value	0.59	0.77	0.82
Q42_SQ425value	0.86	0.95	0.96
Q42_SQ426value	0.71	0.83	0.88
Q43_SQ431value	0.61	0.72	0.74
Q43_SQ433value	0.67	0.76	0.76
Q43_SQ434value	0.59	0.69	0.78
Q43_SQ435value	0.55	0.70	0.78
Q43_SQ436value	0.54	0.69	0.79
Q43_SQ437value	0.79	0.92	0.93
Q44_SQ441value	0.43	0.65	0.73
Q44_SQ442value	0.82	0.88	0.88
Q44_SQ444value	0.17	0.45	0.53
Q44_SQ445value	0.40	0.62	0.72

Appendix 9

Sample of Ch-ICAR Items

Table 14. Specimen of Each Item Type of the Ch-ICAR (Loe et al., 2018; The International Cognitive Ability Resource Team, 2014)

Verbal Reasoning If the day after tomorrow is two days before Thursday then what day is it today? (1) Friday (2) Monday (3) Wednesday (4) Saturday (5) Tuesday (6) Sunday (7) None of these (8) I don't know	Figural Analogies
Matrix Reasoning	Number Series 34,150,35,160,36,170, ...

This table was taken from the Appendix of: Dutry, M., Vereeck, A., Duyck, W., Deraus, E., Schelfhout, S., Szmalec, A., Woumans, E., Schittekatte, M., Debeer, D., & Dirix, N. (2025). Validation of the Children’s International Cognitive Ability Resource (Ch-ICAR). *Behavior Research Methods*, 57(2), no. 66.

Appendix 10

Information Letter for Schools

INFORMATIEBRIEF VOOR DE SCHOOL (2021-2022)

Geachte mevrouw, meneer

Hartelijk dank voor uw medewerking aan mijn onderzoek. In dit document heb ik getracht alle relevante informatie te bundelen zodat de uitvoering van de studie zo vlot mogelijk kan verlopen. Allereerst vindt u onder A de algemene gegevens van de studie: de titel, de betrokken personen en instellingen. Vervolgens worden onder B en C het tijdspad en de metingen uiteengezet. Daarmee zijn de krijtlijnen van de studie getekend. Onder D is een beknopte verklaring opgenomen in verband met wetenschapsethiek en de inachtneming van de privacy van de deelnemers.

De volgende punten behandelen de praktische kant van de zaak. Punt E heeft betrekking op het inplannen van de tests. Punt F "Taken voor de school" biedt een hopelijk verhelderend overzicht van wat er concreet van de kant van de school wordt verwacht. Punt G bevat een belangrijke opmerking over hoe er tegenover leerlingen en ouders naar de studie dient te worden verwezen, lees deze opmerking dus zeker aandachtig. Onder H, ten slotte, worden mogelijke vergoedingen voorgesteld.

Ik verheug me op een vruchtbaar onderzoek in samenwerking met u.

Met vriendelijke groeten

Alexandra Vereeck

A. ALGEMENE GEGEVENS VAN DE STUDIE

- Titel: Non scholae sed vitae? Een empirische studie naar de cognitieve transfer effecten van het studeren van klassieke talen in het Vlaams secundair onderwijs
- Onderzoeker:

Dra. Alexandra Vereeck
FWO-aspirant fundamenteel onderzoek
Vakgroep Taalkunde UGent
alexandra.vereeck@ugent.be
- Promotoren:
 - prof. dr. Mark Janse, Vakgroep Taalkunde UGent;
 - prof. dr. Katja De Herdt, Vakgroep Letterkunde UGent;
 - prof. dr. Wouter Duyck, Vakgroep Experimentele Psychologie UGent.
- Betrokken onderzoeksinstituten:
 - Fonds Wetenschappelijk Onderzoek Vlaanderen (FWO);
 - Universiteit Gent.

B. TIJDPAD VAN DE STUDIE

De studie strekt zich uit over drie schooljaren, met één meetmoment per schooljaar. In het huidige schooljaar richt het onderzoek zich op de leerlingen van het **tweede middelbaar**; dit betreft zowel de leerlingen klassieke talen als de leerlingen in andere studierichtingen. Deze groep leerlingen zal daarna nog twee jaar worden opgevolgd, d.w.z. voor zover zij doorstromen naar het derde en vierde middelbaar op dezelfde school. Bij ieder meetmoment gebeuren telkens de onder C opgesomde metingen.

2021-2022	september: vragenlijst voor leerkrachten klassieke talen september: documentatie voor leerlingen en hun ouders oktober-november: metingen op school (2 ^e middelbaar)
2022-2023	tweede meetmoment (3 ^e middelbaar)
2023-2024	derde meetmoment (4 ^e middelbaar)

C. METINGEN PER MEETMOMENT

- **vragenlijst aanpak in de lessen klassieke talen**
Dit is de enige meting die niet plaatsvindt bij de leerlingen maar bij de leerkrachten. De leerkrachten klassieke talen die lesgeven in het deelnemende leerjaar wordt een korte vragenlijst voorgeschied met een aantal stellingen die betrekking hebben op hun lesaanpak.
- **vragenlijst consciëntieusheid en schoolmotivatie**
In de documentatie voor leerlingen en hun ouders zal de leerling worden doorverwezen naar een digitale vragenlijst die peilt naar consciëntieusheid (volharding en zelfcontrole) en schoolmotivatie. Het is de bedoeling dat de leerling deze vragenlijst thuis invult.
- **test Nederlandse taalvaardigheid**
Het eerste luik van de metingen die op school plaatsvinden, behelst Nederlandse taalvaardigheid. De afname van de test gebeurt digitaal; er is dus nood aan een computerlokaal of laptops. De test neemt **twee opeenvolgende lesuren** in beslag.
- **test algemeen-cognitieve vaardigheden**
Het tweede luik van de metingen die op school plaatsvinden, behelst algemeen-cognitieve vaardigheden (oftewel IQ). De afname van de test gebeurt digitaal; er is dus nood aan een computerlokaal of laptops. De test duurt **één lesuur**.
- **schoolresultaten**
Naast de scores op de tests zullen ook schoolresultaten in de analyse worden betrokken. Graag zou het onderzoeksteam van de deelnemende leerlingen de resultaten van het eerste trimester opvragen voor de volgende vakken:

- Nederlands;
- moderne vreemde talen;
- wiskunde;
- wetenschapsvakken.

D. VERKLARING BETREFFENDE ETHIEK EN PRIVACY

Het onderzoek wordt uitgevoerd in overeenstemming met de geaccepteerde maatstaven voor wetenschappelijk en ethisch gedrag. Voor de aanvang van de studie is bovendien het advies ingewonnen van een Ethische Commissie van de Universiteit Gent.

Van de deelnemende leerlingen zullen verschillende soorten persoonsgegevens worden verzameld en verwerkt. Hierbij wordt voldaan aan de richtlijnen rond *informed consent*: de ouder / voogd wordt eerst om nadrukkelijke toestemming gevraagd via het daartoe voorziene informatie- en toestemmingsformulier. Alle persoonsgegevens zullen nauwgezet worden beschermd en geanonimiseerd in overeenstemming met de nationale en Europese privacy-wetgeving. Voor meer details, zie het informatie- en toestemmingsformulier voor de ouder / voogd.

E. TESTPLANNING

De testplanning wordt opgesteld in gezamenlijk overleg tussen de onderzoeker en de school. Deze planning moet zowel efficiënt als flexibel zijn en uiteraard in overeenstemming met de agenda van beide partijen. Vermits dit logistiek geen eenvoudige operatie is, is enige duiding op zijn plaats.

Een vijftal scholen neemt deel aan de studie. Het afnemen van alle tests over de scholen heen betekent hoe dan ook meerdere weken fulltime veldwerk voor de onderzoeker, vandaar is een goede planning essentieel voor het slagen van het onderzoek. Hierbij dient rekening te worden gehouden met de volgende factoren (telkens toegepast op het tweede middelbaar):

- het aantal klassen;
- de openingsuren van de school en de exacte indeling van lesuren & pauzes;
- de ICT-voorzieningen op school, met name het inlogsysteem en de mogelijkheid tot het organiseren van paralleltests;
- eventuele dagen in oktober en november die niet beschikbaar zijn voor tests vanwege schoolactiviteiten.

Na het inwinnen van deze informatie zal de onderzoeker een **voorbeeldplanning** ter goedkeuring voorleggen. Deze voorbeeldplanning blijft beperkt tot een aanduiding van welke momenten geschikt zouden zijn voor het klaarzetten – hier dient apart tijd voor te worden ingecalculerd – en afnemen van de tests – taaltests in blokken van twee lesuur, algemeen-cognitieve tests in een enkel lesuur. Het staat de school volledig vrij om te beslissen welke

klasgroep op welke dag aan de beurt komt. Eenzelfde klas hoeft overigens niet de taaltest en de algemeen-cognitieve test op eenzelfde dag af te leggen.

F. TAKEN VOOR DE SCHOOL

- (1) Voor het tweede middelbaar bevestigt de school de **leerlingenaantallen** per studierichting en biedt ze een overzicht van het **curriculum** per studierichting. Deze gegevens zijn van belang voor het onderzoek.
- (2) De school verschaft de hoger genoemde informatie die noodzakelijk is voor de opstelling van de **testplanning**.
- (3) De **leerkrachten Latijn en Grieks** die in het tweede middelbaar lesgeven, vullen de **vragenlijst** in die hun wordt bezorgd. Het is mogelijk dat hun daarna wordt gevraagd om hun antwoorden mondeling toe te lichten.
- (4) De school ziet erop toe dat aan het begin van het schooljaar de **documentatie voor leerlingen en hun ouders** wordt uitgedeeld alsook tijdig terug opgehaald en bezorgt de ingevulde formulieren aan de onderzoeker. De documentatie bestaat uit een informatie- en toestemmingsformulier voor de ouder / voogd, een formulier om achtergrondgegevens van de leerling te verzamelen, en een begeleidende brief. Het achtergrondgegevens-formulier komt in twee verschillende versies: type A voor de leerlingen die geen Latijn doen en type B voor de leerlingen die wel Latijn doen.
- (5) De onderzoeker wordt ontvangen om op voorhand kennis te maken met de **ICT-voorzieningen** en de **plattegrond** van de school. Bij deze gelegenheid kunnen eveneens de ingevulde formulieren worden afgegeven (cf. F.4) en kan de attentie voor de leerkrachten in ontvangst worden genomen (cf. H).
- (6) De school roostert voor alle tweedejaars de **taaltests & algemeen-cognitieve tests** in op de overeengekomen momenten en licht de begeleidende leerkrachten (cf. F.8) in over het onderzoek.
- (7) De school voorziet **computers** voor de testafnames en stelt deze aan de onderzoeker ter beschikking voor het klaarzetten van de tests. De school voorziet eveneens **ICT-ondersteuning** in geval van problemen.
- (8) Tijdens de testafnames is er steeds een **leerkracht** aanwezig die het klasmanagement voor zijn/haar rekening neemt en toezicht houdt.
- (9) De school levert de opgevraagde **schoolresultaten** aan in een courant bestandsformaat.

G. VERWIJZEN NAAR DE STUDIE TEGENOVER LLN. EN OUDERS

In eerder mailverkeer tussen het onderzoeksteam en leerkrachten en/of directieleden is de inhoud en bedoeling van de studie in alle transparantie toegelicht. Ten aanzien van de deelnemende leerlingen en hun ouders is dergelijke transparantie echter niet wenselijk vanwege de kans op *stereotype threat* en *test anxiety*. Om ongewenste effecten voorgebracht

door stereotypes te vermijden, is het van belang om tegenover leerlingen en ouders niet in te gaan op de eigenlijke inhoud en bedoeling van de studie. In de documentatie voor leerlingen en hun ouders is het onderwerp van het onderzoek in algemene termen geherformuleerd als “de **cognitieve ontwikkeling van leerlingen** in het Vlaams secundair onderwijs”. Deze formulering kan indien nodig door leerkrachten worden overgenomen, bijvoorbeeld wanneer zij alsnog de vraag zouden krijgen waar de studie over gaat. Voorts is het aangewezen om tegenover leerlingen en ouders zo min mogelijk aandacht te schenken aan het plaatsvinden van de tests.

H. VERGOEDING

In ruil voor de medewerking van de school en haar leerkrachten & leerlingen stelt het onderzoeksteam de volgende vergoedingen voor.

- Onder de deelnemende leerlingen kunnen in samenspraak met de school jaarlijks enkele geschenkbonnen worden verloot.
- Het onderzoeksteam verbindt zich ertoe om jaarlijks een smakelijke attentie te voorzien voor in de leraarskamer.
- Promotor Mark Janse verklaart zich bereid om een lezing te geven, bijvoorbeeld in het kader van een Griekse dag.
- De school wordt op de hoogte gehouden van alle gepubliceerde onderzoeksbevindingen.
- De school kan tot en met 31 mei op aanvraag de scores van de leerlingen op de taaltests en algemeen-cognitieve tests van dat schooljaar verkrijgen bij de onderzoeker. Deze data mogen strikt en alleen voor onderwijsdoeleinden worden aangewend.

Appendix 11

Cover Letter for Pupils & Parents

**FACULTEIT LETTEREN
EN WIJSBEGEERTE**

VAKGROEP TAALKUNDE
ONDERZOEKSGROEP ΔIALING

Dra. Alexandra Vereeck
FWO-aspirant fundamenteel onderzoek

E alexandra.vereeck@ugent.be
M +32 479 448 440

Campus Boekentoren
Rozier 44
9000 Gent
Belgie

www.ugent.be
research.fw.ugent.be/en/alexandra.vereeck

Bijlagen:
Document 1: Informatie- en toestemmingsformulier voor de ouder / voogd (5 blz.)
Document 2: Achtergrondgegevens van de leerling (2 blz.)

DATUM
1 september 2021

PAGINA
1/1

Deelname wetenschappelijke studie

Beste leerling
Beste ouder / voogd

_____ werkt mee aan een wetenschappelijke studie naar de cognitieve ontwikkeling van leerlingen in het Vlaams secundair onderwijs, onder leiding van ondergetekende. Dergelijk onderzoek komt natuurlijk niet ver zonder deelnemende leerlingen. Daarom hoop ik van harte dat ik ook op jullie medewerking kan rekenen. Onder de deelnemers zullen bovendien mooie prijzen worden verloot.

Bij deze brief horen twee bijlagen. Gelieve beide documenten grondig door te nemen. **Document 1** bevat alle belangrijke informatie over de wetenschappelijke studie: dit is bedoeld voor de ouder / voogd om te lezen en bij goedkeuring de laatste pagina te **ondertekenen**.

Document 2 is reeds deel van het onderzoek en dient om achtergrondgegevens over de leerling te verzamelen. Het bestaat uit enerzijds een gedeelte dat de leerling zelf kan **invullen** onder begeleiding van een ouder / voogd en anderzijds opnieuw een gedeelte dat de ouder / voogd moet **ondertekenen**.

Vervolgens geeft de leerling de laatste pagina van document 1 en het volledige document 2 tegen de afgesproken datum **terug af op school**.

Ik dank jullie alvast voor jullie inspanning!

Met vriendelijke groeten



Alexandra Vereeck

**UNIVERSITEIT
GENT**

Appendix 12

Informed Consent Documents for Pupils & Parents

1. INFORMATIE- EN TOESTEMMINGSFORMULIER VOOR DE OUDER / VOOGD

Uw kind wordt uitgenodigd om deel te nemen aan een wetenschappelijke studie. Neem voldoende tijd om de informatie in dit document aandachtig te lezen voordat u beslist uw kind aan deze studie te laten deelnemen. Aarzel niet om contact op te nemen met de onderzoeker als er onduidelijkheden zijn of als u meer duiding wenst. Indien u hierna instemt met de deelname van uw kind, wordt u vriendelijk verzocht om de verklaring van geïnformeerde toestemming achteraan dit document te ondertekenen.

INFORMATIE OMTRENT DE WETENSCHAPPELIJKE STUDIE

Algemene gegevens

- Onderwerp: cognitieve ontwikkeling van leerlingen in het Vlaams secundair onderwijs
- Onderzoeker:

Dra. Alexandra Vereeck
FWO-aspirant fundamenteel onderzoek
Vakgroep Taalkunde UGent
alexandra.vereeck@ugent.be
- Promotoren:
 - prof. dr. Mark Janse, Vakgroep Taalkunde UGent;
 - prof. dr. Katja De Herdt, Vakgroep Letterkunde UGent;
 - prof. dr. Wouter Duyck, Vakgroep Experimentele Psychologie UGent.
- Betrokken onderzoeksinstituten:
 - Fonds Wetenschappelijk Onderzoek Vlaanderen (FWO);
 - Universiteit Gent.

Ethiek

Het onderzoek wordt uitgevoerd in overeenstemming met de geaccepteerde maatstaven voor wetenschappelijk en ethisch gedrag. Alle persoonsgegevens zullen nauwgezet worden beschermd en geanonimiseerd in overeenstemming met de nationale en Europese privacy-wetgeving, zie ook "informatie omtrent de privacy en persoonsgegevens". Voor de aanvang van de studie is het advies ingewonnen van de Ethische Commissie van de Faculteit Psychologie en Pedagogische Wetenschappen van de Universiteit Gent.

INFORMATIE OMTRENT DEELNAME

Wat houdt deelname in?

De studie loopt over drie opeenvolgende schooljaren, te beginnen in 2021-2022. Ieder schooljaar is er één meetmoment voorzien. Per meetmoment zal uw kind een vragenlijst en testen doorlopen. De vragenlijst wordt thuis ingevuld, zie hiervoor de tweede bijlage. De testen worden op school afgenomen. De school maakt hiervoor tijd vrij tijdens de reguliere uren. Het is dus in geen geval zo dat uw kind lessen mist of langer op school aanwezig moet zijn.

De vragenlijst en testen hebben betrekking op de volgende thema's:

- consciëntieusheid (volharding en zelfcontrole);
- schoolmotivatie;
- Nederlandse taalvaardigheid;
- algemeen-cognitieve vaardigheden (IQ).

Daarnaast zal de onderzoeker (een deel van) zijn/haar schoolresultaten opvragen. Hierbij is het goed om te weten dat de data zullen worden gepseudonimiseerd, zoals verderop toegelicht onder "informatie omtrent privacy en persoonsgegevens".

Waarom doet deelname ertoe?

Dit onderzoek naar de cognitieve ontwikkeling van leerlingen in het Vlaams secundair onderwijs is het eerste in zijn soort. Wat de studie zo bijzonder maakt, is dat er een groot aantal leerlingen uit verschillende scholen en studierichtingen meerdere jaren zullen worden gevolgd. Door drie jaar op rij te meten, eerder dan slechts één meetmoment te organiseren, kunnen evoluties in de cognitieve ontwikkeling heel precies in kaart worden gebracht. Het grote aantal leerlingen is belangrijk om geldige conclusies uit het onderzoek te kunnen trekken. Hoe meer leerlingen meedoen, hoe betrouwbaarder de resultaten zullen zijn. Anders gezegd, iedere deelname telt.

Is deelname verplicht?

Deelname aan deze studie is volledig vrijwillig en bijgevolg kan er op geen enkele manier sprake zijn van dwang. U kunt weigeren dat uw kind deelneemt aan de studie. Ook wanneer u in eerste instantie instemt met de deelname, behoudt u het recht om uw kind op eender welk moment terug te trekken uit de studie zonder opgave van reden. Noch niet-deelname, noch terugtrekking uit de studie heeft op enige manier een invloed op de verdere relatie met de school, de onderzoeker of de betrokken onderzoeksinstellingen.

Wat wordt aangeboden in ruil voor deelname?

Om de leerlingen te bedanken voor hun inzet en hun onmisbare bijdrage aan het onderzoek, zullen jaarlijks onder alle deelnemende leerlingen van de school enkele geschenkbonnen worden verloot.

U hebt recht op een samenvatting van de onderzoeksbevindingen nadat de studie is afgerond en de resultaten bekend zijn. Een dergelijke samenvatting kan via mail worden aangevraagd bij de onderzoeker (alexandra.vereeck@ugent.be). Daarenboven kan u indien gewenst na afloop van het derde meetmoment en ten laatste op 31 mei 2024 een persoonlijk cognitief rapport van uw kind verkrijgen met zijn/haar taalscores en IQ.

INFORMATIE OMTRENT PRIVACY EN PERSOONSGEGEVENS

Wettelijk kader

Het wettelijk kader voor de verwerking van persoonsgegevens en vertrouwelijke informatie wordt bepaald door:

- de Europese Algemene Verordening Gegevensbescherming 2016/679 van 27 april 2016, die van kracht is sinds 25 mei 2018 (AVG, Engels: GDPR);
- de Belgische Wet betreffende de bescherming van natuurlijke personen met betrekking tot de verwerking van persoonsgegevens van 30 juli 2018.

Welke persoonsgegevens worden verzameld en verwerkt?

De volgende categorieën van persoonsgegevens zullen worden verzameld en verwerkt:

- identificatiedata (naam, school, klas e.d.);
- subjectvariabelen (leeftijd, gender, socio-economische status e.d.);
- scores op de hoger vermelde vragenlijsten en testen;
- schoolresultaten.

Hoe worden de persoonsgegevens verwerkt?

Na de dataverzameling zal er pseudonimisatie worden toegepast: de identificatiedata zullen zo snel mogelijk van de andere onderzoeksdata worden gesplitst. Het sleutelbestand dat de link bevat tussen de identificatiedata en de andere onderzoeksdata zal apart worden bewaard en beveiligd. De toegang tot het sleutelbestand zal strikt worden beperkt tot de onderzoeker zelf en haar promotoren. Voor de dagelijkse onderzoeksactiviteiten zal worden gewerkt met de gepseudonimiseerde dataset. Alle rapportering over de onderzoeksdata in verslagen, artikels en dergelijke zal volstrekt anoniem gebeuren, zodat de identiteit van uw kind op geen enkele manier af te leiden valt.



3/5

Waarom worden de persoonsgegevens verwerkt?

De verwerking van de persoonsgegevens gebeurt in het kader van een meerjarige observationele studie van leerlingen in het Vlaams secundair onderwijs.

Op welke rechtsgrond worden de persoonsgegevens verwerkt?

Voor de verwerking van de persoonsgegevens van uw kind wordt om uw expliciete toestemming gevraagd. U kan die geven door de verklaring van geïnformeerde toestemming achteraan dit document te ondertekenen. U kan te allen tijde uw toestemming weer intrekken door dit via mail te melden aan de onderzoeker (alexandra.vereeck@ugent.be).

Wie is verantwoordelijk voor de verwerking van de persoonsgegevens?

De Universiteit Gent is de verwerkingsverantwoordelijke. Dit wil zeggen dat de Universiteit Gent bepaalt voor welke doeleinden en met welke middelen de persoonsgegevens worden verwerkt.

Transfer van gegevens

Er zullen gegevens worden getransfereerd naar ontvangers buiten het onderzoeksteam van de wetenschappelijke studie in kwestie, zoals beschreven in deze paragraaf. Een belangrijk onderdeel van de wetenschappelijke studie zijn de testen die peilen naar Nederlandse taalvaardigheid en algemeen-cognitieve vaardigheden. Met het oog op deze testafnames werden aan de onderzoeker enkele meetinstrumenten ter beschikking gesteld die zijn ontwikkeld door andere onderzoekers binnen en buiten de Universiteit Gent. In ruil voor het kosteloos gebruik van hun meetinstrumenten zullen gepseudonimiseerde onderzoeksdata worden gedeeld met de ontwikkelaars van deze instrumenten. Zij kunnen deze data dan verwerken in het kader van ander onderzoek, voor zover dit verband houdt met de verdere verfijning van de genoemde meetinstrumenten. Bij de transfers worden alle nodige maatregelen genomen om de vertrouwelijkheid van de persoonsgegevens te garanderen.

Vragen en klachten

Indien u vragen heeft aangaande het gebruik van de persoonsgegevens van uw kind of indien u een klacht wilt indienen over de manier waarop de persoonsgegevens van uw kind worden gebruikt, dan kunt u contact opnemen met de functionaris voor gegevensbescherming van de Universiteit Gent via privacy@ugent.be of T 09 264 95 17. U kunt ook een klacht indienen bij de Gegevensbeschermingsautoriteit (contact@apd-gba.bep) en/of bij de Vlaamse Toezicht-commissie (contact@toezichtscommissie.be).

VERKLARING VAN GEÏNFORMEERDE TOESTEMMING

Deze pagina moet na ondertekening terug op school worden afgegeven.

Ik, ondergetekende,(naam) verklaar hierbij dat ik in de hoedanigheid van ouder / voogd van(naam), een minderjarige participant aan een studie onder leiding van Alexandra Vereeck, als FWO-aspirant verbonden aan de Universiteit Gent,

- (1) de uitleg over de aard van de vragen en de taken die tijdens deze studie zullen worden aangeboden heb gekregen en dat mij de mogelijkheid is geboden om bijkomende informatie te verkrijgen;
- (2) mijn kind volledig uit vrije wil laat deelnemen aan de wetenschappelijk studie;
- (3) de toestemming geef aan het onderzoeksteam om de resultaten op gepseudonimiseerde wijze te bewaren en te verwerken en om de resultaten anoniem te rapporteren;
- (4) op de hoogte ben van de mogelijkheid om de deelname van mijn kind aan de studie op ieder moment stop te zetten en dit zonder opgave van reden;
- (5) weet dat niet deelnemen of de deelname aan het onderzoek stopzetten op geen enkele manier negatieve gevolgen heeft voor mijn kind;
- (6) weet dat ik op aanvraag een samenvatting van de onderzoeksbevindingen kan krijgen nadat de studie is afgerond en de resultaten bekend zijn;
- (7) toestemming geef dat gepseudonimiseerde data worden gebruikt voor verdere analyse door andere onderzoekers;
- (8) weet dat UGent de verantwoordelijke eenheid is met betrekking tot de tijdens de studie verzamelde persoonsgegevens en dat de functionaris voor gegevensbescherming me meer informatie kan verschaffen over de bescherming van de persoonsgegevens van mijn kind (contact: privacy@ugent.be of T 09 264 95 17).

Gelezen en goedgekeurd op(datum)

Handtekening:

Appendix 13

Background Data Form for Pupils & Parents

FORMULIER TYPE A

2. ACHTERGRONDGEGEVENS VAN DE LEERLING

Nadat dit document is ingevuld en ondertekend, moet het in zijn geheel terug op school worden afgegeven.

GEDEELTE VOOR DE LEERLING

In te vullen door de leerling onder begeleiding van een ouder of voogd.

1. **Voornaam:**.....
2. **Familienaam:**.....
3. **Geboortedatum:**.....
4. **School:**
5. **Klas:**.....
6. **Klasnummer:**.....
7. **Studierichting:**
8. **Geslacht:** ☐ M ☐ V ☐ X
9. **Hoogst behaalde opleidingsniveau van de moeder / voogd:**
 - ☐ geen
 - ☐ lager onderwijs
 - ☐ secundair onderwijs
 - ☐ bachelor (of equivalent)
 - ☐ master (of equivalent)
 - ☐ doctor
 - ☐ niet van toepassing
10. **Hoogst behaalde opleidingsniveau van de vader / voogd:**
 - ☐ geen
 - ☐ lager onderwijs
 - ☐ secundair onderwijs
 - ☐ bachelor (of equivalent)
 - ☐ master (of equivalent)
 - ☐ doctor
 - ☐ niet van toepassing



1/2

FORMULIER TYPE A

11. Wat is je moedertaal? ☐ Nederlands ☐ een andere taal, namelijk
12. Heb je een taalstoornis (bv. dyslexie)? ☐ ja ☐ nee
13. Neem je medicatie die een effect heeft op je aandacht, concentratievermogen, alertheid of dergelijke? ☐ ja ☐ nee
14. Heb je in de afgelopen 12 maanden een IQ-test ondergaan? ☐ ja ☐ nee
15. **Vragenlijst:** Vul de vragenlijst in die je terugvindt op onderstaand webadres. Typ het webadres over in je browser en volg de instructies op je scherm. Als je helemaal klaar bent met de vragenlijst, zet dan een vinkje in het lege vakje hieronder.

Webadres	Klaar?
https://bit.ly/3iQ6elP	

GEDEELTE VOOR DE OUDER / VOOGD

Ik, ondergetekende,(naam) verklaar hierbij dat ik in de hoedanigheid van ouder / voogd van(naam), een minderjarige participant aan een studie onder leiding van Alexandra Vereeck, als FWO-aspirant verbonden aan de Universiteit Gent,

- (1) de onder "gedeelte voor de leerling" ingevulde gegevens heb nagekeken en instea voor de correctheid ervan;
- (2) erop heb toegezien dat mijn kind de onder punt 15 vermelde vragenlijst naar best vermogen heeft ingevuld.

Gelezen en goedgekeurd op(datum)

Handtekening:

Appendix 14

Research Flyer 2023–2024

Met de
groeten van
Alexandra Vereeck,
uit dank voor uw
medewerking aan
mijn onderzoek.
Laat het
smaken!



BEDANKT!



Alexandra Vereeck
Master in de taal- en
letterkunde (Latijn & Grieks)
Specifieke lerarenopleiding
talen (Latijn & Grieks)
FWO-aspirant fundamenteel
onderzoek, Vakgroep Taalkunde &
Vakgroep VTC, Universiteit Gent
alexandra.vereeck@ugent.be
0479 448 440



Het onderzoek
? De cognitieve ontwikkeling
van leerlingen in het Vlaams
secundair onderwijs
! De eerste empirische studie
in Vlaanderen waarbij een groot
aantal leerlingen uit verschillende
scholen en studierichtingen
meerdere jaren wordt gevolgd om
hun cognitieve ontwikkeling in
kaart te brengen



Metingen op scholen
3 jaar, 1 meting per jaar:
van het tweede tot en met het
vierde middelbaar, 2021-2024
4 scholen, ca. 950 leer-
lingen bij de eerste meting
Onder meer getest op
moedertaalvaardigheid in het
Nederlands en algemeen-
cognitieve vaardigheden (IQ)



Niet op de openbare weg gooien a.u.b.
V.U.: Universiteit Gent, Faculteit Letteren en Wijsbegeerte,
Campus Boekentoren



Appendix 15

Individual Feedback Report



FEEDBACKRAPPORT ONDERZOEK
NAAR COGNITIEVE ONTWIKKELING
O.L.V. DRA. ALEXANDRA VEREECK

De school van uw kind werkte de voorbije schooljaren mee aan een studie naar de cognitieve ontwikkeling van jongeren in het Vlaamse algemeen secundair onderwijs. De deelnemende leerlingen legden jaarlijks twee tests af: een taaltest Nederlands en een algemeen-cognitieve test. De taaltest Nederlands peilde onder meer naar woordenschat, werkwoordspelling, zinspelling (dictee) en begrijpend lezen. De algemeen-cognitieve test peilde naar vloeiende intelligentie, gekristalliseerde intelligentie, visuospatieële vaardigheid en verbaal geheugen. In dit rapport vindt u een beschrijving van hoe uw kind op elk van deze variabelen heeft gescoord. Deze feedback is louter informatief. Bedenk enerzijds dat de resultaten van deze tests niet hetzelfde zijn als de uitslag van een klinische IQ-test, en anderzijds dat de steekproef van leerlingen niet noodzakelijk representatief is voor heel Vlaanderen.

Woordenschat

De subtest woordenschat be vraagt de kennis van woordbetekenissen, zowel los als in context.

Persoonlijke score

Bij het eerste meetmoment (2021-2022) beantwoordde uw kind **60%** van de vragen correct.

Bij het tweede meetmoment (2022-2023) beantwoordde uw kind **50%** van de vragen correct.

Bij het derde meetmoment (2023-2024) beantwoordde uw kind **75%** van de vragen correct.

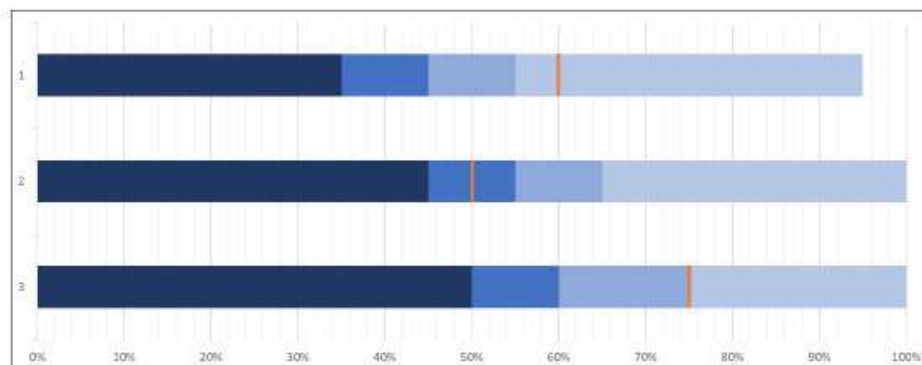
Ter vergelijking: alle deelnemende leerlingen

Bij het eerste meetmoment (2021-2022) lag de gemiddelde score in het tweede middelbaar rond **47%**.

Bij het tweede meetmoment (2022-2023) lag de gemiddelde score in het derde middelbaar rond **55%**.

Bij het derde meetmoment (2023-2024) lag de gemiddelde score in het vierde middelbaar rond **60%**.

Hieronder kan u per meetmoment zien hoe de score van uw kind zich verhoudt tot die van de anderen. De oranje lijn geeft zijn/haar persoonlijke score aan. Ieder blauw blokje stelt een kwart van de leerlingen voor; dat wil zeggen dat telkens een kwart van de leerlingen een score heeft die lager ligt dan of gelijk is aan de rechtergrens van het blokje.



Werkwoordspelling

De subtest werkwoordspelling be vraagt de vervoeging van werkwoorden in zinscontext.

Persoonlijke score

Bij het eerste meetmoment (2021-2022) beantwoordde uw kind **53%** van de vragen correct.

Bij het tweede meetmoment (2022-2023) beantwoordde uw kind **29%** van de vragen correct.

Bij het derde meetmoment (2023-2024) beantwoordde uw kind **35%** van de vragen correct.

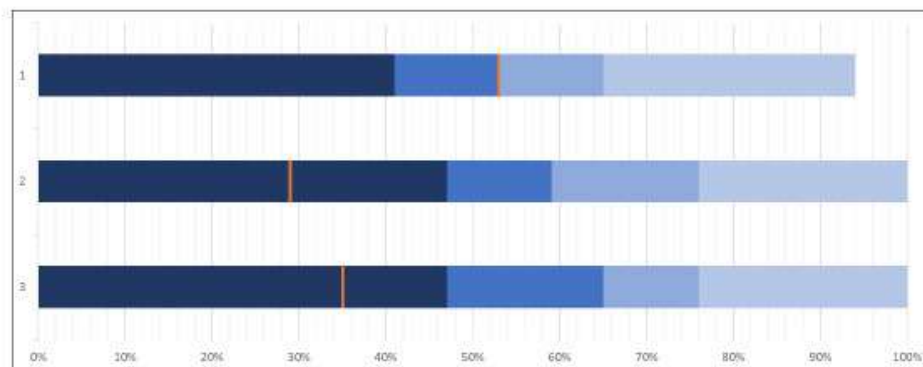
Ter vergelijking: alle deelnemende leerlingen

Bij het eerste meetmoment (2021-2022) lag de gemiddelde score in het tweede middelbaar rond **53%**.

Bij het tweede meetmoment (2022-2023) lag de gemiddelde score in het derde middelbaar rond **60%**.

Bij het derde meetmoment (2023-2024) lag de gemiddelde score in het vierde middelbaar rond **61%**.

Hieronder kan u per meetmoment zien hoe de score van uw kind zich verhoudt tot die van de anderen. De oranje lijn geeft zijn/haar persoonlijke score aan. Ieder blauw blokje stelt een kwart van de leerlingen voor; dat wil zeggen dat telkens een kwart van de leerlingen een score heeft die lager ligt dan of gelijk is aan de rechtergrens van het blokje.



Zinspelling (dictee)

De subtest zinspelling be vraagt de spelling van woorden in voorgelezen zinnen.

Persoonlijke score

Bij het eerste meetmoment (2021-2022) beantwoordde uw kind **33%** van de vragen correct.

Bij het tweede meetmoment (2022-2023) beantwoordde uw kind **44%** van de vragen correct.

Bij het derde meetmoment (2023-2024) beantwoordde uw kind **72%** van de vragen correct.

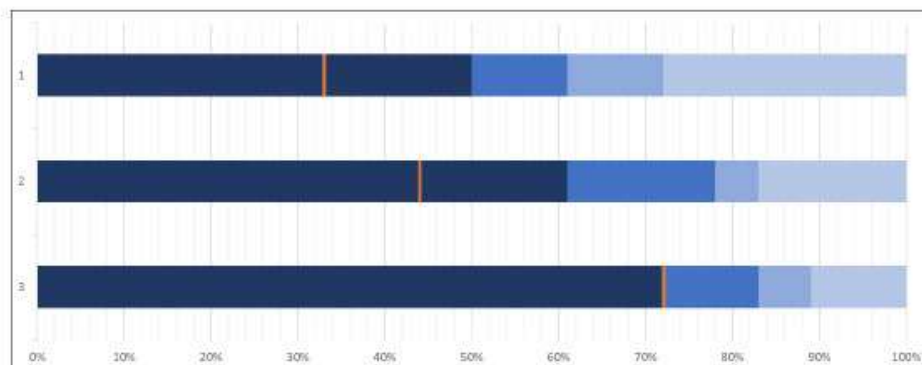
Ter vergelijking: alle deelnemende leerlingen

Bij het eerste meetmoment (2021-2022) lag de gemiddelde score in het tweede middelbaar rond **60%**.

Bij het tweede meetmoment (2022-2023) lag de gemiddelde score in het derde middelbaar rond **73%**.

Bij het derde meetmoment (2023-2024) lag de gemiddelde score in het vierde middelbaar rond **78%**.

Hieronder kan u per meetmoment zien hoe de score van uw kind zich verhoudt tot die van de anderen. De oranje lijn geeft zijn/haar persoonlijke score aan. Ieder blauw blokje stelt een kwart van de leerlingen voor; dat wil zeggen dat telkens een kwart van de leerlingen een score heeft die lager ligt dan of gelijk is aan de rechtergrens van het blokje.



Begrijpend lezen

De subtest begrijpend lezen be vraagt begrip van en inzicht in een gelezen tekst.

Persoonlijke score

Bij het eerste meetmoment (2021-2022) beantwoordde uw kind **0%** van de vragen correct.

Bij het tweede meetmoment (2022-2023) beantwoordde uw kind **67%** van de vragen correct.

Bij het derde meetmoment (2023-2024) beantwoordde uw kind **67%** van de vragen correct.

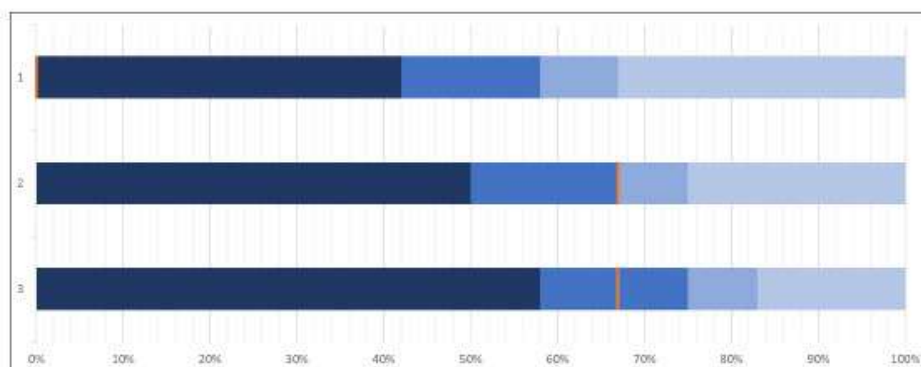
Ter vergelijking: alle deelnemende leerlingen

Bij het eerste meetmoment (2021-2022) lag de gemiddelde score in het tweede middelbaar rond **54%**.

Bij het tweede meetmoment (2022-2023) lag de gemiddelde score in het derde middelbaar rond **62%**.

Bij het derde meetmoment (2023-2024) lag de gemiddelde score in het vierde middelbaar rond **69%**.

Hieronder kan u per meetmoment zien hoe de score van uw kind zich verhoudt tot die van de anderen. De oranje lijn geeft zijn/haar persoonlijke score aan. Ieder blauw blokje stelt een kwart van de leerlingen voor; dat wil zeggen dat telkens een kwart van de leerlingen een score heeft die lager ligt dan of gelijk is aan de rechtergrens van het blokje.



Vloeiende intelligentie

De subtest vloeiende intelligentie test redeneervaardigheid, door participanten cijferreeksen te laten aanvullen.

Persoonlijke score

Bij het eerste meetmoment (2021-2022) beantwoordde uw kind **44%** van de vragen correct.

Bij het tweede meetmoment (2022-2023) beantwoordde uw kind **63%** van de vragen correct.

Bij het derde meetmoment (2023-2024) beantwoordde uw kind **77%** van de vragen correct.

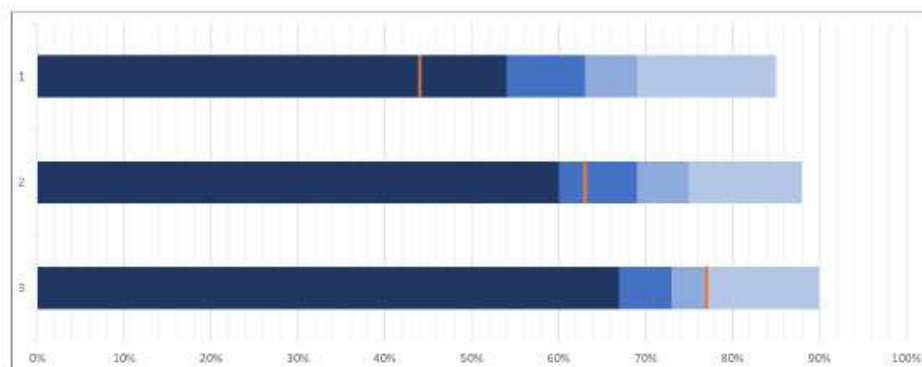
Ter vergelijking: alle deelnemende leerlingen

Bij het eerste meetmoment (2021-2022) lag de gemiddelde score in het tweede middelbaar rond **62%**.

Bij het tweede meetmoment (2022-2023) lag de gemiddelde score in het derde middelbaar rond **67%**.

Bij het derde meetmoment (2023-2024) lag de gemiddelde score in het vierde middelbaar rond **71%**.

Hieronder kan u per meetmoment zien hoe de score van uw kind zich verhoudt tot die van de anderen. De oranje lijn geeft zijn/haar persoonlijke score aan. Ieder blauw blokje stelt een kwart van de leerlingen voor; dat wil zeggen dat telkens een kwart van de leerlingen een score heeft die lager ligt dan of gelijk is aan de rechtergrens van het blokje.



Gekristalliseerde intelligentie

De subtest gekristalliseerde intelligentie test verbale kennis, door participanten het beste antoniem van een gegeven woord te laten aanduiden.

Persoonlijke score

Bij het eerste meetmoment (2021-2022) beantwoordde uw kind **60%** van de vragen correct.

Bij het tweede meetmoment (2022-2023) beantwoordde uw kind **67%** van de vragen correct.

Bij het derde meetmoment (2023-2024) beantwoordde uw kind **62%** van de vragen correct.

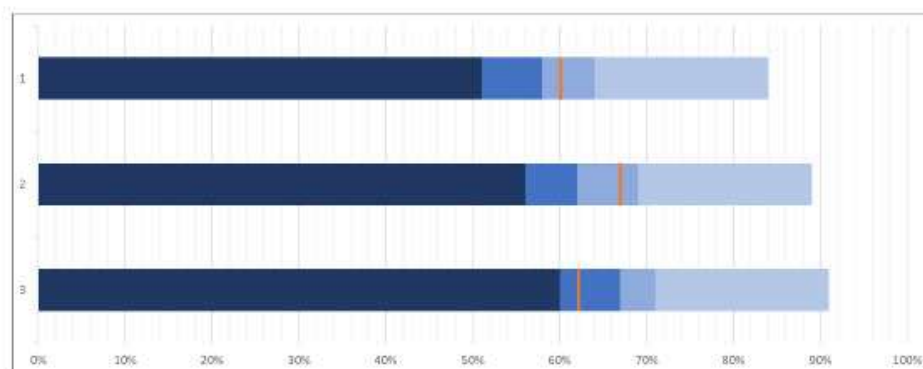
Ter vergelijking: alle deelnemende leerlingen

Bij het eerste meetmoment (2021-2022) lag de gemiddelde score in het tweede middelbaar rond **57%**.

Bij het tweede meetmoment (2022-2023) lag de gemiddelde score in het derde middelbaar rond **62%**.

Bij het derde meetmoment (2023-2024) lag de gemiddelde score in het vierde middelbaar rond **66%**.

Hieronder kan u per meetmoment zien hoe de score van uw kind zich verhoudt tot die van de anderen. De oranje lijn geeft zijn/haar persoonlijke score aan. Ieder blauw blokje stelt een kwart van de leerlingen voor; dat wil zeggen dat telkens een kwart van de leerlingen een score heeft die lager ligt dan of gelijk is aan de rechtergrens van het blokje.



Visuospatiële vaardigheid

De subtest visuospatiële vaardigheid test ruimtelijk inzicht, door participanten te laten aanduiden welke figuren rotaties zijn van een gegeven figuur.

Persoonlijke score

Bij het eerste meetmoment (2021-2022) beantwoordde uw kind **50%** van de vragen correct.

Bij het tweede meetmoment (2022-2023) beantwoordde uw kind **47%** van de vragen correct.

Bij het derde meetmoment (2023-2024) beantwoordde uw kind **53%** van de vragen correct.

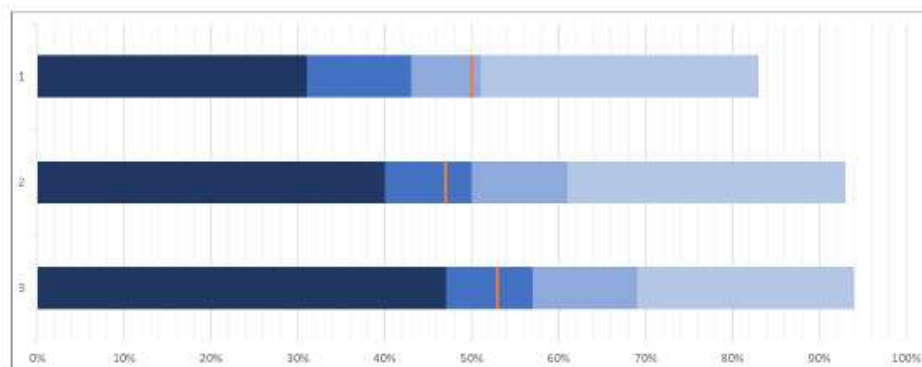
Ter vergelijking: alle deelnemende leerlingen

Bij het eerste meetmoment (2021-2022) lag de gemiddelde score in het tweede middelbaar rond **42%**.

Bij het tweede meetmoment (2022-2023) lag de gemiddelde score in het derde middelbaar rond **50%**.

Bij het derde meetmoment (2023-2024) lag de gemiddelde score in het vierde middelbaar rond **56%**.

Hieronder kan u per meetmoment zien hoe de score van uw kind zich verhoudt tot die van de anderen. De oranje lijn geeft zijn/haar persoonlijke score aan. Ieder blauw blokje stelt een kwart van de leerlingen voor; dat wil zeggen dat telkens een kwart van de leerlingen een score heeft die lager ligt dan of gelijk is aan de rechtergrens van het blokje.



Verbaal geheugen

De subtest verbaal geheugen test het kortetermijngeheugen, door participanten woorden van buiten te laten leren en vervolgens te laten reproduceren.

Persoonlijke score

Bij het eerste meetmoment (2021-2022) beantwoordde uw kind **38%** van de vragen correct.

Bij het tweede meetmoment (2022-2023) beantwoordde uw kind **33%** van de vragen correct.

Bij het derde meetmoment (2023-2024) beantwoordde uw kind **40%** van de vragen correct.

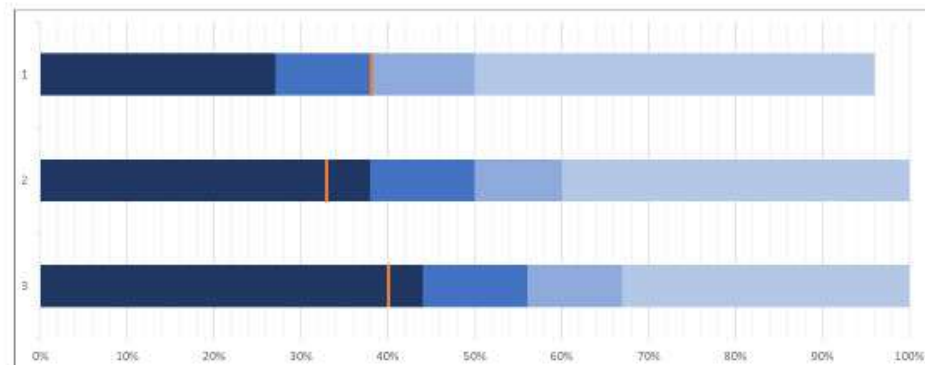
Ter vergelijking: alle deelnemende leerlingen

Bij het eerste meetmoment (2021-2022) lag de gemiddelde score in het tweede middelbaar rond **38%**.

Bij het tweede meetmoment (2022-2023) lag de gemiddelde score in het derde middelbaar rond **49%**.

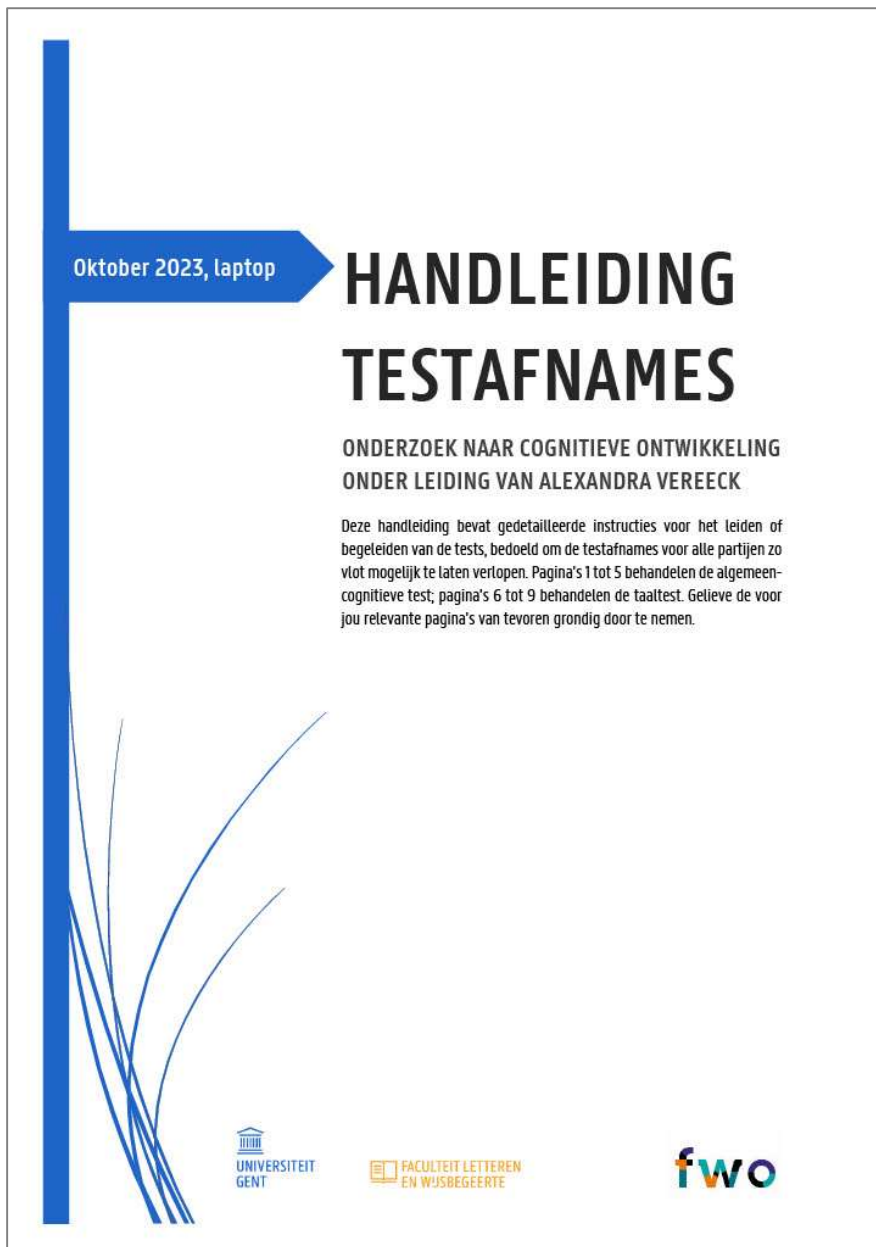
Bij het derde meetmoment (2023-2024) lag de gemiddelde score in het vierde middelbaar rond **54%**.

Hieronder kan u per meetmoment zien hoe de score van uw kind zich verhoudt tot die van de anderen. De oranje lijn geeft zijn/haar persoonlijke score aan. Ieder blauw blokje stelt een kwart van de leerlingen voor; dat wil zeggen dat telkens een kwart van de leerlingen een score heeft die lager ligt dan of gelijk is aan de rechtergrens van het blokje.



Appendix 16

Testing Manual, Version October 2023 (Laptop)



ALGEMEEN-COGNITIEVE TEST



VOORAF

De eerste meting bestaat uit een test algemeen-cognitieve vaardigheden. Er zijn vier oefeningen: Puntreeksen, Tegenstellingen, Gedraaide Figuren en Verbaal Geheugen. De oefeningen zijn getimed; na een vast aantal minuten wordt er automatisch overgegaan naar de volgende oefening. De test kan niet afgebroken en later hervat worden.

De onderzoeker neemt de algemeen-cognitieve test af; de leerkracht ter plaatse **begeleidt de afname**, d.w.z. dat hij/zij de onderzoeker ondersteunt bij het opstarten en afronden van de test en het oplossen van problemen die zich voordoen.

Welk materiaal is er nodig?

Voor de leerlingen: een laptop en een hoofdtelefoon / oortjes.

Voor de testleiders: **reserveoortjes** (op school aanwezig of door de onderzoeker voorzien) en **instructiekaarten** (door de onderzoeker voorzien).

Waaraan moet de testomgeving voldoen?

De testomgeving moet aan de volgende punten voldoen:

- Alle leerlingen beschikken over een eigen toestel met eigen hoofdtelefoon / oortjes.
- De leerlingen gebruiken **geen kladpapier en schrijfgerief**.
- De gsm van de leerlingen staat uit of op stil en zit in hun boekentas of dergelijke. Andere mogelijke afleidingen zoals achtergrondgeluid worden tot een minimum beperkt.
- Er zijn geen derden aanwezig in de ruimte.
- De ruimte is comfortabel georganiseerd: comfortabele stoel en tafel, goede verlichting en verluchting enzovoort.

Waarop moet ik letten tijdens het toezicht houden?

Zie erop toe dat de testafname correct verloopt. Let in het bijzonder hierop:

- De testafnames gebeuren **individueel**. Afkijken of samenwerken zijn dus uit den boze. Als leerkracht ondersteun je de leerlingen alleen bij de besturing van hun toestel indien nodig, niet bij de inhoud van de test.
- Hulpmiddelen zijn eveneens niet toegestaan, d.w.z. geen andere tabbladen, browsers of applicaties, geen gsm of smartwatch, geen kladpapier en schrijfgerief. De leerlingen mogen enkel de test voor zich hebben.
- Het is ten strengste **verboden om inhoudelijke elementen van de test te kopiëren** en in bezit te nemen, bijvoorbeeld door screenshots te maken.

OPSTART VAN DE TEST

Bij het inleiden van de test zijn de volgende twee waarschuwingen cruciaal:



"Heel erg belangrijk om te weten is dat deze **test zichzelf afsluit**. Wanneer je straks klaar bent met het invullen van de laatste oefening, dan **wacht** je gewoon totdat er op je scherm verschijnt dat de test is afgelopen. Wat je zeker **NIET** mag doen, is je browser meteen sluiten, want dan worden je antwoorden niet opgeslagen. Je blijft dus gewoon **wachten** tot de test zegt dat het gedaan is." "Daarenboven mag je terwijl je bezig bent met de test de browser **niet refreshen**, want dan ga je oefeningen missen."

Laat daarna de leerlingen de volgende stappen doorlopen:

1. Open Google Chrome.



2. Surf naar www.covat-chc.be.
3. Log in op het testplatform. Let op, je moet twee keer inloggen!
 - a) Log in met de **algemene daglogin**.

Datum	Gebruikersnaam	Wachtwoord

- b) Klik op "Start testafname".

c) Log in met de **persoonlijke login**.

- De **gebruikersnaam** bestaat uit *achternaam + eerste letter voornaam*, bv. Jan Jansens → *jansensj*. Eventuele spaties en speciale tekens zoals trema's, accenten, koppeltekens en apostrofes vallen weg. De gebruikersnaam is niet hoofdlettergevoelig.

Indien er een andere deelnemer is met dezelfde familienaam en initialen, komt achteraan er nog een '1' of '2' bij. Samengestelde familienamen worden ingekort wanneer ze de maximale lengte overschrijden. Indien nodig kan de persoonlijke login worden opgezocht in het **aparte overzicht** in bijlage bij deze handleiding.

- Het **wachtwoord** is 12345678.

- Voor de zekerheid zijn er steeds enkele **reserveaccounts** voorzien. Zie hiervoor het **aparte overzicht** in bijlage. Als een leerling gebruik zou maken van zo'n (anonieme) reserveaccount, moet er goed worden bijgehouden om wie het gaat.

d) Klik op "Start testafname". De test begint.

TROUBLESHOOTING

Wat als het geluid niet werkt of de video's niet willen afspelen?

Per oefening zijn er **instructievideo's** met gesproken uitleg. Wanneer leerlingen aangeven dat het geluid niet werkt, controleer dan eerst of de hoofdtelefoons / oortjes goed zijn ingestoken, of de juiste geluidsbron is geselecteerd, en of het volume hoog genoeg staat. Als het probleem niet zo eenvoudig te verhelpen is en niet alle leerlingen de instructievideo's kunnen bekijken en beluisteren, moeten de instructies mondeling worden gegeven. Zie hiervoor de **instructiekaarten**. De testleider leest de instructiekaarten voor met "**basisinstructie**", de deelnemers volgen op de instructiekaarten met "**voorbeelden basisinstructie**".

Wat als er een pop-upvenster verschijnt?

Wanneer er teveel fouten worden gemaakt in de probeefase van een oefening, zal er een **pop-upvenster** verschijnen.

De eerste maal dat een leerling dit voorheeft bij een bepaalde oefening, voer je in het venster de code **14** in en klik je op “**Verder oefenen**”. Vervolgens heeft de leerling de mogelijkheid om het ofwel dadelijk nog een keer te proberen ofwel eerst een extra instructievideo te bekijken. In geval van problemen met de video's moet deze instructie mondeling worden gegeven. Zie hiervoor de instructiekaarten met “extra instructie” en “voorbeelden extra instructie”.

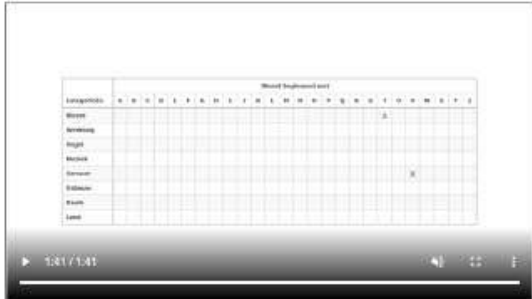
Indien diezelfde leerling een tweede maal een pop-upvenster krijgt bij dezelfde oefening, dan voer je in het venster opnieuw de code **14** in, maar klik je dit keer op “**Taak overslaan**”. Als dit zich voordoet, noteer je de naam van de leerling en breng je de onderzoeker op de hoogte.

Wat als iemand niet verder kan klikken naar de volgende opgave?

Bij de vierde oefening gebeurt het soms dat na afloop van de instructievideo de knop om de oefening te starten niet onmiddellijk wordt geactiveerd. De enige oplossing hier is om een paar seconden te wachten totdat “Start de taak” vanzelf zwart wordt. Klik **NIET** op de refresh-knop, want dan wordt de oefening overgeslagen.

Verbaal Geheugen (1)

Bekijk eerst het filmpje en start daarna met de taak.



Start de taak

Wat als iemand per ongeluk een oefening heeft overgeslagen?

Als in de loop van de test de browser wordt gesloten of de pagina opnieuw wordt geladen, dan wordt de huidige oefening overgeslagen. (Let op: met “oefening” wordt verwezen naar één van de vier grote oefeningen, niet naar een opgave binnen een oefening. Evenmin wordt het geval bedoeld waarin het systeem automatisch naar een volgende oefening springt, ook al zijn nog niet alle opgaves van de vorige oefening beantwoord, omdat de tijdslimiet is verstreken; dit is perfect normaal en gewenst.)

Als een leerling per ongeluk een oefening heeft overgeslagen door de test te **sluiten** of te **refreshen**, laat hem/haar dan eerst de eventuele resterende oefeningen afwerken. Zodra de test zich na de vierde oefening correct heeft afgesloten, doorloop je terug de loginprocedure en log je in op een **ongebruikte reserveaccount** (zie apart overzicht van de persoonlijke logins). Dan leid je de leerling naar de oefening die hij/zij nog moet doen. Reeds afgelegde oefeningen kunnen nu worden overgeslagen door doelbewust de refresh-knop te benutten.

Wat als iemand naar het toilet moet?

Onderbrekingen, bijvoorbeeld voor een toiletbezoek, zijn enkel mogelijk **tussen oefeningen** in vanwege de **tijdslimieten** per oefening. Vraag de leerling dus om eerst de oefening af te maken waar hij/zij mee bezig is. Wanneer de leerling aankomt bij het startscherm van de volgende oefening, is eventueel een korte onderbreking mogelijk.

AFRONDEN VAN DE TEST

Op het einde van de laatste oefening moet er tegen de tijd een raster worden ingevuld.

Zet voor ieder woord dat je nog weet een kruisje op de juiste plaats. Werk verder tot de tijd om is.

Categorieën	Woord beginnend met:																										
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
Bloem																											
Werktuig																											
Vogel																											
Muziek																											
Vervoer																											
Geboort																											
Naam																											
Land																											

Wat moet ik doen?

Dikwijls zijn leerlingen al klaar met het invullen van het raster voordat de tijd op is. In dat geval moet ervoor worden gezorgd dat ze nog **niet hun browser dichtdoen maar wachten** tot de test vanzelf afsluit en meldt dat ze klaar zijn. Anders worden hun antwoorden niet opgeslagen. Na het laatste scherm zal opnieuw de inlogpagina verschijnen.

Je bent nu helemaal klaar met alle taken. Je hebt goed gewerkt.

Je antwoorden werden opgeslagen. Je bent nu uitgelogd.

Al naargelang hoe vlot de opstart van de test verlopen is, kan het voorvallen dat leerlingen nog met de laatste oefening bezig zijn **wanneer de bel gaat**. Laat hen dan wat langer **doorwerken**; het is een kwestie van hooguit enkele minuten. Benadruk dat het niet lang meer zal duren en dat ze het **tabblad open moeten laten staan** totdat de test zichzelf heeft afgesloten.

 Dringende vraag? Bel 0479 448 440.

 Minder dringende vraag? Mail naar alexandra.vereeck@ugent.be.

TAALTEST MET VRAGENLIJSTEN



VOORAF

De tweede meting bestaat uit een test Nederlandse taalvaardigheid, gemengd met vragenlijsten rond persoonseigenschappen zoals motivatie en consciëntieusheid. De vragen zijn ongetimed. De test kan niet afgebroken en later hervat worden.

De onderzoeker neemt de taaltest af; de leerkracht ter plaatse **begeleidt de afname**, d.w.z. dat hij/zij de onderzoeker ondersteunt bij het opstarten van de test en het oplossen van problemen die zich voordoen.

Welk materiaal is er nodig?

Voor de leerlingen: een laptop en een hoofdtelefoon / oortjes.

Voor de testleiders: **reserveoortjes** (op school aanwezig of door de onderzoeker voorzien).

Waarom moet de testomgeving voldoen?

De testomgeving moet aan de volgende punten voldoen:

- Alle leerlingen beschikken over een eigen toestel met eigen hoofdtelefoon / oortjes.
- De leerlingen gebruiken **geen kladpapier en schrijfgierief**.
- De gsm van de leerlingen staat uit of op stil en zit in hun boekentas of dergelijke. Andere mogelijke afleidingen zoals achtergrondgeluid worden tot een minimum beperkt.
- Er zijn geen derden aanwezig in de ruimte.
- De ruimte is comfortabel georganiseerd: comfortabele stoel en tafel, goede verlichting en verluchting enzovoort.

Waarom moet ik letten tijdens het toezicht houden?

Zie erop toe dat de testafname correct verloopt. Let in het bijzonder hierop:

- De testafnames gebeuren **individueel**. Afkijken of samenwerken zijn dus uit den boze. Als leerkracht ondersteun je de leerlingen alleen bij de besturing van hun toestel indien nodig, niet bij de inhoud van de test.
- Hulpmiddelen zijn eveneens niet toegestaan, d.w.z. geen andere tabbladen, browsers of applicaties, geen gsm of smartwatch, geen kladpapier en schrijfgierief. De leerlingen mogen enkel de test voor zich hebben.
- Het is ten strengste **verboden om inhoudelijke elementen van de test te kopiëren** en in bezit te nemen, bijvoorbeeld door screenshots te maken.

OPSTART VAN DE TEST

Laat de leerlingen de volgende stappen doorlopen:

1. Open Google Chrome.

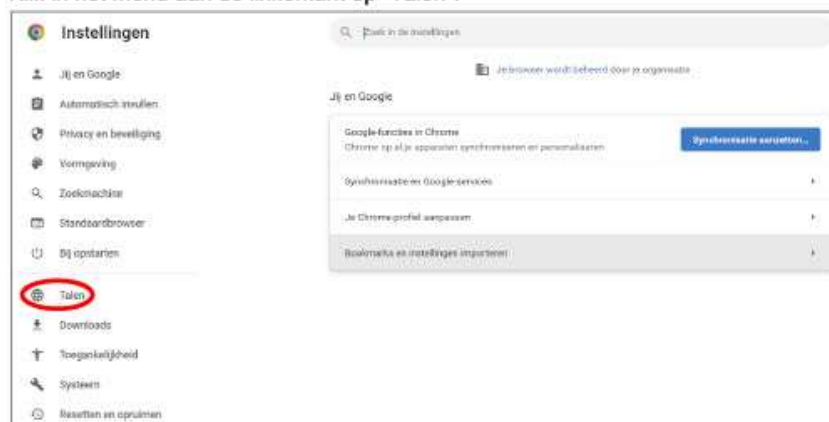


2. Schakel de spellingcontrole uit.

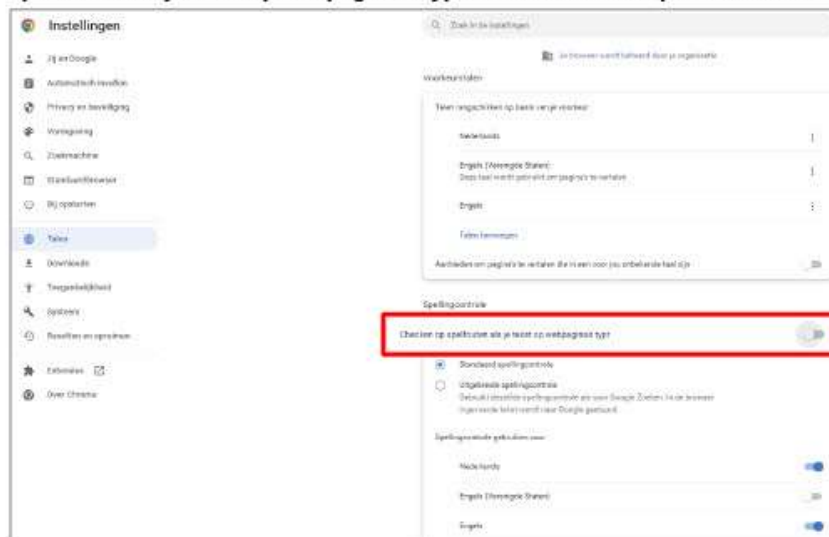
a) Klik op de drie puntjes rechts bovenaan en ga naar "Instellingen".



b) Klik in het menu aan de linkerkant op "Talen".



c) Halverwege de pagina zie je "Spellingcontrole" staan, en daaronder "Checken op spelfouten als je tekst op webpagina's typt". Zet de schuifknop daarnaast uit.



3. Surf naar bit.ly/taal2324 en volg de instructies op je scherm.
(De long url is <https://tpsSurvey.ugent.be/limesurvey315/index.php/893118?lang=nl>)

TROUBLESHOOTING

Wat als het geluid niet werkt?

De eerste oefening van de taaltest is een dictee. Voorafgaand daaraan is er een instructievideo met achtergrondmuziek, die fungeert als audiocheck. Wanneer leerlingen aangeven dat het geluid niet werkt, controleer dan eerst of de hoofdtelefoons / oortjes goed zijn ingestoken, of de juiste geluidsbron is geselecteerd, en of het volume hoog genoeg staat. Als het probleem niet zo eenvoudig te verhelpen is en niet alle leerlingen het dictee kunnen beluisteren, moet het worden voorgelezen. Het dictee staat uitgeschreven in het kader hieronder. Lees het dictee precies zo voor als in het kader staat en doe dit slechts één maal.

Zin 1

- Gelukkig biedt Sams antigifmiddel de oplossing tegen die slangenbeet. (3 sec. wachten)
- Gelukkig biedt Sams antigifmiddel (5 sec. wachten)
- de oplossing tegen die slangenbeet (5 sec. wachten)
- Gelukkig biedt Sams antigifmiddel de oplossing tegen die slangenbeet. (7 sec. wachten)

Zin 2

- De leraar geschiedenis vertelt dat hij de belangrijkste Parijse bezienswaardigheden heeft bezocht. (3 sec. wachten)
- De leraar geschiedenis vertelt (5 sec. wachten)
- dat hij de belangrijkste Parijse bezienswaardigheden heeft bezocht (10 sec. wachten)
- De leraar geschiedenis vertelt dat hij de belangrijkste Parijse bezienswaardigheden heeft bezocht. (7 sec. wachten)

Zin 3

- Ik belandde in het ziekenhuis met geschaafde knieën en een arm- en beenbreuk. (3 sec. wachten)
- Ik belandde in het ziekenhuis (5 sec. wachten)
- met geschaafde knieën (5 sec. wachten)
- en een arm- en beenbreuk (5 sec. wachten)
- Ik belandde in het ziekenhuis met geschaafde knieën en een arm- en beenbreuk.

Wat als iemand de kluts kwijt is bij het dictee?

Als een leerling door omstandigheden (achtergrondlawaai, verwarring, vertraging bij het inpluggen van de oortjes et cetera) een deel van het dictee heeft gemist, kan hij/zij de test **herstarten**. Let op: sta dit enkel toe indien er een goede reden voor is. Het is in principe niet de bedoeling dat het dictee meermaals wordt beluisterd. Om de taaltest te herstarten, klik op "Annuleren, verwijder alle ingevulde antwoorden" in de rechterbovenhoek en daarna op "Herstart deze enquête". De leerling vult de sectie "Personalia" opnieuw in en kan weer verder met de test.

Annuleren, verwijder alle ingevulde antwoorden

0%

Antwoorden verwijderd

[Herstart deze enquête](#)

Wat als iemand een vraag heeft bij de test?

Het is toegestaan om:

- ✓ opgaves te parafraseren;
- ✓ te helpen met de besturing van het toestel, bv. een trema typen;
- ✓ de betekenis van de stellingen in de **vragenlijsten** te verduidelijken.

Het is **niet** toegestaan om:

- X te helpen met de **inhoud van de taal oefeningen** zelf, bv. woorden verklaren.

 Dringende vraag? Bel 0479 448 440.

 Minder dringende vraag? Mail naar alexandra.vereeck@ugent.be.

Appendix 17

Empathy Scales, Flemish Version

Affectieve empathie

Als mijn moeder blij is, voel ik me ook blij.

Ik voel me vaak triest wanneer ik naar een trieste film kijk.

Wanneer een vriend van streek is, voel ik me ook van streek.

Wanneer een vriend huilt, moet ik zelf huilen.

Als iemand in mijn familie verdrietig is, voel ik me erg slecht.

Ik voel me vreselijk wanneer twee mensen ruziemaken.

Cognitieve empathie

Wanneer een vriend boos is, weet ik gewoonlijk waarom.

Als een vriend verdrietig is, begrijp ik meestal waarom.

Als een vriend huilt, begrijp ik vaak wat er is gebeurd.

Antwoordopties

Niet waar (1) – een beetje waar (2) – waar (3)

Appendix 18

Reading Motivation Scale

Stellingen

Ik lees graag boeken.

Ik lees graag in tijdschriften.

Ik ga geregeld naar de bibliotheek.

Antwoordopties

Participanten duiden aan hoezeer ze het met de stellingen eens zijn op een schaal van 1 tot 5.

Appendix 19

Descriptive Statistics of the Non-Cognitive Variables

Descriptives

	Latinist	Time	N	Mean	SD	Minimum	Maximum
Consistency	0	0	254	10.5	2.42	4.00	18.0
		1	255	10.2	2.20	4.00	17.0
		2	257	10.5	2.48	4.00	18.0
	1	0	200	11.5	2.74	6.00	19.0
		1	200	10.8	2.64	4.00	18.0
		2	202	10.5	2.81	4.00	19.0
Perseverance	0	0	254	13.6	2.26	4.00	19.0
		1	255	13.2	2.29	8.00	20.0
		2	257	13.3	2.42	5.00	20.0
	1	0	200	14.6	2.46	7.00	20.0
		1	200	13.7	2.68	7.00	20.0
		2	202	13.5	2.67	6.00	20.0
SelfControl	0	0	254	39.7	7.23	21.00	56.0
		1	255	37.9	6.96	20.00	57.0
		2	257	38.0	7.55	18.00	59.0
	1	0	200	42.1	7.16	26.00	63.0
		1	200	40.3	7.44	15.00	60.0
		2	202	39.8	8.01	16.00	58.0
ControlledMot	0	0	254	23.1	5.89	8.00	39.0
		1	255	24.5	5.78	10.00	40.0
		2	257	24.9	5.59	8.00	40.0
	1	0	200	23.4	5.78	8.00	38.0
		1	200	22.8	6.28	8.00	37.0
		2	202	23.1	5.94	8.00	36.0

Descriptives

	Latinist	Time	N	Mean	SD	Minimum	Maximum
AutoMot	0	0	254	20.4	6.27	8.00	38.0
		1	255	19.0	6.35	8.00	40.0
		2	257	18.9	6.17	8.00	39.0
	1	0	200	23.1	6.61	9.00	38.0
		1	200	20.8	6.42	8.00	38.0
		2	202	21.4	6.23	8.00	38.0
ControlledMotLatin	0	0	0	NaN	NaN	NaN	NaN
		1	0	NaN	NaN	NaN	NaN
		2	0	NaN	NaN	NaN	NaN
	1	0	199	16.8	6.62	8.00	38.0
		1	69	17.6	5.47	8.00	31.0
		2	202	16.6	6.19	8.00	34.0
AutoMotLatin	0	0	0	NaN	NaN	NaN	NaN
		1	0	NaN	NaN	NaN	NaN
		2	0	NaN	NaN	NaN	NaN
	1	0	199	27.1	7.49	8.00	40.0
		1	69	21.2	7.61	8.00	37.0
		2	202	23.2	7.11	8.00	40.0

Appendix 20

Data Storage Fact Sheets

Data Storage Fact Sheet – main study on cognitive transfer

Doctoral dissertation: Non scholae sed vitae? The Cognitive Transfer Effects of Studying Classical Languages in Flemish Secondary Education

Author: Alexandra Vereeck

Date: 8th of April, 2025

1. Contact details

=====

1a. Main researcher

- name: Alexandra Vereeck
- address: Groot-Brittanniëlaan 45, 9000 Ghent
- e-mail: alexandra.vereeck@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Evy Woumans
- address: Groot-Brittanniëlaan 45, 9000 Ghent
- e-mail: evy.woumans@ugent.be

If a response is not received when using the above contact details, please contact the Data Steward of the Faculty of Arts and Philosophy by sending an email to rdm.support@ugent.be.

2. Information about the datasets to which this sheet applies

=====

* Reference of the publication in which the datasets are reported:
as of yet unpublished

* Which datasets in that publication does this sheet apply to?
all data in Chapters 8 and 9 of the dissertation

3. Information about the files that have been stored

=====

3a. Raw data

* Have the raw data been stored by the main researcher? ☒ YES / ☐ NO
If NO, please justify:

* On which platform are the raw data stored?

- ☒ researcher PC
- ☒ research group file server
- ☒ other (specify): external drives

* Who has direct access to the raw data (i.e., without intervention of another person)?

- ☒ main researcher
- ☒ responsible ZAP
- ☐ all members of the research group
- ☐ all members of UGent
- ☐ other (specify):

3b. Other files

* Which other files have been stored?

- ☐ file(s) describing the transition from raw data to reported results. Specify:
- ☒ file(s) containing processed data. Specify: SPSS files
- ☐ file(s) containing analyses. Specify:
- ☐ files(s) containing information about informed consent
- ☐ a file specifying legal and ethical provisions
- ☐ file(s) that describe the content of the stored files and how this content should be interpreted.
Specify:
- ☐ other files. Specify:

* On which platform are these other files stored?

- ☒ individual PC
- ☒ research group file server
- ☒ other: external drives

* Who has direct access to these other files (i.e., without intervention of another person)?

- ☒ main researcher
- ☒ responsible ZAP
- ☐ all members of the research group
- ☐ all members of UGent
- ☐ other (specify):

4. Reproduction

=====

* Have the results been reproduced independently? ☐ YES / ☒ NO

* If yes, by whom (add if multiple):

- name:
- address:
- affiliation:
- e-mail:

Data Storage Fact Sheet – complementary study on teaching method

Doctoral dissertation: Non scholae sed vitae? The Cognitive Transfer Effects of Studying Classical Languages in Flemish Secondary Education
Author: Alexandra Vereeck
Date: 8th of April, 2025

1. Contact details

=====

1a. Main researcher

- name: Alexandra Vereeck
- address: Groot-Brittanniëlaan 45, 9000 Ghent
- e-mail: alexandra.vereeck@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Katja De Herdt
- address: Blandijnberg 2, 9000 Ghent
- e-mail: katja.deherdt@ugent.be

If a response is not received when using the above contact details, please contact the Data Steward of the Faculty of Arts and Philosophy by sending an email to rdm.support@ugent.be.

2. Information about the datasets to which this sheet applies

=====

* Reference of the publication in which the datasets are reported:

De Herdt, K., & Vereeck, A. (2024). Van grammaticaal naar genietend lezen? Recente tendensen in de Vlaamse lespraktijk Latijn. *Lampas*, 57(2), 125–151.

* Which datasets in that publication does this sheet apply to?
all data in Chapter 7 of the dissertation

3. Information about the files that have been stored

=====

3a. Raw data

* Have the raw data been stored by the main researcher? ☒ YES / ☐ NO
If NO, please justify:

* On which platform are the raw data stored?

- ☒ researcher PC
- ☒ research group file server
- ☒ other (specify): external drives

* Who has direct access to the raw data (i.e., without intervention of another person)?

- ☒ main researcher
- ☒ responsible ZAP
- ☐ all members of the research group
- ☐ all members of UGent
- ☐ other (specify):

3b. Other files

* Which other files have been stored?

- ☐ file(s) describing the transition from raw data to reported results. Specify:
- ☐ file(s) containing processed data. Specify:
- ☒ file(s) containing analyses. Specify: OneNote files
- ☐ files(s) containing information about informed consent
- ☐ a file specifying legal and ethical provisions
- ☐ file(s) that describe the content of the stored files and how this content should be interpreted.
Specify:
- ☐ other files. Specify:

* On which platform are these other files stored?

- ☒ individual PC
- ☒ research group file server
- ☒ other: external drives

* Who has direct access to these other files (i.e., without intervention of another person)?

- ☒ main researcher
- ☒ responsible ZAP
- ☐ all members of the research group
- ☐ all members of UGent
- ☐ other (specify):

4. Reproduction

=====

* Have the results been reproduced independently? ☐ YES / ☒ NO

* If yes, by whom (add if multiple):

- name:
- address:
- affiliation:
- e-mail:

