Eindrapportage Effectstudie Rekenfaculteit

Maart - 2026

STICHTING
DE VERRE
BERGEN



Inhoudsopgave

1.	**Samenvatting**	**4**
2.	**Achtergrond**	**6**
2.1	Omvang en aard van onderwijsachterstanden in Rotterdam	6
2.2	(High Dosage) Tutoring: een bewezen effectieve interventie	7
2.3	Nederlandse toepassing van het *Saga Education Program*	8
2.4	Doelstellingen huidig onderzoek	9
3.	**Werkwijze Rekenfaculteit**	**11**
3.1	Doelgroep	11
3.2	Aanpak	12
4.	**Onderzoeksmethode**	**15**
4.1	Onderzoeksopzet	15
4.2	Leerlingen en hun scholen	16
4.3	Meetinstrumenten	17
4.4	Dataverzameling/afnamecondities	21
4.5	Statistische analyses	22
4.6	Goedkeuring ethische commissie	27
5.	**Resultaten**	**28**
5.1	Beschrijvende statistieken	28
5.2.	BSA-resultaten	29
5.3.	LVS-Resultaten	35
5.4.	Resultaten rekenangst en rekenattitude	38
5.5.	Resultaten leerlingkenmerken: voor wie werkt de Rekenfaculteit?	42
5.6.	Resultaten programmakenmerken	44
6.	**Discussie**	**45**
6.1.	Het effect van de Rekenfaculteit op rekenvaardigheid	45
6.2.	Het effect van de Rekenfaculteit in vergelijking met andere rekeninterventies	45

6.3. Het effect van de Rekenfaculteit op rekenattitude en rekenangst 46
6.4. Invloed leerling- en programmakenmerken op de groei van rekenvaardigheid 48
6.5. Sterke punten en beperkingen 49
6.6. Conclusie 50

7. Literatuurlijst 52

8. Bijlagen 55

8.1 Bijlage 1 - Onderzoeksvragen 55
8.2 Bijlage 2 - Statistische achtergrond en toelichting 57

1. Samenvatting

Onderwijsachterstanden vormen een structureel en urgent probleem in Rotterdam, met name in Rotterdam-Zuid, waar bijna vier op de tien kinderen een onderwijsachterstand heeft. Deze achterstanden hangen sterk samen met sociaaleconomische factoren en belemmeren kinderen in het benutten van hun cognitieve potentieel. Vooral op het gebied van rekenen zijn de verschillen met het landelijk gemiddelde aanzienlijk.

Om kinderen in Rotterdam-Zuid beter te leren rekenen, werd in 2015 de Rekenfaculteit opgericht. Geïnspireerd door een succesvol programma in de Verenigde Staten biedt de Rekenfaculteit sindsdien *High Dosage Tutoring* (HDT) rekenlessen aan op inmiddels tien basisscholen. Dit zijn intensieve, extracurriculaire tutorlessen tijdens schooltijd, gericht op het verbeteren van rekenvaardigheden. In 2023 is het onderzoek naar de effectiviteit van de Rekenfaculteit gestart. Deze eindevaluatie doet verslag van de resultaten van de onderzoeksjaren 2023-2024 en 2024-2025.

In totaal namen ruim 1200 leerlingen uit groep 5 en 7 deel aan het onderzoek. Het onderzoek had een quasi-experimenteel design met twee groepen: leerlingen van scholen die deelnamen aan de Rekenfaculteit (interventiegroep) en leerlingen van vergelijkbare scholen zonder Rekenfaculteit (controlegroep). Het onderzoek liep over twee schooljaren (2023–2025), waarbij rekenvaardigheid op drie momenten in het schooljaar is gemeten: bij de start (T0), halverwege (T1) en aan het einde (T2).

Het evaluatieonderzoek richt zich op het vaststellen van het effect van de Rekenfaculteit op de rekenresultaten van deelnemende leerlingen. Dit effect is gedefinieerd als de extra groei in rekenvaardigheid, gemeten met de Basic Skills Assessment (BSA) en Leerlingvolgsysteem (LVS)-toetsen (Cito & IEP), die leerlingen tijdens hun deelname aan de Rekenfaculteit maken. Er is specifiek gekeken naar de 'extra groei' gedurende één schooljaar bovenop de verwachte vooruitgang van vergelijkbare leerlingen die niet aan de Rekenfaculteit hebben deelgenomen.

De resultaten laten zien dat de Rekenfaculteit in beide schooljaren (2023–2024 en 2024–2025) een duidelijk positief effect had op de rekenvaardigheid van leerlingen in zowel groep 5 als groep 7. Zowel uit de BSA-metingen als uit de LVS-resultaten (Cito en IEP) blijkt dat leerlingen die deelnamen aan de Rekenfaculteit gemiddeld aanzienlijk meer vooruitgang boekten dan hun leeftijdsgenoten in de vergelijkingsgroepen, waarmee het programma substantieel bijdraagt aan het verkleinen van rekenachterstanden.

De effecten zijn bovendien groter dan doorgaans gevonden wordt in andere onderwijsinterventies. Voor de BSA varieerden de *effect sizes* van 0,60 tot 0,82, wat neerkomt op middelgrote tot grote effecten. Ter vergelijking: overzichtsstudies naar interventies voor kinderen met leerachterstanden rapporteren meestal *effect sizes* rond 0,20 tot 0,30. Ook binnen de categorie tutoringprogramma's

zijn de resultaten van de Rekenfaculteit bovengemiddeld. Daarmee lijkt de Rekenfaculteit een krachtige interventie om rekenachterstanden in het primair onderwijs te verkleinen.

In tegenstelling tot het duidelijke effect op de groei van rekenvaardigheid, is het effect van de Rekenfaculteit op veranderingen in rekenangst en rekenattitude minder eenduidig. In groep 5 nam rekenangst tijdens beide schooljaren sterker af bij leerlingen van de Rekenfaculteit dan bij leerlingen van de vergelijkingsscholen, maar in groep 7 werd dit effect niet gevonden. Voor rekenattitude werden in beide school- en leerjaren geen significante verschillen gevonden tussen de Rekenfaculteit en de vergelijkingsscholen. Het is mogelijk dat de Rekenfaculteit slechts een beperkt of geen eenduidig effect heeft op veranderingen in rekenangst en rekenattitude. Een andere verklaring is dat het gebruikte meetinstrumentarium niet voldoende aansloot bij de doelgroep, waardoor subtiele veranderingen bij deze jonge leerlingen met beperkte taalvaardigheid mogelijk minder goed zijn gemeten.

Aanvullend bleek de invloed van de in deze studie geïncludeerde leerling- en programmakenmerken op het effect van de Rekenfaculteit klein: verschillen in groei in rekenvaardigheid konden met factoren als aanwezigheid tijdens de lessen of tutorkwaliteit slechts beperkt worden verklaard.

De resultaten van de Rekenfaculteit zijn voor beide leerjaren veelbelovend. Er blijven echter nog een aantal vragen onbeantwoord. Zo is het nog onduidelijk of de effecten op de lange termijn behouden blijven en hoe verbeteringen in rekenvaardigheid samenhangen met de taalontwikkeling van leerlingen. Daarnaast hebben de observaties van de tutorlessen veel waardevolle informatie opgeleverd. Hiermee kan worden onderzocht in hoeverre de werkzame elementen van de Rekenfaculteit daadwerkelijk in de praktijk worden gebracht, en welke onderdelen van het programma het meest bijdragen aan de groei in rekenvaardigheid.

2. Achtergrond

2.1 Omvang en aard van onderwijsachterstanden in Rotterdam

Een onderwijsachterstand verwijst naar een achterblijvende ontwikkeling van kinderen ten opzichte van leeftijdsgenoten en hangt vaak samen met sociaaleconomische factoren, zoals het opleidingsniveau van ouders, het gezinsinkomen en de thuissituatie (Ledoux, 2021). Kinderen met een onderwijsachterstand presteren onder hun cognitieve potentieel, waardoor hun talenten niet volledig tot ontwikkeling komen. Deze achterstanden zijn doorgaans moeilijk in te halen en bemoeilijken zowel het behalen van de leerdoelen in het basisonderwijs als de doorstroom naar passend vervolgonderwijs (Jepma et al., 2024; Zumbuehl & Dilling, 2020).

Onderwijsachterstanden vormen een hardnekkig probleem in Rotterdam: 31% van de Rotterdamse leerlingen heeft een onderwijsachterstand, tegenover 25% gemiddeld in de G4 en 15% landelijk (Centraal Bureau voor de Statistiek, z.d.; Gemeente Rotterdam, z.d.). Rekenprestaties zijn daarbij een belangrijk aandachtspunt: Rotterdamse leerlingen scoren aantoonbaar lager dan hun leeftijdsgenoten in de rest van Nederland[1] (Gemeente Rotterdam, z.d.).

In Rotterdam-Zuid is de situatie nog zorgwekkender. Volgens de *Monitor scholen op Zuid* heeft daar naar schatting 39% van de kinderen een onderwijsachterstand (De Boom et al., 2024). Deze cijfers illustreren de bredere problematiek van structurele (kansen)ongelijkheid in Rotterdam-Zuid en vergelijkbare stedelijke gebieden, waar sociaaleconomische factoren de kans op onderwijsachterstanden aanzienlijk vergroten (Centraal Bureau voor de Statistiek, z.d.; De Boom et al., 2024).

Om achterstanden in rekenvaardigheid terug te dringen en de onderwijskansen van kinderen in Rotterdam-Zuid te vergroten, werd in 2015 de Rekenfaculteit opgericht. De Rekenfaculteit biedt *High Dosage Tutoring* (HDT): een intensieve vorm van tutoring in kleine groepjes, geïntegreerd in de schooldag, met een hoge frequentie en een langdurige inzet. Leerlingen volgen bij de Rekenfaculteit gedurende een schooljaar drie tot vijf extra rekenlessen per week, in duo's begeleid door één tutor. De lessen vinden onder schooltijd plaats, aanvullend op het bestaande rekenaanbod[2], en worden gegeven door speciaal getrainde tutoren. Het programma richt zich op het versterken van de rekenvaardigheid van basisschoolleerlingen (groep 5 en 7) in Rotterdam-Zuid.

1 10% haalt het referentie basisniveau (1F) niet, tegenover 8% landelijk (Staat van het Onderwijs Rotterdam, 2021-2022).

2 Bij het merendeel van de scholen vonden de rekenlessen extracurriculair plaats: buiten het reguliere lesprogramma, maar wel onder schooltijd (90%). Bij een klein deel van de scholen werden de lessen (deels) geïntegreerd in het reguliere rooster (10%).

2.2 (High Dosage) Tutoring: een bewezen effectieve interventie

Tutoring[3] is een bewezen effectieve onderwijsinterventie gericht op het versterken van de basisvaardigheden in rekenen en taal. Recente, internationale meta-analyses laten zien dat de *effect sizes* (zie Kader A voor de uitleg van *effect sizes*) van tutoring substantieel zijn. Zo analyseerden Nickow et al. (2024) 89 studies naar uiteenlopende vormen van tutoring, variërend in intensiteit, frequentie, duur en tutorprofiel[4]. Zij rapporteerden een gemiddelde *effect size* van 0,29[5], en 0,27 specifiek voor rekenen. De positieve effecten op rekenvaardigheid bleken bovendien consistent: verschillende programma's leverden vergelijkbare effecten op, terwijl taalprogramma's meer variatie in uitkomsten lieten zien. Kraft et al. (2024) rapporteerden zelfs nog grotere effecten in een meta-analyse van 265 tutoringprogramma's, uitgevoerd in het K–12 onderwijs (5–18 jaar) en gericht op lezen of wiskunde: gemiddeld 0,42[4], en 0,39 specifiek voor rekenen.

Onderzoek toont daarnaast aan dat specifieke varianten van tutoring zich onderscheiden door een hogere effectiviteit. Dietrichson et al. (2021) voerden een meta-analyse[6] uit van 205 studies naar uiteenlopende interventies in verschillende vakgebieden[7] voor leerlingen met risico op leerachterstanden. Zij concludeerden dat HDT en *peer-assisted instruction* (waarbij leerlingen elkaar tutoren) tot de meest effectieve aanpakken behoren. Ook in een meta-analyse specifiek gericht op rekeninterventies, blijkt tutoring het meest effectief. Pellegrini et al. (2021) analyseerden 87 studies naar verschillende typen interventies om rekenvaardigheden te verbeteren, waaronder tutoring, docentprofessionalisering en digitale hulpmiddelen. Tutoringprogramma's kwamen daarbij naar voren als het meest effectieve type rekeninterventie.

Kader A – Effect sizes

Een *effect size* is een gestandaardiseerde maat om de grootte van een verschil of verband aan te geven. In het huidige rapport worden effect sizes gebruikt om het verschil in rekenprestaties tussen de Rekenfaculteit en de vergelijkingsgroep uit te drukken. Omdat *effect sizes* in standaarddeviaties worden uitgedrukt, zijn ze goed vergelijkbaar tussen studies en interventies.

Volgens de klassieke richtlijnen van Cohen (1988) geldt: 0,20–0,49 = klein effect, 0,50–0,79 = middelgroot effect, en 0,80 of hoger = groot effect. In het onderwijs liggen deze grenzen doorgaans lager, vanwege de complexiteit van het veld en de beperkte schaalbaarheid van interventies. Effecten van ongeveer 0,20 worden daarom al als groot en beleidsrelevant beschouwd (Kraft, 2020).

3 In deze tekst wordt verwezen naar HDT wanneer uit de betreffende studies duidelijk blijkt dat het om deze specifieke vorm van tutoring gaat. Indien dit niet met zekerheid vastgesteld kan worden, wordt de bredere term 'tutoring' gehanteerd.

4 Tutoring door leraren, paraprofessionals, vrijwilligers, ouders of via *peer-to-peer* (kinderen onderling).

5 Hedges' g.

6 De effectgrootte is berekend met verschillende methoden, zoals Hedges' g, Cohen's d en Glass's δ. In de meta-analyse is rekening gehouden met deze variatie in berekeningswijzen.

7 Zoals taal, rekenen en algemene schoolvaardigheden.

In dit rapport wordt in de tekst waarin (inter)nationale literatuur wordt besproken uitgegaan van *Cohen's d* als standaardmaat voor effect sizes, tenzij anders vermeld. Om *Cohen's d* te berekenen wordt het verschil in gemiddelde groei tussen een experimentele en een vergelijkingsgroep gedeeld door de gemiddelde standaarddeviatie van alle leerlingen samen.

In het huidige onderzoek naar het effect van de Rekenfaculteit wordt gebruikgemaakt van *Glass' δ* als maat voor effect sizes. Deze maat lijkt sterk op *Cohen's d*, maar om *Glass' δ* te berekenen wordt het verschil in gemiddelde groei gedeeld door de standaarddeviatie van alléén de vergelijkingsgroep. Die standaarddeviatie is doorgaans betrouwbaarder, vooral in de vergelijkingssgroep van dit onderzoek die een grote landelijke populatie leerlingen betreft. Hierdoor is *Glass' δ* minder gevoelig voor meetfout dan *Cohen's d*.

2.3 Nederlandse toepassing van het *Saga Education Program*

Een internationaal voorbeeld van een effectief HDT-rekenprogramma is het *Saga Education Program* uit de Verenigde Staten (Cook et al., 2014), waarop de Rekenfaculteit is gebaseerd. Uit verschillende grootschalige *randomized controlled trials* (RCT's) blijkt dat dit programma effectief is, met *effect sizes* van 0,16 tot 0,37 voor de rekenprestaties van middelbare scholieren (Guryan et al., 2023). Het is echter weinig onderzocht of dergelijke effecten ook te realiseren zijn bij basisschoolleerlingen.

Een voorbeeld van een HDT-interventie in Nederland is het programma *The Bridge*, dat in Amsterdam en Haarlem wordt uitgevoerd. In Amsterdam werd het programma (tussen 2019-2022) via een RCT onderzocht op acht basisscholen onder groep-7-leerlingen met een rekenachterstand (Tijms & De Ree, 2023). De voltijdse variant (vijf dagen per week, een schooljaar lang) liet in het tweede onderzoeksjaar[8] substantiële effecten zien op de Cito-toets rekenen/wiskunde (*effect sizes* van 0,62 in jaar 2; overeenkomend met drie tot zes maanden extra schooltijd[9]).

In Haarlem is *The Bridge* geëvalueerd (tussen 2018 en 2020) met een cohortvergelijking op vijf basisscholen (Paulle, De Ree, Epema, & Speelman, 2020). Ook daar werden duidelijke leerwinsten gevonden (gemiddelde *effect size* 0,40), waarbij 40% van de achterstand op het landelijk gemiddelde werd ingehaald. Daarnaast verbeterden de schooladviezen van leerlingen aanzienlijk: het aandeel leerlingen met een advies lager dan VMBO-GT ging van 70–80% naar 40%.

8 In het eerste onderzoeksjaar, waarin COVID-19 een rol speelde, rapporteren de onderzoekers een effectgrootte van 0,26; in het derde jaar, met een lichtere driedaagse variant, was dit 0,24. Beiden liggen boven ons gehanteerde significantieniveau van $\alpha = 0{,}05$ en worden daarom niet als significant beschouwd.

9 Deze inschatting van leerwinst in maanden is gebaseerd op het principe van de didactische leeftijdsequivalent (DLE). Hoewel DLE's met voorzichtigheid moeten worden geïnterpreteerd (zie Evers & Resing, 2007), geven zij wel een bruikbare indicatie van de toegevoegde waarde van een interventie (zie o.a. Melis, Oosterveld & Schokker, 2012).

Voortbouwend op het succesvolle voorbeeld van het *Saga Education Program* is in 2015 het HDT-programma van de Rekenfaculteit opgezet, in samenwerking met een aantal scholen in de wijk Pendrecht (Rotterdam-Zuid). Het programma startte met het aanbieden van tutoring in groep 7, het laatste volledige schooljaar voor de eindtoets. In de jaren daarna is het programma stap voor stap verder verfijnd op basis van voortschrijdend inzicht praktijkervaring en nieuwe bevindingen uit literatuur en onderzoek. Toen de Rekenfaculteit uitbreidde naar nieuwe scholen, zijn de tutorlessen ook in groep 5 geïntroduceerd, omdat in dit schooljaar een belangrijke basis wordt gelegd in rekenvaardigheid voor de rest van de basisschoolloopbaan. Daarnaast is er een eigen curriculum ontwikkeld. Tutoren gebruiken informatie uit rekentoetsen om leerplannen te maken, die ze uitvoeren met Rekenfaculteit-lesmateriaal. Deze ontwikkeling is vormgegeven op basis van eigen praktijkervaring en met voorbeelden van *best practices* uit de sector. Het curriculum legt een basis om op grotere schaal en in een setting met meer leerlingen per tutor (1 op 4) kwaliteit te kunnen borgen.

In de ontwikkelingen van de Rekenfaculteit zijn verschillende werkzame elementen uit internationaal onderzoek naar HDT verwerkt. Zo is zowel de inzet van leerkrachten of paraprofessionals (zoals onderwijsassistenten) als tutor effectief[10], evenals een zo vroeg mogelijk startmoment (groep 5), een intensief aanbod van minimaal drie keer per week en het geven van de lessen onder schooltijd (Nickow et al., 2024). Onderzoek van Kraft et al. (2024) laat bovendien zien dat rekentutoring in het basisonderwijs vooral effectief is in groep 5 tot en met 7 (*effect size* 0,44 tegenover 0,33 in de onderbouw).

2.4 Doelstellingen huidig onderzoek

In 2023 is een tweejarig onderzoek gestart naar de effectiviteit van de Rekenfaculteit. Deze eindevaluatie doet verslag van de resultaten van zowel het eerste onderzoeksjaar (2023-2024) als het tweede onderzoeksjaar (2024-2025).

Het doel van deze evaluatie is tweeledig. Enerzijds biedt het onderzoek inzichten waarmee het tutoringprogramma van de Rekenfaculteit verder kan worden geoptimaliseerd. Anderzijds levert het, vanuit wetenschappelijk perspectief, nieuwe kennis op over de werking en effectiviteit van HDT. Want hoewel de eerdere (meta-)studies waardevolle inzichten geven in de effectiviteit van (reken)tutorinterventies, bestaan er nog een aantal kennislacunes, te weten:

Beperkte aandacht voor passende meetinstrumenten: eerder onderzoek maakte veelal gebruik van gestandaardiseerde toetsen. Hoewel deze zeker waardevolle informatie bieden, is het ook relevant om effecten te meten met instrumenten die beter aansluiten bij de doelen en doelgroep van de interventie. Voor dit onderzoek is daarom een toetsinstrumentarium ontwikkeld dat past bij het niveau en de belevingswereld van leerlingen met het risico op een onderwijsachterstand.

10 De effecten van tutoring in kleine groepjes zijn vergelijkbaar of de tutor nu een leerkracht is of een onderwijsassistent. Dit betekent dat een tutor geen bevoegde leerkracht hoeft te zijn.

Onvoldoende kennis over niet-cognitieve uitkomstmaten: eerder onderzoek richtte zich vooral op cognitieve uitkomstmaten, zoals rekenvaardigheid. Er mag echter verwacht worden dat HDT ook effect heeft op niet-cognitieve factoren, zoals rekenattitude en rekenangst (De Ree et al., 2023). In dit onderzoek worden beide typen uitkomstmaten meegenomen.
Geen zicht op langetermijneffecten: er is bewijs voor positieve langetermijneffecten van HDT, maar dit betreft voornamelijk taalvaardigheid. Voor rekenen is hierover nog weinig bekend (Dietrichson et al., 2021). In dit onderzoek wordt ook gekeken naar de effecten op langere termijn, al zijn deze nog niet opgenomen in de huidige rapportage[11].
Weinig Nederlandse evidentie: tot nu toe is er slechts beperkt Nederlands onderzoek uitgevoerd. Omdat de effecten van tutoringprogramma's kunnen verschillen per context (Kraft et al., 2024), is het belangrijk de effectiviteit ook op andere Nederlandse scholen te onderzoeken.

Dit onderzoek draagt bij aan het verkleinen van deze kennislacunes. De volgende (hoofd)vraag staat centraal:

In hoeverre heeft de Rekenfaculteit effect op de rekenresultaten van deelnemende leerlingen?

Als de Rekenfaculteit effectief blijkt, kan dat belangrijke implicaties hebben: rekenvaardigheid is immers een belangrijke voorspeller van schoolsucces en doorstroommogelijkheden (NRO, z.d.). Uit een omvangrijke reeks onderzoeken blijkt bovendien dat betere onderwijsprestaties leiden tot meer economische zelfstandigheid, een hoger inkomen, minder criminaliteit en verslaving, meer gezondheid en zelfs een langere levensduur (Centraal Planbureau, 2018). Voor kinderen in een achterstandspositie is onderwijs daarmee een belangrijk middel om hun kansen te vergroten. Het verkleinen van rekenachterstanden bij kinderen in Rotterdam-Zuid kan dus niet alleen hun schoolsucces versterken, maar ook bijdragen aan grotere kansengelijkheid op de lange termijn.

Naast de hoofdonderzoeksvraag richt deze evaluatie zich ook op andere uitkomsten, zoals het effect van de Rekenfaculteit op veranderingen in rekenattitude en rekenangst bij leerlingen. Daarnaast wordt onderzocht welke leerling- en programmakenmerken samenhangen met de effectiviteit van de Rekenfaculteit. Een volledig overzicht van alle onderzoeksvragen is te vinden in Bijlage 1.

11 Deze resultaten volgen in 2026

3. Werkwijze Rekenfaculteit

3.1 Doelgroep

De Rekenfaculteit richt zich op scholen waar veel leerlingen het risico lopen op een onderwijsachterstand. In de schooljaren 2023-2025 verzorgde de Rekenfaculteit op tien basisscholen in Rotterdam-Zuid rekentutorlessen. Het programma werd in 16 groepen onder schooltijd uitgevoerd en bereikte in totaal ruim 650 kinderen van groep 5 en 7. De deelnemende scholen hebben een schoolweging van 35 of hoger (Kader B), wat wijst op een verhoogd risico op onderwijsachterstanden.

Alleen leerlingen in groep 5 en groep 7 van de deelnemende scholen krijgen tutorlessen van de Rekenfaculteit. In 2015 begon de Rekenfaculteit met (alleen) lessen voor groep 7, omdat deze groep het beste aansloot bij de doelgroep van een vergelijkbaar programma in de Verenigde Staten. Vanaf 2020 werd groep 5 ook toegevoegd, omdat op dit niveau cruciale leermomenten ('rekendrempels') plaatsvinden die essentieel zijn voor de verdere rekenontwikkeling van de leerlingen[12].

Kader B - Schoolweging

De schoolweging (20-40) wordt als maat gebruikt door de Inspectie van het Onderwijs om vast te stellen hoe de leerlingpopulatie eruitziet. Het Centraal Bureau voor de Statistiek (CBS) berekent de schoolweging van een school op basis van de volgende kenmerken:

- het opleidingsniveau van de ouders
- het gemiddeld opleidingsniveau van alle moeders op school
- het land van herkomst van de ouders
- de verblijfsduur van de moeder in Nederland
- of ouders in de schuldsanering zitten

Hoe hoger de schoolweging, hoe groter het risico op onderwijsachterstanden. De meeste scholen hebben een schoolweging rond het gemiddelde, dat iets onder 30 ligt. Scholen met een zeer lage of juist zeer hoge schoolweging komen relatief weinig voor. In totaal hebben 679 Nederlandse basisscholen een schoolweging van 35 of hoger; dat is ongeveer 10% van het totaal (Inspectie van het Onderwijs). Ongeveer de helft van de Rotterdamse basisscholen heeft een schoolweging van 35 of hoger (openbare onderwijsdata DUO).

12 Volgens de conceptkerndoelen Rekenen en Wiskunde 2025 (SLO) is groep 5 een sleutelmoment in de rekenontwikkeling. Leerlingen maken hier de overgang van concreet rekenen naar abstractere strategieën, werken met grotere getallen, complexe bewerkingen en verhoudingen, en leggen de basis voor breuken, meten en meetkunde. Dit vormt het fundament dat alle leerlingen in het primair onderwijs zouden moeten beheersen.

3.2 Aanpak

Opzet programma Rekenfaculteit

De Rekenfaculteit is een HDT-programma, waarbij leerlingen gedurende één schooljaar intensieve rekenlessen volgen. Deze lessen worden aan de hele klas aangeboden, dus ook aan leerlingen die geen achterstand hebben. Ze krijgen drie tot vijf rekenlessen per week, waarbij ze één-op-twee[13] begeleiding krijgen van een vaste tutor. Een tutorles duurt 45 minuten in groep 5 en 1 uur in groep 7. De tutorelessen zijn aanvullend op het reguliere rekenonderwijs en vinden plaats onder schooltijd zonder andere belangrijke vakken, zoals taal, te vervangen. Wanneer de tutorlessen plaatsvinden verschilt per school: de meeste scholen organiseren dit via verlengde schooltijd, anderen maken ruimte binnen het bestaande rooster, bijvoorbeeld door thematisch te werken waardoor tijdwinst ontstaat. De Rekenfaculteit streeft ernaar de lessen volledig buiten de reguliere lessen rekenen aan te bieden (geheel extra-curriculair)[14] om het aantal uren rekenen voor de leerlingen te maximaliseren.

De Rekenfaculteit vormt in overleg met de basisschoolleerkrachten zorgvuldig de tutor-leerlingkoppels, waarbij rekening wordt gehouden met overeenkomende rekenvaardigheden en gedragscompatibiliteit. Elke tutor begeleidt gemiddeld vier tot zes groepjes. Om voor continuïteit en stabiliteit in de begeleiding te zorgen worden wisselingen in groepjes en tutoren tijdens het schooljaar zoveel mogelijk vermeden.

Inhoud programma Rekenfaculteit

Tutoren van de Rekenfaculteit ontwerpen tutorlessen zelfstandig, binnen een vaste lesopbouw en met behulp van een vaste structuur (Kader C). Tutoren gebruiken lesmateriaal dat door de Rekenfaculteit is aangeboden om de lessen te ontwerpen. De lessen richten zich op vier rekendomeinen: getallen basis, meten, verhoudingen en verbanden, en zijn in overeenstemming met de kerndoelen van het basisonderwijs. Tutoren hebben toegang tot een breed scala aan materialen, zoals rekenrekjes, fiches en getallenlijnen, om het leren zo concreet mogelijk te maken. Ze leren leerlingen vaste strategieën, zoals het gebruik van de vijf- en tienstructuur voor getalbegrip en het toepassen van de getallenlijn bij optellen met overschrijding (zoals van het tiental). Iedere leerling werkt met een eigen leerplan, die tutoren in samenwerking met de groepsleerkracht opstellen.

---

13 Dit is het uitgangspunt. In de praktijk komt het soms voor dat er één-op-drie gewerkt wordt.

14 In groep 5 is het streven om de Rekenfaculteit volledig buiten de reguliere lestijd (geheel extra-curriculair) aan te bieden; in groep 7 wordt de Rekenfaculteit bij een minderheid (5 scholen) zowel tijdens als buiten reguliere lestijd (gedeeltelijk extra-curriculair) aangeboden.

Kader C – Lesopbouw Rekenfaculteit

1. **Lesvoorbereiding:** de tutor selecteert lesstof die aansluit bij het niveau van de leerling, zodat deze voldoende wordt uitgedaagd.
2. **Vliegende start:** de les begint met vijf minuten zelfstandig werken aan een opdracht om de voorkennis van de leerlingen te activeren.
3. **Nabespreking** 'vliegende start': de tutor bespreekt de vliegende start, stelt zo nodig het lesplan bij, en stelt per leerling het lesdoel vast.
4. **De les:** de leerlingen werken -zowel zelfstandig als in samenwerking- aan opdrachten die zijn afgestemd op het lesdoel.
5. **Finish:** de les eindigt met een korte zelfstandige opdracht die de tutor inzicht geeft in de voortgang van de leerling en richting geeft aan de volgende les

De werkzame elementen die in elke les aanwezig moeten zijn staan beschreven in Kader D. Tutoren kunnen informatiekaarten (zogenoemde 'kwaliteitskaarten') raadplegen voor een overzicht van de richtlijnen en praktische tips per werkzaam element.

Kader D – Werkzame elementen Rekenfaculteit

- **Lesstructuur:** elke les heeft een opbouw volgens vaste elementen
- **Maatwerk:** lessen worden afgestemd op individuele leerplannen (differentiatie), met stapsgewijze ondersteuning (scaffolding) om leerlingen zelfstandig verder te laten groeien.
- **Effectieve leertijd:** maximale benutting van lestijd door een strak lesritme en minimale verstoringen.
- **Tutor-leerlingband:** empathische en consistente aanpak, respectvolle communicatie en samenwerken vanuit het idee dat je rekenvaardigheid kunt ontwikkelen door inzet en doorzettingsvermogen (growth mindset).
- **Positieve rekenattitude:** rekenen aantrekkelijk maken door nieuwsgierigheid te prikkelen, succeservaringen te bieden, en reflectie te stimuleren.

Uitvoering

Het tutorenteam bestaat grotendeels uit parttimers met uiteenlopende achtergronden en opleidingsniveaus. Het streven ligt bij een hbo-denkniveau, maar er zijn ook tutoren met een mbo-diploma of tutoren die direct na het voortgezet onderwijs instromen. De selectie bestaat uit een rekentoets en een proefles, waarmee zowel het vakinhoudelijke niveau als het contact met kinderen worden beoordeeld. Nieuwe tutoren worden intern opgeleid middels een twee weken durende training en krijgen gedurende het jaar aanvullende scholing via studiedagen, thematische werkgroepen en feedbackgesprekken. De resultaten hiervan worden vastgelegd en vormen de basis voor een gesprekkencyclus om ambities en aandachtspunten te bespreken. Samenwerking en kennisdeling vinden plaats via *peer-to-peer* sessies en zogenoemde '*morning meetings*'.

Teamleiders van de Rekenfaculteit spelen een cruciale rol door tutoren aan te sturen, feedback te geven, en te fungeren als eerste aanspreekpunt voor scholen. Ook zijn zij de schakel tussen tutoren en de programmadirectie, wat de consistentie en kwaliteit van het tutoronderwijs waarborgt.

Binnen de Rekenfaculteit werken tutoren, teamleiders, leerkrachten en ouders nauw met elkaar samen. Tutoren worden intensief ondersteund door hun teamleiders via regelmatige observaties, feedback, intervisie, en gezamenlijke lesvoorbereiding. Ouders worden tweewekelijks telefonisch door tutoren geïnformeerd en hun inzichten worden meegenomen om lessen beter af te stemmen op de behoeften van hun kind. Continu contact tussen tutoren en leerkrachten zorgt voor adequate afstemming en doorlopende voortgangsmonitoring.

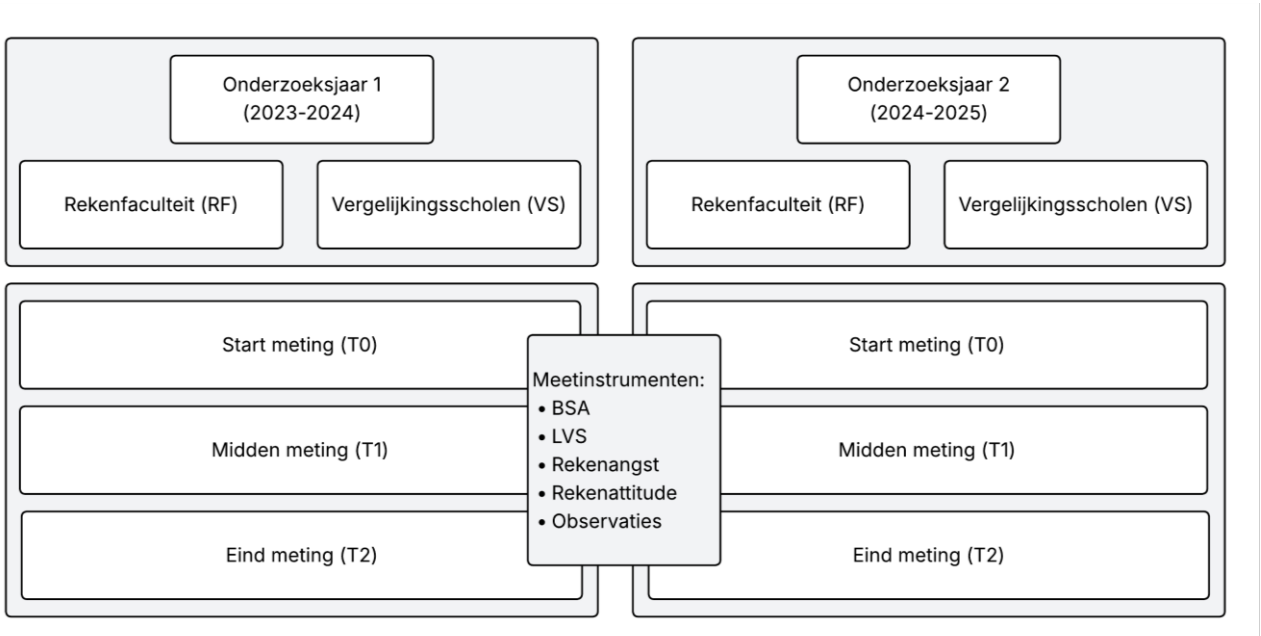
4. Onderzoeksmethode

4.1 Onderzoeksopzet

Het onderzoek is opgezet als een quasi-experimentele studie, waarbij twee groepen leerlingen met elkaar worden vergeleken:

- Rekenfaculteit: bestaande uit leerlingen op scholen die deelnamen aan de Rekenfaculteit – oftewel de 'interventiegroep';
- Vergelijkingsscholen: bestaande uit leerlingen van vergelijkbare scholen die niet deelnamen aan de Rekenfaculteit – oftewel de 'controlegroep'.

Er is dus *geen* sprake van willekeurige toewijzing (zoals bij een volledige experimentele studie) aan de groepen, maar wel van een systematische vergelijking tussen beide groepen. Het onderzoek liep over twee schooljaren (2023-2025), waarbij in het eerste jaar ruimte was om de meetinstrumenten te ontwikkelen en valideren. Rekenvaardigheid, rekenangst en rekenattitude zijn gemeten op drie momenten gedurende het schooljaar: aan het begin (T0), halverwege (T1) en aan het einde van het schooljaar (T2). De onderzoeksopzet is weergegeven in Figuur 1. De meetinstrumenten worden beschreven in paragraaf 4.3.



Figuur 1. *Overzicht van de onderzoeksopzet*

4.2 Leerlingen en hun scholen

Rekenfaculteit

Ruim 650 leerlingen van de Rekenfaculteit deden mee aan het onderzoek gedurende twee schooljaren. Zie Bijlage 2, sectie 1 voor een gedetailleerd overzicht van het aantal leerlingen per groep en schooljaar. De tutorlessen werden verzorgd op tien basisscholen in Rotterdam, verspreid over drie wijken in Rotterdam-Zuid: de Afrikaanderwijk, Bloemhof en Pendrecht. Bij het merendeel van de scholen vonden de rekenlessen extracurriculair plaats – buiten het reguliere lesprogramma maar wel onder schooltijd (90%). Bij een klein aantal scholen werden de lessen deels geïntegreerd in het reguliere rooster (10%). Leerlingen uit groep 5 en groep 7 ontvingen gedurende het hele schooljaar rekentutorlessen in een 1-op-2 setting (70%), een 1-op-3 setting (27%) of een 1-op-4 setting (3%)[15]. Gemiddeld heeft 91% van de geplande tutorlessen plaatsgevonden[16]. In het schooljaar 2024-2025[17] hebben de leerlingen in groep 5 gemiddeld 162 tutorlessen (SD = 37,7) aangeboden gekregen, en in groep 7 ontvingen de leerlingen gemiddeld 141 lessen (SD = 17,2)[18].

Vergelijkingsscholen

In totaal namen ruim 550 leerlingen van zes vergelijkbare basisscholen in Rotterdam deel aan het onderzoek. Deze vergelijkingsscholen zijn aangedragen door twee Rotterdamse schoolbesturen — Stichting BOOR en PCBO — op basis van overeenkomsten in leerlingpopulatie en ligging in Rotterdam-Zuid. Hierbij werden de volgende criteria gebruikt: 1) De schoolweging[19] kwam overeen met die van de Rekenfaculteitscholen[20] en 2) De scholen gebruikten hetzelfde leerlingvolgsysteem (LVS)[21] als de Rekenfaculteitscholen. Op basis van deze criteria zijn zes vergelijkingsscholen geselecteerd.

Exclusiecriteria

Er zijn vooraf geen specifieke exclusiecriteria vastgesteld voor leerlingen of klassen, voor zowel de Rekenfaculteitscholen als de vergelijkingsscholen. Alle klassen zijn opgenomen in het onderzoek, ongeacht eventuele problemen op de scholen. Dit is belangrijk om verschillende redenen:

---

15 De gewenste groepssamenstelling is 1-op-2, maar door praktische omstandigheden komt het soms voor dat groepen kleiner (1-op-1) of groter (1-op-3 of 1-op-4) zijn. Dit biedt de mogelijkheid om te onderzoeken of groepsgrootte invloed heeft op de effecten van de interventie. In dit rapport is groepsgrootte daarom als programmakenmerk meegenomen in de analyses om te bekijken in hoeverre het de effecten van de Rekenfaculteit op rekenvaardigheid kan verklaren. In vervolgonderzoek zal dit structureel worden onderzocht met een RCT, waarin 1-op-2 wordt vergeleken met 1-op-4.

16 Lessen kunnen uitvallen door studiedagen, vieringen, uitjes en afwezigheid van leerkrachten. De Rekenfaculteit heeft als doelstelling dat 95% van de geplande lessen doorgaan.

17 Deze cijfers bleken voor het schooljaar 2023-2024 niet accuraat en betrouwbaar uit te rekenen; daarom worden hier alleen de cijfers uit 2024-2025 gerapporteerd.

18 In groep 7 krijgen de leerlingen 4 keer 1 uur les en in groep 5 krijgen de leerlingen 5 keer 45 minuten les. Hierdoor valt het gemiddeld aantal lessen voor groep 7 lager uit.

19 De schoolweging (20-40) wordt als maat gebruikt door de Inspectie van het Onderwijs om vast te stellen hoe de leerlingpopulatie eruitziet. Hoe hoger de schoolweging, des te groter het risico op onderwijsachterstanden.

20 De schoolwegingen van de Rekenfaculteit scholen lagen in de onderzoeksjaren tussen de 33 en 40. Eén school had een weging van onder de 35. Het gemiddelde en de mediaan lag bij een weging van 37.

21 Cito of IEP (Bureau ICE).

1. De uitdagingen op deze scholen zijn typerend voor de context waarin de Rekenfaculteit werkt en moet werken;
2. Er is minder inzicht in de omstandigheden op de vergelijkingsscholen. Mogelijk spelen daar vergelijkbare problemen als op de Rekenfaculteit, maar die blijven vaak onzichtbaar door beperkte aanwezigheid van de onderzoekers aldaar;
3. Ook vergelijkingsscholen die andere rekeninterventies uitvoerden zijn niet uitgesloten van deelname. Voor zover bekend betrof dit vooral kleinschalige ondersteuning, zoals *remedial teaching* voor zwakker presterende leerlingen. Er wordt verwacht - gezien de intensiteit van het programma - dat de resultaten van de Rekenfaculteit ondanks andere rekeninterventies alsnog beter zullen zijn.

Wanneer er sprake is van ernstige problematiek op een school waardoor het Rekenfaculteitprogramma niet naar behoren kan worden uitgevoerd, wordt de school toch meegenomen in de hoofdanalyses. Dit gebeurt volgens het *intention-to-treat*-principe, waarbij alle deelnemers worden meegenomen ongeacht de mate van uitvoering. Deze keuze is bewust gemaakt, omdat dergelijke problematiek — zoals een lerarentekort of onrust op school — een realistische afspiegeling vormt van de schoolcontext. Om te toetsen of deze situatie invloed heeft op de hoofdresultaten, worden alle analyses herhaald zonder de leerlingen van scholen waar het programma niet naar behoren kan uitgevoerd worden (zie paragraaf 4.2.4: Sensitiviteitanalyses).

4.3 Meetinstrumenten

Basic Skills Assessment (BSA)

De primaire uitkomstmaat voor rekenvaardigheid is de *Basic Skills Assessment* (BSA). De BSA bestaat uit drie toetsen die verspreid over het schooljaar worden afgenomen (start, midden, eind). De BSA is ontwikkeld door de Rekenfaculteit en gevalideerd om het startniveau en de voortgang van leerlingen met een relatief laag rekenvaardigheidsniveau nauwkeurig te volgen. De BSA is goed geschikt voor de doelgroep dankzij het eenvoudige en weinig talige karakter van de toets, en de met veel aandacht ontwikkelde voorbeelden die aansluiten bij de belevingswereld van deze leerlingen.

De BSA biedt betrouwbare informatie over zowel het beginniveau als de ontwikkeling van rekenvaardigheid gedurende het schooljaar. Het instrument wordt bij alle leerlingen afgenomen (in tegenstelling tot verschillende toetsen bij het LVS – Cito/IEP), waardoor er een uniforme meting van rekenvaardigheid voor alle leerlingen is.

De BSA bevat zowel meerkeuze- als open vragen, waarbij het type en de moeilijkheidsgraad van de vragen consistent zijn over de drie toetsen. De BSA bestaat uit deel A en deel B. Hoe deze delen tot stand zijn gekomen wordt uitgelegd in Kader E. Leerlingen van de Rekenfaculteit maakten beide delen, waarbij deel A een tijdslimiet heeft (maximaal 35 minuten voor groep 5 en 45 minuten voor

groep 7) en deel B zonder tijdslimiet is. Leerlingen van de vergelijkingsschool maakten alleen deel A.

De BSA werd gevalideerd door SDVB in samenwerking met Stichting BOOR en CitoLab. Dat betekent dat de toetsresultaten betrouwbaar inzicht geven in de groei van rekenvaardigheid. In Bijlage 2, sectie 2 wordt de validatie van de BSA nader toegelicht. Voor het ijken van de drie BSA-toetsen werd gebruikgemaakt van *Item Response Theory* (IRT). Daarmee kunnen toetsresultaten, ongeacht de versie van de toets, op één gemeenschappelijke schaal worden geplaatst, waardoor groei over tijd betrouwbaar kan worden gemeten (zie Bijlage 2, sectie 3). Door deze ijking zijn de scores onafhankelijk van de moeilijkheidsgraad van de afgenomen toets en weerspiegelen zij uitsluitend de vaardigheid van de leerling. De BSA-scores konden daardoor over de verschillende meetmomenten en over verschillende versies met elkaar worden vergeleken.

Leerlingvolgsystemen (LVS)

Naast de BSA heeft de Rekenfaculteit ook gegevens uit het LVS verzameld. Basisscholen zijn verplicht om de voortgang van leerlingen in taal en rekenen te volgen via een LVS, waarbij zij kunnen kiezen uit verschillende aanbieders. De deelnemende scholen gebruikten Cito (Leerling in Beeld) of IEP (Inzicht in Eigen Profiel), waardoor leerlingen één van beide toetsen maakten.

Hoewel dit leidt tot minder data per toets en dus minder statistische power[22], bieden de LVS-scores waardevolle aanvullende informatie. Ze versterken de resultaten door inzicht te geven in de consistentie van effecten tussen BSA en LVS. Bovendien zijn de LVS-scores relevant voor de terugkoppeling naar scholen (omdat zij hier zelf mee werken), omdat deze uiteindelijk bepalend zijn voor de verdere schoolloopbaan van de leerlingen.

Zowel Cito als IEP beschikken over uitgebreide landelijke normgroepen, waardoor het mogelijk was om de prestaties van Rekenfaculteit-leerlingen te vergelijken met die van een gematchte groep (op basis van vaardigheidsniveau en achterstandsscore).

Vragenlijsten over rekenattitude en rekenangst

Voor het meten van rekenangst en attitude zijn zorgvuldig geselecteerde, gevalideerde vragenlijsten gebruikt die zo goed mogelijk aansluiten bij de leeftijd en het ontwikkelingsniveau van de leerlingen.

Rekenattitude

De rekenattitude van de leerlingen werd gemeten met de *Math and Me Survey* (Adelson & McCoach, 2011) en werd op dezelfde momenten als de (BSA-)toets voor rekenvaardigheid afgenomen. Deze vragenlijst is in een eerdere studie door een expert vertaald naar het Nederlands (Jansen et al., 2013) en bestaat uit 18 items die elk een specifieke rekensituatie beschrijven. Deze items zijn verdeeld over twee subschalen: rekenzelfbeeld (8 items) en rekenplezier (10 items).

22 Statistische power is de kans dat een statistische toets een werkelijk bestaand effect detecteert.

Leerlingen beantwoordden de items op een 5-puntschaal, waarbij ze aangaven in hoeverre ze het eens waren met de beschreven situatie (1 = helemaal oneens, 5 = helemaal eens). Er werd een totaalscore berekend door de scores van de 18 items op te tellen, wat resulteerde in een minimale totaalscore van 18 en een maximale totaalscore van 90. Voor de analyses werd deze totaalscore gebruikt.

Rekenangst

Rekenangst bij leerlingen werd gemeten met de *Mathematics Anxiety Scale for Children* (MASC-NL), zoals vertaald door Jansen et al. (2013) en werd op dezelfde momenten als de (BSA-)toets voor rekenvaardigheid afgenomen. Deze vragenlijst bevat 23 items die elk een specifieke rekensituatie beschrijven. Leerlingen gaven op een 4-puntschaal aan hoe nerveus ze zich voelen in de geschetste situatie, waarbij 1 staat voor 'niet nerveus' en 4 voor 'heel erg nerveus'.

Na het invullen van de vragenlijst werd een totaalscore berekend door de scores op alle vragen op te tellen, met een mogelijke score variërend van 23 (minimale angst) tot 92 (maximale angst).

Observatie-instrument kwaliteit tutoring

Een observatie-instrument is door de Rekenfaculteit in samenwerking met iPabo ontwikkeld en gevalideerd (zie Bijlage 2, sectie 4) om de kwaliteit van de tutoring te beoordelen en vast te stellen in hoeverre tutoren de werkzame elementen van het programma adequaat uitvoeren (programma-integriteit).

Het 'observatie-instrument kwaliteit tutoring' werd vanaf schooljaar 2024-2025 door teamleiders van de Rekenfaculteit gebruikt om de kwaliteit van de tutorlessen te observeren en door het onderzoeksteam om programma-integriteit te meten.

Het gevalideerde observatie-instrument bestaat uit 36 items, verdeeld over vijf categorieën, die overeenkomen met de werkzame elementen van de Rekenfaculteit: Maatwerk (8 items), Effectieve leertijd (7 items), Band tussen tutor en leerling (6 items), Rekenattitude (8 items), en Structuur van de les (7 items). Elk item wordt beoordeeld met behulp van een 4-punts Likertschaal, waarbij de volgende kwalificaties worden gehanteerd: A = gedrag wordt niet of nauwelijks geobserveerd terwijl de situatie er wel om vraagt, B = gedrag wordt geobserveerd maar is meer zwak dan sterk, C = gedrag wordt geobserveerd maar is meer sterk dan zwak, en D = gedrag wordt geobserveerd en is sterk. Er is ook een optie 'Niet van toepassing' beschikbaar wanneer een item niet gescoord kan worden, bijvoorbeeld omdat het niet van toepassing is op de desbetreffende les.

Voor de eindscore zijn de beoordelingen van A t/m D omgezet naar een numerieke schaal van 1 t/m 10[23]. Voor elke tutorobservatie wordt een gemiddelde score berekend over de 35 items van het

23 De oorspronkelijke schaal voor tutorvaardigheid liep van 1 tot 4. De Rekenfaculteit heeft deze omgezet naar een schaal van 1 tot 10, omdat deze voor beoordelaars herkenbaarder is (vergelijkbaar met schoolcijfers). Dit betekent dat alleen de waarden 1, 4, 7 en 10 voorkomen.

observatie-instrument. De kwaliteit van de tutoring[24] wordt uitgedrukt op leerlingniveau. Voor elke tutor wordt eerst het gemiddelde over alle observaties in het schooljaar berekend. Wanneer een leerling door twee tutoren wordt begeleid, wordt voor die leerling het gemiddelde van beide tutoren gebruikt.

Ontwikkelingen meetinstrumentarium tussenevaluatiejaar (2023-2024)

Tijdens het eerste onderzoeksjaar (2023-2024) hebben zich enkele veranderingen en doorontwikkelingen voorgedaan in de gebruikte meetinstrumenten. Deze aanpassingen waren gericht op het verbeteren van de bruikbaarheid van de instrumenten en het vaststellen van de validiteit. Zo is de BSA in twee delen gesplitst, zijn er aanvullende analyses uitgevoerd om de methodologische beperkingen van de IEP-toets te ondervangen en is het observatie-instrument kwaliteit tutoring gevalideerd. In Kader E worden deze ontwikkelingen toegelicht.

Kader E – Ontwikkelingen tijdens het tussenevaluatiejaar (2023-2024)

1. Aanpassing in BSA: opsplitsen in deel A en B

Tijdens het tussenevaluatiejaar bleek dat de afnamecondities van de BSA verschilden tussen de Rekenfaculteit-scholen en de vergelijkingsscholen. Op de vergelijkingsscholen werd de toets afgenomen door een onderzoeker, die in overleg met de leerkracht een vaste tijdslimiet hanteerde. Op de Rekenfaculteit-scholen gebeurde de afname door de tutor, die zoveel tijd kon nemen als nodig was. Waarschijnlijk heeft dit verschil in afnamecondities ertoe bijgedragen dat leerlingen in de vergelijkingsgroep (bij de start- en middenmeting) minder vragen beantwoordden en lagere scores haalden, onafhankelijk van hun daadwerkelijke rekenvaardigheid.

Om dit verschil te ondervangen, is in 2023-2024 (bij de start- en middenmeting) bij de analyses alleen de eerste helft van de BSA meegenomen (deel A). In dit deel waren de verschillen in onbeantwoorde vragen kleiner en, -hoewel de bijbehorende resultaten met enige voorzichtigheid geïnterpreteerd moeten worden- bood dit deel een goede weerspiegeling van de moeilijkheidsgraad en leerdoelen van de BSA.

Vanaf de eindmeting van 2023–2024 is de BSA gesplitst in twee delen. Deel A, de ‘korte BSA’, wordt onder dezelfde tijdsdruk bij zowel de Rekenfaculteit- als de vergelijkingsscholen afgenomen en vormt de basis voor de groepsvergelijkingen. Deel B wordt alleen door de Rekenfaculteit-leerlingen gemaakt, zodat samen met deel A een ‘volledige BSA-toets’ ontstaat waarmee de groei van Rekenfaculteit-leerlingen individueel kan worden gevolgd.

Bij het opsplitsen van de toets is zorgvuldig en onderbouwd te werk gegaan. De items voor deel A zijn geselecteerd met een inhoudelijke afweging en een evenwichtige verdeling over de verschillende rekendomeinen. Uit de validiteitstoetsing door Cito blijkt dat deel A rekenvaardigheid goed meet. Bovendien is de correlatie tussen deel A en de volledige toets (deel A + B) hoog (r = 0.9), wat bevestigt dat deel A een betrouwbare afspiegeling is van de totale toets.

[24] De tutorobservaties bevatten nog aanzienlijk meer informatie die extra tijd en aanvullende analyses vergen. Dit valt buiten de scope van het huidige onderzoek. In vervolgonderzoek wordt dit nog verder uitgewerkt.

2. Methodologische beperking IEP-toets ondervangen
Tijdens het eerste onderzoeksjaar bleek dat de LVS-data van de IEP-toets uit twee afzonderlijke reeksen bestaat die niet op elkaar zijn geijkt. Dit betekent dat de scores niet op een gemeenschappelijke meetschaal staan en daardoor niet één-op-één te vergelijken zijn. De eerste reeks loopt van groep 3 tot midden groep 6, de tweede van eind groep 6 tot groep 8. Eenzelfde score heeft daardoor in beide reeksen een andere betekenis en is alleen binnen dezelfde reeks goed vergelijkbaar. Vooral in groep 7, waar veel leerlingen overstappen van de eerste naar de tweede reeks, bemoeilijkt dit betrouwbare groeimetingen. In overleg met IEP zijn daarom aanvullende analyses uitgevoerd, waarbij alleen leerlingen zijn meegenomen die volledig binnen één toetsreeks bleven. De IEP-resultaten van groep 7 zijn gebaseerd op een kleine en specifieke groep leerlingen, voornamelijk met relatief hogere rekenvaardigheid. Deze resultaten moeten met enige voorzichtigheid geïnterpreteerd worden.

3. Validatie observatie-instrument kwaliteit tutoring
Het onderzoeksteam heeft samen met de Rekenfaculteit en een rekenvakdidacticus vanaf januari 2024 een validatiestudie uitgevoerd naar de kwaliteit van het observatie-instrument. Het instrument is op een aantal punten aangepast en blijkt valide en betrouwbaar. Meer informatie over de validatie van het observatie-instrument is te vinden in Bijlage 2, sectie 4.

4.4 Dataverzameling/afnamecondities

De dataverzameling op de Rekenfaculteitscholen werd grotendeels uitgevoerd door de tutoren van het programma, die de BSA-toetsen en de vragenlijsten over rekenattitude en rekenangst afnamen. Omdat het verzamelen van gegevens onderdeel was van het reguliere werkproces van de Rekenfaculteit, verliep dit proces efficiënt en geïntegreerd. Daarnaast werd het observatie-instrument (vanaf schooljaar 2024-2025) afgenomen door teamleiders van de Rekenfaculteit. De afname van de LVS-toetsen (Cito en IEP) werd door de scholen zelf gedaan en zij stelden de data beschikbaar aan het onderzoeksteam.

Bij de vergelijkingsscholen werd de volledige dataverzameling verzorgd door het onderzoeksteam van Stichting De Verre Bergen. Scholen hadden vaak extra ondersteuning nodig bij het afnemen van toetsen en vragenlijsten. De aanwezigheid van onderzoekers tijdens de dataverzameling droeg daarnaast bij aan een zo hoog mogelijke respons. Binnen het onderzoeksteam coördineerde de programmasecretaris de dataverzameling en het contact met de scholen, met ondersteuning van de onderzoekers om het proces soepel te laten verlopen.

Afnamecondities

De afname van de toetsen en vragenlijsten werd zorgvuldig georganiseerd om een betrouwbare vergelijking tussen de Rekenfaculteitscholen en vergelijkingsscholen mogelijk te maken. Zo werd er gebruikgemaakt van een afnameprotocol om ervoor te zorgen dat alle leerlingen de toetsen en vragenlijsten onder vergelijkbare omstandigheden maakten. Hierin is de manier waarop de toetsen en vragenlijsten worden afgenomen vastgelegd, met duidelijke instructies voor de leerkrachten, tutoren en onderzoekers. Dit zijn de belangrijkste afnamecondities:

- Alle toetsen en vragenlijsten werden online afgenomen via het WOOTS-platform.
- De afname van de vragenlijsten over rekenattitude en rekenangst vond vóór de afname van de BSA-rekenvaardigheidstoets plaats[25], zodat er geen negatieve beïnvloeding van de toetsing was op de perceptie van rekenen.
- Uit een pilotstudie bleek dat sommige woorden in de vragenlijsten over rekenattitude en rekenangst onduidelijk waren voor de leerlingen. Om dit te verhelpen, werden deze woorden voorzien van extra uitleg. Bijvoorbeeld, het woord 'nerveus' werd verduidelijkt met 'spannend'.
- Er werd een tijdslimiet ingesteld voor alle leerlingen bij het maken van de BSA-toets (Kader E). Bij Rekenfaculteit-leerlingen werd genoteerd bij welke vraag van de BSA toets zij zijn binnen dit tijdslimiet.

Bij de Rekenfaculteit vonden de toets- en vragenlijstafnames plaats in kleine tutorgroepjes en werden uitgevoerd door de eigen tutor van de leerlingen. Bij de vergelijkingsscholen daarentegen werden de toets en vragenlijsten klassikaal afgenomen door de onderzoekers, met aanwezigheid van de docent. In groep 5 werden bij de vergelijkingsscholen de vragenlijsten (rekenangst en rekenattitude) mondeling doorgenomen. Bij de Rekenfaculteit werd dit niet gedaan, maar konden leerlingen wel laagdrempeliger vragen stellen wanneer zij een vraag niet begrepen.

4.5 Statistische analyses

Groei in BSA-rekenvaardigheid meten (Latent Change Score model)

Om de groei tussen de toetsmomenten te meten en om de prestaties van de Rekenfaculteit te vergelijken met die van de vergelijkingsgroep, is het *Latent Change Score* (LCS-)model gebruikt (McArdle, 2009; Newsom, 2015).

In het LCS-model worden veranderingen gemeten als 'latente veranderingsscores', die laten zien hoe de BSA-rekenvaardigheidsscores van leerlingen veranderen tussen toetsen. Met dit model werd het effect van de Rekenfaculteit op rekenvaardigheid gemeten, door de groeiscores van de Rekenfaculteit-leerlingen te vergelijken met die van de vergelijkingsleerlingen. Een toelichting op de voordelen van het LCS-model is opgenomen in Bijlage 2, sectie 5. Voor de analyse van de groei in rekenvaardigheid zijn drie mogelijke LCS-groeimodellen met elkaar vergeleken; in Bijlage 2, sectie 6 staat deze aanpak uitgebreid beschreven.

De LCS-modellen werden getoetst met een eenzijdige hypothesetoets, waarbij verondersteld werd dat de leerlingen van de Rekenfaculteit sterker zullen groeien dan de leerlingen van de vergelijkingsscholen. Het significantieniveau is vastgesteld op $\alpha = 0{,}05$.

[25] In het schooljaar '23-'24 was dit niet voor alle klassen op alle meetmomenten het geval, voor het schooljaar '24-'25 is deze afspraak aangescherpt.

Sensitiviteitsanalyses BSA-rekenvaardigheid

Om de betrouwbaarheid van de onderzoeksresultaten te waarborgen, is het belangrijk om te toetsen of de conclusies standhouden onder verschillende aannames en analysemethoden. Dit wordt ook wel het toetsen van de *robuustheid* van de resultaten genoemd.

Daarom zijn aanvullend enkele sensitiviteitsanalyses uitgevoerd: analyses die controleren of de bevindingen veranderen wanneer de data op een iets andere manier worden geanalyseerd. Zo wordt voorkomen dat de conclusies afhankelijk zijn van één specifieke analysemethode. De uitgevoerde analyses worden in Kader F toegelicht.

Kader F - Sensitiviteitsanalyses

1. Covariaatanalyse

Sommige factoren, zoals start rekenvaardigheid, rekenangst of rekenattitude, kunnen mogelijk invloed hebben op de groei in rekenvaardigheid. Deze factoren worden in het LCS-model opgenomen als zogenoemde covariaten: variabelen waarvan we verwachten dat ze samenhangen met de uitkomst. Zo is gecontroleerd of het effect van de Rekenfaculteit zichtbaar blijft als voor deze verschillen tussen leerlingen wordt gecorrigeerd.

2. Multilevelanalyse

Leerlingen zitten in klassen, en leerlingen uit dezelfde klas lijken over het algemeen meer op elkaar dan op leerlingen uit andere klassen. De data heeft hierdoor een hiërarchische ('multilevel') structuur. Als deze structuur genegeerd wordt, kunnen uitkomsten vertekend raken. Daarom is een analysemodel gebruikt dat rekening houdt met deze hiërarchie, waarbij de metingen van kinderen zijn gegroepeerd binnen klassen. In een gewone analyse wordt één gemiddelde startwaarde (in dit geval: niveau voor rekenvaardigheid) geschat voor alle leerlingen samen. Dat is het gemiddelde van de totale steekproef, maar is niet per se representatief voor het gemiddelde in individuele klassen. In een multilevelmodel wordt daarom voor elke klas een eigen gemiddeld startniveau geschat, en voor elke leerling een individueel startniveau. Zo wordt rekening gehouden met verschillen tussen leerlingen binnen dezelfde klas én met verschillen tussen klassen onderling.

3. Per-protocolanalyse

Bij één school is de uitvoering van de Rekenfaculteit niet verlopen zoals gepland. Om te beoordelen of dit invloed heeft op de uitkomsten, is de LCS-analyse herhaald zonder deze school. Hiermee wordt vastgesteld of de resultaten ook standhouden wanneer deze afwijkende situatie buiten beschouwing wordt gelaten.

4. Matchinganalyse (propensity score matching)

Leerlingen zijn niet willekeurig toegewezen aan de Rekenfaculteit of de vergelijkingsgroep. Daardoor kunnen er bij de start al verschillen bestaan, bijvoorbeeld in rekenvaardigheid. Om de vergelijking eerlijker te maken, is een matchinganalyse uitgevoerd waarbij voor elke leerling uit de Rekenfaculteit een leerling uit de vergelijkingsgroep is gezocht met een vergelijkbare beginscore op de BSA. Dit is gedaan met nearest neighbour matching: een methode waarbij voor elke Rekenfaculteit leerling de meest vergelijkbare controle 'match' wordt geselecteerd. De matching van de Rekenfaculteit leerlingen is per schooljaar

uitgevoerd (2023/2024 en 2024/2025), waarbij vergelijkingsleerlingen uit beide jaren gezamenlijk zijn gebruikt om de kans op een passende match te vergroten. [26]

5. Controle LCS groeimodellen en uitschieters
Tot slot werd gecontroleerd of de keuze tussen de drie LCS-groeimodellen (Bijlage 2, sectie 6) invloed heeft op de resultaten, en in hoeverre uitschieters[27] (outliers) een effect hebben op de uitkomsten.

Groei in LVS-rekenvaardigheid meten

Voor de analyse van de leerlingvolgsystemen Cito en IEP is dezelfde methode toegepast, die hieronder wordt toegelicht. In beide gevallen is onderzocht of leerlingen van de Rekenfaculteit meer leerwinst hebben geboekt dan vergelijkbare leerlingen uit de landelijke database van Cito of IEP. Eerst zijn de resultaten van leerlingen van de Rekenfaculteit vergeleken met die van de volledige landelijke normgroepen (zie Kader G voor aantallen). Vervolgens zijn aanvullende analyses uitgevoerd met specifiek gematchte deelgroepen uit deze landelijke populatie.

Kader G – Aantallen leerlingen per normgroep

Voor de analyses zijn gegevens gebruikt uit de landelijke normgroepen van Cito en IEP. De omvang van deze normgroepen verschilt per toetsaanbieder, leerjaar en schooljaar. Onderstaand overzicht toont het aantal leerlingen per normgroep.

Schooljaar	LVS	Groep 5	Groep 7
2023/2024	Cito	25.578	27.876
2023/2024	IEP	18.333	13.508
2024/2025	Cito	73.250	75.898
2024/2025	IEP	21.617	15.500

Er is gebruikgemaakt van een stapsgewijze matchingbenadering, waarbij Rekenfaculteit-leerlingen werden vergeleken met leerlingen uit de landelijke database met een vergelijkbaar beginniveau in rekenvaardigheid. Bij elke stap werd de vergelijking iets preciezer zodat het effect van de Rekenfaculteit bij verschillende niveaus van vergelijkbaarheid kon worden onderzocht (Kader H).

26 Dit resulteerde in een 100% match voor groep 5 en een 99% match voor groep 7 in schooljaar 2023–2024. In schooljaar 2024–2025 bedroeg de match voor groep 5 98% en voor groep 7 99%.

27 Uitschieters werden geïdentificeerd op basis van de *Mahalanobis distance*, een multivariate maat voor de afstand van een datapunt tot het centrum van de verdeling.

Kader H – Stappen in de matchingprocedure

1. **Geen matching** – vergelijking met alle leerlingen in de landelijke database.
2. **Vaardigheidsniveaus I–V** – indeling in vijf brede vaardigheidsniveaus (van hoogste 20% naar 20% laagste scores).
3. **Decielscores** – fijnere indeling op basis van tien vaardigheidsdecielen (van laagste 10% tot hoogste 10%).
4. **Didactische Leeftijdsequivalent (DLE)** – vergelijking op basis van het aantal maanden onderwijs dat een leerling heeft gevolgd.
5. **Vaardigheidsscore** – meest precieze methode, koppeling op basis van de exacte startscores van leerlingen.

Bij een tweede reeks analyses is ook de achterstandsscore van scholen meegenomen. Deze score, berekend door het CBS en gebruikt door OCW bij de verdeling van het onderwijsachterstandenbudget, geeft de verwachte onderwijsachterstand van een school weer. Door deze mee te nemen, sluiten de vergelijkingsleerlingen ook qua schoolachtergrond beter aan bij de Rekenfaculteit-leerlingen.

De uitgebreide uitleg over de matchingsprocedure is opgenomen in Bijlage 2- sectie 7.

Bij elke matchingsstap is een steekproef getrokken uit de landelijke Cito- of IEP-database, waarbij het aantal leerlingen gelijk is aan het aantal leerlingen binnen de Rekenfaculteit. Een grove manier van matching bij Cito is bijvoorbeeld dat van rekenniveaus I t/m V. Stel dat er binnen de rekenfaculteit 15 leerlingen op niveau I zitten, 20 op niveau II, 30 op niveau III, 20 op niveau IV en 15 op niveau V, dan zijn exact dezelfde aantallen leerlingen met die specifieke rekenniveaus uit de landelijke Cito-groep getrokken.

Voor elke steekproef is de gemiddelde leerwinst berekend: de toename in vaardigheidsscore tussen begin en eind van het schooljaar. Die gemiddelde leerwinst werd vervolgens vergeleken met de gemiddelde leerwinst van alle Cito- of IEP-leerlingen in de landelijke database. Dit gebeurde met een *one-sample t-toets*, die aangeeft of de steekproef gemiddeld afwijkt van de landelijke groep (de ‘populatie’).

Door dit proces 1000 keer te herhalen, ontstond een verdeling van t-waardes: een beeld van welke t-waardes je zoal kunt verwachten als de Rekenfaculteit-leerlingen willekeurige Cito- of IEP-leerlingen zouden zijn. Eenzelfde *one-sample t-toets* is ook uitgevoerd voor de Rekenfaculteit leerlingen in vergelijking tot de landelijke database. Door de t-waarde van de leerwinst Rekenfaculteit vervolgens in de verdeling van t-waardes te plaatsen, konden we zien of deze t-waarde uitzonderlijk groter was dan de t-waardes van de willekeurige leerlingen die geen Rekenfaculteit hebben genoten. Als deze t-waarde groter is dan de bovengrens van het 95%-betrouwbaarheidsinterval van de steekproeven ligt, duidt dit op een grotere leerwinst dan bij vergelijkbare leerlingen uit de landelijke groep.

Rekenattitude en rekenangst meten

Voor rekenattitude en rekenangst is de aanpak vergelijkbaar met de analyses die zijn uitgevoerd om de groei in (BSA-)rekenvaardigheid te meten. Eerst is de samenhang tussen rekenangst, rekenattitude en rekenvaardigheid in kaart gebracht met behulp van een correlatieanalyses. Vervolgens is het effect van de Rekenfaculteit op rekenangst en rekenattitude onderzocht met LCS-modellen, waarbij de scores van de Rekenfaculteit-leerlingen zijn vergeleken met die van leerlingen van de vergelijkingsscholen.

Het toetsen van de LCS-modellen is uitgevoerd met een eenzijdige hypothesetoets, waarin is verondersteld dat leerlingen van de Rekenfaculteit een sterkere groei laten zien in rekenattitude en een sterkere afname in rekenangst dan leerlingen van de vergelijkingsscholen (met significantieniveau van $\alpha = 0{,}05$).

Om de robuustheid van de resultaten te toetsen, zijn de uitkomsten van drie LCS-modellen vergeleken (Bijlage 2, sectie 6) en zijn de analyses herhaald zonder uitschieters (‘*outliers*’) - uitzonderlijke (hoge of lage) meetwaarden die de resultaten onevenredig kunnen beïnvloeden. Wanneer alle uitkomsten significant én consistent zijn, duidt dit op een robuust effect; wanneer dat niet zo is, is het effect niet eenduidig.

Leerling- en programmakenmerken meten

Leerlingkenmerken

De volgende leerlingkenmerken zijn onderzocht: beginscores op rekenattitude, rekenangst en rekenvaardigheid, geslacht en aanwezigheid bij de lessen[28]. Om te bepalen of deze kenmerken de groei in rekenvaardigheid kunnen voorspellen, is een LCS-model gebruikt. In dit model zijn alle bovengenoemde kenmerken opgenomen als voorspellers van groei in rekenvaardigheid (gemeten met de volledige BSA-toets29). Niet-significante voorspellers zijn vervolgens stapsgewijs verwijderd. Daarna zijn de significante voorspellers en de verklaarde variantie gerapporteerd (R2). De verklaarde variantie is een maat om aan te geven welk deel van de verschillen (in groei in rekenvaardigheid) kan worden toegeschreven aan een bepaalde voorspeller.

Programmakenmerken

De volgende programmakenmerken zijn onderzocht op leerlingniveau:

- Dosering: het aantal aangeboden lessen (ongeacht aanwezigheid)[30];
- Ratio: groepsgrootte, gedefinieerd als het aantal leerlingen per tutorgroep;

[28] Aanwezigheid is berekend als het aantal bijgewoonde lessen gedeeld door het totaal van bijgewoonde en gemiste lessen. Lessen die uitvallen om structurele redenen (zoals schoolkamp) of lessen waarvan de registratie ontbreekt, zijn hierbij niet meegenomen.

[29] Deze analyse is uitgevoerd met de volledige BSA-toets (deel A+B), omdat deze toets de meeste items bevat en daardoor een betrouwbaardere meting geeft.

[30] De variatie in dosering hangt mogelijk samen met roostertechnische omstandigheden.

- Tutorkwaliteit, gemeten met observaties[31]. Het gemiddelde van alle observaties vormt de vaardigheidsscore per tutor. Wanneer een leerling door twee tutoren werd begeleid, is het gemiddelde van beide scores genomen.

De analyse is, net als bij de leerlingkenmerken, uitgevoerd met een LCS-model. Daarbij is op dezelfde manier getoetst of deze programmakenmerken significante voorspellers zijn van groei in rekenvaardigheid.

Poweranalyse

Voorafgaand aan het onderzoek is een poweranalyse uitgevoerd, om voldoende statistische kracht (power) te waarborgen. Statistische power geeft de waarschijnlijkheid weer dat een werkelijk effect in de populatie ook in de steekproef wordt gevonden. Vooraf is een *effect size* van 0,15 gehanteerd als ondergrens voor het te verwachten effect. Op basis daarvan kon worden vastgesteld dat het op dat moment beschikbare aantal Rekenfaculteit-leerlingen (ongeveer 120 per leerjaar) voldoende was om een effect van deze grootte betrouwbaar te kunnen detecteren[32]. Ook hield de analyse rekening met externe effecten, zoals school- en tutoreffecten, die ruis kunnen veroorzaken. Ondanks deze externe effecten liet de poweranalyse zien dat bij een *effect size* van 0,15 met 90% zekerheid een verschil in BSA-groei over het jaar kan worden vastgesteld.

4.6 Goedkeuring ethische commissie

Voor het onderzoek is via het Nationaal Ethiek Overleg Sociale en Gedragswetenschappen (Nethics) ethische goedkeuring gevraagd. De facultaire ethische commissie van de Universiteit Utrecht heeft het onderzoeksprotocol en de bijbehorende documentatie beoordeeld en goedgekeurd. Het bijbehorende correspondentienummer is: FETC25-00001[33].

[31] In 2023/2024 is tutorkwaliteit geschat op basis van opleidingsniveau (1 = startkwalificatie, 2 = hbo afgerond, 3 = wo afgerond) en aantal jaren ervaring, omdat er toen nog geen observatiedata beschikbaar waren.

[32] Het gevonden effect bleek uiteindelijk substantieel hoger uit te vallen (zie Hoofdstuk 5).

[33] Nethics/ de Universiteit Utrecht is niet betrokken bij de planning en uitvoering van het onderzoek en draagt geen verantwoordelijkheid voor de wijze waarop het onderzoek wordt uitgevoerd.

5. Resultaten

5.1 Beschrijvende statistieken

Om de vergelijkbaarheid van de Rekenfaculteit en de vergelijkingsscholen bij aanvang van de schooljaren (2023/2024 & 2024/2025) te beoordelen, zijn de twee groepen vergeleken op een aantal achtergrondkenmerken, waaronder rekenvaardigheid (BSA-startscore), rekenangst, rekenattitude, geslacht en gemiddeld aantal leerlingen per klas (Tabel 1).

In groep 5 werden op sommige achtergrondkenmerken significante verschillen gevonden. Leerlingen van groep 5 (2023/2024) van de Rekenfaculteit scoorden significant hoger op de startmeting op rekenangst dan de vergelijkingsgroep ($p < 0,05$). Leerlingen van de Rekenfaculteit scoorden bij aanvang van het schooljaar 2024/2025 lager op rekenvaardigheid (BSA) dan hun leeftijdsgenoten op de vergelijkingsscholen ($p < 0,05$). Voor de overige kenmerken (rekenattitude, geslacht en klasgrootte) werden geen significante verschillen gevonden.

In groep 7 scoorden leerlingen bij de start van schooljaar 2024/2025 hoger op rekenangst dan de vergelijkingsgroep ($p < 0,05$). Rekenfaculteit leerlingen verschilden niet significant van leerlingen van de vergelijkingsscholen op de overige onderzochte achtergrondkenmerken. Er waren dus geen significante verschillen in BSA-startscore, rekenattitude, geslacht en gemiddeld aantal leerlingen per klas (alle p-waarden > 0,05).

Tabel 1. Beschrijvende statistieken startmeting over schooljaren 2023/2024 en 2024/2025

	2023/2024			2024/2025		
Groep 5	**Rekenfaculteit n = 136**	**Vergelijkingsscholen n = 145**	**p waarde[2]**	**Rekenfaculteit n = 141**	**Vergelijkingsscholen n = 150**	**p waarde[2]**
Rekenvaardigheid BSA (Gem, SD)	214 (18,7)	217 (19,5)	0,252	207 (17,1)	212 (18,6)	**0,018**
Rekenangst (Gem, SD)	44,7 (12,4)	41,7 (11,8)	**0,042**	43,4 (11,2)	41,8 (13,2)	0,260
Rekenattitude (Gem, SD)	63,1 (17,4)	65,9 (16,0)	0,244	63,5 (16,6)	65,4 (17,9)	0,368
Meisjes[1] (%)	43%	47%	0,512	57%	46%	0,102
Leerlingvolgsysteem (%)						
Cito	44%	77%		52%	81%	
IEP	56%	23%		48%	19%	
Aantal klassen (n)	9	8		7	8	
Leerlingen per klas (Gem, SD)	15,7 (4,9)	16,5 (7,6)	0,796	20,6 (3,0)	19,5 (5,0)	0,619
Ratio tutor-leerling (%)						
1-op-1	1%			<1%		
1-op-2	79%			68%		
1-op-3	20%			32%		
1-op-4				<1%		
Aanwezigheid (%)	92%			95%		

	2023/2024			2024/2025		
Groep 7	**Rekenfaculteit n = 172**	**Vergelijkingsscholen n = 140**	**p waarde[2]**	**Rekenfaculteit n = 213**	**Vergelijkingsscholen n = 117**	**p waarde[2]**
Rekenvaardigheid BSA (Gem, SD)	231 (19,9)	227 (21,0)	0,127	229 (18,2)	227 (19,4)	0,214
Rekenangst (Gem, SD)	40,38(11,4)	41,2 (12,3)	0,743	43,3 (11,1)	39,8 (11,7)	**0,009**
Rekenattitude (Gem, SD)	66,8 (14,2)	65,5 (15,7)	0,155	66,8 (14,9)	65,7 (16,3)	0,559
Meisjes[1] (%)	49%	57%	0,386	54%	53%	0,999
Leerlingvolgsysteem (%)						
Cito	58%	90%		67%	84%	
IEP	42%	10%		33%	16%	
Aantal klassen (n)	8	7		10	6	
Leerlingen per klas (Gem, SD)	21,1 (7,3)	20,1 (8,0)	0,809	21,4 (3,7)	20,7 (3,7)	0,709
Ratio tutor-leerling (%)						
1-op-1	-					
1-op-2	66%			68%		
1-op-3	34%			32%		
Aanwezigheid (%)	93%			93%		

n: aantal observaties; Gem: Gemiddelde; SD: Standaarddeviatie. [1] Op basis van openbare onderwijsdata vanuit DUO. [2] Verschillen tussen Rekenfaculteit en Vergelijkingsscholen zijn getoest met een independent samples t-test voor continue variabelen en een chi-kwadraat toest voor categorische variabelen.

5.2. BSA-resultaten

In dit deelhoofdstuk wordt het effect van de Rekenfaculteit op de rekenvaardigheid van leerlingen gepresenteerd, gemeten met de BSA-toets. Eerst wordt het algemene effect beschreven, waarna wordt ingezoomd op mogelijke verschillen per vaardigheidsniveau en per klas. Tot slot worden de resultaten van de sensitiviteitsanalyses besproken, waarin wordt nagegaan of de bevindingen standhouden onder verschillende aannames en analysemethoden.

Voor de analyses zijn alleen leerlingen meegenomen die een startmeting én minimaal één vervolgmeting hebben gedaan. Leerlingen die al bij de start- of middenmeting de maximale score behaalden, zijn buiten beschouwing gelaten. Dit komt doordat zij een ‘plafondeffect’ vertonen: ze

kunnen op de toets niet verder groeien, waardoor hun leerontwikkeling niet goed zichtbaar is. Het aantal leerlingen per meetmoment en per uitkomstmaat, evenals het totaal aantal leerlingen dat is meegenomen in de analyses, staat in Bijlage 2, sectie 1.

Interventie effect

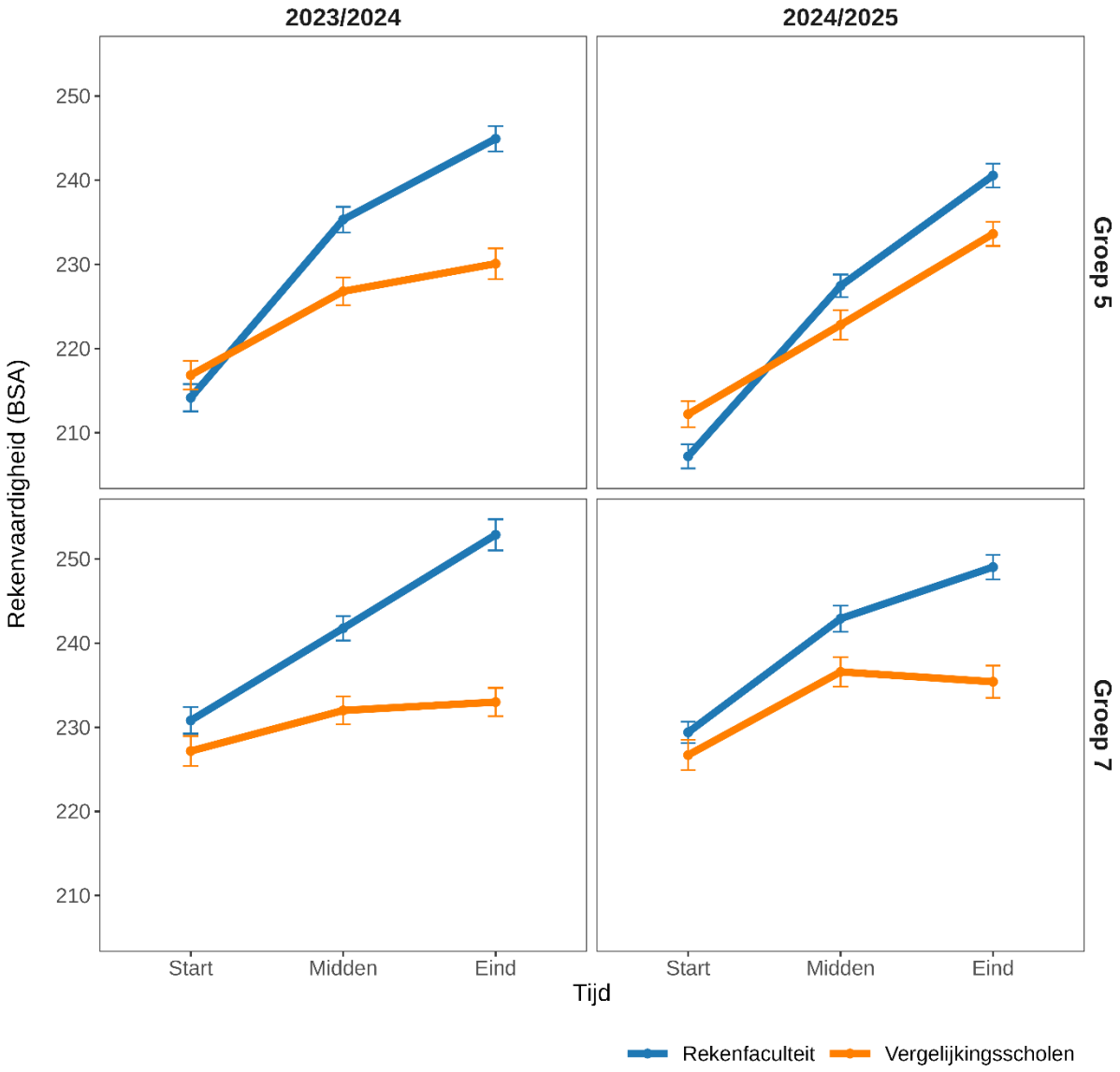
De effectiviteit van de Rekenfaculteit is onderzocht aan de hand van de groei in rekenvaardigheid (BSA-scores) over het schooljaar.

In Figuur 2 is de groei in rekenvaardigheid (BSA-score) voor beide school- en leerjaren van de startmeting tot de eindmeting weergegeven, waarbij een vergelijking wordt gemaakt tussen leerlingen van de Rekenfaculteit en die van de vergelijkingsscholen. In alle combinaties van schooljaar en leerjaar groeiden leerlingen van de Rekenfaculteit significant meer in rekenvaardigheid dan leerlingen van de vergelijkingsscholen (Tabel 2: alle p-waarden < 0,001). De bijbehorende *effect sizes* waren substantieel: voor groep 5 varieerden deze van 0,65 tot 0,83, en voor groep 7 van 0,60 tot 0,81.

De leerwinst wordt gedefinieerd als de extra toename in vaardigheidsscore van Rekenfaculteit-leerlingen ten opzichte van leerlingen op de vergelijkingsscholen, gemeten tussen het begin en het einde van het schooljaar. Leerlingen van vergelijkingsscholen groeien gemiddeld 15–21 BSA-punten in groep 5 en 5–8 punten in groep 7 over een periode van negen maanden. Leerlingen van de Rekenfaculteit groeien aanzienlijk sterker: 12–16 punten méér in groep 5 en 11–17 punten méér in groep 7. Deze extra leerwinst[34] komt overeen met een inhaalslag van ongeveer 5–10 maanden extra rekenles in groep 5 en ruim 12 maanden[35] in groep 7.

34 Berekening aantal maanden leerwinst= (groei Rekenfaculteit – groei vergelijkingsscholen) / (groei vergelijkingsscholen/ 9 maanden)

35 12 maanden in 2023/2024 en 31 maanden in 2024/2025. In 2024/2025 valt het aantal maanden extra leerwinst hoog uit, omdat de vergelijkingsleerlingen weinig groeien. De vertaling naar aantal maanden leerwinst is gebaseerd op het principe van de didactische leeftijdsequivalent (DLE) en moet met voorzichtig worden geïnterpreteerd (zie Evers & Resing, 2007). De vertaling naar DLE's geeft een bruikbare indicatie van de toegevoegde waarde van een interventie (zie o.a. Melis, Oosterveld & Schokker, 2012), maar betekent niet dat de leerlingen daadwerkelijk dit aantal maanden extra groeien.



***Figuur 2**. Effect van de Rekenfaculteit op rekenvaardigheid voor beide leer- en schooljaren. De foutmarges in het figuur geven de standaardfout (standard error) van het gemiddelde weer. Zie Bijlage 2 – sectie 8, Figuur B2 voor een versie van hetzelfde figuur waarin het 95% betrouwbaarheidsinterval is weergegeven.*

Tabel 2. Effect van de Rekenfaculteit op BSA-rekenvaardigheid

	Rekenfaculteit (RF)			Vergelijkingsscholen (VS)			Interventie effect		
	n	Gem (SD)	Groei[1] (SD)	n	Gem (SD)	Groei[1] (SD)	Verschil in groei (RF - VS)	Effect size[2] [95% CI]	p waarde
Groep 5									
2023/2024									
BSA (start)	134	214 (18,8)		129	217 (19,6)				
BSA (midden)	134	235 (16,9)	21 (13,6)	121	228 (18,6)	11 (12,2)	10		
BSA (eind)	127	245 (16,6)	31 (17,1)	117	232 (19,9)	15 (13,2)	16	0,83 [0,58 - 1,08]	**<0,001**
2024/2025									
BSA (start)	138	208 (16,9)		144	212 (18,7)				
BSA (midden)	138	228 (15,9)	20 (15,9)	131	223 (20,5)	11 (14,1)	9		
BSA (eind)	137	241 (16,9)	33 (17,7)	141	234 (17,5)	21 (13,2)	12	0,65 [0,40 - 0,90]	**<0,001**
Groep 7									
2023/2024									
BSA (start)	154	230 (19,4)		134	227 (20,7)				
BSA (midden)	154	242 (18,2)	11 (13,1)	126	232 (18,8)	5 (15,3)	6		
BSA (eind)	151	253 (22,5)	22 (16,7)	124	233 (19,0)	5 (17,6)	17	0,81 [0,56 - 1,06]	**<0,001**
2024/2025									
BSA (start)	196	228 (17,1)		110	227 (18,7)				
BSA (midden)	196	241 (20,2)	13 (12,9)	105	237 (17,6)	9 (12,7)	4		
BSA (eind)	194	248 (19,6)	19 (14,9)	101	236 (19,0)	8 (15,0)	11	0,60 [0,35 - 0,85]	**<0,001**

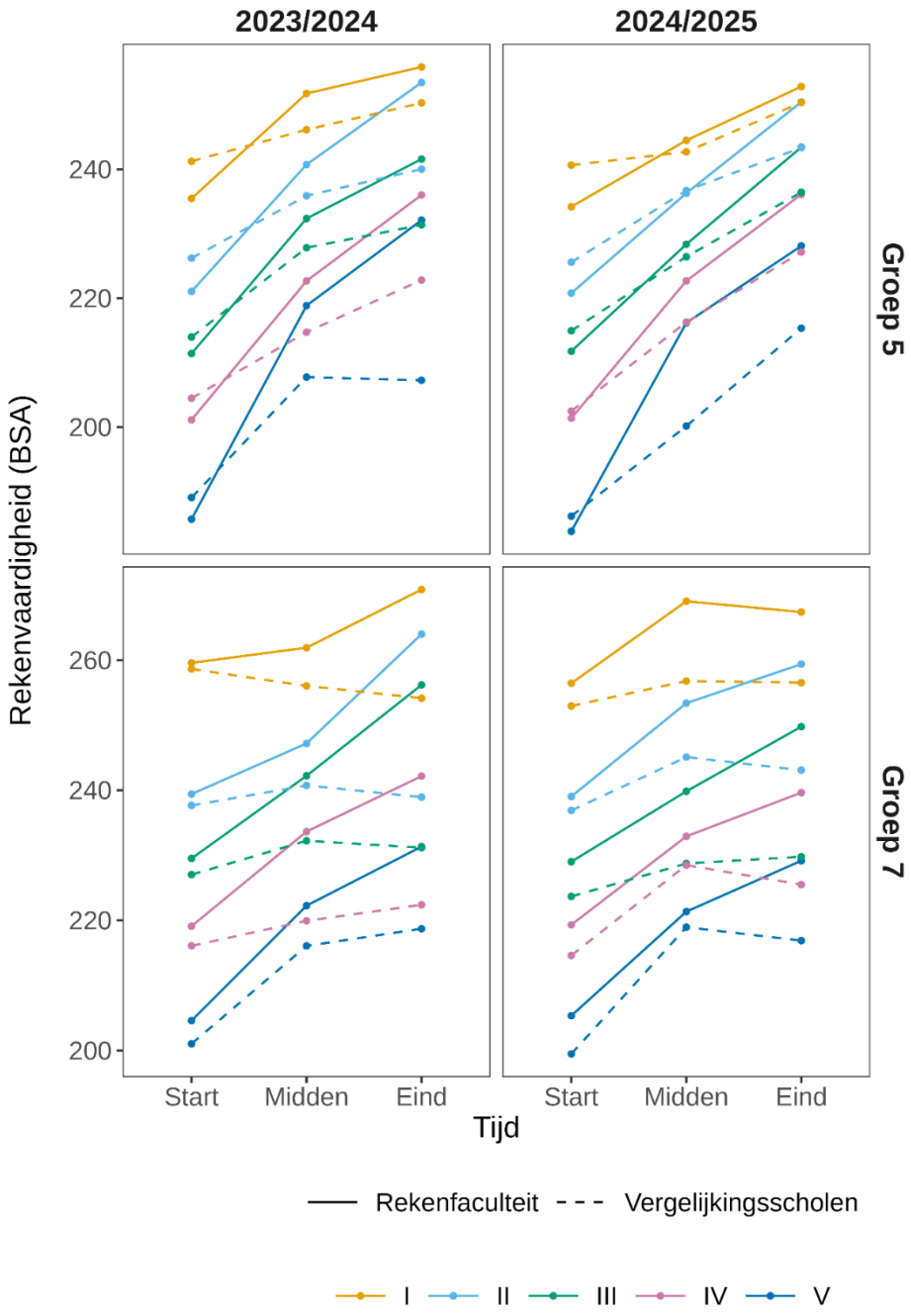
RF: Rekenfaculteit; VS: Vergelijkingsscholen; n: aantal observaties; Gem: Gemiddelde; SD: Standaarddeviatie; CI: Confidence (betrouwbaarheids) interval. [1] De gemiddelde groei is berekend binnen de groepen, zowel van middenmeting ten opzichte van de startmeting als van eindmeting ten opzichte van de startmeting. [2] De effectsize is berekend met Glass's Delta door het verschil in gemiddelde groei tussen de rekenfaculteit (RF) en de vergelijkingsscholen (VS) te delen door de standaarddeviatie van de VS-leerlingen bij de startmeting (T0).

Effect per vaardigheidsniveau

Om te onderzoeken of de Rekenfaculteit even effectief is voor leerlingen met verschillende niveaus van startrekenvaardigheid, is de leerwinst ook bekeken per niveaugroep.

Leerlingen zijn hiervoor onderverdeeld in vijf even grote niveaugroepen (20% per groep), op basis van hun startscores op de BSA-toets. Deze indeling is apart gemaakt voor de Rekenfaculteit-leerlingen en voor de leerlingen van de vergelijkingsscholen. Groep A bevat daarbij steeds de 20% sterkste leerlingen binnen de betreffende groep, en groep E de 20% zwakste leerlingen.

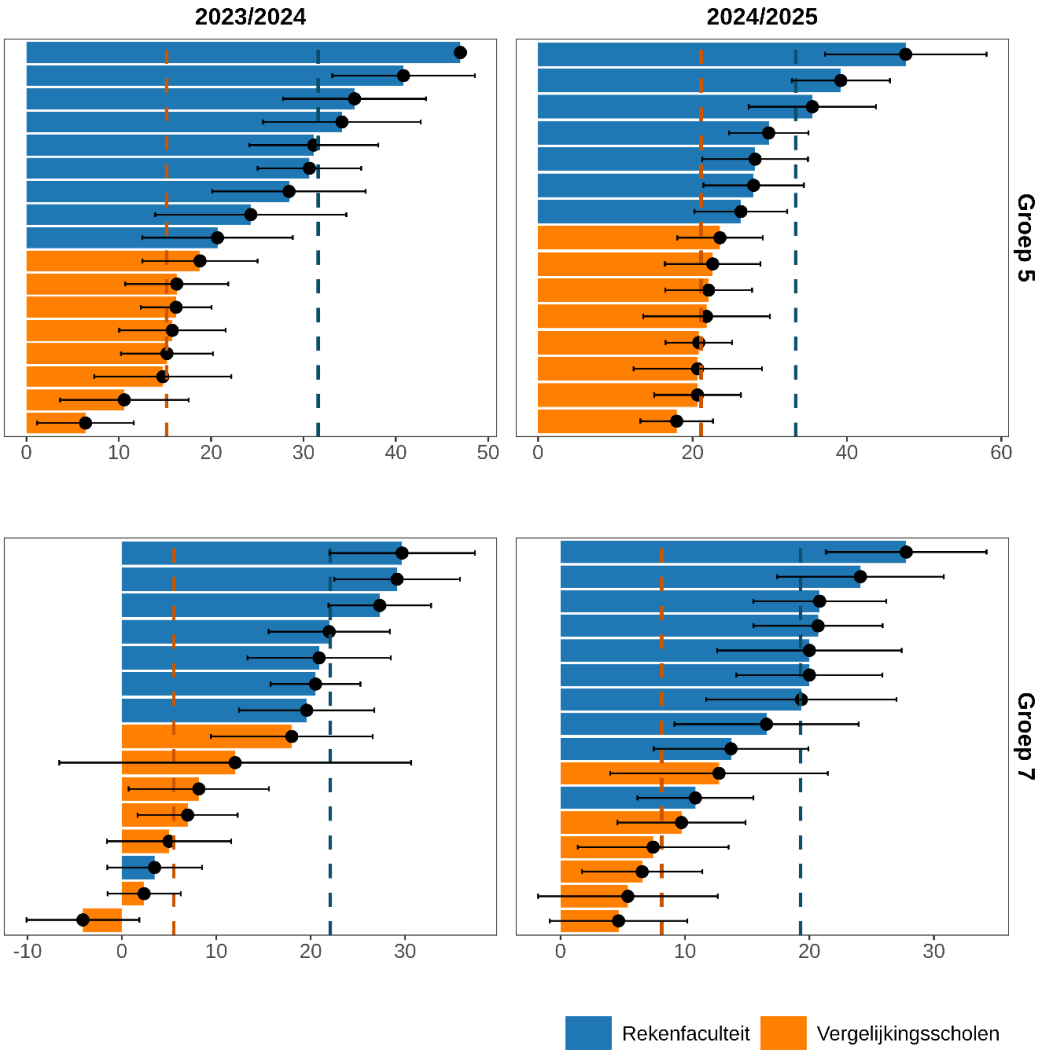
In alle vaardigheidsniveaus is groei zichtbaar bij de leerlingen van de Rekenfaculteit, zowel in groep 5 als in groep 7 (Figuur 3). Dit betekent dat de interventie effect heeft voor zowel zwakkere als sterkere leerlingen. In beide schooljaren en voor beide leerjaren scoren Rekenfaculteit-leerlingen aan het eind van het schooljaar vaak een vaardigheidsniveau hoger dan leerlingen van de vergelijkingsscholen, of komen zij dicht bij dit niveau in de buurt. Daarbij valt op dat leerlingen die bij aanvang op een lager vaardigheidsniveau starten, over het algemeen meer groei laten zien dan leerlingen die bij start al sterker presteerden. Zo scoorden groep 5- en groep 7-leerlingen van schooljaar 2023/2024 die bij aanvang in niveau E startten, na een jaar Rekenfaculteit gemiddeld even hoog als de vergelijkingsschoolleerlingen uit niveau C — een sprong van twee niveaus.



Figuur 3. *Groei in BSA-scores per vaardigheidsniveau in beide school- en leerjaren. Groei in rekenvaardigheid is uitgesplitst in vijf subgroepen die verschillen in start vaardigheidsniveau (I t/m V). Groep I is de 20% leerlingen die het hoogst scoren op de BSA bij start (sterkere rekenaars) en groep V is de 20% laagst scorende leerlingen (zwakkere rekenaars).*

Effect per klas

Om te laten zien hoe de leerwinst zich vertaalt naar het niveau van de klas, is de gemiddelde groei in rekenvaardigheid per klas weergegeven in Figuur 4. In deze figuren staat elke balk voor één klas. De gemiddelde groei over alle klassen is gemarkeerd met een verticale lijn: blauw voor de Rekenfaculteit en oranje voor de vergelijkingsscholen.



Figuur 4. *Groep 5: Gemiddelde groei (van T0 naar T2) in rekenvaardigheid (BSA-score) per klas. Elke balk vertegenwoordigt één klas (blauw voor Rekenfaculteit klassen en oranje voor klassen van de vergelijkingsscholen). Voor elke klas is de gemiddelde groei op de BSA weergegeven (zwart bolletje) en de 95% betrouwbaarheidsinterval (horizontale streepje rondom het gemiddeld). De gemiddelde groei over alle klassen is gemarkeerd met een verticale lijn voor de Rekenfaculteit (blauw) en de vergelijkingsscholen (oranje).*

In groep 5 is er in beide schooljaren in alle klassen van de Rekenfaculteit meer groei in rekenvaardigheid te zien dan in de vergelijkingsschoolklassen.

In groep 7 groeit in 2023/2024 het merendeel van de Rekenfaculteit-klassen harder dan de klassen van de vergelijkingsscholen, met uitzondering van één klas. In 2024/2025 is het patroon vergelijkbaar: de meeste klassen van de Rekenfaculteit boeken meer vooruitgang, met uitzondering van één vergelijkingsschoolklas die relatief sterk vooruitgaat.

Sensitiviteitanalyses

Om te toetsen of de resultaten robuust zijn, zijn meerdere sensitiviteitsanalyses uitgevoerd. In al deze aanvullende analyses bleven de bevindingen overeind: leerlingen van de Rekenfaculteit groeiden significant meer in rekenvaardigheid dan leerlingen van de vergelijkingsscholen (Tabel 3). De verschillen in *effect sizes* waren klein en deden geen afbreuk aan de hoofdconclusie. Ook als een ander LCS-model gekozen werd of uitschieters uit de data werd verwijderd[36], deed dit geen afbreuk aan de resultaten. Voor keuzes m.b.t. de LCS-modellen, zie Bijlage 2, sectie 6.

Tabel 3. Resultaten sensitiviteitsanalyses

	Interventie effect		
	Verschil in groei[1] (RF - VS)	Effect size[2]	p waarde
Groep 5			
2023/2024			
covariaten analyse	16	0,83	**<0,001**
multilevel analyse	16	0,83	**<0,001**
per protocol analyse	-	-	-
matching analyse	13	0,86	**<0,001**
2024/2025			
covariaten analyse	12	0,65	**<0,001**
multilevel analyse	12	0,65	**<0,001**
per protocol analyse	-	-	-
matching analyse	10	0,67	**<0,001**
Groep 7			
2023/2024			
covariaten analyse	17	0,81	**<0,001**
multilevel analyse	17	0,81	**<0,001**
per protocol analyse	16	0,76	**<0,001**
matching analyse	13	0,95	**<0,001**
2024/2025			
covariaten analyse	11	0,60	**<0,001**
multilevel analyse	11	0,60	**<0,001**
per protocol analyse	11	0,59	**<0,001**
matching analyse	11	0,63	**<0,001**

[1] De gemiddelde groei van de eindmeting ten opzichte van de start.

[2] De effectsize is berekend met Glass's Delta door het verschil in gemiddelde groei tussen de Rekenfaculteit (RF) en de vergelijkingsscholen (VS) te delen door de standaarddeviatie van de VS-leerlingen bij de startmeting (T0).

5.3. LVS-Resultaten

In dit deelhoofdstuk wordt het effect van de Rekenfaculteit op de LVS-scores (Cito en IEP) onderzocht. De toetsscores van Rekenfaculteit-leerlingen zijn vergeleken met die van de landelijke normgroep. Daarnaast zijn aanvullende analyses uitgevoerd met gematchte subgroepen uit die normgroep, om een zo eerlijk mogelijke vergelijking te maken.

36 In 2023/2024 zijn 9 uitschieters uitgesloten (5 bij de Rekenfaculteit, 4 bij de vergelijkingsscholen); in 2024/2025 waren dit er 11 (5 en 6 respectievelijk).

Resultaten: Rekenfaculteit versus landelijke normgroep

In beide schooljaren boekten leerlingen van de Rekenfaculteit meer leerwinst dan leerlingen uit de landelijke normgroepen van Cito en IEP (Tabel 4). Dit gold voor zowel groep 5 als groep 7, en voor beide toetsaanbieders (alle p-waarden ≤ 0,009).

De *effect sizes* in de Cito-analyses varieerden van 0,22 tot 0,51 voor groep 5 en van 0,31 tot 0,47 voor groep 7. Bij de IEP-analyses lagen de *effect sizes* nog iets hoger: van 0,36 tot 0,43 in groep 5 en van 0,54 tot 0,68 in groep 7. Dit wijst op een consistent voordeel in leerwinst voor Rekenfaculteit-leerlingen ten opzichte van hun leeftijdsgenoten in de landelijke populatie.

De *effect sizes* voor groep 5 in 2024–2025, op basis van de Cito-metingen, lijken lager te liggen dan in 2023–2024, en wordt nader besproken in Hoofdstuk 6 - Discussie.

Tabel 4. LVS-resultaten Rekenfaculteit t.o.v. landelijke normgroep

	Rekenfaculteit (RF)			Landelijke normgroep			Interventie effect		
	n	Gem (SD)	Groei[1] (SD)	N	Gem (SD)	Groei[1] (SD)	Verschil in groei (RF - VS)	Effect size[2] [95% CI][3]	p waarde
Groep 5									
2023/2024									
Cito (start)	58	174 (28,5)		25.578	190 (30,5)				
Cito (eind)	56	216 (25,8)	42 (21,8)	25.578	216 (26,8)	26 (21,0)	15,7	0,51 [0,25 - 0,77]	**<0,001**
IEP (start)	70	30 (3,3)		18.333	32 (4,6)				
IEP (eind)	70	48 (4,3)	17 (2,9)	18.333	47 (6,2)	15 (4,0)	2	0,43 [0,21 - 0,67]	**<0,001**
2024/2025									
Cito (start)	70	173 (32,4)		73.250	192 (31,0)				
Cito (eind)	74	206 (21,2)	32 (23,4)	73.250	216 (26,7)	25 (21,0)	6,8	0,22 [-0,01 - 0,45]	**0,009**
IEP (start)	61	30 (4,1)		21.617	32 (4,3)				
IEP (eind)	61	46 (4,6)	17 (3,8)	21.617	48 (6,1)	15 (4,1)	2	0,36 [0,11 - 0,61]	**0,002**
Groep 7									
2023/2024									
Cito (start)	94	230 (25,4)		27.876	241 (23,4)				
Cito (eind)	94	264 (27,2)	33 (13,4)	27.876	267 (23,8)	26 (16,1)	7,2	0,31 [0,11 - 0,51]	**<0,001**
IEP (start)[4]	31	56 (9,2)		13.508	63 (10,2)				
IEP (eind)[4]	31	75 (9,7)	19 (8,0)	13.508	75 (11,0)	12 (9,5)	7	0,68 [0,33 - 1,03]	**<0,001**
2024/2025									
Cito (start)	135	230 (29,6)		75.898	242 (23,8)				
Cito (eind)	141	267 (29,5)	37 (16,7)	75.898	268 (24,1)	26 (16,4)	11,1	0,47 [0,21 - 0,73]	**<0,001**
IEP (start)[4]	38	55 (9,5)		15.500	62 (10,3)				
IEP (eind)[4]	38	75 (11,0)	20 (11,2)	15.500	77 (10,5)	14 (9,0)	6	0,54 [0,22 - 0,86]	**0,004**

RF: Rekenfaculteit; VS: Vergelijkingsscholen; N: aantal observaties; Gem: Gemiddelde; SD: Standaarddeviatie; CI: Confidence (betrouwbaarheids) interval. [1] De gemiddelde groei is berekend binnen de groepen, van de start tot eind meting. [2] De effectsize is berekend met Glass's Delta door het verschil in gemiddelde groei tussen de rekenfaculteit (RF) en de LVS normgroepen te delen door de SD van de LVS normgroep-leerlingen bij de startmeting (T0). [3] De betrouwbaarheidsintervallen van de effectgrootte zijn conservatiever dan de p-waarde, omdat de groei is afgezet tegen de SD van de startmeting in plaats van de SD van de groeiscores. Hierdoor kan het voorkomen, zoals bij Cito groep 5 in 2024/2025, dat de p-waarde significant is terwijl het betrouwbaarheidsinterval de nul omvat. [4] Groep 7 IEP Leerlingen zijn geselecteerd die bij de start- en eindmeting dezelfde toestversie hebben gemaakt (referentieniveautoets). Dit betreft 56% van de RF-leerlingen in schooljaar 2023/2024 en 43% in schooljaar 2024/2025.

Resultaten: Rekenfaculteit versus gematchte groepen

In schooljaar 2023/2024 boekten leerlingen van de Rekenfaculteit, zowel bij de Cito- als de IEP-toetsen en in zowel groep 5 als groep 7, meer leerwinst dan vergelijkbare leerlingen uit de gematchte controlegroepen (Tabel 5). Dit verschil bleef zichtbaar bij elke stap in het matchingproces (zie voor meer informatie Bijlage 2, sectie 7), en gold zowel wanneer wél als wanneer géén rekening werd gehouden met de achterstandsscore van de school.

In schooljaar 2024/2025 was het beeld gemengd. In groep 5 werden bij de Cito-toets geen significante effecten gevonden, terwijl deze bij de IEP-toets wel aanwezig waren. Voor groep 7 gold het omgekeerde: daar werden significante effecten gevonden bij de Cito-toets, maar niet bij de IEP-resultaten. Deze gemengde uitkomsten hangen wellicht samen met de beperkingen van de analyses en specifieke uitdagingen bij twee klassen (zie ook Hoofdstuk 5, Discussie).

Tabel 5. LVS-resultaten: Rekenfaculteit t.o.v. gematchte leerlingen

	Matching zonder achterstand score				Matching met achterstand score			
	niveauscore	decielscore	DLE	vaardigheidsscore	niveauscore	decielscore	DLE	vaardigheidsscore
Groep 5								
2023/2024								
Cito	**<0,001**	**<0,001**	**<0,001**	**0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**
IEP	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**
2024/2025								
Cito	0,444	0,433	0,632	0,574	0,227	0,245	0,514	0,449
IEP	**0,001**	**<0,001**	**0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**
Groep 7								
2023/2024								
Cito	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**0,007**	**0,002**	**0,002**	**0,003**
IEP[1]	**<0,001**	**<0,001**	**0,002**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**
2024/2025								
Cito	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**
IEP[1]	0,062	0,088	0,084	0,056	0,107	0,090	0,117	0,103

DLE: Didactisch Leeftijd Equivalent. [1] Groep 7 IEP Leerlingen zijn geselecteerd die bij de start- en eindmeting dezelfde toestversie hebben gemaakt (referentieniveautoets). Dit betreft 56% van de RF-leerlingen in schooljaar 2023/2024 en 43% in schooljaar 2024/2025.

5.4. Resultaten rekenangst en rekenattitude

In dit deelhoofdstuk wordt het effect van de Rekenfaculteit op veranderingen in rekenangst en rekenattitude van leerlingen gepresenteerd. Eerst wordt de samenhang tussen rekenangst, rekenattitude en rekenvaardigheid besproken. Vervolgens wordt ingezoomd op de effecten van de Rekenfaculteit op de verandering in rekenangst en rekenattitude, door de scores van de Rekenfaculteit-leerlingen op de bijbehorende vragenlijsten te vergelijken met die van leerlingen van de vergelijkingsscholen. Hierbij is telkens nagegaan of de bevindingen standhouden bij verschillende statistische modellen (Bijlage 2 – sectie 6) en na het verwijderen van uitschieters.

Het verband tussen rekenangst, rekenattitude en rekenvaardigheid

Aan de hand van een correlatiecoëfficiënt (*r*)[37] is het verband tussen rekenangst en rekenvaardigheid op de beginmeting onderzocht. De correlatiecoëfficiënt loopt van -1 tot 1. Hoe dichter *r* bij -1 of 1 ligt, hoe sterker het verband; *r* rond de 0 betekent dat er geen verband is.

Tabel 6. Correlatietabel rekenvaardigheid, rekenangst en rekenattitude

Groep 5					
		BSA start	Angst start	Attitude start	Groei BSA schooljaar
BSA start		-			
Angst start	23/24	-0,29***	-		
	24/25	-0,21*	-		
Attitude start	23/24	0,29***	-0,17**	-	
	24/25	0,30***	-0,10	-	
Groei BSA schooljaar	23/24	-0,43***	0,12	-0,01	-
	24/25	-0,29***	0,11	-0,05	-
Groep 7					
		BSA start	Angst start	Attitude start	Groei BSA schooljaar
BSA start		-			
Angst start	23/24	-0,27***	-		
	24/25	-0,28***	-		
Attitude start	23/24	0,40***	-0,35***	-	
	24/25	0,45***	-0,21**	-	
Groei BSA schooljaar	23/24	-0,03	-0,16	0,13	-
	24/25	-0,14*	0,06	0,00	-

* p < 0,05; ** p < 0,01; *** p < 0,001.

Voor beide leergroepen blijkt er een negatieve samenhang te zijn tussen rekenangst en rekenvaardigheid (*r* varieert tussen -0,21 en -0,29, Tabel 6) op de startmeting. Oftewel, leerlingen die veel rekenangst rapporteren presteren gemiddeld minder goed op de rekentoets, terwijl leerlingen met minder rekenangst een hogere rekenvaardigheid hebben (gemeten aan het begin van het schooljaar).

Daarnaast is er voor beide schooljaren en leergroepen een positieve samenhang tussen rekenattitude en rekenvaardigheid (*r* varieert tussen 0,29 en 0,45, Tabel 6). Leerlingen met een

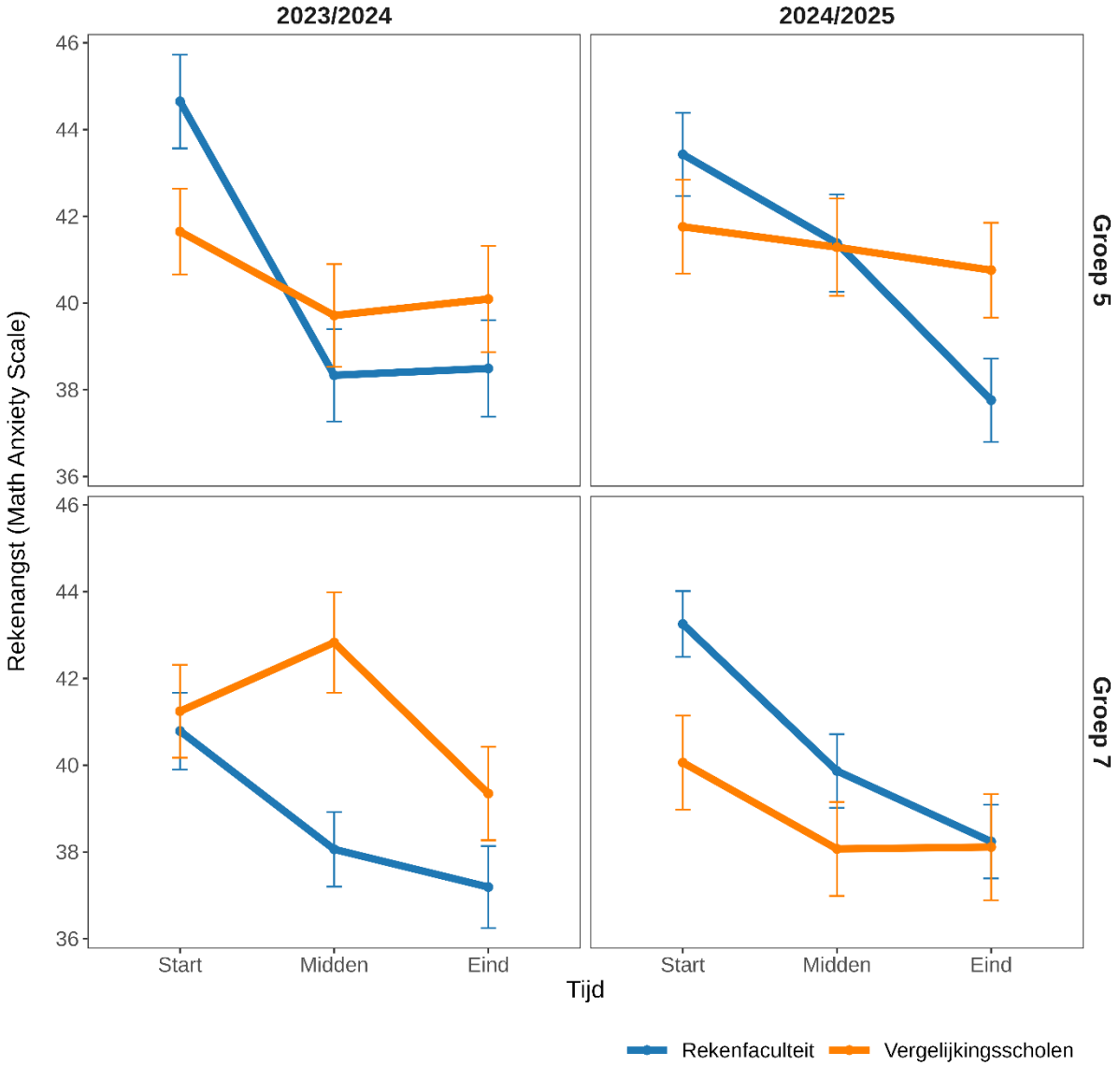
37 De gerapporteerde correlatiecoëfficiënt betreft een Pearson-correlatie.

positievere houding tegenover rekenen hebben dus gemiddeld hogere rekenvaardigheidsscores, terwijl een negatieve attitude samenhangt met lagere scores bij de start van het schooljaar.

In groep 7 (beide schooljaren) en groep 5 (2023/2024) is er daarnaast een negatieve samenhang tussen rekenangst en rekenattitude (*r* varieert tussen -0,17 en –0,21). Leerlingen met hogere angstniveaus hebben doorgaans een minder positieve attitude ten aanzien van rekenen, en omgekeerd.

Effect van Rekenfaculteit op rekenangst

Voor beide schooljaren en leergroepen is een afname van rekenangst zichtbaar gedurende het schooljaar. Dit geldt voor leerlingen van de Rekenfaculteit en voor leerlingen van de vergelijkingsscholen. Figuur 5 toont deze afname van start- tot eindmeting, waarbij leerlingen van de Rekenfaculteit worden vergeleken met leerlingen van de vergelijkingsscholen.



Figuur 5. *Effect van de Rekenfaculteit op rekenangst voor beide school- en leerjaren. De foutmarges in het figuur geven de standaardfout (standard error) van het gemiddelde weer.*

Leerlingen van de Rekenfaculteit in groep 5 lieten in beide schooljaren een significant sterkere afname in rekenangst zien dan leerlingen van de vergelijkingsscholen (*effect size* tussen 0,34 en 0,45, $p < 0{,}015$; Tabel 7). Dit resultaat bleef overeind bij herhaling van de analyses met verschillende LCS-modellen en na het verwijderen van uitschieters.

Ook in groep 7 is sprake van een afname in rekenangst gedurende het schooljaar. Hoewel leerlingen van de Rekenfaculteit gemiddeld een significant grotere afname in rekenangst lieten zien dan leerlingen van de vergelijkingsscholen (Tabel 7), hield dit verschil geen stand na uitsluiting van uitschieters.

Tabel 7. Effect van Rekenfaculteit op Rekenangst

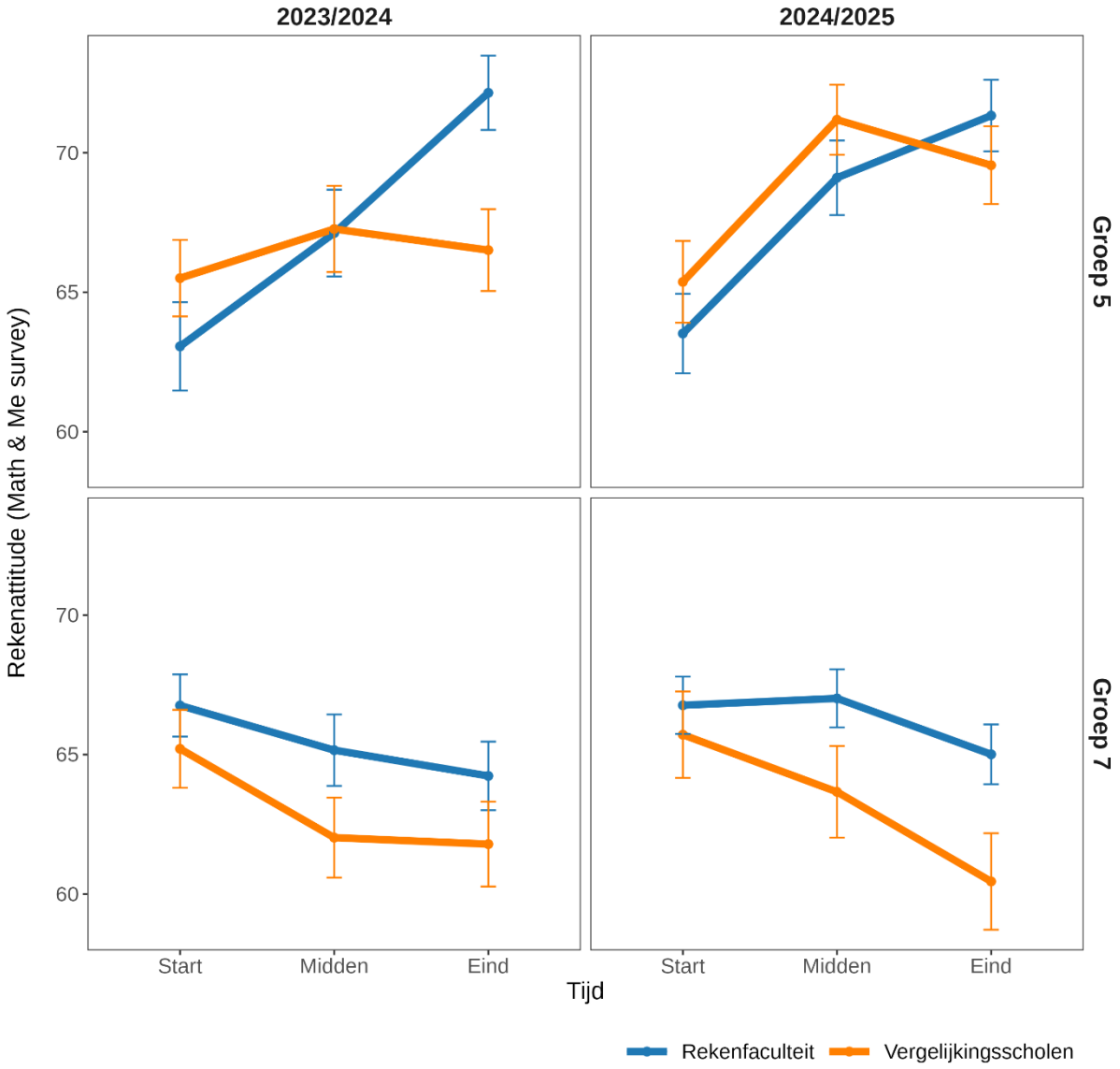
	Rekenfaculteit (RF)			Vergelijkingsscholen (VS)			Interventie effect		
	n	Gem (SD)	Afname[1] (SD)	n	Gem (SD)	Afname[1] (SD)	Verschil in groei (RF - VS)	Effect size[2] [95% CI]	p waarde
Groep 5									
2023/2024									
Rekenangst (start)	130	44.8 (12.3)		137	41,9 (11,8)				
Rekenangst (midden)	130	38.5 (12.6)	-6.3 (12.9)	124	40,1 (13,5)	-1,6 (12,1)			
Rekenangst (eind)	124	38.5 (12.8)	-6.3 (14.3)	123	40,4 (14,0)	-1,0 (12,7)	-5,3	0,45 [0,20 - 0,70]	**0,006**
2024/2025									
Rekenangst (start)	137	44,7 (12,6)		134	41,4 (11,2)				
Rekenangst (midden)	137	38,3 (12,5)	-6,4 (12,8)	121	39,7 (13,3)	-1,5 (12,1)			
Rekenangst (eind)	131	38,5 (12,9)	-6,3 (14,0)	123	40,4 (14,0)	-1,0 (12,7)	-5,3	0,34 [0,10 - 0,58]	**0,015**
Groep 7									
2023/2024									
Rekenangst (start)	158	40.5 (11.4)		127	41.5 (12.3)				
Rekenangst (midden)	157	38.0 (11.1)	-2.5 (11.5)	118	42.8 (13.2)	1.1 (9.9)			
Rekenangst (eind)	156	36.7 (11.8)	-3.9 (11.5)	119	39.1 (12.1)	-2.0 (11.6)	-1,9	0,15 [-0,09 -0,39]	**0,017**
2024/2025									
Rekenangst (start)	210	43,2 (11,1)		111	39,7 (11,7)				
Rekenangst (midden)	206	40,0 (12,4)	-3,3 (10,6)	102	38,0 (11,8)	-0,9 (10,1)			
Rekenangst (eind)	206	38,3 (12,3)	-5,0 (11,9)	102	37,9 (12,7)	-1,6 (9,7)	-3,4	0,27 [0,03 - 0,51]	**0,028**

RF: Rekenfaculteit; VS: Vergelijkingsscholen; n: aantal observaties; Gem: Gemiddelde; SD: Standaarddeviatie; CI: Confidence (betrouwbaarheids) interval.

[1] De gemiddelde afname is berekend binnen de groepen, zowel van middenmeting ten opzichte van de startmeting als van eindmeting ten opzichte van de startmeting. De berekeningen zijn uitgevoerd op alle beschikbare data. Door ontbrekende waarden (missings) kunnen de berekende verschilscores daardoor licht afwijken van de verschillen die op basis van de absolute aantallen verwacht zouden worden.[2] De effectsize is berekend met Glass's Delta door het verschil in gemiddelde groei tussen de rekenfaculteit (RF) en de vergelijkingsscholen (VS) te delen door de standaarddeviatie van de VS-leerlingen bij de startmeting (T0).

Effect van Rekenfaculteit op rekenattitude

De resultaten met betrekking tot veranderingen in rekenattitude tijdens het schooljaar zijn niet consistent en eenduidig. Over het algemeen lieten leerlingen in groep 5 een toename zien in hun rekenattitude gedurende het schooljaar. Opvallend is dat de rekenattitude bij leerlingen in groep 7 gemiddeld juist afnam (Figuur 6). Dit geldt zowel voor leerlingen van de Rekenfaculteit als voor leerlingen van de vergelijkingsscholen.



***Figuur 6**. Effect van de Rekenfaculteit op rekenattitude voor beide school- en leerjaren. De foutmarges in het figuur geven de standaardfout (standard error) van het gemiddelde weer.*

Voor groep 5 in schooljaar 2023/2024 was de toename in rekenattitude bij leerlingen van de Rekenfaculteit significant groter dan bij leerlingen van de vergelijkingsscholen (*effect size* 0,51; $p < 0,001$; zie Tabel 8). Dit effect bleef bestaan bij herhaling van de analyses met verschillende LCS-modellen en na verwijderen van de uitschieters.

Leerlingen in groep 5 in schooljaar 2024/2025 lieten een vergelijkbaar patroon zien, maar de effecten waren kleiner en bereikten geen statistische significantie (effect size 0,26; $p = 0,076$). De

aanvullende analyses met verschillende LCS-modellen en na verwijdering van uitschieters gaven een wisselend beeld (p-waarden tussen 0,001 en 0,068).

Voor groep 7 werden geen eenduidige effecten gevonden. In schooljaar 2023/2024 nam de rekenattitude bij alle leerlingen af, ongeacht deelname aan de Rekenfaculteit, en was er geen significant verschil tussen de groepen. In schooljaar 2024/2025 leek de afname in attitude bij Rekenfaculteit-leerlingen iets minder groot dan bij leerlingen van de vergelijkingsscholen (p = 0,008), maar dit verschil verdween na verwijdering van de uitschieters (p = 0,063).

Tabel 8. Effect van Rekenfaculteit op Rekenattitude

	Rekenfaculteit (RF)			Vergelijkingsscholen (VS)			Interventie effect		
	n	Gem (SD)	Verschil[1] (SD)	n	Gem (SD)	Verschil[1] (SD)	Verschil in groei (RF - VS)	Effect size[2] [95% CI]	p waarde
Groep 5									
2023/2024									
Rekenattitude (start)	120	63,3 (17,3)		132	65,4 (16,0)				
Rekenattitude (midder	118	66,8 (17,8)	3,2 (18,8)	119	67,1 (17,0)	1,0 (13,4)			
Rekenattitude (eind)	114	71,8 (15,1)	8,5 (16,7)	118	66,3 (16,8)	0,4 (14,1)	8,1	0,51 [0,26 - 0,76]	**<0,001**
2024/2025									
Rekenattitude (start)	133	63,5 (16,7)		145	65,9 (17,5)				
Rekenattitude (midder	133	68,9 (15,7)	5,4 (18,5)	131	71,2 (14,8)	5,2 (15,2)			
Rekenattitude (eind)	129	71,3 (15,2)	7,7 (19,4)	140	69,3 (16,9)	3,4 (17,6)	4,3	0,26 [0,02 - 0,50]	0,076
Groep 7									
2023/2024									
Rekenattitude (start)	157	66,9 (14,3)		123	65,1 (15,7)				
Rekenattitude (midder	155	65,2 (16,5)	-2,0 (11,8)	115	62,5 (16,0)	-2,3 (11,1)			
Rekenattitude (eind)	153	64,7 (16,0)	-2,0 (12,2)	114	62,4 (17,5)	-2,6 (12,7)	0,6	0,04 [-0,20 - 0,28]	0,707
2024/2025									
Rekenattitude (start)	209	66,7 (15,0)		106	66,1 (16,4)				
Rekenattitude (midder	209	67,0 (15,3)	0,2 (11,1)	102	64,0 (17,4)	-2,8 (11,7)			
Rekenattitude (eind)	204	65,0 (15,5)	-1,5 (11,5)	97	60,5 (18,0)	-4,7 (11,3)	3,2	0,19 [-0,05 - 0,42]	**0,008**

RF: Rekenfaculteit; VS: Vergelijkingsscholen; n: aantal observaties; Gem: Gemiddelde; SD: Standaarddeviatie; CI: Confidence (betrouwbaarheids) interval. [1] De gemiddelde groei is berekend binnen de groepen, zowel van middenmeting ten opzichte van de startmeting als van eindmeting ten opzichte van de startmeting. De berekeningen zijn uitgevoerd op alle beschikbare data. Door ontbrekende waarden (missings) kunnen de berekende verschilscores daardoor licht afwijken van de verschillen die op basis van de absolute aantallen verwacht zouden worden.[2] De effectsize is berekend met Glass's Delta door het verschil in gemiddelde groei tussen de rekenfaculteit (RF) en de vergelijkingsscholen (VS) te delen door de standaarddeviatie van de VS-leerlingen bij de startmeting (T0).

5.5. Resultaten leerlingkenmerken: voor wie werkt de Rekenfaculteit?

In dit deelhoofdstuk wordt onderzocht welke leerlingkenmerken samenhangen met groei in rekenvaardigheid (gemeten met de volledige BSA-toets). Daarbij is gekeken naar de beginscores op rekenattitude, rekenangst en rekenvaardigheid, evenals naar geslacht en hoe vaak leerlingen aanwezig waren bij de lessen. Er is onderzocht welke kenmerken de groei in rekenvaardigheid significant voorspellen, wanneer alle factoren gezamenlijk in het LCS-model worden meegenomen.

Leerlingkenmerken die groei in rekenvaardigheid voorspellen

In beide school- en leerjaren blijkt de beginscore op rekenvaardigheid (BSA-startscore) een significante voorspeller van de groei in rekenvaardigheid gedurende het schooljaar. De relatie is negatief: leerlingen die aan het begin lager scoorden op de BSA-toets lieten gemiddeld sterkere groei zien dan leerlingen met hogere beginscores. De vraag is echter of deze grotere groei in rekenvaardigheid bij lagere startscores een werkelijk inhaaleffect weerspiegelt of (deels) het gevolg is van meetfout. Dit onderscheid wordt nader toegelicht in Kader I.

De verklaarde variantie laat zien welk deel van de verschillen in groei in rekenvaardigheid kan worden toegeschreven aan een bepaalde voorspeller (hier: leerlingkenmerken). Voor groep 5 (schooljaar 2023/2024) verklaarde de BSA-startscore en geslacht gemiddeld[38] 7% van de groei in rekenvaardigheid over het gehele schooljaar. Hierbij groeiden jongens iets meer dan meisjes. In schooljaar 2024/2025 voorspelde alleen rekenvaardigheid de groei met een gemiddelde verklaarde variantie van 13%.

Voor groep 7 gold in 2023/2024 dat naast de beginscore op rekenvaardigheid, ook rekenangst een significante voorspeller was: leerlingen met meer angst groeiden minder sterk in hun rekenvaardigheid. Samen verklaarden startrekenvaardigheid en startrekenangst respectievelijk 3% van de verschillen in groei over het gehele schooljaar. In 2024/2025 bleek alleen de beginscore op rekenvaardigheid een significante voorspeller (verklaarde variantie = 2,5%).
Tot slot, uit de analyses bleek dat het grootste deel van individuele verschillen in de groei in rekenvaardigheid niet verklaard kon worden door de gemeten leerlingkenmerken (zie ook Discussie sectie 6.4).

Kader I – Is de grotere groei bij leerlingen met lagere startscores betekenisvol?

Uit de analyses blijkt dat leerlingen met een lagere beginscore op de BSA-toets gemiddeld meer groei laten zien in rekenvaardigheid. De vraag is echter of dit verschil wijst op een daadwerkelijk nivellerend effect (waarbij zwakkere leerlingen echt meer vooruitgaan), of dat dit het gevolg is van toeval en meetfout, een verschijnsel dat bekendstaat als *regression to the mean*.

Regression to the mean houdt in dat extreme scores (heel hoog of heel laag) bij een volgende meting vaak minder extreem zijn. Dit komt doordat iedere toets tot op zeker hoogte een meetfout bevat: het verschil tussen de geobserveerde score en de werkelijke vaardigheid van een leerling. Wanneer dezelfde toets wordt gebruikt om groei te meten (zoals bij BSA-groei = eindscore – startscore), is de meetfout van de startscore ook onderdeel van de groeiscore. Heeft een leerling bijvoorbeeld een extreem lage startscore door toevallige meetfout, dan lijkt deze leerling sterk te groeien richting de eindscore, terwijl de 'ware' score eigenlijk al dichter bij de eindscore lag. Hierdoor ontstaat een negatief verband tussen startniveau en groei, zonder dat er daadwerkelijk sprake is van inhalen.

38 De hieronder gerapporteerde cijfers hebben betrekking op de gemiddelde verklaarde variantie van de eerste en tweede helft van het schooljaar.

Om te onderzoeken of de lagere scorende leerlingen daadwerkelijk meer leerden, is ook gekeken naar de relatie tussen LVS-startscores (Cito of IEP) en de groei op de BSA-toets. Omdat de LVS-metingen onafhankelijk zijn van de BSA-groei, wijzen negatieve correlaties hier wél op een echt (nivellerend) effect.

De resultaten geven een wisselend beeld. Alleen in groep 5 (2023/2024) is er aanwijzing voor een nivellerend effect: binnen de Cito/ IEP groep correleert de LVS-startscore negatief met de groei op de BSA ($r = -.234$). In 2024/2025 is geen verband gevonden, en in groep 7 zien we zelfs een omgekeerd (denivellerend) patroon: leerlingen met lagere LVS-scores groeien iets minder ($r = .265$). Hierbij moet wel worden opgemerkt dat een groot aantal leerlingen is uitgesloten vanwege het gebruik van de IEP-toets (42% in 2023/2024 en 33% in 2024/2025).

Conclusie: alleen in groep 5 (2023/2024) lijkt sprake van een 'inhaalslag' voor zwakkere leerlingen; in andere groepen is hiervan geen overtuigend bewijs. Echter, de groepen per LVS-type zijn relatief klein, waardoor de analyses onvoldoende statistische power hebben om kleine verbanden betrouwbaar te detecteren.

5.6. Resultaten programmakenmerken

In dit deelhoofdstuk is onderzocht in hoeverre programmakenmerken de groei in rekenvaardigheid (mee) voorspellen. Daarbij is gekeken naar: het aantal lessen per leerling, de tutorvaardigheid, en de groepsgrootte. In paragraaf 4.2 is een overzicht gegeven van de groepsgroottes (aantal leerlingen per tutor, zie ook Tabel 1) en het gemiddeld aantal gevolgde lessen per leerjaar. Hieruit blijkt dat de meeste leerlingen in een 1-op-2-verband worden getutord, maar dat er ook groepen zijn met een 1-op-3- of incidenteel een 1-op-4-verhouding.) De gemiddelde tutorkwaliteit, gemeten met het observatie-instrument kwaliteit tutoring, bedraagt in groep 5: 6,85 (SD = 1,02) en in groep 7: 7,04 (SD = 1,12) op een schaal van 1 tot 10.

Programmakenmerken en de groei in rekenvaardigheid

In groep 5 (beide schooljaren) blijken geen van de programmakenmerken significante voorspellers van groei in rekenvaardigheid. Dit suggereert dat leerlingen in deze groep vergelijkbare groeiverwachtingen hebben, ongeacht verschillen in tutorkwaliteit, groepsgrootte of het aantal gevolgde lessen. Daarbij moet worden opgemerkt dat de variatie binnen deze kenmerken beperkt is: het aantal lessen dat leerlingen krijgen aangeboden is schoolafhankelijk met weinig spreiding tussen leerlingen binnen scholen, en de tutoren scoren over het algemeen tussen gemiddeld en goed op tutorvaardigheid, zonder uitschieters naar beneden. Ook kan pas in vervolgonderzoek meer inzicht worden verkregen in het effect van groepsgrootte, wanneer de verhouding 1-op-2 en 1-op-4 structureel zal worden gevarieerd.

Ook in groep 7 (2023/2024) blijken geen van de programmakenmerken significante voorspellers van groei in rekenvaardigheid. In groep 7 (2024/2025) is tutorvaardigheid daarentegen wel een significante voorspeller van groei in rekenvaardigheid ($p = 0,013$). Hoewel het effect positief is: een hogere tutorkwaliteit hangt samen met sterkere leerwinst bij leerlingen in groep 7, is het ook erg klein (verklaarde variantie = 1%; zie ook Discussie sectie 6.4). De overige programmakenmerken laten geen significant verband zien met groei in rekenvaardigheid.

6. Discussie

6.1. Het effect van de Rekenfaculteit op rekenvaardigheid

Het HDT-programma van de Rekenfaculteit heeft in zowel schooljaar 2023/2024 als 2024/2025 aantoonbaar effect gehad op de rekenvaardigheid van deelnemende leerlingen. Uit de BSA-metingen blijkt dat leerlingen van de Rekenfaculteit in groep 5 en groep 7 gemiddeld meer vooruitgang boekten dan leerlingen die niet deelnamen aan het programma. Hun extra leerwinst komt overeen met een inhaalslag van ongeveer 5 tot ruim 12 maanden extra rekenles, bovenop de leerwinst van vergelijkbare leerlingen zonder tutoring.

Ook uit de LVS-metingen blijken positieve effecten. Zowel op de CITO- als de IEP-toetsen, en in beide school- en leerjaren, is de gemiddelde leerwinst van leerlingen van de Rekenfaculteit groter dan die van de landelijke normgroepen.

Hoewel alle effecten positief zijn, laat de vergelijking met een gematchte leerlingpopulatie in groep 5 een gemengd beeld zien. Waar in 2023/2024 bij de Cito-toets nog significante effecten zichtbaar waren, is dat in 2024/2025 niet het geval. De IEP-toetsen laten daarentegen wél significante positieve effecten zien. Belangrijk om te vermelden is dat de Cito-analyses gebaseerd zijn op een kleine groep leerlingen (vier klassen). Daardoor kunnen omstandigheden in de klas of school een relatief grote invloed hebben op de uitkomsten. Mogelijke verklaringen liggen in de context. Zo kan een zogenoemd cohorteffect spelen: de samenstelling van de groep leerlingen in 2024/2025 verschilt van schooljaar 2023/2024. Ook waren er in sommige groepen onrustige omstandigheden, bijvoorbeeld door nieuwe leerlingen of het vertrek van een leerkracht. Uit onderzoek blijkt dat docentwisselingen een negatief effect kunnen hebben op leerprestaties, vooral bij leerlingen die al lager scoren (Ronfeldt et al., 2012). Zulke contextuele factoren kunnen mogelijk verklaren waarom de resultaten van de vergelijking met gematchte leerlingpopulaties in het schooljaar 2024/2025 niet eenduidig zijn. Voor dit onderzoek zijn dergelijke contextuele factoren echter niet systematisch in kaart gebracht.

6.2. Het effect van de Rekenfaculteit in vergelijking met andere rekeninterventies

De effecten van de Rekenfaculteit zijn groter dan die van andere onderwijsinterventies. Voor de BSA, de uitkomstmaat die het meest aansluit bij het doel van de Rekenfaculteit, variëren de *effect sizes* (voor beide leerjaren) van 0,60 tot 0,83: middelgrote tot grote effecten. In overzichtsstudies van onderwijsinterventies worden doorgaans minder grote effecten gerapporteerd. Zo vonden Dietrichson en collega's (2021) in een meta-analyse naar interventies voor kinderen met (risico op) leerachterstanden een gemiddelde *effect size* van 0,30. Ook een meta-analyse naar specifieke

rekeninterventies laat gemiddeld kleinere effecten zien: in deze studie varieerden de *effect sizes* van 0,00 tot 0,20, afhankelijk van het type interventie (Pelligrini et al., 2022).

In de inleiding van dit rapport is al opgemerkt dat tutoringprogramma's gemiddeld effectiever zijn dan andere onderwijsinterventies. De huidige resultaten suggereren dat de Rekenfaculteit ook binnen deze categorie bovengemiddeld scoort. Voor tutoringprogramma's gericht op rekenen rapporteren internationale overzichtsstudies gemiddelde *effect sizes* van 0,27 (Nickow et al., 2024) tot 0,39 (Kraft et al., 2024; Dietrichson et al., 2021). Ook in onderzoek naar SAGA, het programma waarop de Rekenfaculteit is gebaseerd, worden kleinere *effect sizes* gevonden: 0,16 tot 0,37. Dit verschil kan overigens ook samenhangen met de leeftijd van de doelgroep: SAGA richt zich op adolescenten, de Rekenfaculteit op basisschoolleerlingen. Het leerproces van deze groepen leerlingen verschilt aanzienlijk: de aard van de leerstof is anders net zoals het type en de hardnekkigheid van misconcepties, de grootte en aard van eventuele achterstanden en de manier waarop leerlingen leren.

Wanneer de huidige effecten vergeleken worden met de resultaten van HDT-interventie The Bridge (*effect sizes* 0,30–0,60 op de Cito-toets in groep 7), blijkt dat de effecten bij de Cito-resultaten van de Rekenfaculteit-leerlingen in dezelfde orde van grootte liggen (0,31–0,47). Bij de BSA-resultaten zijn de effecten met 0,60–0,81 zelfs nog een stuk hoger.

Kortom, de effecten van de Rekenfaculteit zijn groter dan die van andere onderwijs- en tutorinterventies. Dit duidt erop dat de Rekenfaculteit een effectieve interventie is om de rekenvaardigheid te vergroten van leerlingen met verhoogd risico op leerachterstanden, en zo rekenachterstanden in het primair onderwijs te verminderen.

6.3. Het effect van de Rekenfaculteit op rekenattitude en rekenangst

In groep 5 werd tijdens het schooljaar een grotere afname in rekenangst gevonden bij leerlingen die deelnamen aan Rekenfaculteit dan bij leerlingen van de vergelijkingsscholen. Dit patroon werd consistent gevonden in beide schooljaren, ondanks de beperkingen van de gebruikte meetinstrumenten (zie paragraaf 6.3.1.). In het schooljaar 2023-2024 werd de rekenattitude tijdens het schooljaar bovendien significant hoger bij leerlingen in groep 5 van de Rekenfaculteit dan bij hun leeftijdsgenoten op de vergelijkingsscholen. In 2024-2025 werden geen consistente veranderingen in rekenattitude gevonden.

In groep 7 werden geen significante verschillen in de vermindering van rekenangst gevonden tussen de groepen. Ook werden er voor rekenattitude geen eenduidige resultaten gevonden; de rekenattitude van leerlingen ging achteruit, maar dit gold voor beide groepen, waardoor er geen significant verschil tussen de Rekenfaculteit en vergelijkingsscholen werd gevonden.

Deze bevindingen kunnen op twee manieren worden geïnterpreteerd. Mogelijk is er slechts een beperkt effect van de tutorlessen op de rekenattitude en mate van rekenangst van leerlingen (in groep 7). Een alternatieve verklaring is dat dergelijke effecten wel bestaan, maar dat deze met het gehanteerde meetinstrumentarium niet betrouwbaar konden worden vastgesteld. In de volgende paragraaf worden mogelijke verklaringen besproken voor de beperkte en niet-eenduidige resultaten op deze uitkomstmaten.

Mogelijke verklaringen voor gevonden resultaten

Dat de resultaten voor rekenattitude en rekenangst geen duidelijk beeld gaven, kan verschillende oorzaken hebben. Zo zijn de gebruikte vragenlijsten (*Math and Me Survey*; MASC-NL) mogelijk gevoelig voor ongewenst invulgedrag. Tijdens de afnames is geobserveerd dat sommige kinderen de vragen niet altijd serieus namen of zich lieten beïnvloeden door klasgenoten. Dit kan hebben geleid tot sociaal wenselijke of anderszins onnauwkeurig ingevulde vragenlijsten[39]. Dit speelt vaker bij jonge kinderen, bij wie taalbegrip en concentratie nog in ontwikkeling zijn (Borgers, de Leeuw & Hox, 2000). Hoewel beide vragenlijsten een hoge interne consistentie laten zien (Cronbach's α MASC-NL = 0,92; Cronbach's α Math and Me Survey = 0,93), is de validiteit (de mate waarin de vragenlijsten daadwerkelijk meten wat zij beogen te meten) beperkt onderzocht voor beide instrumenten. Zo beoordeelt de *Education Endowment Foundation* de psychometrische kwaliteit van de *Math and Me Survey* als laag, mede vanwege het beperkte validiteitsonderzoek.

Een andere mogelijke verklaring ligt in de leeftijd en cognitieve ontwikkeling van de doelgroep. Tijdelijke ervaringen (zoals een vervelende rekenles of een naderende toets) kunnen de antwoorden van jonge kinderen sterk kleuren, waardoor ingevulde vragenlijsten vooral de beleving van dat moment weerspiegelen. Daarnaast staan kinderen in deze leeftijdsfase aan het begin van de overgang van dichotoom (ja/nee) denken naar het kunnen onderscheiden van meer nuance. Een vijfpuntschaal kan voor hen daardoor nog relatief complex zijn, en het is aannemelijk dat sommige leerlingen vereenvoudigde antwoordstrategieën gebruiken (Borgers, de Leeuw & Hox, 2000; Mari, 2024).

Daarbij komt dat de leerlingen in dit onderzoek sociaal en taalkundig kwetsbaar zijn. Observaties tijdens de afnames laten zien dat sommige begrippen, zoals 'voor de lol', 'verwarrend' en 'nerveus', lastig te begrijpen waren, met name voor de groep 5 leerlingen. Tijdens de vragenlijst afnames worden deze begrippen toegelicht. Desondanks, is het mogelijk dat sommige kinderen antwoorden geven zonder de vraag volledig te begrijpen. Deze ontwikkelings- en taalgerelateerde factoren maken de resultaten niet ongeldig, maar beperken wel de zekerheid waarmee de resultaten m.b.t. rekenattitude en rekenangst kunnen worden geïnterpreteerd.

Alles afwegend kunnen de gevonden resultaten niet worden geïnterpreteerd als bewijs dat de Rekenfaculteit geen effect heeft op rekenattitude of rekenangst. Wel schetsen de data een voorlopig beeld: tutorlessen lijken een consistent en robuust effect te hebben op de vermindering

39 Bijvoorbeeld overal hoog of laag scoren zonder de items goed te lezen.

van rekenangst in groep 5, maar er is geen aanwijzing voor (grote) effecten op het vergroten van de rekenattitude van leerlingen (in groep 5 en 7) of het verminderen van rekenangst in groep 7. De onderzoeksopzet en steekproefomvang waren voldoende om effecten te detecteren; deze zijn echter niet gevonden. Dat is in lijn met de verwachting dat mogelijke effecten voor rekenangst en rekenattitude eerder subtiel zijn, en met de gehanteerde meetinstrumenten lastiger aan te tonen.

Een preciezer inzicht in deze niet-cognitieve factoren vraagt om een bredere onderzoeksaanpak, bijvoorbeeld door observaties van het gedrag van leerlingen in de klas en aanvullende informatie hierover vanuit tutoren en leerkrachten te combineren met de huidige leerlingvragenlijsten. Daarnaast kan kwalitatief onderzoek inzicht bieden in hoe rekenattitude en rekenangst bij leerlingen tot uiting komen en hoe deze in de loop van de tijd al dan niet veranderen.

6.4. Invloed van leerling- en programmakenmerken op de groei van rekenvaardigheid

Om beter te begrijpen waardoor verschillen in leergroei ontstaan, is onderzocht in hoeverre kenmerken van leerlingen en van het programma samenhangen met de groei in rekenvaardigheid. Hierbij is gekeken naar leerlingkenmerken zoals de beginscores op rekenvaardigheid, rekenattitude en rekenangst, evenals naar geslacht en aanwezigheid bij de lessen. Daarnaast zijn programmakenmerken meegenomen, zoals het aantal gevolgde lessen, de tutorkwaliteit en de groepsgrootte.

Binnen de Rekenfaculteit is zichtbaar dat leerlingen in uiteenlopende mate groeien in rekenvaardigheid: sommige leerlingen boeken duidelijke vooruitgang, terwijl anderen minder sterk groeien. Uit de analyses blijkt echter dat de invloed van de gemeten leerling- en programmakenmerken hierop beperkt is. De modellen verklaren slechts een klein deel van de verschillen in groei. Dit kan deels samenhangen met meetfouten in het instrument dat groei vaststelt (de BSA), en deels met de manier waarop bepaalde kenmerken zijn gemeten: zoals de beperkingen van de vragenlijsten voor rekenangst en rekenattitude (zie paragraaf 6.3). Daarnaast is het goed mogelijk dat factoren die buiten de reikwijdte van dit onderzoek vielen, zoals familiestructuur, ouderlijke stress of gezinsconflict en educatieve verwachtingen, ook van invloed zijn op de verschillen in groei tussen leerlingen (Wang et al., 2023).

In groep 7 was wel een verband tussen tutorkwaliteit en leerwinst zichtbaar: hogere tutorkwaliteit hing samen met sterkere groei in rekenvaardigheid. De verklaarde variantie van dit model was echter zeer klein, waardoor de vraag blijft in hoeverre dit resultaat ook inhoudelijk relevant is. Vervolgonderzoek is nodig om de beschikbare data over tutorkwaliteit optimaal te benutten (in dit rapport is uitsluitend de gemiddelde observatiescore gebruikt om tutorkwaliteit te meten). Daarmee kan in detail worden nagegaan in hoeverre de werkzame elementen van de Rekenfaculteit daadwerkelijk terugkomen in de tutorlessen. Ook kunnen hiermee afzonderlijke onderdelen van het programma worden gekoppeld aan groei in rekenvaardigheid en kan worden onderzocht hoe de

kwaliteit van tutorlessen zich gedurende het schooljaar ontwikkelt. Aanvullende analyses per programmaonderdeel worden nog verder uitgewerkt.

Bij het interpreteren van de resultaten over de invloed van programmakenmerken is het belangrijk te bedenken dat de programmakenmerken niet losstaan van de scholen waar de Rekenfaculteit is uitgevoerd. Dit betekent dat de bevindingen niet eenvoudig kunnen worden gegeneraliseerd naar andere scholen. Vooral de variabelen groepsgrootte en dosering zijn sterk afhankelijk van schoolcontext: bij scholen met grotere klassen is er vaker gekozen voor het werken in trio's, en de dosering hangt deels samen met organisatorische mogelijkheden. In een vervolgstudie waarin groepsgrootte (bijvoorbeeld 1-op-2 en 1-op-4) binnen scholen willekeurig wordt gevarieerd en de dosering tussen scholen op een gestructureerde manier wordt aangepast, kan in meer detail worden onderzocht wat het effect is van deze programmakenmerken op de groei in rekenvaardigheid.

Kortom, het ontbreken van een duidelijk verband tussen leerlingkenmerken en leergroei laat zien dat de Rekenfaculteit voor alle leerlingen effectief kan zijn, ongeacht geslacht, rekenvaardigheidsniveau, rekenangst, attitude of aanwezigheid bij de lessen. De invloed van programmakenmerken op de groei in rekenvaardigheid wordt verder onderzocht in vervolgonderzoek.

6.5. Sterke punten en beperkingen

Om de bevindingen over de effectiviteit van de Rekenfaculteit goed te kunnen duiden, worden hieronder de belangrijkste sterke punten en beperkingen van het onderzoek besproken:

- Een sterk punt van dit onderzoek is de combinatie van een grote onderzoekspopulatie (ruim 1.200 leerlingen), dataverzameling over twee schooljaren en de grondige analyses op zowel BSA- als LVS-gegevens, inclusief normgroepen. Hierdoor ontstaat een stevig beeld van de ontwikkeling van leerlingen en de effecten van de interventie.
- Dit onderzoek is uitgevoerd met een quasi-experimenteel design. Dit brengt mee dat er mogelijk systematische verschillen tussen groepen bestaan, wat invloed kan hebben op de effectgroottes. Eerder onderzoek laat zien dat effectgroottes in quasi-experimenteel onderwijsonderzoek doorgaans iets hoger zijn dan in RCT-designs (Cheung & Slavin, 2016). Hoewel een quasi-experimenteel design minder controle biedt over mogelijke groepsverschillen, laten de aanvullende sensitiviteitsanalyses zien dat de resultaten robuust zijn. Bovendien is er in twee schooljaren (cohorten) data verzameld, waarbij de resultaten consistent bleven.
- Over het geheel genomen laten de resultaten robuuste en positieve effecten zien van de Rekenfaculteit op de rekenvaardigheid van leerlingen. Er zijn echter aanzienlijke verschillen tussen scholen: niet alle scholen boeken in dezelfde mate vooruitgang. Voor deze verschillen is nog geen sluitende verklaring gevonden op basis van de huidige

onderzoeksresultaten. Het systematisch in kaart brengen van schoolcontexten kan helpen om verschillen tussen scholen te duiden.

- De resultaten geven een globaal beeld van het effect van de Rekenfaculteit op veranderingen in rekenattitude en rekenangst, maar laten geen grote of consistente effecten zien. Dit sluit echter niet uit dat de Rekenfaculteit invloed heeft: de gebruikte vragenlijsten zijn momentopnamen en kennen beperkingen. Hoewel de MASC-NL en de *Math and Me Survey* betrouwbaar zijn, is hun validiteit voor deze specifieke doelgroep mogelijk beperkt door factoren als taalbegrip, context en invulgedrag. Ze bieden daardoor een bruikbare, maar nog grove indicatie. Voor een vollediger en preciezer beeld is een combinatie van meetmethoden wenselijk, zoals observaties in de klas en aanvullende informatie van tutoren en leerkrachten.
- De invloed van leerling- en programmakenmerken op groei in rekenvaardigheid is uitsluitend onderzocht voor de factoren die in het onderzoek zijn opgenomen.
- Tot slot, hoewel dit onderzoek overtuigend aantoont dat de Rekenfaculteit een effectieve interventie is om de rekenvaardigheden van basisschoolleerlingen te vergroten, vormen de relatief hoge kosten een belangrijk obstakel voor verdere opschaling van het programma. Momenteel wordt daarom een vervolgonderzoek opgezet om te onderzoeken of kostenreductie met behoud van effectiviteit mogelijk is. Dit onderzoek richt zich op de vraag of 1-op-4-tutoring (met/zonder minder lesuren) even effectief en uitvoerbaar is als de huidige 1-op-2-variant.

6.6. Conclusie

Het HDT-programma van de Rekenfaculteit laat in zowel 2023–2024 als 2024–2025 overtuigende effecten zien op de rekenvaardigheid van deelnemende leerlingen. De behaalde leerwinst komt overeen met een aanzienlijke inhaalslag en is groter dan bij veel andere onderwijsinterventies. Ook in vergelijking met internationale studies naar tutoringprogramma's vallen de effecten vergelijkbaar of zelfs hoger uit. Daarmee is de Rekenfaculteit een krachtige interventie om rekenachterstanden in het primair onderwijs te verkleinen.

Voor de aanvullende onderzochte (niet-cognitieve) uitkomstmaten komt een gemengd beeld naar voren. Bij Rekenfaculteit-leerlingen in groep 5 nam rekenangst sterker af dan bij leerlingen van de vergelijkingsscholen. In groep 7 werden dergelijke verschillen echter niet gevonden, en ook de effecten op rekenattitude bleken over beide leerjaren beperkt en niet eenduidig. Daarnaast was de invloed van leerling- en programmakenmerken beperkt; de modellen konden de groei in rekenvaardigheid maar voor een klein deel verklaren. Het ontbreken van een duidelijk verband tussen leerlingkenmerken en leergroei laat zien dat de Rekenfaculteit voor alle leerlingen effectief kan zijn, ongeacht geslacht, rekenvaardigheidsniveau, rekenangst, attitude of aanwezigheid bij de lessen.

De resultaten van de Rekenfaculteit zijn voor beide leerjaren veelbelovend. Er blijven nog een aantal vragen onbeantwoord. Zo is het nog onduidelijk of de effecten op de lange termijn behouden blijven en hoe verbeteringen in rekenvaardigheid samenhangen met de taalontwikkeling van leerlingen. De observatiedata over tutorkwaliteit bevatten daarnaast nog veel waardevolle informatie voor vervolgonderzoek.

7. Literatuurlijst

Adelson, J. L., & McCoach, D. B. (2011). Development and psychometric properties of the Math and Me Survey. Measurement and Evaluation in Counseling and Development, 44(4), 225–247. https://doi.org/10.1177/0748175611418522

Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. Bulletin de Méthodologie Sociologique, 66(1), 60–75. https://doi.org/10.1177/075910630006600106

De Boom, J., Roode, A. L., Van Wensveen, P., & Vogel, I. (2024). Monitor scholen op Zuid: Het primair onderwijs, 2023. Risbo, Erasmus Universiteit; OBI, gemeente Rotterdam.

Centraal Bureau voor de Statistiek. (z.d.). Onderwijsachterstanden [Dashboard]. Geraadpleegd op 15-08-2024, van https://dashboards.cbs.nl/v5/onderwijsachterstanden/

Centraal Planbureau. (2018, 6 juni). Effectieve interventies leerachterstanden in het primair onderwijs (CPB-notitie). Den Haag: Centraal Planbureau.

Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. Educational Researcher, 45(5), 283–292. https://doi.org/10.3102/0013189X16656615

Cook, P. J., Dodge, K., Farkas, G., Fryer, R. G., Jr., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., & Steinberg, L. (2014). The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth: Results from a randomized experiment in Chicago (NBER Working Paper No. 19862). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w19862/w19862.pdf

Dietrichson, J., Filges, T., Seerup, J. K., Klokker, R. H., Viinholt, B. C., Bøg, M., & Eiberg, M. (2021). Targeted school-based interventions for improving reading and mathematics for students with or at risk of academic difficulties in grades K–6: A systematic review. Campbell Systematic Reviews, 17(2), e1152. https://doi.org/10.1002/cl2.1152

Evers, A., & Resing, W. (2007). Het drijfzand van didactische leeftijdsequivalenten. De Psycholoog, 42(9), 466–472.

Geenen, R. (2023). Validatie van de MASC-NL voor het meten van rekenangst bij kwetsbare basisschoolleerlingen [Bachelorscriptie, Erasmus Universiteit Rotterdam].

Gemeente Rotterdam. (z.d.). Staat van het onderwijs: Primair onderwijs. Geraadpleegd op 15-08-2025, van https://onderzoek010.nl/mosaic/staat-van-het-onderwijs/primair-onderwijs

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M., Dodge, K., ... Stoddard, G. (2023). Not too late: Improving academic outcomes among adolescents. American Economic Review, 113(3), 738–765. https://doi.org/10.1257/aer.20200513

Inspectie van het Onderwijs. (z.d.). De schoolweging: maat voor de leerlingenpopulatie. Ministerie van Onderwijs, Cultuur en Wetenschap. Geraadpleegd op 16 augustus 2024, van https://www.onderwijsinspectie.nl/onderwerpen/onderwijsresultaten-primair-onderwijs/de-schoolweging-maat-voor-de-leerlingenpopulatie

Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H. G., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. Learning and Individual Differences, 24, 190–197.

Jepma, I., Vander Heyden, K., Waaijer, C., Welles, C., & Hoonhoud, E. (2024, december). Kansen op een goede start: Onderzoek naar het vergroten van onderwijskansen bij de start in het basisonderwijs (Sardes & SEO Economisch Onderzoek). Sardes. https://sardes.nl/wp-content/uploads/2025/03/Sardes-SEO-Kansen-op-een-goede-start-17012025-definitief-1.pdf
Kraft, M. A. (2020). Interpreting effect sizes of education interventions. Educational researcher, 49(4), 241-253.

Kraft, M. A., Schueler, B. E., & Falken, G. (2024). What impacts should we expect from tutoring at scale? Exploring meta-analytic generalizability (EdWorkingPaper No. 24-1031). Annenberg Institute for School Reform at Brown University.

Ledoux, G. (2021, februari). Wat zijn onderwijsachterstanden? Nationaal Regieorgaan Onderwijsonderzoek (NRO). Onderwijskennis.nl. https://www.onderwijskennis.nl/sites/onderwijskennis/files/media-files/Wat%20zijn%20onderwijsachterstanden%20-%20DEF%20-%20PDF.pdf

Mari, A. (2024). Making conventional data collection more child-friendly: Questionnaires with young students. African Evaluation Journal, 12(1), Article a690. https://doi.org/10.4102/aej.v12i1.690

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data.

Melis, G., Oosterveld, P., & Schokker, J. (2012). Het verantwoord gebruik van didactische leeftijdsequivalenten. Tijdschrift voor Remedial Teaching, 20(2), 30–33.

Newsom, S. (2015). Longitudinal structural equation modeling (p. 248). Routledge.

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence.

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The promise of tutoring for PreK–12 learning: A systematic review and meta-analysis of the experimental evidence. American Educational Research Journal, 61(1), 74-107.

NRO. (z. d.). In de klas aan de slag met reken-wiskunde | Onderwijskennis. Geraadpleegd op 04-09-2025, van https://www.onderwijskennis.nl/themas/rekenen-wiskunde-in-groep-3-5

Paulle, B., De Ree, J., Epema, T., & Speelman, I. (2020, 23 juni). Rekenen op de toekomst: Een evaluatie van twee jaar Bridge High Dosage Tutoring in het basisonderwijs te Haarlem Schalkwijk en Haarlem Oost [Rapport]. Universiteit van Amsterdam.

Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. Aera Open, 7, 2332858420986211.

De Ree, J., Maggioni, M. A., Paulle, B., Rossignoli, D., Ruijs, N., & Walentek, D. (2023). Closing the income-achievement gap? Experimental evidence from high-dosage tutoring in Dutch primary education. Economics of Education Review, 94(94), 102383.

Reise, S. P., & Revicki, D. A. (2014). Handbook of item response theory modeling. New York, NY, USA: Taylor & Francis.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. American educational research journal, 50(1), 4-36.

Tijms, J., & De Ree, J. (2023, 24 januari). Reken op Zuidoost: Een evaluatie van drie jaar Bridge High-Dosage Tutoring in het Amsterdamse basisonderwijs [Rapport]. Universiteit van Amsterdam.

Wang, X. S., Perry, L. B., Malpique, A., & Ide, T. (2023). Factors predicting mathematics achievement in PISA: a systematic review. Large-Scale Assessments in Education, 11(1), 24.

Wijzer over de Basisschool. (z.d.). Cito-volgsysteem: Wat je moet weten. Geraadpleegd op 15-08-2024, van https://wijzeroverdebasisschool.nl/uitleg/cito-volgsysteem

Zumbuehl, M., & Dillingh, R. (2020, december). Ongelijkheid van het jonge kind (CPB Notitie). Centraal Planbureau.https://www.cpb.nl/sites/default/files/omnidownload/CPB-Notitie-Ongelijkheid-van-het-jonge-kind.pdftekst begint hier]

8. Bijlagen

8.1 Bijlage 1 - Onderzoeksvragen

Onderzoeksvraag 1: In hoeverre heeft de Rekenfaculteit effect op de rekenresultaten van deelnemende leerlingen?
1a: Wat is het effect van de Rekenfaculteit gedurende het schooljaar op rekenvaardigheid (BSA) van hun leerlingen in vergelijking met leerlingen die niet hebben deelgenomen aan de Rekenfaculteit?
1b: Wat is het effect van de Rekenfaculteit gedurende het schooljaar op de rekenvaardigheid (BSA) van de Rekenfaculteit-leerlingen?
1c: In hoeverre verschilt de groei in rekenvaardigheid gedurende het schooljaar van Rekenfaculteit-leerlingen met de groei in rekenvaardigheid van hun landelijke tegenhangers op de Cito- en IEP-toetsen?
1d: In hoeverre blijft de (mogelijke) extra groei gedurende het schooljaar bestendigd na deelname aan de Rekenfaculteit?

Onderzoeksvraag 2: In hoeverre heeft de Rekenfaculteit effect op de rekenattitude en rekenangst van deelnemende leerlingen?
2a: Wat is het effect van de Rekenfaculteit gedurende het schooljaar op de rekenattitude en rekenangst van hun leerlingen in vergelijking met leerlingen die niet hebben deelgenomen aan de Rekenfaculteit?
2b: Wat is het effect van de Rekenfaculteit gedurende het schooljaar op de rekenattitude en rekenangst van de Rekenfaculteit-leerlingen?

Onderzoeksvraag 3: Welke leerlingkenmerken dragen bij aan de effectiviteit van de Rekenfaculteit?
3a: In hoeverre dragen de volgende kenmerken bij aan de effectiviteit van de Rekenfaculteit: start rekenvaardigheid, startrekenattitude, startrekenangst, geslacht en mate van aanwezigheid tijdens tutorlessen?

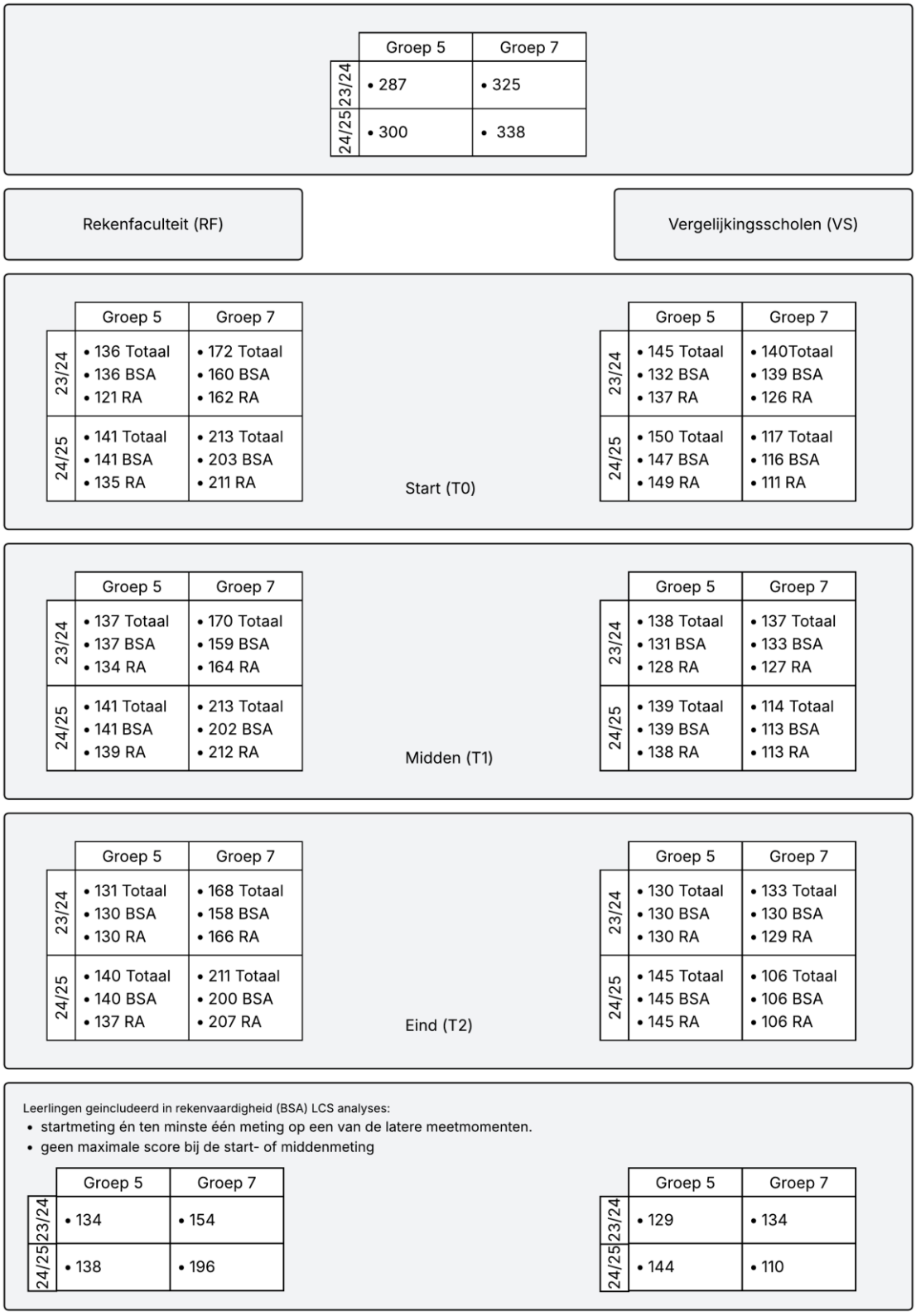
Onderzoeksvraag 4: Welke programmakenmerken dragen bij aan de effectiviteit van de Rekenfaculteit?
4a: In hoeverre hebben de uitvoeringscondities van de Rekenfaculteit invloed op de effectiviteit van de Rekenfaculteit op de rekenresultaten van deelnemende leerlingen? *Uitvoeringscondities zijn omstandigheden die de uitvoering van het programma mogelijk beïnvloeden, zoals de groepsgrootte, het opleidingsniveau en het aantal jaren ervaring van de tutor, en of de lessen geheel of gedeeltelijk buiten het reguliere curriculum plaatsvinden*
4b: In hoeverre heeft de programma-integriteit invloed op de effectiviteit van de Rekenfaculteit op de rekenresultaten van de deelnemende leerlingen? *Programma-integriteit is de mate waarin de tutoring daadwerkelijk volgens de werkwijze van de Rekenfaculteit wordt uitgevoerd.*

Onderzoeksvraag 5: In hoeverre heeft de Rekenfaculteit effect op de taalvaardigheid van deelnemende leerlingen?

5a: Wat is het effect van de Rekenfaculteit gedurende het schooljaar op taalvaardigheid (LVS) van hun leerlingen in vergelijking met leerlingen die niet hebben deelgenomen aan de Rekenfaculteit?

8.2 Bijlage 2 - Statistische achtergrond en toelichting

1. Flowchart aantal leerlingen per meetmoment (Rekenfaculteit vs. Vergelijkingsscholen)



Figuur B1. *Overzicht van leerlingaantallen per meetmoment per leerjaar en per conditie. BSA = Basic Skills Assessment (rekenvaardigheid); RA = rekenangst en rekenattitude.*

2. Validatie Basic Skills Assessment (BSA)

De BSA is ontwikkeld en gevalideerd over een periode van drie jaar, met als doel een betrouwbaar instrument te creëren om rekenvaardigheid te meten, met een specifieke focus op leerlingen in Rotterdam-Zuid die vaak lager scoren op de Cito of IEP-toetsen. Het instrument bestaat uit 3 parallelle toetsen met verschillende items die d.m.v. een statistisch model geijkt zijn zodat de scores vergelijkbaar zijn en ook hetzelfde meten.

De BSA omvat de vier rekendomeinen[40]: 1. Getallen, 2. Meten, 3. Verbanden, 4. Verhoudingen. Het validatieproces vond plaats bij zowel de Rekenfaculteit als bij andere scholen van BOOR, met de deelname van:

- *145 leerlingen uit groep 5 en 163 leerlingen uit groep 7 van de Rekenfaculteit*
- *322 leerlingen uit groep 5 en 327 leerlingen uit groep 7 van BOOR-scholen*

Tijdens dit proces werden uiteindelijk 533 vragen ontwikkeld en in meerdere rondes beoordeeld door vakdidactici om ervoor te zorgen dat ze geschikt waren voor de doelgroep van de Rekenfaculteit.

De BSA is gevalideerd met behulp van een Item Response Theory (IRT-)schaal. Dit houdt in dat de moeilijkheidsgraad van elke vraag is bepaald op basis van hoe leerlingen deze hebben beantwoord. Dit is belangrijk omdat de rekenvaardigheid van leerlingen, uitgedrukt in een vaardigheidsscore (variërend van 150 tot 300), nauwkeuriger kan worden vastgesteld wanneer de toets vragen van verschillende moeilijkheidsniveaus bevat. Bijvoorbeeld, als alle opgaven relatief moeilijk zijn, presteren uitsluitend de sterkere leerlingen goed. Dit maakt het onmogelijk om binnen de overige groep betekenisvolle verschillen in rekenniveau te onderscheiden. Daarnaast maakt de IRT-schaal het mogelijk om verschillende versies van de BSA-toets te ontwikkelen, waarbij elke versie deels andere vragen bevat, terwijl de vaardigheden van leerlingen consistent gemeten blijven. Dit betekent dat een leerling dezelfde vaardigheidsscore zou behalen, ongeacht welke versie van de BSA-toets hij of zij maakt.

De BSA bleek een valide meetinstrument om rekenvaardigheid te meten. Dankzij deze validatie kan het effect van de Rekenfaculteit op de groei in rekenvaardigheid van haar leerlingen betrouwbaar gemeten worden.

3. Het IRT-model nader uitgelegd

Voor het meten van rekenvaardigheid (BSA-score) is gebruikgemaakt van een *Item Response Theory* (IRT-)model (Reise & Revicki, 2014). Dit statistische model beschrijft de relatie tussen de rekenvaardigheid bij een leerling en de kans dat deze een toetsvraag correct beantwoordt. Een IRT-model is geschikt voor het meten van ‘latente variabelen’—eigenschappen die niet direct zichtbaar zijn, zoals rekenvaardigheid.

40 Zoals deze ook in het PO-curriculum en in de referentieniveaus rekenen gehanteerd worden.

Het IRT-model stelt vast hoe groot de kans is dat een leerling met een bepaalde mate van rekenvaardigheid een BSA-toetsvraag goed beantwoordt. Daarbij houdt het model rekening met twee belangrijke eigenschappen van de vraag: moeilijkheidsgraad en onderscheidend vermogen, oftewel hoe goed de vraag onderscheid maakt tussen leerlingen met verschillende niveaus van rekenvaardigheid.

In de analyse is het IRT-model stap voor stap verfijnd, totdat deze optimaal aansloot bij de verzamelde testgegevens. Dit proces zorgt ervoor dat de BSA-scores de werkelijke rekenvaardigheid van leerlingen zo goed mogelijk weerspiegelen.

Het IRT-model schat de vaardigheid van leerlingen en de moeilijkheid van toetsvragen op basis van de kans dat een leerling een vraag correct beantwoordt. Als een leerling dezelfde vaardigheid heeft als de moeilijkheidsgraad van de vraag, zou de kans op een correct antwoord 50% moeten zijn.
Een goede manier om het IRT-model te begrijpen, is door het te vergelijken met het Elo-systeem dat in schaken wordt gebruikt om de vaardigheid van spelers te bepalen. In het Elo-systeem krijg je punten wanneer je een schaakwedstrijd wint en verlies je punten wanneer je verliest. Hoeveel punten je wint of verliest, hangt af van het verschil in Elo-punten tussen jou en je tegenstander aan het begin van de wedstrijd. Als je wint van een sterkere tegenstander, verdien je meer punten dan wanneer je wint van een zwakkere tegenstander. Dit systeem kwantificeert schaakvaardigheid op basis van prestaties tegen verschillende tegenstanders.

In het IRT-model is het vergelijkbaar, maar dan ‘spelen’ leerlingen niet tegen elkaar, maar tegen de toetsvragen. Als een leerling moeilijkere vragen correct beantwoordt, krijgt hij of zij meer punten. Omgekeerd verliest een leerling punten als hij of zij een gemakkelijkere vraag fout beantwoordt. Omdat alle leerlingen in een BSA-toets dezelfde vragen maken, is het totaal aantal behaalde punten bepalend. Op deze manier wordt de rekenvaardigheid gemeten op basis van hoeveel en welke vragen de leerlingen correct hebben beantwoord.

Het IRT-model gebruikt zogenaamde ‘anker-items’ om scores van verschillende toetsmomenten vergelijkbaar te maken. Deze vaste vragen zorgen ervoor dat, zelfs wanneer verschillende populaties of toetsversies worden gebruikt, de scores gestandaardiseerd kunnen worden. Hierdoor kan de groei van leerlingen over tijd nauwkeurig worden vastgesteld, ondanks variaties in de toetsvragen.

4. Validatie observatie-instrument

De validatie van het observatie-instrument vond plaats in vier fasen tussen september 2023 en mei 2024:

1. Eerste betrouwbaarheidsanalyse (maart 2023): In de eerste fase werd de betrouwbaarheid van de oorspronkelijke vragenlijst onderzocht. De betrouwbaarheid van het volledige instrument was zeer hoog (α = 0.906).

2. Vakdidactische kwaliteitscheck (jan-maart 2024): Een vakdidacticus beoordeelde de vragenlijst, wat leidde tot aanpassingen in de items en een herziening van het observatie-instrument, resulterend in 36 items verdeeld over vijf categorieën: maatwerk, effectieve leertijd, band tussen tutor en leerling, rekenattitude en structuur van de les.

3. Observatortrainingen (maart 2024): Er werden twee trainingen met observatoren georganiseerd om het nieuwe observatie-instrument te testen en af te stemmen, wat leidde tot verdere verfijningen in items en scoringsmethoden. De nieuwe versie van het observatie-instrument maakt gebruik van een 4-punts Likertschaal, waarbij cijfers zijn omgezet naar letters.

4. Definitieve validatieanalyses (april-mei 2024): 26 tutorlessen werden geobserveerd door 6 observatoren in koppels om de intercodeurbetrouwbaarheid (ICB) vast te stellen. De betrouwbaarheid van de vernieuwde lijst verbeterde (α = 0.934), en de structuur van de vijf categorieën werd bevestigd door een *Confirmatory Factor Analyse* (CFA). De uiteindelijke ICB op categorieniveau was voldoende voor gebruik in het evaluatieonderzoek. Items met de meeste mismatches werden in een laatste training met observatoren besproken.

5. Voordelen van het Latent Change Score (LCS)-model

Het LCS-model maakt het mogelijk om conclusies te trekken over de groei in rekenvaardigheidsscores en de factoren die deze groei verklaren. Dit zijn de voordelen van het LCS-model:

- **Eenvoudige interpretatie:** het LCS-model levert direct een score op die laat zien hoeveel groei er per groep (Rekenfaculteit vs. vergelijkingsgroep) is en of deze groei verschilt tussen de groepen. Dit maakt de resultaten eenvoudig te interpreteren.
- **Eenvoudige analyse van covariaten:** het LCS-model maakt het eenvoudig om te onderzoeken hoe aanvullende factoren, zoals leerlingkenmerken (o.a. startrekenvaardigheid) en programmakenmerken (o.a. groepsgrootte) de groei in rekenvaardigheid beïnvloeden.
- **Toetsing groeiverschillen onder verschillende aannames:** het LCS-model maakt het mogelijk om de verschillen tussen de Rekenfaculteit en vergelijkingsscholen te toetsen onder verschillende aannames over groei. Bijvoorbeeld, is de groei verschillend tussen de start- en middenmeting en tussen de midden- en eindmeting?

6. Vergelijking van LCS-groeimodellen

Om te bepalen welk model het beste aansluit bij onze data, hebben we drie groeimodellen met elkaar vergeleken.

Gelijkmatige groei

In dit eenvoudigste model wordt aangenomen dat de groei tussen de eerste en tweede meting (η_1) gelijk is aan de groei tussen de tweede en derde meting (η_2). Er is dus sprake van een constant groeitempo gedurende de onderzoeksperiode.

Verschillende groeiperioden
Met dit model wordt onderzocht of het groeitempo tussen beide perioden verschilt ($\eta_1 \neq \eta_2$). Dit model houdt er rekening mee dat leerlingen in de eerste periode sneller of juist langzamer kunnen groeien dan in de tweede periode.

Feedbackmodel
In dit model wordt gekeken of de behaalde score op een eerder meetmoment van invloed is op de groei in de daaropvolgende periode. De startscore voorspelt hierbij de groei in de eerste periode (η_1), en de score op het tweede meetmoment voorspelt de groei in de tweede periode (η_2). Dit noemen we ook wel de *feedbackparameter*: we onderzoeken hiermee in hoeverre eerdere prestaties de verdere ontwikkeling beïnvloeden. In dit model worden verschillen in groeitempo opgenomen wanneer deze blijken uit het vorige groeimodel (verschillende groeiperioden).

Na het vergelijken van deze modellen is vastgesteld welk model het beste past bij de onderzoeksdata. Vervolgens is met dit model onderzocht of leerlingen van de Rekenfaculteit significant meer groei in rekenvaardigheid laten zien dan leerlingen van de vergelijkingsscholen.

7. Stappen in de matchinganalyse (Cito en IEP)
Om de leerwinst van Rekenfaculteit-leerlingen zo goed mogelijk te vergelijken met die van vergelijkbare leerlingen uit de landelijke Cito- en IEP-databases, zijn verschillende niveaus van matching toegepast. Daarbij is steeds gekeken hoe nauwkeurig de vergelijkingsgroep overeenkomt met de Rekenfaculteit-leerlingen op basis van hun beginniveau.

Er is bewust gekozen om te werken met meerdere niveaus van nauwkeurigheid in de matching. Wanneer leerlingen van de Rekenfaculteit steeds worden gematcht met exact dezelfde (kleine groep) vergelijkingsleerlingen, kan een vertekend beeld ontstaan dat niet goed te generaliseren is naar de bredere populatie. De meest nauwkeurige matching is daarom niet per definitie de beste. Door met verschillende matchingsniveaus te werken, ontstaat een robuuster beeld van de leerwinst — dat wil zeggen: een beeld dat minder afhankelijk is van één specifieke vergelijking en daarmee betrouwbaarder is.

De analyse is uitgevoerd in de volgende stappen:
Géén matching
In eerste instantie is de gemiddelde leerwinst van Rekenfaculteit-leerlingen vergeleken met de gemiddelde leerwinst van alle leerlingen in de Cito- of IEP-database. Dit gebeurde met een *one sample t-test*. In deze stap is dus geen matching toegepast.

1. **Matching op basis van niveauscore I t/m V**
   Bij deze stap worden Rekenfaculteit-leerlingen gekoppeld aan vergelijkbare leerlingen uit de landelijke database met hetzelfde vaardigheidsniveau (een grovere indeling dan de decielscores hieronder). Deze niveaus lopen van I (hoogste scores) tot en met V (laagste scores). Elk niveau bestaat uit twee decielen, bijvoorbeeld: niveau I bevat de 20% leerlingen

met de hoogste scores, en niveau V de 20% met de laagste scores. Zo ontstaat per Rekenfaculteit-leerling een vergelijkingsgroep met ongeveer hetzelfde startniveau.

2. **Matching op basis van decielscore**
   In deze stap vindt de matching plaats op een fijner vaardigheidsniveau dan bij stap 2. In plaats van vijf brede niveaus wordt gekeken naar de exacte decielscore van een leerling: de relatieve positie van zijn of haar score ten opzichte van de landelijke normgroep. Het eerste deciel bevat de 10% leerlingen met de laagste vaardigheidsscores, het tiende deciel de 10% met de hoogste scores. Zo wordt elke Rekenfaculteit-leerling vergeleken met landelijke leerlingen die qua startscore vrijwel identiek zijn.

3. **Matching op basis van Didactische Leeftijdsequivalent (DLE)**
   In deze stap wordt gematched op basis van het aantal onderwijsmaanden dat een leerling heeft genoten (DLE), uitgaande van tien maanden per schooljaar (Anstadt, 2019). Zo wordt een Rekenfaculteit-leerling vergeleken met landelijke leerlingen die ongeveer even ver zijn in hun schoolloopbaan.

4. **Matching op basis van vaardigheidsscore**
   Dit is de meest precieze matchingsmethode, waarbij wordt gematched op de vaardigheidsscore van een leerling op de toets. Rekenfaculteit-leerlingen worden zo vergeleken met landelijke leerlingen die hetzelfde scoorden bij de startmeting.

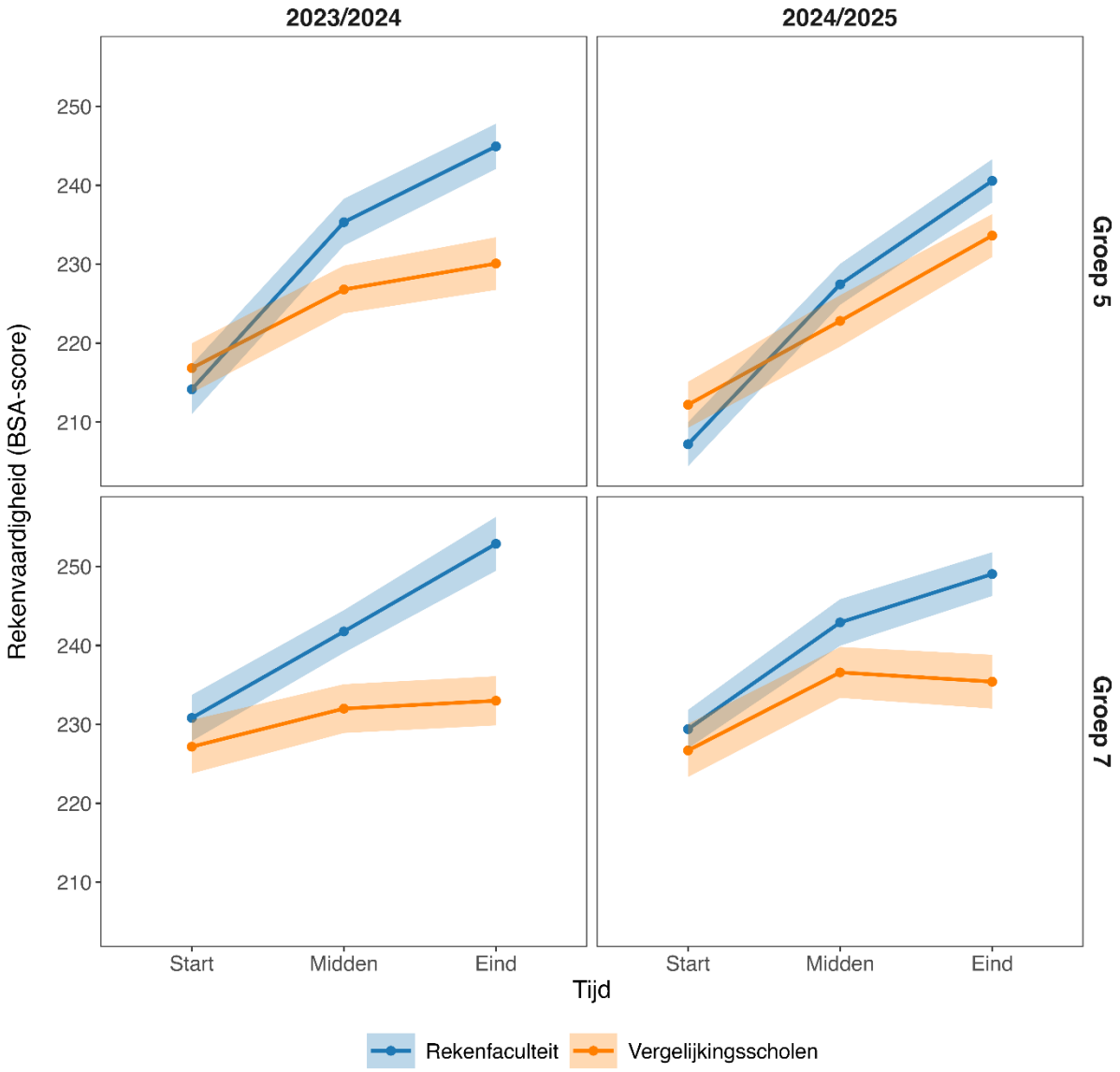
In een tweede reeks analyses is in elk van deze stappen ook de achterstandsscore van 2024 meegenomen, om tot een nog nauwkeurigere vergelijking te komen. Hieronder staat beschreven hoe dit in zijn werk ging.

Achterstandsscore

De achterstandsscore is een maat voor de verwachte onderwijsachterstand op een school. Het ministerie van Onderwijs, Cultuur en Wetenschap (OCW) gebruikt deze score om het onderwijsachterstandenbudget over scholen te verdelen. Het Centraal Bureau voor de Statistiek (CBS) berekent deze scores op basis van de onderwijsachterstandenindicator. Daarbij wordt eerst per leerling een onderwijsscore vastgesteld, die vervolgens wordt opgeteld tot een totaalscore per basisschool (CBS, 2024).

In dit onderzoek wordt de exacte achterstandsscore niet exact meegenomen, omdat dit de matching vrijwel onmogelijk zou maken. De kans is dan groot dat er dan geen voldoende vergelijkbare leerlingen overblijven. In plaats daarvan is ervoor gekozen de achterstandsscore van de scholen af te ronden op het dichtstbijzijnde honderdtal. Deze afgeronde score wordt vervolgens gebruikt als extra criterium bij het kiezen van vergelijkingsleerlingen, naast een van de eerdergenoemde matchingsmethoden. Zo wordt een vergelijkingsleerling gekozen die qua achtergrond nóg beter aansluit bij een leerling uit de Rekenfaculteit.

8. Aanvullende Figuren



Figuur B2. Effect van de Rekenfaculteit op rekenvaardigheid (BSA). De donkergekleurde lijn toont de gemiddelde groei in BSA over tijd voor zowel de Rekenfaculteit als de vergelijkingsscholen. De lichtgekleurde band rond de lijn geeft het 95% betrouwbaarheidsinterval weer.

Colofon

Dit onderzoek is opgezet en uitgevoerd door J. Leerssen, M. Ogwuru, K. Brandes, M. Speet en G. Moenandar (Stichting De Verre Bergen) in samenwerking met A. Hofman (ZZP onderzoeker/ Universiteit van Amsterdam) en K. Lek (CitoLab). In opdracht van Stichting De Verre Bergen hebben CitoLab en Bureau ICE analyses uitgevoerd op de gegevens uit het leerlingvolgsysteem (LVS); en heeft CitoLab, in samenwerking met Stichting BOOR en de Rekenfaculteit, een validatieonderzoek uitgevoerd naar de BSA rekentoets. Het onderzoek kwam verder tot stand dankzij deskundig advies vanuit de leden van de adviescommissie (R. C. W. Feskens (CitoLab), R. Keijzer (Hogeschool IPABO), M. J. S. Brinkhuis (Universiteit Utrecht), E. S. Smit (Universiteit van Amsterdam), K. Brandes (NOLAI)). Wij danken de Rekenfaculteit (S. Hartveld, S. Lodder, M. van Maarseveen, K. Steekelenburg), alle teamleiders en tutoren, en het team van Stiching De Verre Bergen (M. Pires, J. de Reus, M. van Leuven, S. Fikrat) voor hun bijdrage aan dit onderzoek. De Rekenfaculteit heeft geen invloed gehad op de analyses en resultaten van dit onderzoek.

Vormgeving

Evers + de Gier

Contact

Vragen over dit onderzoek? Neem contact op via onderzoek@sdvb.com