

Evaluatie-rapport

Reken op Zuidoost

Een evaluatie van het loopjaar 2019/20 Bridge High Dosage Tutoring op vijf basisscholen in Amsterdam Zuidoost

31 december 2020

Bowen Paulle (Universiteit van Amsterdam)

Joppe de Ree (Erasmus Universiteit Rotterdam, JDR Analytics)



UNIVERSITEIT
VAN AMSTERDAM

Samenvatting

De Onderwijsinspectie (2016; 2017) constateerde een aantal jaren achter elkaar dat de kansengelijkheid in het onderwijs toenam. Een gedeeltelijke verklaring hiervoor is het groeiende ‘schaduwonderwijs’, waarvan vooral kansrijke groepen profiteren (Elffers & Jansen, 2019). Deze bevindingen hebben veel losgemaakt in Nederland en hebben geleid tot meer initiatieven en samenwerking op het gebied van kansen(on)gelijkheid in het onderwijs. Kansengelijkheid is ook een belangrijk speerpunt van het beleid van de huidige wethouder onderwijs in Amsterdam, Marjolein Moorman. Het college van burgemeester en wethouders van de gemeente Amsterdam zet in op het structureel aanpakken van de kansengelijkheid in de stad.

In dit rapport beschrijft het onderzoeksteam van de Universiteit van Amsterdam de resultaten van een gerandomiseerd veldexperiment (een *randomized controlled trial* (of RCT)). RCT's zijn de gouden standaard in evaluatieonderzoek (en bijvoorbeeld medisch onderzoek) omdat door de goede vergelijkbaarheid van twee groepen dit de meest betrouwbare manier is om causale relaties tussen een interventie en effecten aan te tonen. In dit rapport presenteren we de resultaten van het eerste loopjaar (2019/20) Bridge High Dosage Tutoring (Bridge HDT) in Amsterdam Zuidoost. De interventie heeft plaatsgevonden op vijf basisscholen in Zuidoost en in nauwe samenwerking met de deelnemende basisscholen, de Gemeente Amsterdam, de betrokken schoolbesturen en de uitvoerende organisatie, Stichting The Bridge Learning Interventions (The Bridge).

Bij Bridge HDT krijgen leerlingen in groep 7 een schooljaar lang vijf dagen in de week extra rekenondersteuning op maat onder schooltijd in een vast team van één tutor en twee leerlingen. Daarnaast onderhoudt de tutor wekelijks telefonisch contact met ouders/verzorgers en is er nauw contact met het schoolpersoneel. De tutores worden dagelijks begeleid en ondersteund door de *Site Director* (projectleider). Bridge HDT richt zich in eerste instantie op leerlingen die beneden gemiddeld scoren op rekentoetsen en met name op groepen met een verhoogde kans op onderpresteren.

De doelgroep voor de interventie zijn leerlingen die halverwege groep 6 op, of onder het landelijk gemiddelde scoorden op de Cito LVS-toetsen Rekenen/Wiskunde. In totaal hebben in 56 leerlingen deelgenomen aan Bridge HDT. 49 leerlingen uit dezelfde klassen vormden de controlegroep. De effectmeting door het onderzoeksteam richt zich hoofdzakelijk op de rekenprestaties, gemeten aan de hand van de Cito LVS-toetsen Rekenen/Wiskunde. Verder hebben we onderzocht of deelname aan Bridge HDT negatieve effecten heeft op het Begrijpend Lezen. Leerlingen missen door Bridge HDT dagelijks 50 minuten van de reguliere les, waardoor zij voornamelijk talige vakken missen.

We vinden statistisch significante effecten na een half schooljaar Bridge HDT van +0.30 populatiestandaarddeviaties¹ op de Cito LVS-toets Rekenen/Wiskunde halverwege groep 7.

¹ In dit rapport drukken we de effecten van Bridge HDT uit in termen van ‘populatiestandaarddeviaties’. De standaarddeviatie is een maat voor de spreiding van een uitkomst, in dit geval scores op de Cito LVS toetsen

Projecties op basis van een relevante doelgroep laten zien dat +0.30 populatiestandaarddeviaties betekent dat ongeveer 45% van de leerlingen een half niveau hoger scoort op de rekentoetsen en 15% een volledig niveau hoger (bijvoorbeeld van vmbo-gt naar havo-niveau).

Door de landelijke sluiting van de scholen in het tweede deel van het schooljaar 2019/20 is de uitvoering van de interventie Bridge HDT niet volgens plan verlopen. Tijdens en ook deels na de eerste lockdown hebben de tutores van Bridge HDT de contacten en de (online) tutorlessen beperkt kunnen voortzetten. In deze periode heeft men zich gericht op herhaling van de rekenstof. Dit heeft invloed gehad op de resultaten van het onderzoek. We vinden specifiek hieraan gerelateerd dat de effecten richting het eind van het jaar stabiliseren en niet verder doorontwikkelen (zoals verwacht). Wanneer Bridge HDT het hele jaar zonder onderbrekingen had kunnen doorgaan, verwachten we op basis van onze bevindingen en veronderstellingen gebaseerd op literatuur, effecten van rond de +0.40 populatiestandaarddeviaties. Een dergelijk effect betekent dat ongeveer 40% van de leerlingen een half niveau hoger scoort op de rekentoetsen en 40% een volledig niveau hoger. Deze effecten zouden logischerwijs kunnen leiden tot een hoger schooladvies voor de middelbare school.

Op basis van deze projecties laten we zien in hoofdstuk 4.2 dat met effecten van +0.30 en +0.40 populatiestandaarddeviaties een belangrijk deel van de achterstanden in de rekenprestaties tussen leerlingen in Amsterdam Zuidoost en de rest van Amsterdam kunnen worden weggewerkt.

We vinden geen statistisch significante (positieve of negatieve) effecten op de Cito LVS-toets Begrijpend Lezen.

Recent onderzoek schat de effecten van de *learning-loss* door de consequenties van de lockdown in Nederland op -0.08 tot -0.12 standaarddeviaties (Engzell, Freyd & Verhagen, 2020). Dergelijke achterstanden in de rekenprestaties kunnen met Bridge HDT dus gemakkelijk worden weggewerkt. Hierbij merken we op dat de lockdown waarschijnlijk ook de ontwikkeling van allerlei andere vaardigheden remt.

Een tweede loopjaar (2020/21) op basis van dezelfde interventie wordt op dit moment uitgevoerd. In een derde loopjaar verwacht The Bridge - in opdracht van de gemeente - een soortgelijke, maar minder intensieve interventie te implementeren, tegen lagere kosten per leerling per jaar. The Bridge is op dit moment ook al actief bij de implementatie van een model op basis van een 'halve dosering' (2,5 in plaats van 5 dagen per week) in Amsterdam. Dit model wordt geïmplementeerd in het VO in Amsterdam Nieuw-West en in Noord

Rekenen/Wiskunde en Begrijpend Lezen. Met de toevoeging 'populatie' refereren aan het feit dat standaarddeviatie waarin de effecten worden uitgedrukt, betrekking heeft op de gehele populatie aan (Nederlandse) basisschoolleerlingen. +0.30 populatiestandaarddeviaties is dus 30% van een standaarddeviatie in de vaardigheidsscores van een volledige leeftijdsgroep in het basisonderwijs. Het is in principe ook mogelijk om de effecten van interventies in andere grootheden uit te drukken, bijvoorbeeld in termen van controlegroep standaarddeviaties.

Een aantal scholen en schoolbestuurders in Amsterdam Zuidoost hebben aangegeven geïnteresseerd te zijn in dit programma. Ook het tweede en het eventuele derde loopjaar zullen door het onderzoeksteam van de Universiteit van Amsterdam worden geëvalueerd met behulp van RCT's. Bij de evaluatie van het tweede loopjaar Bridge HDT, verwachten we (op basis van de resultaten uit dit onderzoek) richting het einde van het schooljaar, en zonder grote onderbrekingen door COVID-19, effecten van rond de +0.40 populatiestandaarddeviaties te kunnen meten. Met dergelijke effecten kunnen ook substantiële stappen worden gezet richting mogelijk belangrijke beleidsdoelstellingen: het wegwerken van de onderwijsachterstanden tussen Amsterdam Zuidoost en de rest van de stad. Belangrijk hierbij is om de nauwe samenwerking tussen de betrokken partijen voort te zetten, de uitvoering en de resultaten te blijven volgen, en samen na te denken over toekomstige modellen.

Dankwoord

Dit onderzoek is mogelijk gemaakt door financiële steun van de Gemeente van Amsterdam. Verder profiteert het onderzoek van de inzet van de leerkrachten en ander personeel en bestuur van de deelnemende vijf basisscholen in Amsterdam Zuidoost. Ook wil het onderzoeksteam het personeel van The Bridge bedanken voor de inzet en uitvoering van het programma. We bedanken Anna Tammens, Hannah van der Ham en Thabiso Epema voor hun inzet bij het mogelijk maken van het onderzoek.

Op basis van de samenwerking tussen wetenschap, overheid, uitvoeringsinstanties en scholen kan de samenleving lessen trekken over de werkzaamheid van Bridge HDT in deze specifieke context. De lessen leveren opties voor het bestrijden van onderwijsachterstanden. Niet alleen in Amsterdam, maar ook in de rest van Nederland. We danken hiervoor de Gemeente Amsterdam en de deelnemende scholen (waaronder de leerkrachten en de leerlingen) die hiermee bijdragen bij aan het zichtbaar maken van de effecten van HDT-interventies onder een breder publiek.

Inhoudsopgave

Samenvatting	1
Dankwoord	4
Inhoudsopgave	5
1. Achtergrond, aanleiding en probleemstelling	6
2. Interventie Bridge HDT	10
2.1 Het programma	10
2.2 De uitvoering van Bridge HDT (in Amsterdam Zuidoost)	13
3. Het onderzoek	15
3.1 De onderzoeksopzet (design) voor de evaluatie van Bridge HDT-Amsterdam PO	16
4. Resultaten van het onderzoek	19
4.1 Projecties van rekenniveaus na een half en een volledig schooljaar Bridge HDT	21
4.2 Projecties en mogelijk relevante beleidsdoelstellingen	24
4.3 Vervolgonderzoek	25
5. Conclusies en aanbevelingen	27
Literatuurlijst	30
Appendix A: Operationalisering van statistische analyses en validiteitsanalyses	32
Appendix B: Effecten van Bridge HDT op de prestaties in het begrijpend lezen	33
Appendix C: Bevindingen uit de literatuur	34
Appendix D: Operationalisatie toetsadviezen	35

1. Achtergrond, aanleiding en probleemstelling

Amsterdam Zuidoost is een deel van Amsterdam met een bovengemiddeld percentage huishoudens met laag tot zeer lage inkomens. Zoals algemeen bekend blijven de schoolprestaties van leerlingen uit gezinnen met lage inkomens vaak achter bij het gemiddelde. Uitdagingen thuis of in de bredere sociale context kunnen hierbij een rol spelen.

In Amsterdam Zuidoost zien we ook dat de resultaten op de basisschool achterblijven bij scholen en groepen met een vergelijkbaar sociaal-economisch profiel. Onderzoek Informatie en Statistiek (OIS) van de Gemeente Amsterdam laat zien dat scholen met vergelijkbare percentages ‘doelgroepleerlingen’, in Zuidoost minder presteren dan bijvoorbeeld in Amsterdam Nieuw-West of Noord.^{2 3}

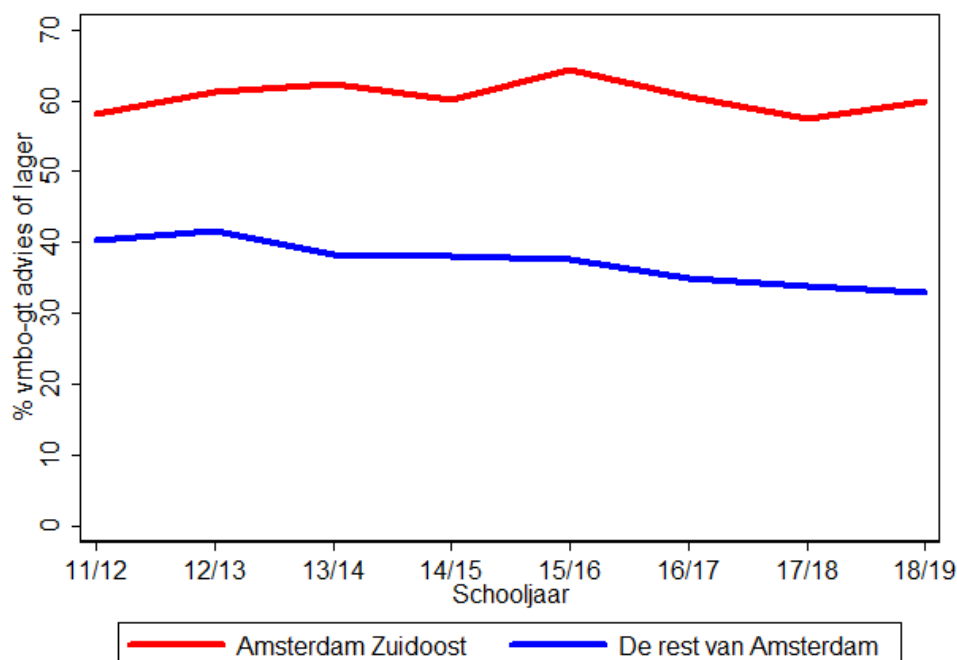
Onze eigen berekeningen laten zien dat de achterstand van Zuidoost ten opzichte van de rest van Amsterdam in de laatste tien jaar is toegenomen, met name door een afname in lagere VO⁴-adviezen in de rest van de stad. Figuur 1 brengt de oplopende achterstanden in beeld. Een kleine tien jaar geleden kregen de leerlingen in Zuidoost anderhalf keer vaker een vmbo-advies (vmbo-gt of lager) dan in de rest van de stad. Deze achterstand is in de afgelopen jaren opgelopen tot ongeveer 1.7 keer vaker.

² OIS en de Gemeente spreken hierbij over ‘doelgroepleerlingen’, leerlingen uit gezinnen die leven van een inkomen op bijstandsniveau, etc. (Achtergronddocument berekening doelgroepleerlingen 2017/’18, Onderzoek, Informatie en Statistiek — Gemeente Amsterdam)

³ OIS rapporteert over de relatieve achterstand in Zuidoost, ten opzichte van leerlingen met een vergelijkbaar profiel in hun inputs voor een klankbordgroep wetenschap en beleid, getiteld *Gelijke kansen in het onderwijs in Zuidoost*.

⁴ VO staat voor voortgezet onderwijs.

Figuur 1: Percentage groep 8-leerlingen met vmbo-gt advies of lager, uitgesplitst naar postcode van de basisschool — in Amsterdam Zuidoost en in de rest van Amsterdam

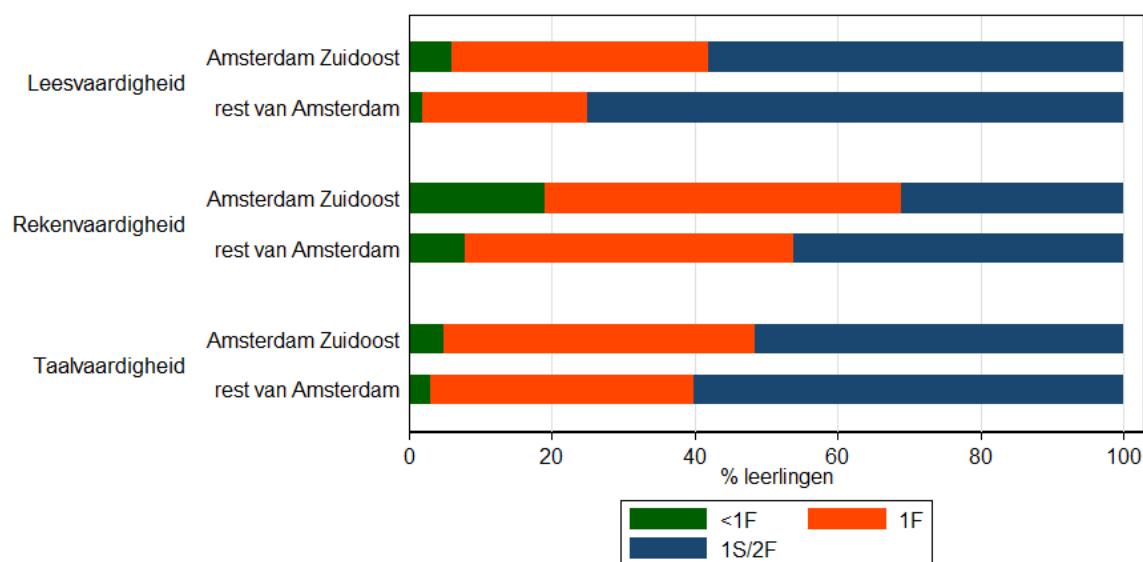


Noot: Figuur 1 is aangemaakt door het onderzoeksteam, op basis van publiek beschikbare gegevens van Dienst Uitvoering Onderwijs op schoolniveau over de gerealiseerde schooladviezen (Bron: https://duo.nl/open_onderwijsdata/databestanden/). Postcodegebied 1109 (Driemond) hebben we voor deze figuur bij 'De rest van Amsterdam' getrokken.⁵

Ook in absolute zin zijn er achterstanden. Afgaande op de referentieniveaus zijn er vooral achterstanden in de rekenvaardigheid (OIS, 2020: 35, figuur 3.2). In Figuur 2 zien we dat tussen de 15 en 20 procent van de leerlingen in Amsterdam Zuidoost aan het einde van de basisschool het eerste fundamentele niveau (1F) niet haalt, ten opzichte van ongeveer 8 procent in de rest van Amsterdam. Het eerste fundamentele niveau (1F) zou aan het einde van de basisschool minimaal moeten worden gehaald om goed te kunnen functioneren in het vervolgonderwijs. De langetermijneffecten voor het individu en de maatschappij van zwakke rekenvaardigheden zijn groot (Gross, et al., 2009). Het is daarom belangrijk om ondersteuningsprogramma's op het gebied van rekenen en wiskunde in te zetten.

⁵ We nemen hiermee een voorschot op de mogelijkheid dat in 2022 Driemond met Weesp een apart stadsdeel zullen vormen.

Figuur 2: Referentieniveaus rekenvaardigheid, taalvaardigheid en leesvaardigheid — in Amsterdam Zuidoost en in de rest van Amsterdam



Noot: Figuur is aangemaakt door het onderzoeksteam, op basis van publiek beschikbare gegevens van Dienst Uitvoering Onderwijs op schoolniveau over de gerealiseerde schooladviezen (https://duo.nl/open_onderwijsdata/databestanden/). Driemond hebben we voor deze figuur bij 'De rest van Amsterdam' getrokken.

In deze context heeft de Gemeente Amsterdam besloten om Bridge HDT in te zetten (op basis van vrijwillige deelname van scholen en leerlingen) in Amsterdam Zuidoost. Eerder onderzoek uit de VS, maar ook uit Nederland, heeft op basis van RCT's laten zien dat HDT-programma's een belangrijke bijdrage kunnen leveren aan het wegwerken van achterstanden (zie bijvoorbeeld www.uva.nl/sepp).

Vanaf het schooljaar 2019/20 wordt Bridge HDT op pilotschaal geïmplementeerd in Amsterdam Zuidoost. Het doel van de pilot is deels om van deze ervaringen te leren op basis van een gerandomiseerd onderzoeksdesign (RCT). Het gebruik van RCT's in de sociale wetenschap heeft onder aanvoering van nobelprijswinnaars Abhijit Banerjee, Esther Duflo en Michael Kremer een sterke ontwikkeling doorgemaakt. In eerste instantie in de ontwikkelingseconomie. De trend heeft zich vanuit de ontwikkelingseconomie doorgezet naar de VS, en nu ook naar Europa. RCT's voor het evalueren van onderwijsinterventies in Nederland zijn schaars. Het CPB doet bijvoorbeeld in de publicatie *Kansrijk Onderwijsbeleid*, niet of nauwelijks een beroep op Nederlands onderzoek. In de VS worden HDT-programma's van Saga Education rigoureus geëvalueerd en vervolgens goed gedocumenteerd. De huidige aandacht voor dergelijke HDT-programma's in de rest van de wereld (inclusief Nederland) kan deels worden toegeschreven aan het commitment van Amerikaanse onderzoekers, NGO's en beleidsmakers aan het opzetten en uitvoeren van onderzoek met hoge kwaliteitseisen (met behulp van RCT's) en het systematisch

documenteren van de resultaten.⁶ HDT-programma's leveren regelmatig positieve effecten op de leerprestaties en zijn dus te beschouwen als programma's die kritisch wetenschappelijk onderzoek kunnen weerstaan (zie bijvoorbeeld Nickow, Oreopoulos en Quan [2020]).

Het commitment van The Bridge en de Gemeente Amsterdam aan het ondersteunen en opbouwen van wetenschappelijke kennis op basis van RCT's heeft dus een voortrekkersrol en een voorbeeldfunctie in Nederland. Zo speelt deze evaluatie een rol bij het ontwikkelen van beleid rondom effectieve en schaalbare onderwijsinterventies voor relatief kansarme leerlingenpopulaties, bijvoorbeeld in voorbereiding op het Masterplan Zuidoost.

⁶ Zie bijvoorbeeld <https://www.povertyactionlab.org/case-study/individualized-tutoring-improve-learning>

2. Interventie Bridge HDT

2.1 Het programma

HDT-programma's zijn in de VS veelvuldig geëvalueerd op basis van gerandomiseerd veldonderzoek. De resultaten laten eenduidig zien dat HDT-interventies een belangrijke bijdrage kunnen leveren aan de ontwikkeling van leerlingen (zie bijvoorbeeld Cook et al., 2014; Cook et al., 2015; Nickow, Oreopoulos en Quan, 2020). Een aantal belangrijke onderzoekers in dit veld zijn duidelijk over hun enthousiasme voor HDT-programma's:

“[HDT can] allow students who have fallen behind grade level to re-engage with regular classroom instruction, likely improving their chances of graduating high school and achieving the many long-term economic benefits that go along with academic success.”
(Ander, Guryan & Ludwig, 2016: 8)

Een belangrijk aspect van HDT-programma's is de mogelijke schaalbaarheid. Naast de resultaten uit kleinschalig en grootschalig RCT's naar Saga HDT in de VS (Cook et al., 2014; Cook et al., 2015), en naast een apart indicatie van de schaalbaarheid van Saga HDT (Davis et al., 2017), zijn er nu een aantal kleinschalig RCT's uitgevoerd in Nederland (zie hiervoor www.uva.nl/sepp).

Een versie van deze aanpak is vanaf 2017 doorontwikkeld en aangepast aan de Nederlandse context door The Bridge.⁷ Bridge HDT is ontwikkeld op basis van het HDT-model van Saga Education uit de Verenigde Staten (VS).⁸ Eerdere projecten van The Bridge hebben veelbelovende resultaten opgeleverd in het VO en PO op het gebied van rekenen wiskunde, o.a. in Nieuw-West (zie hiervoor www.uva.nl/sepp). Het project in Amsterdam Zuidoost is 'Bridge HDT-Amsterdam PO'⁹ genoemd.

Bridge HDT richt zich in eerste instantie op maatschappelijk kwetsbare leerlingen die beneden gemiddeld presteren. Een belangrijk onderdeel van Bridge HDT is de intensiteit, oftewel de *high dosage*. Leerlingen krijgen in het basisprogramma een schooljaar lang, elke dag een uur onder schooltijd rekenbegeleiding op maat in een team van drie personen. De

⁷ Voor meer achtergrond, zie de website van The Bridge: <http://www.tbli.nl>

⁸ Net als The Bridge in Nederland is Saga Education in de VS bezig om goedkopere modellen van HDT te ontwikkelen en laten testen. Echter, zelfs het oorspronkelijke en duurste model van Saga Education HDT is volgens onderzoekers van o.a. University of Chicago uiteindelijk kosteneffectief. In hun eigen woorden (Ander, Guryan & Ludwig, 2016: 6), gebaseerd op hun RCT resultaten in Chicago: *“the authors calculate that the tutorials would be expected to increase participants' adult earnings by between \$700 and \$1,050 each year. Calculating the present discounted value of these gains back to age 15, the authors estimate that the benefits of the tutorial would be roughly 5 to 11 times larger than the costs—suggesting that investment in this tutorial program is economically worthwhile. As an alternative measure, the authors calculate benefit–cost ratios under the extreme assumption that it would be necessary to deliver four years of tutorials to a student to maintain the test score impact in the Chicago study. Even in this extreme case, the authors calculate that the benefits would be 1.3 to 2.9 times as large as the costs. These estimates suggest that this type of tutorial program is a cost-effective way to improve learning and economic outcomes for disadvantaged youth.”*

⁹ PO staat voor primair onderwijs.

teams bestaan uit twee leerlingen en een professionele tutor. De tutores worden dagelijks begeleid en ondersteund door de *Site Director* (projectleider). De teams blijven onder normale omstandigheden bij elkaar gedurende de looptijd van het programma. In deze stabiele omgeving wordt achterstanden weggewerkt, maar wordt ook zelfvertrouwen gekweekt door samen successen te beleven en te vieren. Beter leren rekenen is dus een doel op zich, maar ook een middel voor het ontwikkelen van zelfvertrouwen en een *growth mindset*. Deze vaardigheden en/of eigenschappen kunnen de leerlingen ook in het vervolgetraject helpen, bijvoorbeeld op de middelbare school. Hiernaast onderhoudt de tutor wekelijks telefonisch contact met ouders/verzorgers en is er nauw contact met het schoolpersoneel.

The Bridge draagt de zorg voor training en werving van de tutores en de *Site Directors*, financiële zaken, de dagelijkse gang van zaken, programmaontwikkeling, contact met het onderzoeksteam en kwaliteitswaarborging. Het team van The Bridge dat in Amsterdam Zuidoost actief was bestaat uit zes fulltime tutores, één fulltime *Site Director* (projectleider en eerste aanspreekpunt van de school) en één invaltutor/projectassistent. De teamleden zijn in vaste dienst bij The Bridge en ze krijgen vooraf een intensieve opleiding en worden ook tijdens het werk gecoacht. De tutores geven les in een klaslokaal in de school en werken nauw samen. De *Site Director* wordt gecoacht door het *central office* en de tutores vervolgens door de *Site Director*. The Bridge tracht een hoge kwaliteit van implementatie te waarborgen door continu te werken aan een gezonde en open feedbackcultuur.

Tutores hoeven geen achtergrond in het onderwijs te hebben, maar moeten vooral leergierig zijn omdat ze intensief vooraf en tijdens het lesgeven worden getraind. De hoogte van de salarissen is beperkt voor hun fulltime inzet. Bij het aantrekken van tutores wordt deels aanspraak gemaakt op het maatschappelijke karakter van het werk. Het tutorschap wordt door tutores vaak gezien als een soort ‘maatschappelijke diensttijd’. Echter, (jonge) tutores zien het soms ook als een eerste stap richting een carrière in het onderwijs, als leerkracht bijvoorbeeld. Om elk loopjaar genoeg en kwalitatief goede tutores aan te trekken, zijn er sinds het begin goede arbeidsvoorwaarden opgesteld. Ook moeten ze slagen voor een rekentoets op 2F niveau en goed worden beoordeeld op basis van een proefles met leerlingen.

Het rekencurriculum van Bridge HDT kent vijf domeinen die zijn opgesteld door The Bridge aan de hand van het referentieniveau 1F uit het referentiekader rekenen. Het referentieniveau 1F omvat basisvaardigheden die in principe alle leerlingen minimaal aan het eind van de basisschool zouden moeten beheersen.¹⁰ De vijf rekendomeinen zijn getallen, verhoudingen, tijd en kalender, meten/meetkunde en verbanden. De inhoud van het Bridge HDT-programma is door The Bridge zelf ontwikkeld en samengesteld. Specifieke rekenstrategieën worden echter altijd met de school afgestemd, zodat er geen verwarring ontstaat over bepaalde (oplossings)routines.

Bridge HDT levert maatwerk op het niveau van de leerling. De tutorlessen beginnen aan het begin van het schooljaar met het onderwerp getallen (breuken, etc.) Tutores werken

¹⁰ Aan het begin van groep 7 – de start van de Bridge HDT-interventie in Amsterdam Zuidoost – haalt nog slechts een minderheid van de leerlingen in het basisonderwijs in Nederland het 1F-niveau.

vervolgens in het eigen tempo van de leerlingen door het curriculum. Tutoren stimuleren leerlingen om zoveel mogelijk domeinen volledig te beheersen, in plaats van alle domeinen maar half (volgens het principe van ‘*mastery learning*’). The Bridge gelooft dat het voor de ontwikkeling en het zelfvertrouwen van leerlingen belangrijk is om basisbewerkingen volledig onder de knie te hebben, voordat leerlingen aan nieuwe onderwerpen beginnen. De leerkracht in de klas heeft niet altijd voldoende tijd beschikbaar om specifiek en langdurig op individueel niveau aan bepaalde achterstanden te werken. Het tutorprogramma biedt dus ondersteuning aan de school en leerkracht in de klas. Bridge HDT dient dus verlengstuk van de leerkracht, niet als vervanging van.¹¹ Tegelijkertijd ondersteunen de vaste groepsleerkrachten en intern begeleiders ook The Bridge bij de uitvoering door tips uit te wisselen, informatie over leerlingen te verschaffen en bij het vormen van de leerlingduo’s.

Afhankelijk van het tempo van de individuele leerling zullen niet alle leerlingen het volledige curriculum doorlopen. Wel doorlopen zij het curriculum in de vaste (hierboven genoemde) volgorde. Hoewel het rekencurriculum verhaalsommen bevat, wordt nog extra 20% van de lestijd (één vaste dag per week) besteed aan het oefenen met verhaalsommen. Hier worden speciale stappenplannen en technieken voor gebruikt.

Inherent aan de HDT-methode en het frequente contact pogen de tutoren ook de sociaal-emotionele ontwikkeling van leerlingen te bevorderen. Dit doen zij tijdens de reguliere rekenaartutoring. Zo bieden de tutoren persoonlijke begeleiding vanuit de theorie van de *growth mindset*. Een uitgangspunt is dat The Bridge en dus de tutoren vertrouwen hebben in de ontwikkelingsmogelijkheden van alle leerlingen, ongeacht het aanvangsniveau. Het rekenen wordt als middel gebruikt om leerlingen met de *growth mindset* in aanraking te laten komen. Leerlingen die beneden gemiddeld presteren hebben regelmatig het idee dat ze het ‘niet kunnen’. Door op eigen tempo en met vertrouwen door de stof te gaan is het de bedoeling dat de leerlingen ervaren dat ze ‘het toch wel kunnen’. Dergelijke ervaringen zijn kostbaar, met name voor leerlingen die op school moeilijk mee kunnen komen. Uiteindelijk is de bedoeling dat deze *growth mindset* niet alleen vanuit een externe partij (de tutor) geactiveerd wordt, maar dat de leerling de succeservaringen internaliseert en daarmee in de toekomst ook meer uitdagingen (op school of daarbuiten) aan durft te gaan.¹²

Tutoren bereiden elke dag de lessen per leerling/leerlingduo voor en leveren maatwerk. De tutoren monitoren in detail het niveau en de ontwikkeling van ‘hun’ leerlingen en kunnen op deze manier maatwerk leveren. Het voorbereiden van de lessen gebeurt na de lestijd, tijdens de gezamenlijke en individuele lesvoorbereidingstijd. De *Site Director* heeft hierin een

¹¹ The Bridge is tevens samen met de Gemeente en scholen bezig om een aparte leerkrachtentraining te ontwikkelen waarbij bovenbouwleerkrachten in een aantal sessies meer leren over de reken- en motivatietechnieken bij Bridge HDT. Een pilot zal uitgevoerd worden in schooljaar 2020/2021.

¹² Er is ook expliciet aandacht voor de algemene en sociaal-emotionele ontwikkeling van de leerlingen door middel van de ‘Check-in’-lessen die 4 tot 6 keer per jaar worden georganiseerd tijdens de tutorlessen. De ‘Check-in’-lessen vinden plaats tijdens het reguliere Bridge HDT-programma en dragen bij aan een positieve en productieve sfeer en zijn gebaseerd op het curriculum ‘Kompas’. Er komen thema’s aan bod zoals de *growth mindset*, samenwerken, omgaan met emoties en vertrouwen en de toekomst.

ondersteunende rol. Ook is er een rekenspecialist in dienst als project-assistent en is er een curriculumspecialist in het *central office*.

Bridge HDT blijft niet binnen de school. De tutores onderhouden wekelijks telefonisch contact met ouders over de voortgang van hun kind. Ze zien de ouders hierbij als sparringpartners, waarin samenwerking en gelijkwaardigheid centraal staan. Het lukt The Bridge vaak om meer dan 80% van de ouders/verzorgers te betrekken bij het programma.¹³ De tutor komt naast met kritische feedback, ook vaak met positief nieuws over de ontwikkelingen van hun kind. Zo ontstaat een sfeer van positieve samenwerking en betrokkenheid.¹⁴

Een laatste onderscheidend aspect van Bridge HDT is het commitment aan het (stap voor stap) ontwikkelen van een *evidence base*. Op basis van RCT's, zoals deze in Amsterdam Zuidoost, of op basis van andere onderzoekstechnieken. Het doel van systematisch onderzoek is tweeledig. Niet- of onvoldoende effectieve interventies kunnen snel worden afgeschaald. Bij interventies die succesvol zijn levert het onderzoek handvatten voor beleidsmakers en andere financiers om soortgelijke interventies ook elders in te zetten. Een belangrijk aspect hierbij is dat kennis niet verloren gaat.

2.2 De uitvoering van Bridge HDT (in Amsterdam Zuidoost)

Het eerste loopjaar van Bridge HDT in Amsterdam Zuidoost is gestart in het schooljaar 2019/20. Het plan is om deze - of een aangepaste versie van Bridge HDT - minimaal drie loopjaren te draaien, met na elk jaar een evaluatiemoment. In elk loopjaar wordt het onderzoek verder uitgebreid, zodat na afloop het gehele project kan worden geëvalueerd. Zowel de implementatie door The Bridge als het onderzoek onder leiding van de UvA wordt gefinancierd door de gemeente Amsterdam.

De HDT-projecten van The Bridge worden in principe altijd onderzocht op werkzaamheid. Tot nu toe worden de onderzoeken uitgevoerd door een wetenschappelijk onderzoeksteam van de Universiteit van Amsterdam.¹⁵ Het onderzoek valt niet onder de verantwoordelijkheid van The Bridge en financiële middelen voor uitvoering en onderzoek zijn gescheiden. De verantwoordelijkheid voor de inhoud van Bridge HDT, de continue training en werving van de tutores ligt bij The Bridge.

¹³ Als ouders niet bereikt kunnen worden of bijvoorbeeld een te grote taalachterstand hebben dan zoekt The Bridge naar een ander vast contactpersoon binnen of buiten de familie.

¹⁴ De ervaring van eerdere projecten leert dat ouders het waarderen om op de hoogte te worden gehouden. Ook vinden ouders het fijn dat het contact wederkerig is. De tutores vragen de ouders ook hoe het thuis gaat en kunnen hen om advies vragen over hoe zij hun kind beter kunnen ondersteunen. Naast het wekelijkse contact worden er kennismakingsgesprekken, inlooplessen en ouderbijeenkomsten georganiseerd en komen de tutores bij een aantal leerlingen op huisbezoek, als de ouders dat op prijs stellen. De tutor kan tevens als schakel dienen tussen ouders en de school (Kielman, Paulle, & Van Londen, 2017).

¹⁵ Door de onderzoeksdata in anonieme vorm beschikbaar te stellen voor controle en verder onderzoek.

Saga Education, de organisatie die vergelijkbare HDT-programma's implementeert in de VS, is een vaste consultant van The Bridge. Saga Education adviseert via online gesprekken en legt minimaal één keer per loopjaar een bezoek af waaraan een uitgebreid feedbackrapport en rapportcijfer gekoppeld zijn. Op 18 november 2019 heeft Chris Dupuis van Saga Education een bezoek gebracht aan het project in Zuidoost om de kwaliteit te beoordelen en feedback te geven. Dupuis heeft voor de implementatie in het geheel een 8/10 gegeven. Hij was ook zeer tevreden over de kwaliteit van de tutoren en de samenwerking met de *Site Director*. Een positieve tussenevaluatie van Saga is geen garantie voor verbeteringen in de rekenprestaties, maar het is wel een indicatie dat Bridge HDT in Zuidoost goed op de rails staat.

Saga Education speelt ook een rol bij het adviseren over nieuwe HDT-modellen die in de VS zijn uitgetest en geëvalueerd op basis van gerandomiseerde onderzoeksdesigns. Saga Education loopt een aantal jaren voor en heeft al meerdere HDT-modellen kunnen implementeren en (laten) evalueren.

3. Het onderzoek

De gemeente Amsterdam heeft als financier van Bridge HDT in Amsterdam Zuidoost, de Universiteit van Amsterdam de opdracht gegeven het Bridge HDT-programma te evalueren. Het onderzoeksteam staat onder leiding van Bowen Paulle (Universiteit van Amsterdam). Binnen het team heeft Joppe de Ree (Erasmus Universiteit Rotterdam en JDR Analytics) de leiding over het kwantitatieve gedeelte van het onderzoek. Het onderzoek is geregistreerd in het *social science registry* voor *randomized controlled trials* van de American Economic Association (<https://www.socialscienceregistry.org/trials/4643>).

Het Bridge HDT-programma in Zuidoost wordt - zoals eerder aangegeven - geëvalueerd op basis van een *randomized controlled trial* (RCT's). RCT's zijn de gouden standaard in evaluatieonderzoek. Bij RCT's worden deelnemers willekeurig, oftewel *random*, getrokken uit de doelpopulatie. De consequentie hiervan is dat de geselecteerde groep niet systematisch verschilt van de niet-geselecteerde groep. Systematische verschillen tussen de geselecteerde groep (de treatmentgroep) en de niet-geselecteerde groep (de controlegroep) die we terugzien in de data na afloop van de interventie, kunnen dus redelijkerwijs worden toegeschreven aan de interventie.

Op basis van de hieronder beschreven onderzoeksopzet, verwachtten we een *minimum detectable effect* (MDE) van +0.30 populatiestandaarddeviaties na een geheel schooljaar Bridge HDT (zonder disrupties).¹⁶ Eerder onderzoek naar een vergelijkbare (maar minder intensieve) interventie in het basisonderwijs, heeft effecten opgeleverd van rond de 0.28 populatiestandaarddeviaties na een volledig schooljaar (De Ree en co-auteurs, nog niet gepubliceerd).¹⁷ We verwachtten grotere effecten in Zuidoost onder andere door de grotere intensiteit van het programma.

Bij de registratie van de onderzoeksopzet begin februari 2020 hebben we uiteraard geen rekening gehouden met de (toen nog onbekende) invloed van de COVID-19 pandemie. De lockdown en de sluiting van de scholen in de tweede helft van het schooljaar 2019/20 heeft naar verwachting grote invloed op de bevindingen. The Bridge heeft naar mogelijkheden geprobeerd een deel van het programma online door te zetten, samen met de scholen en in overleg met de Gemeente. Er was echter tegelijkertijd de noodzaak om de scholen niet teveel in de weg te zitten bij het vormgeven van hun eigen methodes voor onderwijs op afstand. Binnen de mogelijkheden heeft The Bridge een '*Keep In Touch Plan*' opgesteld en uitgevoerd. Hierbij heeft The Bridge geprofiteerd van de contacten met ouders en de nauwe samenwerking met de vijf basisscholen.¹⁸ De interventie heeft dus in het tweede deel van het jaar duidelijk minder intensief en effectief kunnen draaien. Dit heeft waarschijnlijk een

¹⁶ Dit wil zeggen dat we (met een $\alpha=0.05$) in 80% van de gevallen een statistisch significant resultaat verwachten, wanneer het werkelijke effect 0.30 populatiestandaarddeviaties is.

¹⁷ Dit HDT-programma is niet geïmplementeerd door The Bridge.

¹⁸ Zie voor meer informatie hierover de inhoudelijke verantwoording van loopjaar 1 door The Bridge.

(negatieve) invloed gehad op de gemeten effecten. We komen hierop terug in de bespreking van de resultaten in het volgende hoofdstuk.

3.1 De onderzoeksopzet (design) voor de evaluatie van Bridge HDT-Amsterdam PO

Vijf basisscholen in Amsterdam Zuidoost zijn - in samenwerking met de Gemeente - voorgeselecteerd op basis van voldoende aantallen doelgroepleerlingen voor het Bridge HDT-programma. De doelgroep van het Bridge HDT-programma in Amsterdam Zuidoost zijn leerlingen die halverwege groep 6 rekenen op de vaardigheidsniveaus III, IV of V. Dit zijn leerlingen die op of onder het landelijk gemiddelde scoren op de Cito LVS-toetsen Rekenen/Wiskunde. Het gaat hierbij dus niet alleen om leerlingen met een (grote) leerachterstand. Echter, leerachterstanden zijn niet gemakkelijk meetbaar. Het kan zijn dat leerlingen met een III-score, d.w.z. een score rond het landelijk gemiddelde, onderpresteren ten opzichte van hun eigen mogelijkheden. Eerder onderzoek heeft laten zien dat ook leerlingen die bovengemiddeld scoren baat kunnen hebben bij HDT-interventies (*intern document met onderzoeksanalyses, nog niet openbaar*). Echter, het is aannemelijk dat onderpresterders zich vooral concentreren bij de lagere niveaus.

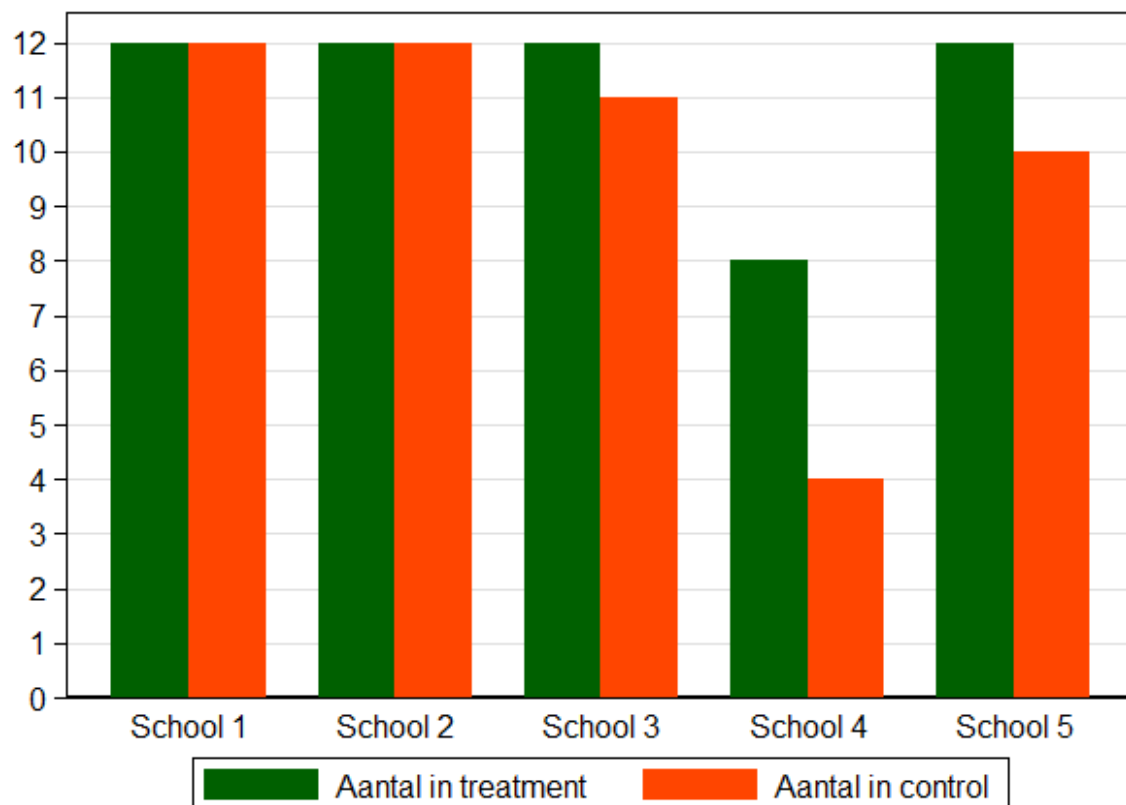
De procedure voor de randomisatie is als volgt opgezet en uitgevoerd.¹⁹ The Bridge heeft van de vijf scholen een overzicht ontvangen met namen en toetsscores van leerlingen die aan de bovengenoemde voorwaarden voldoen. Op basis van deze lijsten zijn de ouders/verzorgers benaderd middels een toestemmingsformulier. Uiteindelijk is voor 105 doelgroepleerlingen toestemming gegeven.²⁰ De data-uitwisseling tussen de scholen, The Bridge en de UvA is op een veilige en ‘AVG-proof’ manier uitgevoerd.

Specifiek is er een *block*-randomisatie procedure uitgevoerd, waarvoor eerst de 105 leerlingen zijn verdeeld in kleinere groepjes, op basis van informatie over de school, de klas, en de score op de Cito LVS-toetsen Rekenen/Wiskunde die halverwege groep 6 zijn afgenomen. Binnen deze ‘*blocks*’, of strata, zijn met behulp van het softwarepakket Stata 56 leerlingen geselecteerd voor de treatmentgroep. De 49 niet-geselecteerde leerlingen vormen automatisch de controlegroep. *Block*-randomisatie zoals hier uitgevoerd heeft het voordeel dat voorafgaand aan de randomisatie meer balans (gelijkheid) wordt geforceerd tussen de treatment- en controlegroep in termen van rekenscores, scholen en klassen binnen de scholen. *Block*-randomisatie procedures leveren over het algemeen preciezere schattingen van effecten van interventies.

¹⁹ Hiervoor zijn vooraf AVG-proof data-overeenkomsten ondertekend.

²⁰ Het gaat hier om een duidelijke meerderheid van het totaal aan doelgroepleerlingen.

Figuur 3. Aantallen leerlingen in treatment- en controlegroep, per school



Figuur 3 laat zien hoe de aantallen leerlingen in de treatment- en controlegroepen zijn verdeeld over de deelnemende scholen. Omdat er zes tutores beschikbaar zijn, bestaat de tutorgroep uit maximaal twaalf leerlingen per school. Om deze reden hebben we er tijdens de randomisatie voor gezorgd dat er per tutorgroep twaalf leerlingen worden geselecteerd, mits er voldoende leerlingen beschikbaar zijn. Voor een van de deelnemende scholen hebben we voor een andere verdeling gekozen door de lagere aantallen doelgroepleerlingen. Het feit dat het aantal leerlingen in de treatment groep groter is dan het aantal leerlingen in de controlegroep heeft geen invloed op de validiteit van de onderzoeksopzet.

Scores op de Cito LVS-toetsen Rekenen/Wiskunde zijn meestal sterk gecorreleerd over tijd. De scores halverwege groep 6 zijn sterk voorspellend voor de scores aan het eind van groep 7, bijvoorbeeld. Kennis over deze correlaties zijn meegenomen bij power-berekeningen. Details over powerberekeningen en, daaraan gerelateerde, ‘*minimum detectable effects*’ of MDE’s, zijn in detail beschreven in het *social science registry* van de American Economic Association.²¹

²¹ <https://www.socialscienceregistry.org/trials/4643>

We meten de effecten van Bridge HDT aan de hand van scores op de Cito LVS-toetsen Rekenen/Wiskunde en Begrijpend lezen.²² We onderzoeken de effecten na een half schooljaar Bridge HDT en een volledige schooljaar na de start van het programma. We ambiëren om ook de scores op de Cito LVS-toetsen halverwege groep 8 te analyseren, anderhalf jaar na de start van de interventie. De toetsen voor dit laatste meetmoment worden begin 2021 afgenomen en zijn dus nu nog niet beschikbaar. Onder normale omstandigheden zullen we deze resultaten meenemen in vervolgrapportages.

Zoals gezegd, maken we bij de effectmetingen gebruik van scores op de Cito LVS-toetsen. De vaardigheidsscores worden eerst gestandaardiseerd ten opzichte van het landelijk gemiddelde en de landelijke spreiding van de betreffende leeftijdsgroep. Vervolgens worden deze gestandaardiseerde vaardigheidsscores gebruikt als afhankelijke variabelen in de regressieanalyses. Als onafhankelijke variabelen gebruiken we een indicator variabele²³ voor de treatment groep en een volledige set aan indicator variabelen voor de verschillende *sampling blocks* of strata. In Appendix A lichten we een aantal details van de analyses verder toe. We hebben ervoor gekozen om hierbij het gebruik van formules te vermijden.

²² Er zijn ook voor implementatie en onderzoek tijdens de *'baseline'* meting een extra sociaal-emotionele vragenlijst (SDQ) en een online taallose rekentoets (TOA) afgenomen aan het begin van groep 7, maar omdat de endline niet kon plaatsvinden in juni zijn hier geen analyses van gemaakt.

²³ Indicator variabelen worden ook wel dummy variabelen of 0/1 variabelen genoemd. Deze variabelen hebben twee mogelijke uitkomsten, een 1 (als aan de voorwaarde wordt voldaan) en een 0 (als niet aan de voorwaarde wordt voldaan).

4. Resultaten van het onderzoek

Tabel 1 kolom (3) en (4) rapporteren de effecten van Bridge HDT (zoals uitgevoerd) op de rekenprestaties. De resultaten in kolom (3) zijn de resultaten na een half schooljaar Bridge HDT, halverwege groep 7. De resultaten in kolom (4) zijn de resultaten na een volledig schooljaar Bridge HDT, inclusief onderbrekingen gerelateerd aan de COVID-19 pandemie. De kolommen (1) en (2) zijn verschillen die zijn geobserveerd voorafgaande aan de start van Bridge HDT. De resultaten in kolom (1) en (2) zijn *baseline* resultaten en dus geen treatmenteffecten. Tabel B1 (in Appendix B) rapporteert de effecten van Bridge HDT (zoals uitgevoerd) op de prestaties in het Begrijpend Lezen.

Tabel 1. Effecten van Bridge HDT op de rekenprestaties op basis van LVS-toetsen Cito R/W (in termen van populatiestandaarddeviaties.)

Kolomnummer	(1)	(2)	(3)	(4)
Meetmoment	Baseline M6	Baseline E6	Midline M7	Endline E7
Baseline resultaten/Treatment effect	-0.02	-0.07	0.29**	0.29
Standaardfout	(0.09)	(0.13)	(0.12)	(0.17)
p-waarde	[0.82]	[0.60]	[0.02]	[0.11]
Aantal observaties	<99>	<100>	<98>	<94>
Noten: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. De effecten van de interventie zijn geschat op basis van een regressiemodel, waarbij de gestandaardiseerde toetscores wordt geregresseerd op stratum/ <i>block effects</i> (dummy/indicator variabelen die de <i>block</i> of stratum aangeven met een 1, of anders 0) en een indicator voor de treatment groep. De geschatte parameter op de indicator voor de treatment groep is gerapporteerd in deze tabel, als schatting voor het treatment effect.				

De *baseline* resultaten van het onderzoek in kolom (1) en (2) laten zien dat er voorafgaand aan de start van Bridge HDT weinig verschil zit tussen de treatment- en de controlegroep. Dit resultaat is een indicatie dat de randomisatie correct is uitgevoerd en dat de niet-geobserveerde toetscores voor een kleine groep deelnemers geen bedreiging is voor de validiteit van het onderzoek. (Appendix A bespreken we nog een extra toets op de validiteit van het onderzoek.)

Kolom (3) beschrijft resultaten van de eerste effectmeting halverwege groep 7, een half schooljaar na de start van de interventie. De leerlingen in de treatment groep scoren gemiddeld rond de +0.3 (afgerond tot 1 decimaal) standaarddeviaties hoger dan de controlegroep. Dit effect is significant op het 5 procent niveau. In kolom (4) zien we dat het

effect stabiliseert richting het einde van het schooljaar.²⁴ De resultaten in kolom (4) meten de effecten van een volledig schooljaar Bridge HDT, maar inclusief onderbrekingen door o.a. de lockdown en andere gerelateerde uitdagingen.

Het feit dat de effecten van Bridge HDT in het tweede deel van groep 7 niet verder worden uitgebouwd is waarschijnlijk toe te schrijven aan de gevolgen van de COVID-19 pandemie. Scholen zijn in maart en april 2020 twee maanden volledig gesloten geweest. Daarna zijn ze tot aan het eind van het schooljaar geleidelijk weer open gegaan. Een korte periode werden externe partijen — waaronder The Bridge — ook na de opening van de scholen og niet toegelaten in de scholen om zo te pogen de besmettingsgraad zo laag mogelijk te houden. In beide periodes heeft The Bridge enige afstand moeten nemen om de scholen de ruimte te geven de nieuwe situatie vorm te geven. Tegelijkertijd heeft The Bridge geprobeerd om zoveel mogelijk ondersteuning aan leerlingen te geven. Daarbij stond herhaling van de rekenstof centraal zodat de opgedane kennis behouden zou blijven.²⁵ Een andere uitdaging was dat veel deelnemers aan het begin van de lockdown niet over internet of computers konden beschikken, wat het contact tussen tutor en leerling heeft bemoeilijkt. De resultaten laten zien dat leerlingen hun eerder geboekte vooruitgang wel vast kunnen houden, maar niet verder hebben kunnen uitbouwen. Deze bevinding is in lijn met bevindingen uit de literatuur (Cascio en Staiger, 2012).²⁶

Omdat het tweede deel van de interventie niet volgens plan is verlopen, kunnen we geen duidelijk inzicht krijgen in de effecten van een volledig loopjaar Bridge HDT in deze specifieke context. Op basis van de resultaten van een half jaar Bridge HDT en conclusies uit de relevante literatuur, verwachten we van een jaar lang ononderbroken Bridge HDT een effect van +0.40 populatiestandaarddeviaties.²⁷ Een dergelijk effect is ook in dezelfde orde van grootte als eerder gevonden effecten van een jaar lang Bridge HDT in Haarlem Oost en Schalkwijk.²⁸ Zonder langdurige lockdowns in het schooljaar 2020/21 hopen we op basis van het tweede loopjaar Bridge HDT in Amsterdam Zuidoost meer te leren over de effecten van een volledig schooljaar Bridge HDT in deze context.

In relevante literatuur worden effecten groter dan 0.20 standaarddeviaties als relatief ‘groot’ gekwalificeerd (Kraft, 2020). De +0.30 en +0.40 standaarddeviaties vallen dus ruim binnen

²⁴ Dit effect is net niet significant op het 10 procent niveau, met een p-waarde van 0.11.

²⁵ Zie de inhoudelijke verantwoording van The Bridge naar de Gemeente voor meer details.

²⁶ Leerlingen die meedoen met Bridge HDT maken vorderingen die ze onder normale omstandigheden niet maken. Met andere woorden, ze presteren op een niveau van leerlingen die normaal gesproken thuis beter worden ondersteund, of die meer in hun mars hebben. Als de ondersteuning (deels) wegvalt, lukt het de leerlingen vaak niet de verworven voorsprong vast te houden. Ze moeten immers op het niveau blijven van de groep die normaal gesproken al beter presteert, waarvan de ondersteuning niet wegvalt. De resultaten in Tabel 1 suggereren dat de ondersteuning van Bridge HDT in de tweede helft van het schooljaar net voldoende was om de voorsprong op pijl te houden.

²⁷ Op basis van literatuuronderzoek rapporteren Andrabi et al. (2011) persistentiefactors van tussen de 0.2 en 0.5. Als we werken met een persistentiefactor van 0.35 en aannames maken over de ontwikkeling van het treatment effect lijkt ons een $0.30+0.35*0.3 = 0.40$ een redelijke schatting, en misschien enigszins conservatief.

²⁸ Bridge HDT in Haarlem Oost en Schalkwijk is geëvalueerd op basis van niet-experimentele methodes. Op basis van deze analyses vonden we effecten van +0.40 standaarddeviaties.

deze categorie.²⁹ Met deze effecten komt Bridge HDT ook duidelijk in de buurt van de 0.37 standaarddeviaties die door Nickow, Oreopoulos en Quan (2020) als gemiddeld effect wordt gerapporteerd in een metastudie naar tutoring interventies. Ook ten opzichte van andere interventies in het onderwijs lijken tutoring-interventies het over het algemeen zeer goed te doen. In een metastudie voornamelijk gebaseerd op RCT's vinden Dietrichson en anderen (2017: 268) de grootste effecten bij tutoring interventies (zie Appendix C voor details over de bevindingen van Dietrichson et al. [2017]).

4.1 Projecties van rekenniveaus na een half en een volledig schooljaar Bridge HDT

Voor de leesbaarheid en toegankelijkheid van dit rapport is het van belang om de effecten ook anders te duiden. In dit hoofdstuk laten we zien welke invloed een half en een heel schooljaar Bridge HDT hebben op de rekenniveaus van de doelgroep.

We nemen hiervoor de de rekenniveaus van 45 leerlingen uit onze controlegroep als uitgangspunt. Deze groep leerlingen heeft (als controlegroep) niet meegegaan met Bridge HDT. Voor deze groep zetten we de scores op de Cito LVS toetsen Rekenen/Wiskunde eerst om in 'toetsadviezen'. Het concept 'toetsadvies' geeft context aan de rekenniveaus van de leerlingen in onze steekproef. Leerlingen met een 'toetsadvies' praktijkonderwijs bijvoorbeeld, rekenen op een niveau dat bij een normale ontwikkeling leidt tot een schooladvies voor praktijkonderwijs. (In Appendix D beschrijven we hoe we het concept 'toetsadvies' voor dit rapport operationaliseren.)

In Figuur 3A en 3B beschrijven we in welke mate de rekenniveaus omhoog gaan na een verbetering in de scores op de Cito LVS-toetsen Rekenen/Wiskunde van van respectievelijk +0.30 en +0.40 populatiestandaarddeviaties. Op de horizontale as staan de huidige rekenniveaus. Op de verticale as staan de projecties, dat wil zeggen de huidige rekenniveaus + het *treatment effect* van +0.30 populatiestandaarddeviaties (Figuur 3A) en +0.40 populatiestandaarddeviaties (Figuur 3B).

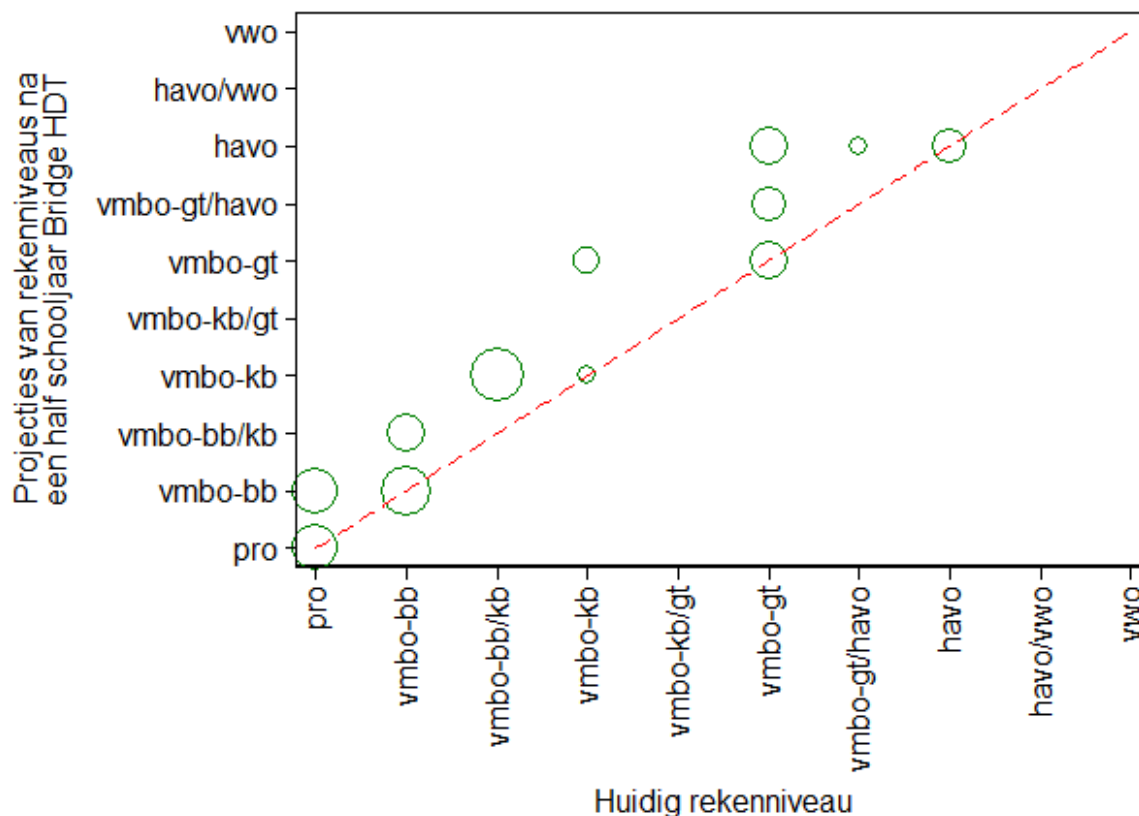
Wanneer de huidige rekenniveaus samenvallen met de projecties, gaan de leerlingen na treatment niet voldoende vooruit om een niveau op te schuiven. De rode stippellijn markeert de situatie waarin de huidige prestaties en de projecties niet verschillen. De omvang van de *bubbles* zijn proportioneel aan het aantal leerlingen met de betreffende niveaus. We zien bijvoorbeeld dat 10 van de 45 leerlingen rekt op een niveau dat past bij een traject naar een advies voor praktijkonderwijs, d.w.z. een toetsadvies praktijkonderwijs.³⁰

²⁹ Kraft (2020) classificeert onderwijsinterventies met een effectgrootte van 0.2 of groter als 'groot'.

³⁰ In de controlegroep scoren halverwege groep 7, 10 van de 45 een toetsadvies praktijkonderwijs, 10 van de 45 een toetsadvies vmbo-basisberoeps, 7/45 een gemengd toetsadvies vmbo-basis/kader, 3/45 een toetsadvies vmbo-kaderberoeps, 11/45 een toetsadvies vmbo-gemengd/theoretisch, 1/45 een gemengd toetsadvies vmbo-gt/havo en 3/45 een toetsadvies havo. In de scholen zijn ook leerlingen die veel beter scoren op toetsen. Deze

In Figuur 3A beschrijven we projecties op basis van een treatment effect van +0.30 populatiestandaarddeviaties, oftewel na een half schooljaar Bridge HDT (zie Tabel 1). Op de verticale as in Figuur 3A staan de geprojecteerde rekenniveaus (de huidige niveaus + een treatment effect van +0.30 populatiestandaarddeviaties). De projecties laten zien dat na een treatment effect van +0.30 populatiestandaarddeviaties een deel van de leerlingen op een hoger niveau rekent. Van de 10 leerlingen met een toetsadvies praktijkonderwijs, heeft de helft na een half schooljaar Bridge HDT een toetsadvies vmbo-basis. Dergelijke verschuivingen vinden plaats bij alle rekenniveaus. Van de 11 leerlingen met een toetsadvies vmbo-gt schuift meer dan de helft een half niveau (van vmbo-gt naar vmbo-gt/havo) of een heel niveau (van vmbo-gt naar havo) op.

Figuur 3A: Projecties rekenprestaties (Cito LVS-toetsen Rekenen/Wiskunde) op basis van een half schooljaar Bridge HDT



De projecties in Figuur 3A laten ook zien dat niet alle leerlingen een half of een heel niveau opschuiven. Dit is een gevolg van het feit dat meerdere vaardigheidsscores samenvallen met één bepaald niveau (of ‘toetsadvies’). Leerlingen kunnen dus beter zijn gaan rekenen zonder direct (formeel) in een hogere categorie te vallen. In onze steekproef zien we bijvoorbeeld

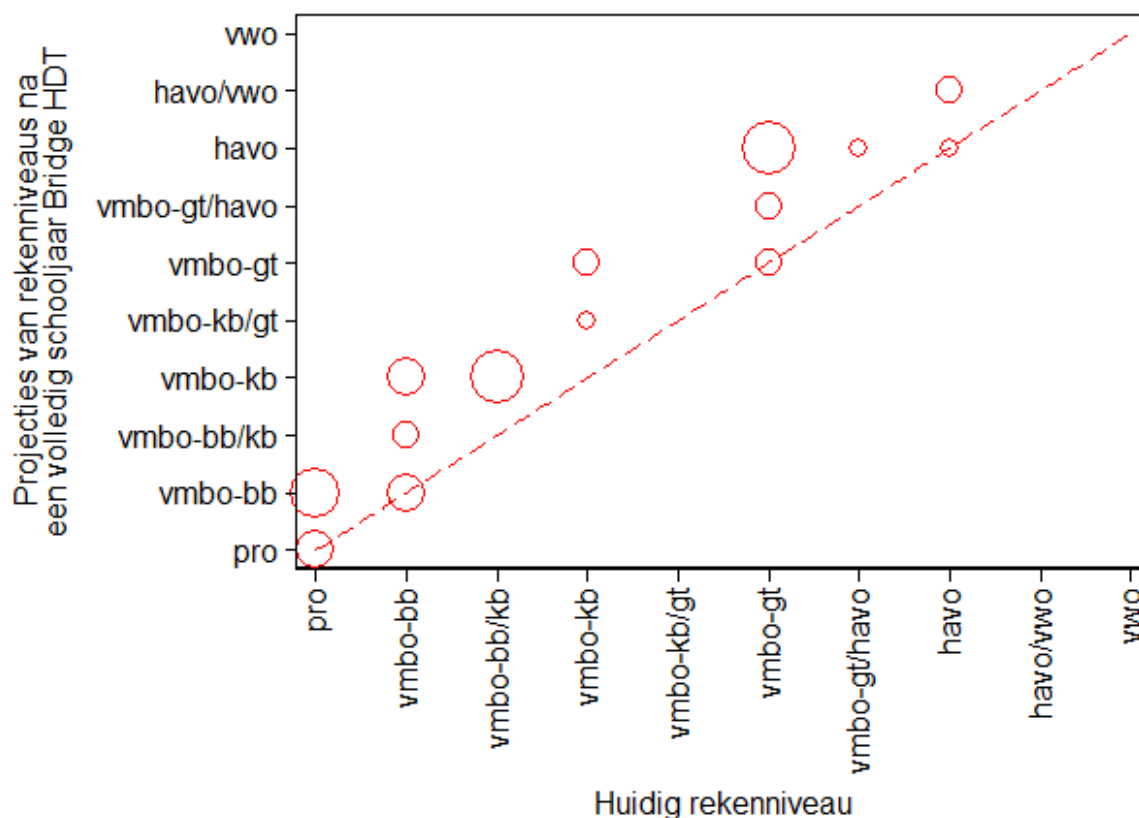
leerlingen zijn echter niet de doelgroep van Bridge HDT en zijn dus ook geen onderdeel van de treatment- of controlegroepen.

relatief veel leerlingen die rekenen op een niveau dat past bij een traject naar een advies voor praktijkonderwijs. Een deel van deze groep scoort ver onder de grenswaardes voor een toetsadvies vmbo-basis. Ze scoren bij aanvang relatief laag voor het praktijkniveau en gaan dus (net) niet genoeg vooruit om een niveau op te kunnen schuiven.

Figuur 3B laat de projecties zien voor een volledig schooljaar ononderbroken Bridge HDT. Bij deze projecties gaan we uit van een treatment effect van +0.40 populatie-standaarddeviaties. De projecties laten substantiële niveauverbeteringen zien.

De Figuren 3A en 3B laten zien dat de effecten van Bridge HDT duidelijk zeer relevant zijn voor de betrokken leerlingen, zelfs al na een half schooljaar. Gemiddeld over alle leerlingen uit de controlegroep zien we dat een half schooljaar Bridge HDT voor 45% van de leerlingen een half niveau winst oplevert en voor 15% een volledig niveau. Een volledig schooljaar ononderbroken Bridge HDT levert voor ongeveer 40% een half niveau winst en voor ongeveer 40% een volledig niveau. De percentages in deze alinea zijn afgerond op 5-tallen.

Figuur 3B: Projecties rekenprestaties (Cito LVS-toetsen Rekenen/Wiskunde) op basis van een volledig schooljaar (ononderbroken) Bridge HDT

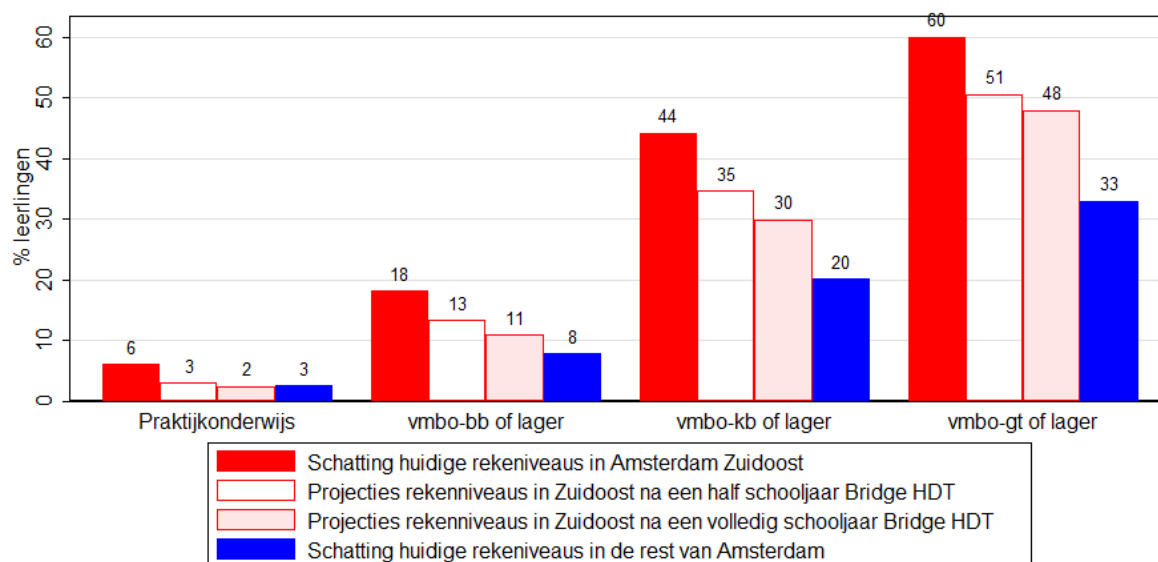


4.2 Projecties en mogelijk relevante beleidsdoelstellingen

Hoofdstuk 4.1 laat zien dat leerwinsten van +0.30 en +0.40 populatiestandaarddeviaties zeer relevant zijn voor de betrokken leerlingen. Met zulke effecten kunnen substantiële stappen worden gezet richting een mogelijk belangrijke beleidsdoelstelling: het wegwerken van de onderwijsachterstanden tussen Amsterdam Zuidoost en de rest van Amsterdam. In dit hoofdstuk maken we dit idee expliciet en koppelen we de projecties uit hoofdstuk 4.1 aan deze mogelijk relevante beleidsdoelstellingen.

Als uitgangspunt nemen we geschatte rekenniveaus in Amsterdam Zuidoost (de rode staven in Figuur 4) en de geschatte rekenniveaus in de rest van Amsterdam (de blauwe staven in Figuur 4). Bij gebrek aan recente beschikbare data over de werkelijke rekenniveaus voor alle scholen en leerlingen in de stad, kiezen we als schatting voor de rekenniveaus de gerealiseerde uitstroompercentages in 2018/19.³¹ De aanname die we hierbij maken is dat de percentages gerealiseerde adviezen, redelijkerwijs overeenkomen met de percentages bijbehorende toetsadviezen Rekenen/Wiskunde. Op basis van publiek beschikbare data op schoolniveau berekenen we dat ongeveer 6% van de leerlingen in Amsterdam Zuidoost de basisschool verlaat met een advies voor praktijkonderwijs, tegen ongeveer 3% in de rest van de stad. We gaan er in het vervolg van dit hoofdstuk dus vanuit dat ongeveer 6% van de leerlingen in Amsterdam Zuidoost rekt op een niveau dat past bij een advies voor praktijkonderwijs, tegen ongeveer 3% in de rest van de stad.

Figuur 4: Projecties en schattingen rekenniveaus geheel Amsterdam in vergelijking met Amsterdam Zuidoost (na een half en volledig jaar Bridge HDT)



De witte en licht rode staven zijn projecties op basis van een half schooljaar Bridge HDT (+0.30 populatiestandaarddeviaties) en een volledig schooljaar ononderbroken Bridge HDT

³¹ Bron: eigen berekeningen op basis van publiek beschikbare data van Dienst Uitvoering Onderwijs (DUO).

(+0.40 populatiestandaarddeviaties). We gebruiken hiervoor de projecties uit het vorige hoofdstuk 4.1.

Op basis van deze projecties concluderen we dat met effecten van +0.30 en +0.40 populatiestandaarddeviaties, en gebaseerd op de hierboven genoemde aannames, de bestaande achterstanden in de rekenniveaus tussen Amsterdam Zuidoost en de rest van Amsterdam voor een substantieel deel kunnen worden weggewerkt. De projecties laten zien dat voor leerlingen die rekenen op een niveau dat past bij een advies voor praktijkonderwijs, de achterstanden volledig kunnen worden gedicht. Bij niveaus van vmbo-basisberoeps of lager³² zien we dat de achterstanden voor ongeveer 70% kunnen worden gedicht.³³ Bij alle vmbo niveaus, d.w.z. vmbo-gt of lager, zien we dat naar schatting (onder bovengenoemde aannames) dat ongeveer de helft van de achterstanden met effecten van +0.40 populatiestandaarddeviaties kan worden weggewerkt.³⁴

Bij de eerdere bespreking van Figuur 1 in dit rapport, hebben we laten zien dat basisschoolleerlingen in Zuidoost tot 1.7 keer vaker een vmbo-advies krijgen dan leerlingen in de rest van de stad. De 1.7 is ook zichtbaar in Figuur 4, waarbij in Amsterdam Zuidoost naar schatting ongeveer 60% van de leerlingen rekt op het niveau van een vmbo-gt advies of lager, ten opzichte van 33% in de rest van Amsterdam ($33\%/60\%=1.7$). Op basis van interventies, zoals Bridge HDT, met een verwacht effect van +0.40 populatiestandaarddeviaties na een volledig schooljaar, kunnen dergelijke achterstanden in de rekenprestaties met ongeveer de helft worden weggewerkt. In Figuur 1 hebben we gezien dat verbeteringen van een dergelijke omvang op een grote schaal in de afgelopen 10 jaar niet zijn voorgekomen. Om achterstanden in het rekenen daadwerkelijk weg te werken zullen interventies met een dergelijk potentieel breed moeten worden ingezet. Ook zal moeten worden gezocht naar interventies die de prestaties in de taalvaardigheid op een vergelijkbare manier verbeteren. Prestaties in de taalvaardigheid worden op de korte termijn niet door Bridge HDT beïnvloedt. Om de achterstanden in de uitstroom volledig of deels in te lopen zal ook de taalbeheersing van leerlingen in Amsterdam Zuidoost moeten verbeteren.

4.3 Vervolgonderzoek

Het onderzoek naar Bridge HDT in Amsterdam Zuidoost duurt in principe drie jaar en dit is de rapportage van het eerste loopjaar (2019/20). Voor het tweede loopjaar (2020/21) worden de implementatie en het onderzoek op dezelfde manier vormgegeven. Uiteindelijk kunnen we

³² De groep vmbo-basisberoeps of lager, omvat de toetsadviezen vmbo-basis en alle toetsadviezen die lager zijn dan vmbo-basisberoeps. In dit geval gaat het dus om de toetsadviezen praktijkonderwijs en vmbo-basisberoeps. Bij de groep vmbo-gt en lager, gaat het dus om de toetsadviezen praktijkonderwijs, vmbo-basisberoeps, vmbo-kaderberoeps en vmbo-gt (en alle tussenliggende, gemengde toetsadviezen).

³³ De huidige onderwijsachterstand tussen Amsterdam Zuidoost en de rest van de stad is $18\%-8\%=10\%$. Een treatment effect van +0.40 populatiestandaarddeviaties leidt in deze projecties tot een afname van $18\%-11\%=7\%$, oftewel $7\%/10\%=0.70$ (een reductie van 70% van de huidige onderwijsachterstand).

³⁴ De Huidige onderwijsachterstand is $60\%-33\%=27\%$. Hiervan wordt $60\%-48\%=12\%$ ingelopen. $12\%/27\%=0.44$, een reductie van ongeveer de helft van de bestaande onderwijsachterstanden.

de data van het eerste en tweede loopjaar combineren om zo met meer zekerheid uitspraken te kunnen doen. Ook gaan we voorsnog uit van een schooljaar 2020/21 met minder COVID-19 gerelateerde problemen, zoals sluitingen van scholen. We hopen dus op basis van een tweede jaar aan experimentele data uitspraken te kunnen doen over een volledig schooljaar Bridge HDT in Amsterdam Zuidoost.

Verder staat er nog een derde loopjaar op de planning. Bridge HDT zal dan waarschijnlijk een andere vorm krijgen. Om de kosten van de interventie te drukken, zal worden geëxperimenteerd met nieuwe, goedkopere (kosten per leerling per jaar) HDT-modellen. Dergelijke modellen zijn nog niet in het basisonderwijs geïmplementeerd, en zijn op dit moment in ontwikkeling. Denk hierbij aan een 1-op-3 setting, en bijvoorbeeld 3 in plaats van 5 uur Bridge HDT op meer basisscholen tegelijk. Hierdoor kunnen meer scholen deelnemen. Het doel is om systematisch te zoeken naar een ‘*sweet spot*’ waarbij met beperkte middelen de beste resultaten worden behaald. Op basis van ervaringen van Saga Education kunnen we concluderen dat het afschalen van het aantal uren tutoring (en dus de kosten per leerling) leidt tot een minder dan proportionele afname van de effectgroottes. Ook in dit derde jaar zal het onderzoeksteam proberen een RCT op te zetten, afhankelijk van de opzet van het nieuwe model.

Tegelijkertijd met de opzet van nieuwe RCT’s hoopt het onderzoeksteam de evaluatie van de resultaten van de loopjaren 1 en 2 verder te verdiepen. Hiervoor zijn extra financiële middelen nodig en extra inspanningen van scholen voor het verzamelen van data. Het onderzoeksteam is momenteel bezig met het zoeken naar extra financiering (via bijvoorbeeld het Kenniscentrum Ongelijkheid). We denken hierbij aan verdiepend onderzoek naar de specifieke aspecten van het rekenen waar leerlingen moeite hebben, en op welke aspecten Bridge HDT de grootste impact heeft. Ook willen we onderzoeken of het mogelijk is om op basis van rekentoetsgegevens op vraagniveau (oftewel op ‘itemniveau’) te bekijken of er aanwijzingen zijn voor effecten van Bridge HDT op niet-cognitieve vaardigheden. Onderzoek van Borghans en Schils (2018)³⁵ en Zamarro et al. (2016)³⁶ dienen hierbij als inspiratie.

³⁵ *The Leaning Tower of Pisa. Decomposing Achievement Test Scores into Measures of Cognitive and Noncognitive Skills*, 2018. Dit onderzoek vindt dat de afname in performance gedurende de toets (van de eerste tot de laatste vraag) normaal gesproken afneemt, en dat de mate van afname in performance iets zegt over niet-cognitieve eigenschappen en vaardigheden, zoals nauwgezetheid (*‘conscientiousness’*) bijvoorbeeld.

³⁶ *Comparing and Validating Measures of Character Skills: Findings from a Nationally Representative Sample*, 2016.

5. Conclusies en aanbevelingen

We kunnen naar aanleiding van dit onderzoek concluderen dat Bridge HDT leidt tot meetbare (statistisch significante) effecten op de rekenprestaties van leerlingen in de treatment groep. Verder is het aannemelijk dat de effecten groter zouden zijn geweest als de implementatie niet belemmerd was door de sluiting van scholen in de tweede helft van het schooljaar 2019/20. Echter, al na een half schooljaar vinden we effecten die in de literatuur als ‘groot’ worden beoordeeld (vgl. Kraft, 2020). Na een half schooljaar Bridge HDT vinden we op de Cito LVS toetsen Reken/Wiskunde treatment effecten van +0.30 populatiestandaarddeviaties. Wanneer Bridge HDT het hele schooljaar ononderbroken had kunnen draaien (dus zonder COVID-gerelateerde uitdagingen), verwachten we op basis van onze bevindingen — en veronderstellingen gebaseerd op literatuur — effecten van rond de +0.40 populatiestandaarddeviaties.

De vooruitgang in de rekenvaardigheden die hoort bij dergelijke leerwinsten zijn zeer relevant voor de betrokken leerlingen. Met effecten van +0.30 populatiestandaarddeviaties scoren 40% van de betrokken leerlingen een half niveau hoger op rekentoetsen (bijvoorbeeld van vmbo-gt naar gemengd vmbo-gt/havo niveau) en 15% een volledig niveau hoger (bijvoorbeeld van vmbo-gt naar havo niveau). Met effecten van +0.40 populatiestandaarddeviaties scoren 40% van betrokken leerlingen een half niveau hoger en 40% een volledig niveau. Dergelijke effecten kunnen bijdragen aan een hogere instroom in het voortgezet onderwijs.

We vinden we het aannemelijk dat een dergelijke *boost* in de leerprestaties ook kan bijdragen aan een toename van het zelfvertrouwen. Dit hebben we echter nog niet formeel kunnen vaststellen.

Met dergelijke effecten kunnen ook substantiële stappen worden gezet richting mogelijk belangrijk beleidsdoelstellingen: het wegwerken van de onderwijsachterstanden tussen Amsterdam Zuidoost en de rest van de stad. Op basis van projecties concluderen we dat met treatment effecten tussen +0.30 en +0.40 populatiestandaarddeviaties achterstanden in de rekenprestaties tussen Amsterdam Zuidoost en de rest van de stad substantieel zouden kunnen worden teruggebracht als Bridge HDT (of een interventie met een vergelijkbaar effect) breder en succesvol wordt opgeschaald.

De projecties laten zien dat een heel jaar ononderbroken Bridge HDT bij leerlingen op het niveau dat past bij praktijkonderwijs, de achterstanden in de rekenprestaties volledig zouden kunnen wegwerken. Bij leerlingen die rekenen op een niveau dat past bij een advies voor vmbo-basisberoeps of lager, kunnen de bestaande achterstanden met 70% worden gereduceerd. Bij leerlingen op het niveau dat past bij vmbo-gt en lager, kan de interventie ongeveer de helft van de achterstanden wegwerken.

Gegeven de huidige interesse in de achterstanden die zijn veroorzaakt door de gevolgen van COVID-19, refereren we hier ook aan in de conclusie. Een half schooljaar Bridge HDT is ruim voldoende om (gemiddelde) achterstanden in de rekenprestaties veroorzaakt door de

schoolsluitingen in het tweede deel van het schooljaar 2019-2020 te kunnen wegwerken. Engzell, Freyd en Verhagen (2020) schatten dat schoolsluitingen in het tweede deel van het schooljaar 2019/20 hebben geleid tot een afname in de toetscores van -0.08 tot -0.12 standaarddeviaties.

We hebben ook onderzocht wat de invloed van Bridge HDT is op de scores op de Cito LVS-toetsen Begrijpend Lezen. We doen dit vooral ter controle omdat leerlingen door Bridge HDT elke schooldag reguliere - eventueel meer talige - lessen moeten missen. In de analyses vinden we geen effecten van Bridge HDT op het Begrijpend Lezen. Om de opgelopen achterstanden in andere domeinen weg te werken lijken dus ook andere (meer op taal gerichte) interventies nodig.

Dit onderzoek — en ook een ander gerelateerd onderzoek naar een HDT programma dat niet wordt uitgevoerd door The Bridge (De Ree en co-auteurs, nog niet gepubliceerd) — geven indicaties dat het mogelijk is om HDT programma's succesvol breder in te zetten, in Amsterdam en ook in andere gemeenten en regio's in Nederland. Het onderzoek draagt dus bij aan een ontwikkeling die we tot nu toe vooral in de VS zien. Hierbij pleiten toonaangevende onderzoekers regelmatig voor systematisch kwantitatief en kwalitatief onderzoek om — in samenwerking met overheden, schoolbesturen, maatschappelijke organisaties en wetenschap — effectieve en schaalbare programma's en interventies voor kansarme leerlingen te ontdekken en verder te ontwikkelen.³⁷ Dergelijke samenwerkingen hebben in de VS laten zien HDT programma's uitgevoerd door organisaties zoals Saga Education succesvol kunnen worden opgeschaald (Cook et al., 2015; Fryer et al., 2014). Op basis van dit rapport en andere evaluatiestudies van ons onderzoeksteam (zie hiervoor www.uva.nl/sepp) komt er ook voor de Nederlandse context meer duidelijkheid dat aanbieders zoals The Bridge — en ook andere organisaties (zie hierover De Ree en co-auteurs, nog niet gepubliceerd) — succesvol zijn geweest in het uitvoeren en deels opschalen van HDT programma's in aandachtswijken zoals Amsterdam Zuidoost.

Naast het behalen van positieve resultaten in Amsterdam Zuidoost heeft The Bridge ook succesvolle HDT programma's gedraaid in Haarlem Oost/Schalkwijk en, met steun van Gemeente Amsterdam, ook in het voortgezet onderwijs in Amsterdam Nieuw-West (Mundus College) en in Amsterdam Noord (Hogelant). We interpreteren de resultaten van dit onderzoek daarom deels ook als als een '*stress test*' voor The Bridge en een indicatie dat deze organisatie in staat is om meerdere interventies tegelijk adequaat te ondersteunen en implementeren.

Om de achterstanden (bijvoorbeeld tussen Amsterdam Zuidoost en de rest van de stad) voor een groot deel te dichten, zullen dergelijke effectieve interventies nog breder moeten worden uitgerold. Het is nu nog de vraag of de kwaliteit van implementatie bij verder opschalen kan worden gewaarborgd. Ook bij succesvol gebleken HDT aanbieders zoals The Bridge zal verdere opschaling binnen een korte tijd een uitdaging zijn.

³⁷ Zie bijvoorbeeld: <https://urbanlabs.uchicago.edu/labs/education> en Ander, Guryan en Ludwig (2016).

Onderzoekers investeren op dit moment veel tijd en energie in het beter begrijpen van de mogelijke problemen die ontstaan bij opschaling. Robert Slavin schreef hier recentelijk in een blogpost over “...*that programs proven to work at small scale can succeed if the key factors that made them work at small scale are implemented with fidelity at large scale.*”³⁸ Hieraan voegt hij het volgende toe:

“[R]ather than spending a lot of taxpayer money and hoping for the best, [the goal should be to] make scale-up successful by using evidence at the beginning, middle, and end of the process, to make sure that this time, we really know what we are doing. We [sh]ould make sure that effective programs remain successful at scale, rather than merely hoping they will.”

Slavin pleit dus voor het gebruik van RCT's tijdens de pilotfase om effectieve HDT programma's en aanbieders te kunnen identificeren. Deze effectieve aanbieders zouden daarna, volgens Slavin, een rol moeten spelen bij de opschaling van hun programma's. Echter, volgens Slavin is het cruciaal om tijdens en zelfs na de opschaalfase de programma's systematisch te blijven monitoren en te toetsen.

Gezien de hier gepresenteerde onderzoeksresultaten en de inzet van de stad om onderwijsongelijkheid in omgevingen zoals Zuidoost te verminderen, zou onze belangrijkste aanbeveling zijn om een pad te volgen dat vergelijkbaar is met het pad dat door Slavin is geschetst in zijn oproepen om op te schalen en door te gaan met de RCT-gebaseerde testen van effectieve HDT-programma's. We benadrukken hierbij het belang van een nauwe samenwerking en commitment van de Gemeente, deelnemende scholen en schoolbesturen en de uitvoerende partij. De eerste indicaties van de substantiële resultaten - die alleen mogelijk waren door een gedeeld commitment- zullen hopelijk betekenisvol zijn voor vele leerlingen in Amsterdam en daarbuiten.

³⁸ Deze blog post kan hier gevonden worden: <https://robertslavinsblog.wordpress.com/2020/12/17/large-scale-tutoring-could-fail-heres-how-to-ensure-it-does-not/>. Zie hiervoor ook een andere recente blog post van Robert Slavin (<https://robertslavinsblog.wordpress.com/2020/12/10/the-details-matter-thats-why-proven-tutoring-programs-work-better-than-general-guidelines/>) en series van blog posts van Mike Goldstein en Bowen Paulle (<https://fordhaminstitute.org/national/commentary/vaccine-makings-lessons-high-dosage-tutoring-part-i>).

Literatuurlijst

Ander, R., Guryan, J. & Ludwig, J. (2016). *Policy Brief 2016-12. Improving Academic Outcomes for Disadvantaged Students: Scaling Up Individualized Tutorials*. Maart 2016.

Andrabi, T., Das, J., Ijaz Khwaja, A. & Zajonc, T. (2011). Do Value-Added Estimates Add Value? Accounting for Learning Dynamics. *American Economic Journal: Applied Economics*. Vol. 3, No. 3, juli 2011.

Bol, T. (2020). *Inequality in homeschooling during the Corona crisis in the Netherlands. First results from the LISS Panel*.

Borghans, L. & Schils, T. (2018). *The Leaning Tower of Pisa. Decomposing achievement test scores into cognitive and noncognitive components*. Draft version 31 oktober 2012, niet gepubliceerd manuscript.

Cascio, E.U. & Staiger, D.O. (2012). *Knowledge, Tests and Fadeout in Educational Interventions*. NBER Working Paper 18038.

Centraal Planbureau. (2016). *Kansrijk onderwijsbeleid*. Den Haag.

Cook, P. J., Dodge, K., Farkas, G., Fryer Jr, R. G., Guryan, J., Ludwig, J. & Steinberg, L. (2014). *The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth: results from a randomized experiment in Chicago* (No. w19862). National Bureau of Economic Research.

Cook, P.J. et al. (2015). Not Too Late: Improving Academic Outcomes for Disadvantaged Youth. Northwestern University Institute for Policy Research, WP 15-01, January 2015. <http://www.ipr.northwestern.edu/publications/papers/2015/ipr-wp-15-01.html>

Davis, J., Guryan, J., Hallberg, K., & Ludwig, J. (2017). The economics of scale-up (No. w23925). National Bureau of Economic Research.

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A. M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research*, 87(2), 243-282.

Elffers, L., & Jansen, D. (2019). *De opkomst van schaduwonderwijs in Nederland: wat weten we en welke vragen liggen nog open?* Universiteit van Amsterdam.

Engzell, P., Freyd, A. & Verhagen, M. (2020, unpublished). *Learning inequality during the COVID-19 pandemic*.

Fryer, R. (2014). Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments. <http://qje.oxfordjournals.org/content/129/3/1355.short>.

Gross, J., Hudson, C., & Price, D. (2009). *The long-term costs of numeracy difficulties*. Every Child a Chance Trust/KPMG.

Kielman, A., Paulle, B. & Van Londen, M. (2017). Rekenen op je ouders: High Dosage Tutoring als middel om ouderbetrokkenheid te stimuleren in Rotterdam-Zuid. *Sociologie*, 13(2/3), 197-222. <https://doi.org/10.5117/SOC2017.2/3.003.KIEL>

Kraft, Matthew A. Interpreting effect sizes of education interventions. *Educational Researcher* 49, no. 4 (2020): 241-253.

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*. NBER Working Paper, (w27476).

Onderzoek, Informatie en Statistiek (2020). *De Staat van het Amsterdamse Primair Onderwijs*. Gemeente Amsterdam.

Onderwijsinspectie (2016). *De Staat van het Onderwijs: onderwijsverslag 2014-2015*. Ministerie van Onderwijs Cultuur en Wetenschap.

Onderwijsinspectie (2017). *De Staat van het Onderwijs: onderwijsverslag 2015-2016*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.

Wise, Steven L. & Lingyun Gao. (2017). A General Approach to Measuring Test-Taking Effort on Computer-Based Tests. *Applied Measurement in Education*, 30 (4), 343-354.

Zamarro, G., Cheng, A., Danish Shakeel, M. & Hitt, C. (2016). *Comparing and Validating Measures of Character Skills: Findings from a Nationally Representative Sample*. EDRE Working Paper 2016-08.

Zamarro, G., Hitt, C. & Mendez, I. (2017). *When Student's Don't Care: Reexamining International Differences in Achievement and Non-cognitive Skills*. EDRE Working Paper 2016-18.

Appendices

Appendix A: Operationalisering van statistische analyses en validiteitsanalyses

De uitkomstmaten op basis waarvan we de effecten van Bridge HDT meten zijn vaardigheidsscores, gebaseerd op de Cito LVS-toetsen Rekenen/Wiskunden en Begrijpend Lezen. De vaardigheidsscores zijn vervolgens geschaald in afwijking van het landelijk gemiddelde van de betreffende leeftijdsgroep en gedeeld door een schatting van de populatiestandaarddeviaties van de betreffende leeftijdsgroep. Het landelijk gemiddelde wordt standaard geleverd. De populatiestandaarddeviatie schatten we door het verschil van het 20e en het 80e percentiel te delen door 1.683. Het 20e en 80e percentiel worden ook standaard geleverd, en zijn respectievelijk de grenzen voor de functioneringsniveaus V-IV en II-I. De factor 1.683 is de afstand in standaarddeviaties tussen het 20e en 80e percentiel in de normale verdeling.

Vervolgens maken we indicator variabelen (of dummyvariabelen) aan. De treatment indicator is 1 voor leerlingen in de treatment groep en 0 voor leerlingen in de controle groep. De *block* indicatoren zijn 1 voor leerlingen in een bepaald *block* of stratum, en anders 0.

De effecten van Bridge HDT worden geschat met een regressiemodel, waarbij de gestandaardiseerde toetsscores worden geregresseerd op de treatment indicator en de volledige set aan *block* of stratum indicatoren. De geschatte parameter bij de treatment indicator is de schatting van het effect van Bridge HDT. De gerapporteerde standaardfouten bij deze schattingen zijn robuust tegen heteroskedasticiteit in de foutterm.

De randomisatie zorgt voor vergelijkbaarheid van de treatment en controlegroep. Echter omdat we niet voor alle 105 leerlingen altijd alle toetsscores hebben kunnen verzamelen, is er een mogelijkheid dat de treatment en controleobservaties van de geobserveerde dataset niet meer vergelijkbaar zijn, in de gebruikelijke zin. We kunnen toetsen voor deze mogelijke verminderde vergelijkbaarheid door een indicator voor een aanwezige toetsscore aan te maken, en deze te regresseren op de treatment indicator, de *block* indicatoren, en een interactie van de treatment indicator en de baseline scores op de Cito LVS-toetsen Rekenen/Wiskunde, gemeten halverwege groep 6, voorafgaand aan de start van Bridge HDT. We vinden hierbij geen evidentie voor verminderde vergelijkbaarheid. Het wel of niet kunnen meten van een toetsscore wordt dus niet significant beïnvloed door deelname aan Bridge HDT. Er is op basis van deze analyses dus geen reden om te twijfelen aan de validiteit van het onderzoek.

Appendix B: Effecten van Bridge HDT op de prestaties in het begrijpend lezen

In Tabel B.1 in Appendix B rapporteren we de resultaten op basis van scores op de Cito LVS-toets voor begrijpend lezen. De geschatte treatmenteffecten zijn niet statistisch significant. We vinden dus geen effecten van Bridge HDT op het begrijpend lezen. Deelnemers aan Bridge HDT scoren gemiddeld lager dan de leerlingen in de controlegroep, maar deze situatie bestond al voorafgaand aan de start van Bridge HDT en is tijdens Bridge HDT niet veranderd.

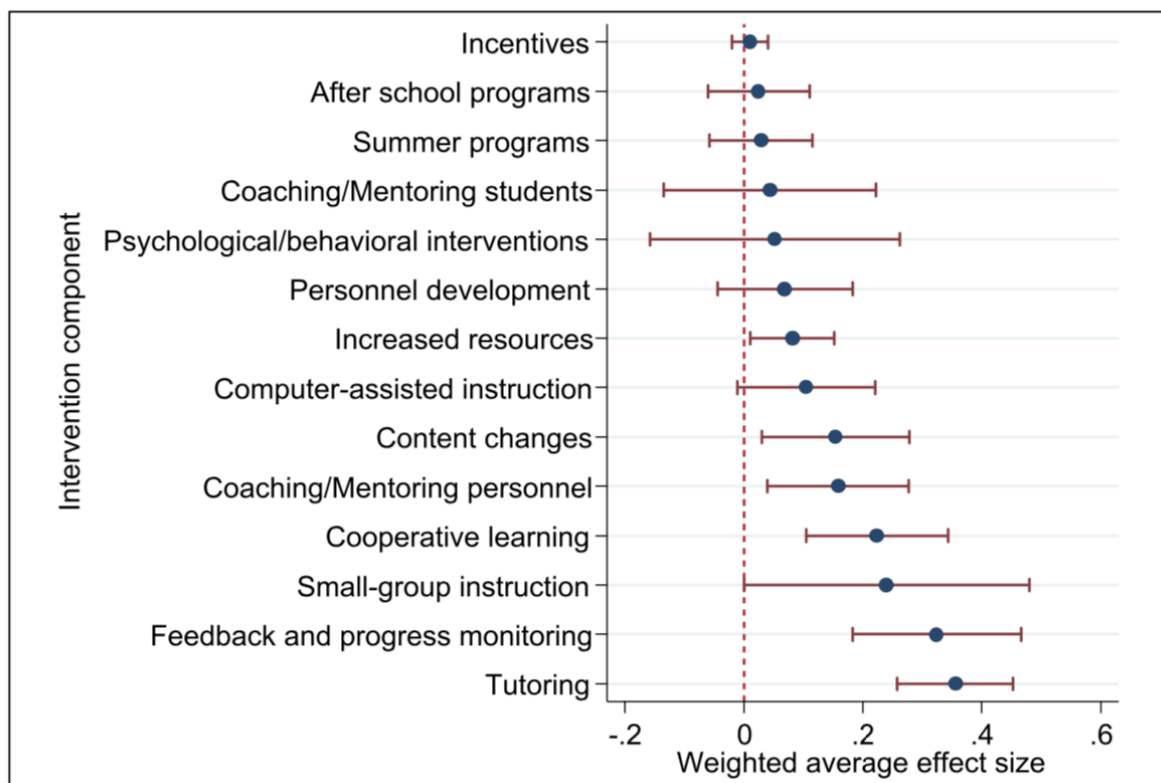
Tabel B.1: Effecten van Bridge HDT op de prestaties in het *begrijpend lezen* op de Cito LVS-toets (in termen van populatiestandaarddeviaties.)

Kolomnummer	(1)	(2)	(3)	(5)
Meetmoment	Baseline M6	Baseline E6	Midline M7	Endline E7
Baseline resultaten/Treatment effect	-0.16	-0.12	-0.12	-0.16
Standaardfout	(0.15)	(0.27)	(0.20)	(0.21)
p-waarde	[0.30]	[0.65]	[0.57]	[0.46]
Aantal observaties	<94>	<79>	<93>	<88>
Noten: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. De effecten van de interventie zijn geschat op basis van een regressiemodel, waarbij de gestandaardiseerde toetsscores wordt geregresseerd op stratum/ <i>block effects</i> (dummy/indicator variabelen die de <i>block</i> of stratum aangeven met een 1, of anders 0) en een indicator voor de treatmentgroep. De geschatte parameter op de indicator voor de treatmentgroep is gerapporteerd in deze tabel, als schatting voor het treatment effect.				

Appendix C: Bevindingen uit de literatuur

Dietrichson et al. (2017: 268) vatten de bevindingen in hun paper samen in het volgende citaat en de figuur hieronder: “Figure 5 shows weighted average effects sizes and confidence intervals graphically. Only effects at the end of intervention are included.”

Figuur C.1: Kopie van figuur “Weighted average effect sizes by component” in Dietrichson et. al. (2017: 268)



Appendix D: Operationalisatie toetsadviezen

Toetsadviezen worden gebruikt om aan te geven welk adviesniveau past bij een bepaalde taal- of rekenvaardigheid. De toetsadviezen die we hier gebruiken zijn gebaseerd op de Cito LVS toetsen. De grenswaardes voor de toetsadviezen kunnen worden uitgedrukt in populatiestandaarddeviaties ten opzichte van het landelijk gemiddelde van de betreffende leeftijdsgroep. Tabel D.1 geeft aan welke gestandaardiseerde toetsscores passen bij een bepaald advies.

De toetsadviezen “bewegen mee” met de leeftijd van de leerling. Om dit te beter te zien werken we met een voorbeeld. Leerlingen met vaardigheidsscores van 0.77 standaarddeviaties boven het landelijk gemiddelde van de leeftijdsgroep, krijgen een toetsadvies (rekenen) vwo. Ook in groep 4 bijvoorbeeld, kunnen leerlingen een toetsadvies vwo krijgen, terwijl in groep 4 verreweg de meeste leerlingen van deze groep nog lang niet aan de eisen voor een vwo advies voldoen. Het toetsadvies is dus relatief ten opzichte van de eigen leeftijdsgroep. Leerlingen die 0.77 standaarddeviaties boven het landelijk gemiddelde scoren in groep 4, zitten dus “op koers” voor een vwo advies in groep 8, maar daarvoor dienen ze wel dat relatief hoge niveau vast te houden.

Tabel D.1: Grenswaardes toetsadviezen, uitgedrukt in populatiestandaarddeviaties ten opzichte van de landelijk gemiddelde vaardigheidsscore van de leeftijdsgroep

Toetsadvies	Ondergrens	Bovengrens
Praktijkonderwijs		-2,29
Vmbo-basisberoeps	-2,29	-1,39
Vmbo-basis/vmbo-kader	-1,39	-1,18
Vmbo-kader	-1,18	-0,77
Vmbo-kader/vmbo-gt	-0,77	-0,73
Vmbo-gt	-0,73	-0,12
Vmbo-gt/havo	-0,12	0,00
Havo	0,00	0,57
Havo/vwo	0,57	0,77
Vwo	0,77	