



UNIVERSITEIT VAN AMSTERDAM

Reken op Zuidoost

Een evaluatie van drie jaar Bridge High-Dosage
Tutoring in het Amsterdamse basisonderwijs.

Jurgen Tijms en Joppe de Ree

24 januari 2023

Voorwoord

Dit onderzoek is uitgevoerd door medewerkers van het *Amsterdam Institute for Social Science Research* (AISSR) van de Universiteit van Amsterdam, in opdracht van de Gemeente Amsterdam. In dit onderzoek staat de volgende onderzoeksvraag centraal: Wat zijn de opbrengsten van Bridge High-Dosage Tutoring (HDT) in het basisonderwijs in Amsterdam Zuidoost?

Inhoud

1	Samenvatting	3
2	Bridge High-Dosage Tutoring	6
3	Onderzoeksopzet	9
4	Resultaten	12
4.1	Operationele bevindingen	12
4.2	De opbrengsten van Bridge HDT	14
5	Opbrengsten van Bridge HDT in de context van onderwijsachterstanden	17
6	Conclusies en aanbevelingen	21
A	Methodologische details	25
B	Opbrengsten uitgedrukt in Didactische Leeftijdsequivalenten (DLE's)	26
C	Robuustheidsanalyses en validiteit	28

1 Samenvatting

Dit rapport bespreekt de eindresultaten van gerandomiseerd onderzoek¹ naar de opbrengsten van twee gerelateerde High-Dosage Tutoring (HDT) programma's van Stichting The Bridge Learning Interventions (The Bridge). De HDT interventies zijn ingezet op acht basisscholen in Amsterdam Zuidoost. De doelgroep van de interventie zijn groep 7 leerlingen met rekenvaardigheidsscores op of onder het landelijk gemiddelde. De rapportage bespreekt de resultaten van drie loopjaren, 2019/20, 2020/21 en 2021/22.

Tijdens de eerste twee loopjaren is een “voltijdsmodel” geïmplementeerd op vijf van de acht deelnemende basisscholen. Deelnemers aan het voltijdsmodel krijgen een schooljaar lang, vijf dagen in de week, 45 minuten per dag, extra rekenondersteuning op maat. Tijdens het derde en laatste loopjaar is een “3-daagsmodel” geïmplementeerd op alle acht deelnemende basisscholen. Het 3-daagse model levert een kostenbesparing op, waardoor met een gelijke investering meer leerlingen kunnen worden bereikt.

Bridge HDT draait onder schooltijd. Een team van zes tutors en een *Site Director* (projectleider) reist hiervoor op een werkdag van school naar school. Aan het begin van het schooljaar worden deelnemende groep 7 leerlingen in groepjes van twee ingedeeld en gekoppeld aan hun (vaste) tutor.² In deze samenstelling wordt een jaar lang intensief gewerkt aan de rekenvaardigheden. Principes als *mastery learning* en een *growth mindset* staan hierbij centraal. Verder onderhouden de tutors regelmatig³ telefonisch contact met de ouders/verzorgers van de deelnemers en is er afstemming met de leerkrachten en ander schoolpersoneel. De tutors worden dagelijks begeleid en ondersteund door de *Site Director*.

Bridge HDT richt zich op leerlingen die beneden gemiddeld scoren op rekentoetsen en met name op leerlingen met een verhoogde kans op onderpresteren. The Bridge is om die

¹Gerandomiseerd onderzoek wordt vaak aangeduid met de Engelse term *randomized controlled trial* of met de afkorting RCT.

²De *matching* wordt gedaan door The Bridge, in samenspraak met de school. Hierbij wordt onder andere rekening gehouden met het rekenniveau van de deelnemers.

³Bij het voltijdsmodel is er wekelijks telefonisch contact. Bij het 3-daagse model is er tweewekelijks telefonisch contact.

reden actief in wijken en scholen met relatief veel huishoudens met o.a. lagere inkomens. In samenspraak met de Gemeente Amsterdam zijn in Amsterdam Zuidoost de deelnemende scholen (deels) geselecteerd op basis van percentages zogenaamde “doelgroepleerlingen”.⁴ The Bridge is ook actief op basis- en middelbare scholen in Haarlem Schalkwijk/Oost, Amsterdam Nieuw-West, Gouda en Zaandam.

De uitvoering van het eerste loopjaar is sterk gehinderd door de gevolgen van de coronapandemie. Toch vinden we relevante effecten op de rekenprestaties, met effectgroottes in de orde van 0.30. Het tweede loopjaar werd in veel mindere mate gehinderd. We vinden voor het tweede loopjaar effectgroottes in de orde van 0.60. Dergelijke substantiële opbrengsten worden zelden gerapporteerd in de literatuur (zie bijvoorbeeld [14]). Voor het derde loopjaar, waarbij het 3-daagse model is uitgetoetst, vinden we effecten in de orde van 0.25-0.30 na een schooljaar. We kunnen deze effectgroottes omrekenen naar de door onderwijsprofessionals veelgebruikte Didactische Leeftijdsequivalenten (DLE's). De effectgroottes van 0.30 komen overeen met ongeveer drie DLE punten. Effectgroottes van 0.60 komen overeen met ongeveer zes DLE punten. De twee toepassingen van Bridge HDT leveren dus verbeteringen in de rekenprestaties die overeenkomen met drie tot zes maanden extra schooltijd.

Met deze opbrengsten kunnen belangrijke beleidsdoelen worden gerealiseerd. In Amsterdam Zuidoost, bijvoorbeeld, zijn leerlingen met lagere rekenniveaus oververtegenwoordigd. Met effectgroottes van 0.3-0.6 kan deze oververtegenwoordiging in een schooljaar worden weggewerkt (zie hoofdstuk 5). Echter, hierbij maken we twee kanttekeningen. Ten eerste brengt het wegwerken van achterstanden op grotere schaal (Zuidoost-breed, bijvoorbeeld) weer nieuwe uitdagingen met zich mee. Het is tijdens het opschalen van belang om de (vooral nog onbekende) kernonderdelen van de hier geëvalueerde HDT toepassing te behouden. Recente literatuur spreekt hierbij van een mogelijke *voltage drop*, waarbij *core components* tijdens het opschalen kunnen verwateren (zie bijvoorbeeld [1] en [16]). Ten tweede zien we

⁴Doelgroepleerlingen zoals gebruikt door zijn leerlingen die o.a. opgroeien in gezinnen met ouders/verzorgers met een laag opleidingsniveau of een inkomen rond het bijstandsniveau. Er zijn echter nog een aantal factoren die een rol spelen. Zie <https://onderzoek.amsterdam.nl/artikel/project-berekening-doelgroepleerlingen-amsterdamse-basisonderwijs-2017-2018>.

in dit rapport dat de opbrengsten na verloop van tijd deels wegzakken. Leerlingen die in eerste instantie van de HDT sessies hebben geprofiteerd, leveren na een half schooljaar ongeveer de helft van de opgedane voorsprong weer in.⁵ Dit resultaat suggereert dat de langetermijnopbrengsten lager zullen zijn dan de opbrengsten die direct na afloop van de interventie konden worden vastgesteld. Een dergelijke *fadeout* van opbrengsten is echter in de literatuur een algemeen verschijnsel en niet specifiek voor HDT interventies. Ook is de interpretatie van dit fenomeen niet eenduidig (zie bijvoorbeeld [2] en [4]).

We trekken we de volgende conclusies:

1. Op de korte termijn levert Bridge HDT zinvolle opbrengsten. Basisschoolleerlingen in Amsterdam Zuidoost gaan beter rekenen na gerichte en intensieve ondersteuning. In samenspel met het reguliere programma op school biedt Bridge HDT dergelijke ondersteuning. Met opbrengsten in de orde van 0.3-0.6 kan de oververtegenwoordiging van lagere rekenniveaus in Amsterdam Zuidoost substantieel worden teruggebracht, zo niet, worden weggewerkt. Een voorwaarde voor het behalen van een dergelijk beleidsdoel is echter, natuurlijk, dat HDT breed moet worden ingezet. In principe op alle basisscholen in Zuidoost en met behoud van effectiviteit.
2. De opbrengsten van HDT lijken samen te hangen met de intensiteit van de interventie. Het voltijdsmodel levert meer op dan het 3-daagse model. De verschillen tussen de modellen zijn echter niet statistisch significant. Omdat de beschikbare literatuur ook een dergelijke samenhang lijkt op te leveren, zijn we toch bereid om deze voorlopige conclusie te trekken (zie bijvoorbeeld [3] en [10]).
3. De langetermijneffecten op de rekenprestaties zijn waarschijnlijk kleiner dan de opbrengsten die direct na afloop van de interventie kunnen worden vastgesteld. Deze bevinding is in lijn met de bestaande literatuur over interventies in het onderwijs (zie bijvoorbeeld [2] en [4]). Om de onderwijsprestaties langdurig te verbeteren is het wellicht zinvol om te zoeken naar “onderhoudsinterventies”, waarbij de doelgroep gedu-

⁵[12] vinden meer persistentie in de effecten van Saga Education in de Verenigde Staten.

rende langere tijd extra wordt ondersteund. Ook kan worden nagedacht over of, en hoe, de aansluiting met het reguliere aanbod op school kan worden verbeterd.

In hoofdstuk 5 laten we zien dat de onderwijsachterstanden in Amsterdam Zuidoost een structureel karakter hebben. Trends in de schooladviezen geven aan dat leerlingen in Zuidoost al ruim een decennium achterblijven, zonder noemenswaardige verbeteringen. Bridge HDT kan onderdeel zijn van een aanpak gericht op dit bredere kader van onderwijsachterstanden. Bridge HDT is complementair aan het reguliere programma op school. Het kan de leerkracht ontlasten om zo systematisch, en met meer energie, grotere slagen te maken. Ook in de Verenigde Staten, waar dit type HDT-programma's voor het eerst zijn ingezet en systematisch geëvalueerd, gaan nu stemmen op om HDT-programma's een vast(er) onderdeel te laten zijn van de schooldag (zie bijvoorbeeld [15]). Niet in plaats van het reguliere instructiepakket, maar als onderdeel ervan.

Zoals ook bij de evaluatie van het eerste loopjaar is opgemerkt, laat recent onderzoek overtuigend zien dat de lockdowns tijdens het schooljaar 2019/2020 de prestaties op school negatief hebben beïnvloed. De orde van grootte van deze *covid learning-loss* lijkt echter (in absolute waarde) veel kleiner dan de effecten van een jaar lang Bridge HDT [9]. Bridge HDT kan dus worden ingezet om dergelijke achterstanden weg te werken, maar moet in eerste instantie worden gezien als onderdeel van een oplossing voor grotere en meer structurele vormen van achterstanden.

2 Bridge High-Dosage Tutoring

High-Dosage Tutoring (HDT) is ontworpen en ontwikkeld in de Verenigde Staten (VS). Een belangrijke aanbieder van HDT in de VS is Saga Education.⁶ Saga HDT is ook herhaaldelijk met behulp van RCT's onderzocht en ook op grotere schaal succesvol gebleken [12]. Een versie van deze aanpak is vanaf 2017 doorontwikkeld en aangepast aan de Nederlandse con-

⁶<https://www.sagaeducation.org/>

text door The Bridge Learning Interventions (The Bridge).⁷ Saga Education is hierbij vanaf het begin betrokken geweest als consultant van The Bridge. In de huidige toepassing richt Bridge HDT zich op maatschappelijk kwetsbare leerlingen die beneden gemiddeld presteren op school. Een belangrijk onderdeel van Bridge HDT is de intensiteit, oftewel de *high dosage*. Leerlingen krijgen in het voltijdsprogramma een schooljaar lang, elke dag een lesuur onder schooltijd, rekenbegeleiding op maat.

The Bridge is gedurende het project in Amsterdam Zuidoost actief met zes fulltime tutors, één fulltime *Site Director* (projectleider en eerste aanspreekpunt van de school) en één invaltutor/projectassistent. De teamleden zijn in vaste dienst bij The Bridge. De tutors hoeven geen achtergrond te hebben in het onderwijs. Ook worden er door The Bridge geen strenge voorwaarden gesteld aan het opleidingsniveau van tutors.⁸ Om in aanmerking te komen moeten tutors minimaal een MBO-3 diploma hebben behaald, slagen voor een rekentoets op 2F niveau en met “goed” worden beoordeeld op basis van een proefles met leerlingen.⁹ De bedoeling is om tutors vooraf en tijdens de interventie voldoende te trainen en te begeleiden, waardoor ze ook zonder specifieke opleiding effectief kunnen zijn. Hierbij wordt de *Site Director* gecoacht door medewerkers van het *central office* en de tutors vervolgens (op dagelijkse basis) door de *Site Director*.

De salarissen voor beginnende tutors liggen rond het minimumloon. Bij het aantrekken van tutors wordt deels aanspraak gemaakt op het maatschappelijke karakter van het werk. Het tutorschap wordt door tutors vaak gezien als een soort “maatschappelijke diensttijd”. Jonge tutors zien het soms ook als een eerste stap richting een carrière in het onderwijs, als leerkracht bijvoorbeeld. Om voldoende en kwalitatief goede tutors aan te trekken, zijn er vanaf het begin goede arbeidsvoorwaarden opgesteld.

Bij de start van het project in Amsterdam Zuidoost worden in elke deelnemende school maximaal 12 groep 7 leerlingen geselecteerd voor deelname aan het HDT programma. Op basis van rekenniveaus (en eventueel andere kenmerken) worden de deelnemers in groep-

⁷<https://www.tbli.nl/>

⁸Zie bijvoorbeeld <https://tbli.nl/onderwijsheld-woorden/> voor voorbeeldvacatures bij The Bridge.

⁹In de “coronatijd” is het niet altijd gelukt om de proefles te realiseren.

jes van twee ingedeeld en gekoppeld aan een vaste tutor. Het is de intentie om gedurende het project in deze vaste samenstelling verder te werken. De stabiliteit faciliteert de opbouw van een band tussen de leerlingen en de tutor. In deze setting worden achterstanden weg-gewerkt, 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*. Hiernaast onderhoudt de tutor wekelijks (bij het voltijdsmodel) of tweewekelijks (bij het 3-daagse model) telefonisch contact met ouders/verzorgers en is er afstemming met het schoolpersoneel. De tutors verzorgen de sessies in een vast klaslokaal op school.

Het rekencurriculum van Bridge HDT kent vijf domeinen die zijn opgesteld aan de hand van het referentieniveau 1F uit het referentiekader rekenen.¹⁰ Het referentieniveau 1F omvat basisvaardigheden die in principe alle leerlingen aan het eind van het basisonderwijs zouden moeten kunnen beheersen. De vijf rekendomeinen zijn getallen, verhoudingen, tijd en kalender, meten/meetkunde en verbanden. Het rekencurriculum is door The Bridge zelf ontwikkeld en samengesteld. Specifieke rekenstrategieën worden altijd door de *Site Director* met de school afgestemd zodat er geen verwarring ontstaat.

Bridge HDT levert maatwerk op het niveau van de leerling. De sessies beginnen aan het begin van het schooljaar met het onderwerp getallen. Tutors werken vervolgens in het eigen tempo van de leerlingen door het curriculum. Tutors stimuleren deelnemers om zoveel mogelijk domeinen volledig te beheersen in plaats van alle domeinen maar half, volgens het principe van *mastery learning*.¹¹ Afhankelijk van het tempo waarin de individuele leerling zich ontwikkelt zullen meer of minder onderwerpen kunnen worden behandeld. De leerkracht in de klas heeft niet altijd voldoende tijd om specifiek en langdurig op individueel niveau aan bepaalde achterstanden te werken. Het HDT programma biedt dus, indirect, ondersteuning aan de school en aan de leerkracht in de klas.

¹⁰Zie bijvoorbeeld <https://www.rijksoverheid.nl/onderwerpen/taal-en-rekenen/referentiekader-taal-en-rekenen>.

¹¹Bij de introductie van het 3-daagse model is niet van dit principe afgeweken. De consequentie hiervan is dat bij de uitvoering van het 3-daagse model over het algemeen minder onderwerpen aan bod zijn gekomen.

Bridge HDT zet ook in op de sociaal-emotionele ontwikkeling van leerlingen. De tutors bieden persoonlijke begeleiding vanuit de theorie van de *growth mindset*. Een uitgangspunt hierbij is dat tutors vertrouwen hebben in de ontwikkelingsmogelijkheden van alle deelnemers, ongeacht het aanvangsniveau. Het rekenen wordt hierbij 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 soms moeilijk meekomen. Uiteindelijk is de bedoeling dat de *growth mindset* wordt geïnternaliseerd en daarmee ook in de toekomst (op school of daarbuiten) van pas kan komen.

Tutors bereiden elke dag elke sessie opnieuw voor. De tutors monitoren in detail het niveau en de ontwikkeling van hun leerlingen en kunnen op deze manier maatwerk leveren. Het voorbereiden van de sessies gebeurt na afloop van de sessies. De *Site Director* heeft hierin een ondersteunende rol. Hierbij kan nog worden opgemerkt dat bij de overgang van het voltijdsmodel naar het 3-daagse model, de *caseload* voor de tutors en de *Site Director* hoger kwam te liggen. Bij het voltijdsmodel is elke tutor verantwoordelijk voor het functioneren van 5 groepjes van 2 deelnemers. Bij het 3-daagse model is elke tutor verantwoordelijk voor 8 groepjes van 2 deelnemers. Bij deze aanpassing is het inspelen op de specifieke eigenschappen en behoeftes van elke deelnemer dus een grotere uitdaging.

3 Onderzoeksopzet

Bridge HDT in Amsterdam Zuidoost wordt geëvalueerd op basis van gerandomiseerd onderzoek.¹² Bij gerandomiseerd onderzoek worden potentiële deelnemers willekeurig (*random*) toegewezen aan een interventiegroep of een controlegroep. De consequentie van de wille-

¹²Het onderzoek is geregistreerd in het *social science registry* voor *randomized controlled trials* van de *American Economic Association* (<https://www.socialscienceregistry.org/trials/4643>).

keurige toewijzing is dat, voorafgaand aan de interventie, de interventiegroep en de controlegroep niet systematisch van elkaar verschillen. Systematische verschillen tussen de interventiegroep en de controlegroep die we na afloop van de interventie terugzien in de data, kunnen om die reden worden beoordeeld als een effect van de interventie. Gerandomiseerd onderzoek wordt gezien als de gouden standaard in het onderzoek naar de opbrengsten van o.a. beleidsinterventies.

Tijdens de eerste twee loopjaren is het voltijdsmodel uitgetoetst op vijf van de acht deelnemende basisscholen. Tijdens het derde loopjaar is het 3-daagse model ingezet, waarbij meer leerlingen konden worden bereikt. Voor het derde loopjaar zijn drie nieuwe scholen toegevoegd.

In groep 6 wordt de doelgroep van de HDT interventie vastgelegd. Leerlingen in groep 6 met rekenscores op, of onder, het landelijk gemiddelde, komen in aanmerking. Meer specifiek gaat het om leerlingen met vaardigheidsniveaus V, IV of III op basis van de Cito LVS-toetsen Rekenen-Wiskunde. Het gaat dus niet alleen om leerlingen met grote leerachterstanden. De doelgroep is vervolgens benaderd met het verzoek om deel te nemen aan het HDT project van The Bridge. Ouders hebben voor deelname actief moeten tekenen. Uiteindelijk heeft dit wervingsproces voor de drie loopjaren een doelgroep opgeleverd van 105, 102 en 137 leerlingen. Het gaat hierbij om een duidelijke meerderheid van de in aanmerking komende leerlingen.

Vanuit de doelgroep is een deel toegewezen aan de interventiegroep en een deel aan de controlegroep. De willekeurige toewijzing is uitgevoerd op basis van een *block* randomisatieprocedure waarbij *sampling blocks* zijn aangemaakt op basis van de rekenscores in groep 6, in het schooljaar voorafgaand aan de HDT-interventie in groep 7. Onderzoek op basis van *block* randomisatie levert over het algemeen preciezer schattingen van de effecten op.

Af en toe weken de rekenscores in groep 6 sterk af van de rekenscores van andere leerlingen uit de doelgroep. Voor het goed functioneren van het onderzoeksdesign hebben we er in een aantal gevallen voor gekozen om leerlingen direct toe te wijzen aan de interventiegroep, of anderszins voor het onderzoek uit te sluiten. Voor de drie loopjaren zijn er uiteindelijk res-

pectievelijk 105, 98 en 126 leerlingen met een kans tussen 0 en 1 toegewezen aan de interventiegroep. Deze $105 + 98 + 126 = 329$ leerlingen zijn dus onderdeel van het gerandomiseerde onderzoek. Tabel 1 beschrijft per loopjaar het totaal aantal leerlingen dat is toegewezen aan de interventiegroep en aan de controlegroep.

Tabel 1: Aantallen leerlingen toegewezen aan de interventiegroep en de controlegroep, per loopjaar

	(1)	(2)
	interventiegroep	controlegroep
Loopjaar 1 (2019/20)	56	49
Loopjaar 2 (2020/21)	52	46
Loopjaar 3 (2021/22)	69	57

Scores op de Cito LVS-toetsen Rekenen-Wiskunde van verschillende meetmomenten zijn vaak sterk gecorreleerd. De scores halverwege groep 6 zijn bijvoorbeeld zeer voorspellend voor de scores aan het einde van groep 7. Kennis over deze correlaties is meegewogen in de powerberekeningen. Details over powerberekeningen en de daaraan gerelateerde *minimum detectable effects* (MDE's) zijn beschreven in het *social science registry* van de *American Economic Association*.¹³ Bij de registratie is alleen het onderzoek naar het eerste loopjaar beschreven en wordt uitgegaan van een MDE van 0.3 (effectgrootte). Met andere woorden, op basis van de powerberekeningen leveren effectgroottes van 0.3 in 80% van de gevallen statistisch significante resultaten op. Effectgroottes van 0.3 zijn op basis van relevante literatuur aannemelijk (zie bijvoorbeeld [7]). Bij de eerste twee loopjaren worden de opbrengsten van het voltijdsmodel onderzocht. Voorafgaand aan dit onderzoek zijn wij er vanuit gegaan dat we data van de eerste twee loopjaren konden combineren, om zo de opbrengsten preciezer (op basis van meer data) te kunnen vaststellen. De coronapandemie en de bijbehorende

¹³<https://www.socialscienceregistry.org/trials/4643>

lockdowns en schoolsluitingen hebben ervoor gezorgd dat de voordelen van het combineren van data van verschillende loopjaren niet langer duidelijk zijn.

We meten de opbrengsten van de interventie af aan de scores op de Cito-toetsen voor de leerlingvolgsystemen (LVS). De Cito LVS-vaardigheidsscores zijn vervolgens gestandaardiseerd ten opzichte van het landelijk gemiddelde en de landelijke spreiding van de betreffende leeftijdsgroep. Vervolgens gebruiken we de gestandaardiseerde scores als afhankelijke variabelen in regressieanalyses. Als onafhankelijke variabelen gebruiken we een indicator variabele voor de interventiegroep en een volledige set aan indicator variabelen voor de verschillende *sampling blocks*. In Appendix A lichten we een aantal technische details van de analyses verder toe.

4 Resultaten

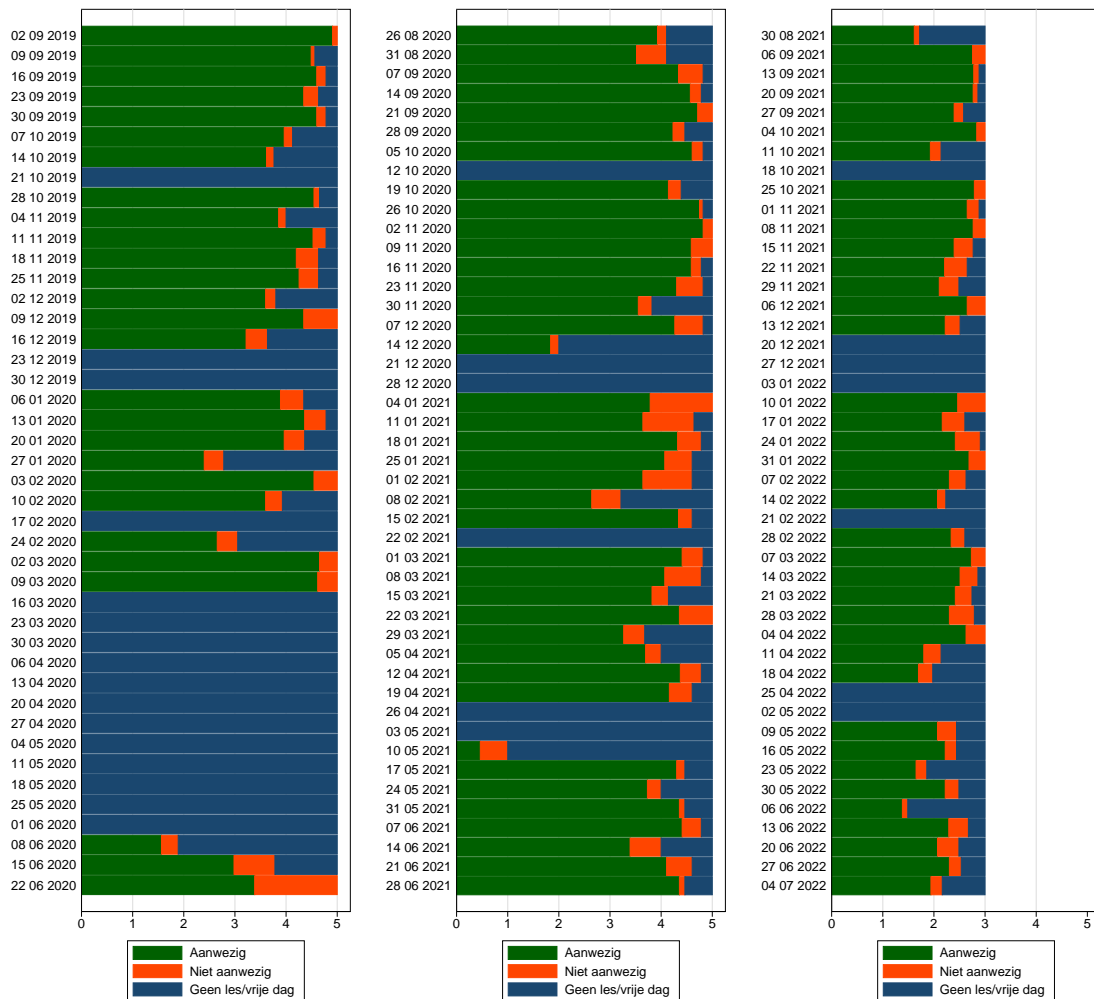
In dit hoofdstuk beschrijven we de resultaten van het gerandomiseerde onderzoek naar de opbrengsten van Bridge HDT in het basisonderwijs in Amsterdam Zuidoost. De interventie en het onderzoek hebben last gehad van de coronapandemie. Met name het eerste loopjaar (schooljaar 2019/20) is hard geraakt door schoolsluitingen etc. In het hoofdstuk 4.1 analyseren we de aanwezigheidsregistraties van The Bridge. Hieruit kan voor een deel worden opgemaakt of de interventie is uitgevoerd zoals vooraf was voorzien. In hoofdstuk 4.2 bespreken we de opbrengsten van het HDT programma, afgemeten aan de scores op de Cito LVS-toetses Rekenen-Wiskunde en Begrijpend Lezen.

4.1 Operationele bevindingen

Figuur 1 beschrijft de aanwezigheid van leerlingen per week en per loopjaar. Van links naar rechts wordt het eerste, tweede en derde loopjaar beschreven. Op basis van de linker figuur, concluderen we dat met name het tweede deel van het eerste loopjaar (2019/20) niet zoals gepland is uitgevoerd. In de figuur zien we duidelijk terug dat in de week van 16 maart 2020 de scholen zijn dichtgegaan. The Bridge heeft in deze fase een plan opgesteld om met week-

taken het HDT programma voort te zetten. Vanaf 13 april zijn tutors gaan videobellen met deelnemers, als aanvulling op de weektaken. Deze online sessies vonden echter een stuk minder frequent plaats dan het oorspronkelijk plan. In juni zijn op vier van de vijf scholen de tutorsessies opnieuw fysiek gestart.

Figuur 1: Aanwezigheid



Noten. De figuur presenteert het gemiddelde aantal sessies aanwezig, per week, per leerling. Van links naar rechts, refereren de figuren aan het eerste, tweede en derde loopjaar. In het eerste en tweede loopjaar zijn er maximaal vijf sessies per week, per leerling. In het derde loopjaar zijn er maximaal drie sessies per week, per leerling.

Het tweede en derde loopjaar hebben geen, of veel minder, last gehad van lockdowns. We zien dat de aanwezigheid bij het tweede en derde loopjaar vrijwel altijd op peil is gebleven. Hoewel tijdens het tweede loopjaar (tussen 14 december en 8 februari) de scholen ook dicht

waren, hebben leerlingen in deze periode de tutorsessies online kunnen volgen. The Bridge en de scholen waren tijdens deze tweede lockdown al veel beter voorbereid op het onderwijs op afstand. De leerlingen hebben dus, ziekte en andere omstandigheden daargelaten, min of meer zoals gepland kunnen deelnemen. Over de invulling van Bridge HDT tijdens de schoolsluitingen heeft The Bridge inhoudelijk verslag gedaan bij de Gemeente Amsterdam.

4.2 De opbrengsten van Bridge HDT

Tabel 2 beschrijft de geschatte opbrengsten van Bridge HDT, voor elk loopjaar afzonderlijk. De opbrengsten zijn uitgedrukt in populatie standaarddeviaties. We zien voor elk loopjaar duidelijke en statistisch significante effecten op de rekenprestaties van deelnemers. We trekken dus de conclusie dat Bridge HDT tot verbeterde rekenprestaties van deelnemers heeft geleid. Meer specifiek zien we dat de opbrengsten van het tweede loopjaar duidelijk substantiëler zijn dan die van het eerste loopjaar. Dit resultaat is mogelijk te verklaren door verschillen in de uitvoering. Net als bij het eerste loopjaar is bij het tweede loopjaar een voltijdsmodel geïmplementeerd. Het eerste loopjaar heeft echter in het tweede deel van het schooljaar veel last gehad van de coronapandemie. Het is aannemelijk dat dit een deel van de verschillen in de resultaten verklaart.

Tabel 2: Resultaten [A] Rekenen-Wiskunden en [B] Begrijpend Lezen. De geschatte opbrengsten zijn uitgedrukt in populatie standaarddeviates.

	(1) <i>midline</i> [M7]	(2) <i>endline</i> [E7]	(3) <i>followup</i> [M8]
A: Rekenen & Wiskunde			
Loopjaar 1	0.30*** (0.11) [99]	0.26* (0.14) [95]	0.13 (0.15) [90]
Loopjaar 2	0.48*** (0.16) [97]	0.62*** (0.15) [96]	0.39** (0.17) [93]
Loopjaar 3	0.28** (0.12) [119]	0.24* (0.14) [116]	
B: Begrijpend Lezen			
Loopjaar 1	-0.05 (0.16) [98]	-0.14 (0.18) [93]	-0.04 (0.16) [90]
Loopjaar 2	0.01 (0.22) [92]	-0.01 (0.21) [95]	0.28 (0.19) [88]
Loopjaar 3	-0.00 (0.17) [119]	0.07 (0.18) [104]	

Noten. Resultaten op basis van het regressiemodel beschreven in Appendix A. ***, **, * voor statistische significantie op het 1, 5, en 10% niveau. Standaardfouten tussen haakjes en steekproefgrootte tussen rechten haken.

We zien ook verschillen in de opbrengsten tussen het tweede en het derde loopjaar. Beide loopjaren zijn niet, of tenminste niet substantieel, door de coronapandemie beïnvloed. Het verschil in opbrengsten kan dus samenhangen met het verschil in de intensiteit van de modellen die zijn uitgeprobeerd. Grosso modo, lijken de opbrengsten enigszins in verhouding te staan tot de intensiteit van het gekozen model. Gerandomiseerd onderzoek naar een online tutorinterventie in Italië vindt ook dat opbrengsten in verhouding staan tot de intensiteit

van de interventie [3] (zie ook [10]).

In kolom 3 rapporteren we de resultaten van de *followup* meting. Deze *followup* meting is nog niet beschikbaar voor het derde loopjaar. We zien bij het eerste en tweede loopjaar, dat leerlingen halverwege groep 8 ongeveer de helft van hun voorsprong weer hebben ingeleverd. Deze bevinding is in lijn met een belangrijk deel van de literatuur. Bijvoorbeeld [4] concluderen op basis van een literatuurstudie dat “*the test score impacts of early educational interventions almost universally ‘fade out’ over time*”.

Voor het begrijpend lezen vinden we in Tabel 2B geen statistisch significante opbrengsten. De parameters in Tabel 2B zijn echter duidelijk minder precies geschat dan de parameters in Tabel 2A. Een verklaring hiervoor is dat de *sampling blocks* minder goede voorspellers zijn voor scores op de taaltoetsen dan voor scores op de rekentoetsen. In Appendix C laten we resultaten zien waarbij we corrigeren voor de rekenscores en taalscores bij *baseline*. Het corrigeren voor het niveau van begrijpend lezen bij *baseline* levert voor het begrijpend lezen preciezer schattingen van de opbrengsten op. Ook op basis van deze modellen zijn de bevindingen voor het begrijpend lezen over het algemeen niet (niet positief en niet negatief) statistisch significant. Dit resultaat komt overeen met eerdere bevindingen in een vergelijkbare context (zie bijvoorbeeld [8], [7] en [12]).

In Appendix C presenteren we ook resultaten op basis van de rekenscores, waarbij we corrigeren voor de scores op rekentoetsen die tijdens de *baseline* periode zijn gemaakt. Het controleren voor de *baseline* scores heeft, zoals verwacht, niet heel veel invloed op de resultaten. Echter, af en toe, vinden we wat stevigere effecten aan het eind van het eerste loopjaar (0.26-0.35) en af en toe iets minder stevige effecten aan het eind van het tweede loopjaar (0.48-0.62).

Samengenomen vinden we duidelijke effecten van Bridge HDT op de rekenprestaties van deelnemers. De verschillen in de opbrengsten tussen de verschillende loopjaren kan ook logisch worden verklaard aan de hand van de verschillen in de uitvoering. De opbrengsten van het tweede loopjaar zijn met name zeer substantieel. Effectgroottes in de orde van 0.5-0.6 worden bij gerandomiseerd onderzoek naar onderwijsinterventies maar zeer zelden gerap-

porteed [13].

Onderwijsprofessionals zijn echter niet altijd vertrouwd met het denken in termen van effectgroottes. Wat betekenen effectgroottes van 0.3 of 0.6 nu in de praktijk? We kunnen de effectgroottes omrekenen naar de veelgebruikte Didactische Leeftijdsequivalenten, oftewel, de DLE's. Een schooljaar duurt in Nederland ongeveer 10 maanden. In deze 10 maanden gaat de gemiddelde basisschoolleerling in Nederland dan ook, per definitie, 10 DLE punten vooruit. Hierbij correspondeert een extra DLE punt met de gemiddelde vaardigheidsgroei tijdens een maand onderwijs. In Appendix B rekenen we de gevonden effectgroottes om naar deze DLE punten en concluderen we dat de resultaten van het eerste en derde loopjaar zich verhouden tot ongeveer drie DLE punten en de resultaten van het tweede loopjaar tot ongeveer zes DLE punten. Met een gemiddelde groei van 10 DLE punten per jaar, zijn de drie en (met name de) zes extra DLE punten in een jaar op school zeer relevant. Het voltijdsmodel van het tweede loopjaar levert dus, in termen van de rekenprestaties, het equivalent van (meer dan) een half jaar extra schooltijd.

In Appendix C rapporteren we resultaten van verschillende valideitschecks en robuustheidsanalyses. We vinden geen aanwijzingen om te twijfelen aan de resultaten die we in de hoofdtekst beschrijven.

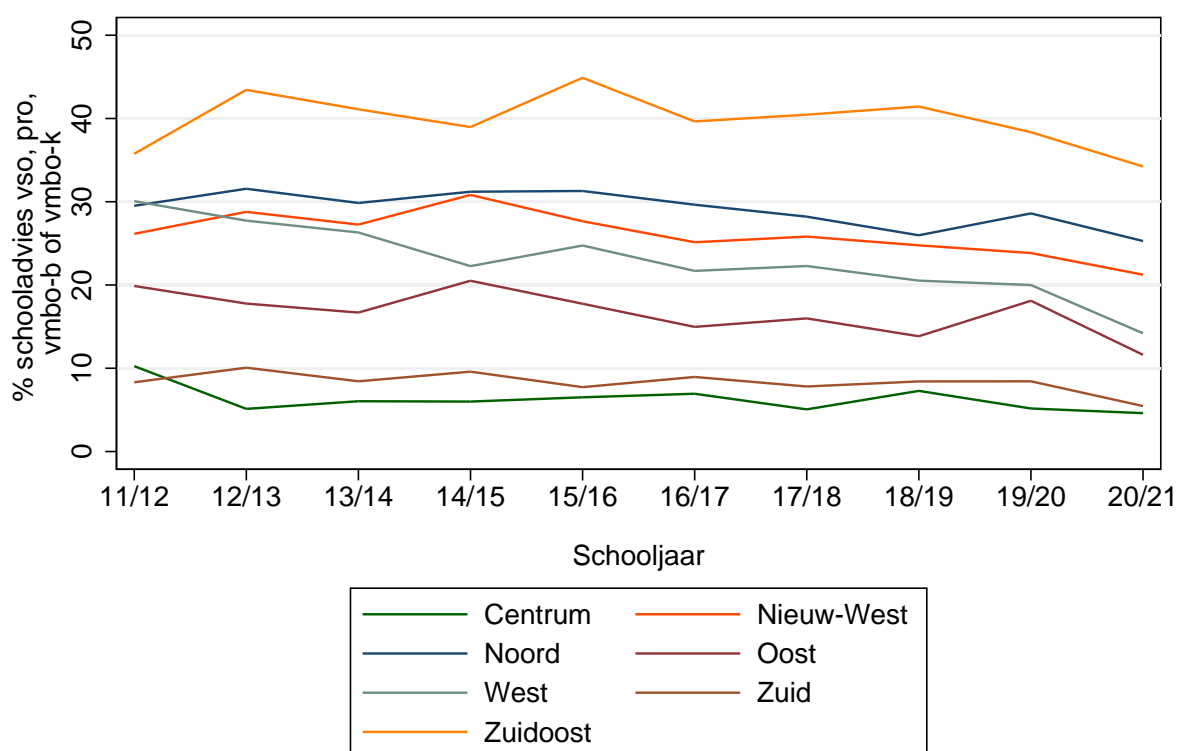
5 Opbrengsten van Bridge HDT in de context van onderwijs-achterstanden

Amsterdam Zuidoost is een stadsdeel in Amsterdam met een bovengemiddeld percentage huishoudens met lage tot zeer lage inkomens [11]. De schoolprestaties van leerlingen uit gezinnen met lage inkomens blijven vaak achter. In Figuur 2 zien we dit patroon terug. In de stadsdelen met meer lage inkomens zien we dat de schoolprestaties relatief achterblijven. Figuur 2 presenteert het percentage groep 8 leerlingen met een relatief laag¹⁴ middelbare

¹⁴We definiëren een laag schooladvies hier als een advies voor het voortgezet speciaal onderwijs, het praktijkonderwijs, vmbo-basisberoepsgerichte leerweg, en/of vmbo-kaderberoepsgerichte leerweg.

schooladvies. In het stadsdeel Amsterdam Zuidoost zijn deze lagere schooladviezen relatief oververtegenwoordigd. Ook zien we dat de verhoudingen tussen de stadsdelen relatief stabiel zijn. Bijvoorbeeld, in het stadsdeel Zuidoost zien we dat ongeveer 40% van de leerlingen de basisschool verlaat met een dergelijk advies, ten opzichte van ongeveer 10% in de stadsdelen Centrum en Zuid. Deze verschillen zijn in het afgelopen decennium niet of nauwelijks veranderd.

Figuur 2: Trends in advisering per stadsdeel



Noot: Eigen berekeningen op basis van data van de Dienst Uitvoering Onderwijs (DUO).¹⁵

De basisvaardigheden taal en rekenen zijn de belangrijkste basis voor het middelbare schooladvies. In dit rapport laten we zien dat de HDT interventie de rekenprestaties van deelnemers positief beïnvloedt. Ook hebben we voorgerekend dat de opbrengsten relatief substantieel zijn in verhouding tot de beschikbare literatuur. De resterende vraag is hoe de opbrengsten zich verhouden tot de bestaande onderwijsachterstanden in de stad. In welke mate, bijvoorbeeld, kan met het breder uitrollen van Bridge HDT, de relatieve achterstand

van basisschoolleerlingen in Amsterdam Zuidoost worden weggewerkt?

In Tabel 3 brengen we de relatieve achterstanden nogmaals in beeld. Voor de tabel maken we gebruik van de CBS Microdata. We gebruiken hiervoor de scores op het rekenonderdeel van de Cito eindtoets van alle groep 8 leerlingen in Nederland. De cijfers in Tabel 3 zijn berekend op basis de Cito eindtoets uit het schooljaar 2013/14. Deze resultaten zijn natuurlijk wat ouder, maar het is gegeven de stabiliteit in de advisering aannemelijk dat ze voldoende representatief zijn voor de huidige situatie.¹⁶

Vervolgens hebben we de scores op het rekenonderdeel van de Cito eindtoets omgeschaald naar niveaus. Hierbij staat het Niveau I voor rekenscores in de landelijk top 20%. Niveau V is een niveau dat past bij de laagst scorende 20% groep 8 leerlingen in Nederland. In de eerste kolom van Tabel 3 presenteren we de situatie in Zuidoost uit het jaar 2013/14, waarbij we aannemen dat deze resultaten redelijkerwijs representatief zijn voor de situatie van vandaag. We zien dat leerlingen met het Niveau V scores in Amsterdam Zuidoost zijn oververtegenwoordigd. In groep 8 scoort ongeveer 32% van de leerlingen in Amsterdam Zuidoost een Niveau V score, ten opzichte van 20% landelijk of ten opzichte van 21% in de rest van Amsterdam.¹⁷ Dergelijke rekenscores passen bij een middelbare schooladvies van vmbo-basis of vmbo-kaderberoepsgerichte leerweg.

¹⁶We kiezen voor het schooljaar 2013/14 als basis, omdat scholen vanaf 2014/15 over kunnen stappen naar andere toetsaanbieders en omdat de eindtoetsen een andere rol hebben gekregen in het adviseringsproces. Het is bijvoorbeeld na 2014/15 niet meer voor elke leerling van belang om goed te scoren op de eindtoets, omdat het voorlopige advies al ruim voor de eindtoets wordt vastgesteld.

¹⁷De bevindingen voor de rest van de stad laten we in dit rapport niet zien. Deze resultaten zijn op aanvraag beschikbaar.

Tabel 3: Rekenprestaties in het basisonderwijs in Amsterdam Zuidoost + projecties

	(1) Amsterdam Zuidoost	(2) Amsterdam Zuidoost +0.3SD	(3) Amsterdam Zuidoost +0.6SD
Niveau I	16	16	16
Niveau II	16	16	20
Niveau III	16	25	32
Niveau IV	19	19	17
Niveau V	32	24	16

Noten. Kolom 1 rapporteert de percentages leerlingen in Amsterdam Zuidoost in de verschillende categorieën (op basis van het rekenonderdeel van de Cito eindtoets uit 2013/14). Kolom 2 is een projectie, waarbij we de invloed simuleren van verbeterde rekenprestaties (+0.3SD in kolom 2 en +0.6SD in kolom 3) van leerlingen met initiële rekenscores op niveau IV en V. In de simulaties zien we dat een deel van de leerlingen doorschuift naar hogere niveaus.

In kolom 2 van Tabel 3 rapporteren we projecties, waarbij we onderzoeken in welke mate de oververtegenwoordiging van Niveau V scores blijft bestaan wanneer, bijvoorbeeld, Bridge HDT succesvol zou kunnen worden uitgerold op alle basisscholen in Amsterdam Zuidoost. Hiervoor rekenen we met opbrengsten van +0.3 (kolom 2) en +0.6 (kolom 3) populatie standaarddeviaties. We nemen aan dat bij deze geprojecteerde uitrol van Bridge HDT, alle leerlingen in Amsterdam Zuidoost, met Niveau V en IV scores van de uitrol profiteren. In de projecties zien we substantiële verschuivingen in de percentages. Met effectgroottes van +0.3, de opbrengsten die we voor het eerste en derde loopjaar hebben kunnen vaststellen, zien we dat een groot deel van de oververtegenwoordiging in Niveau V kan worden weggewerkt. Met de zeer substantiële effectgroottes van +0.6 kan de oververtegenwoordiging van Niveau V scores zelfs (meer dan) volledig worden geëlimineerd.¹⁸

De projecties laten zien dat met effectgroottes in de orde van 0.3-0.6, de relatieve on-

¹⁸De achterstanden in Amsterdam Zuidoost manifesteren zich niet alleen in termen van een overrepresentatie van Niveau V scores, maar ook in termen van een onderrepresentatie van Niveau I scores. Het inzetten van Bridge HDT ter ondersteuning van de (huidige) niveau V en IV leerlingen, heeft in de projecties geen invloed op de onderrepresentatie van Niveau I scores in het stadsdeel.

derwijsachterstanden op een zinvolle manier kunnen worden weggewerkt. Bij een verdere uitrol van Bridge HDT is het van belang om het niveau van de implementatie tijdens de uitrol op peil te houden. Dit is geen sinecure. Recente literatuur spreekt over de mogelijkheid van een *voltage drop*, het idee dat “*treatment effect sizes observed in research studies diminish substantially when the program is rolled out at larger scale*” [1]. Bij een verdere uitrol moet hiernaast ook worden bedacht dat na afloop van de interventie maar een deel van de opbrengsten beklijven. Het lijkt dus zaak om ook na te denken over onderhoud, waarbij de doelgroep wellicht gedurende lagere tijd zal moeten worden ondersteund. Ook zou eventueel de aansluiting met het reguliere programma op school kunnen worden verbeterd. Meer onderzoek in dit domein is hiervoor nodig. Het wegwerken van achterstanden is dus geen eenvoudig project. Echter, dit onderzoek laat zien dat er kansen en mogelijkheden zijn. Bridge HDT kan onderdeel zijn van een oplossing voor het bestrijden van onderwijsachterstanden in de stad (en elders).

6 Conclusies en aanbevelingen

Dit onderzoek laat zien dat met Bridge HDT bruikbare leerwinsten kunnen worden geboekt. Op basis van de in dit rapport geëvalueerde toepassingen van Bridge HDT, schatten we effecten op de rekenprestaties in de orde van 3 tot 6 maanden extra schooltijd. Hierbij lijken de opbrengsten samen te hangen met de intensiteit van de HDT interventie. We verwachten dat het breed opschalen van Bridge HDT in het basisonderwijs in Amsterdam Zuidoost kan leiden tot het volledig wegwerken van relevante onderwijsachterstanden.

De Gemeente Amsterdam laat in de begroting van 2023 weten “programma’s als High-Dosage Tutoring voort” te gaan zetten.¹⁹ Bij verdere opschaling is het van belang om de kernonderdelen van de HDT interventie te behouden. Bij de opschaling kan het bijvoorbeeld een uitdaging zijn om het juiste personeel aan te nemen, waarmee vergelijkbare successen kunnen worden geboekt. Hierbij speelt natuurlijk de krapte op de huidige arbeidsmarkt een

¹⁹<https://www.amsterdam.nl/bestuur-organisatie/financien/begroting-2023/>

belangrijke rol. Ook kunnen de relatief hoge kosten van HDT remmend werken.²⁰ In de wereld van het beleidsonderzoek wordt het succesvol opschalen van veelbelovende pilot-interventies gezien als een belangrijke uitdaging voor de toekomst (zie bijvoorbeeld [6], [1] en [16] die we al eerder aanhaalden).

Het is in onze ogen belangrijk om beleidsdoelen in termen van uitkomsten te formuleren en daarop, vervolgens, te sturen. Een ambitieus (maar gezien dit onderzoek redelijk) beleidsdoel zou kunnen zijn om de overrepresentatie aan laagpresteerders in Amsterdam Zuidoost in vijf jaar weg te werken. Vergelijkbare doelen kunnen worden opgesteld voor andere stadsdelen met onderwijsachterstanden, zoals de stadsdelen Noord of Nieuw-West.

Om op basis van de gestelde doelen te kunnen sturen is het belangrijk om actief te monitoren. Hiermee kan het een en ander in de loop van de tijd worden bijgestuurd, als de verwachte opbrengsten uitblijven. Bij de monitoring kunnen moderne kwantitatieve technieken worden ingezet. Hierbij lijkt het zinvol om, bijvoorbeeld, cohorten en scholen waar HDT verder wordt opgeschaald, te vergelijken met cohorten en scholen waar HDT niet (of nog niet) wordt opgeschaald. Bij dergelijke evaluaties kan, onder voorwaarden, met bestaande data-infrastructuren worden gewerkt zoals bijvoorbeeld het Nationaal Cohortonderzoek²¹ en/of de CBS Microdata. Door van deze infrastructuur gebruik te maken, kan de onderzoekslast bij scholen worden geminimaliseerd.

²⁰Het voltijdsmodel van The Bridge kost ongeveer €3.000 per deelnemer per schooljaar.

²¹<https://www.nationaalcohortonderzoek.nl/>

Literatuur

- [1] Al-Ubaydli, O., J. A. List, and D. Suskind (2019). The science of using science: Towards an understanding of the threats to scaling experiments. *NBER Working Paper Series* (25848).
- [2] Bailey, D. H., G. J. Duncan, F. Cunha, B. R. Foorman, and D. S. Yeager (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest* 21(2).
- [3] Carlana, M. and E. La Ferrara (2021). Apart but connected: Online tutoring and student outcomes during the covid-19 pandemic. *HKS Faculty Research Working Paper Series* 21(001).
- [4] Cascio, E. U. and D. O. Staiger (2012). Knowledge, tests, and fadeout in educational interventions. *NBER* (18038).
- [5] Cito (2015). Wetenschappelijke verantwoording van de lvs-toetsen rekenen-wiskunde tweede generatie: Addendum hernormering september 2013. *Report*.
- [6] Davis, J. M., J. Guryan, K. Hallberg, and J. Ludwig (2017). The economics of scale-up. *NBER working paper* (23925).
- [7] De Ree, J., M. A. Maggioni, B. Paulle, D. Rossignoli, N. Ruijs, and D. Walentek (2021). Closing the income-achievement gap? experimental evidence from high-dosage tutoring in dutch primary education. *SocArXiv*.
- [8] De Ree, J., M. A. Maggioni, B. Paulle, D. Rossignoli, and D. Walentek (2021). High dosage tutoring in pre-vocational secondary education: Experimental evidence from amsterdam. *SocArXiv*.
- [9] Engzell, P., A. Frey, and M. Verhagen (2021). Learning loss due to school closures during the covid-19 pandemic. *Proceedings of the National Academy of Sciences*.

- [10] Fryer Jr., R. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In A. V. Banerjee and E. Duflo (Eds.), *Handbook of Economic Field Experiments*, Chapter 2.
- [11] Gemeente Amsterdam (2021). Amsterdamse armoedemonitor 2021. *Onderzoek, Informatie en Statistiek*.
- [12] Guryan, J., J. Ludwig, M. P. Bhatt, P. J. Cook, J. M. Davis, K. Dodge, G. Farkas, R. G. Fryer Jr., S. Mayer, H. Pollack, L. Steinberg, and G. Stoddard (forthcoming). Not too late: Improving academic outcomes among adolescents. *American Economic Review*.
- [13] Kraft, M. A. (2015). How to make additional time matter: Integrating individualized tutorials into an extended day. *Education Finance and Policy* 10(1).
- [14] Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher* 49(4), 241–253.
- [15] Kraft, M. A. and G. T. Falken (2021). A blueprint for scaling tutoring and mentoring across public schools. *AERA Open* 7(1).
- [16] List, J. (2022). *The Voltage Effect*. Penguin Books Limited.

A Methodologische details

Als afhankelijke variabele in de regressieanalyses gebruiken we gestandaardiseerde Cito LVS-scores voor Rekenen-Wiskunde en Begrijpend Lezen. De parameters worden geschat op basis van de volgende regressievergelijking:

$$Y_{it} = \sum_{\tau} \beta_{0\tau} 1(\tau = t) + \sum_{\tau} \beta_{1\tau} 1(\tau = t, \text{Treatment}_i = 1) + \sum_k 1(\text{block}_{ki}) + e_{it} \quad (1)$$

De subscripten i en t refereren respectievelijk aan de leerling en aan het meetmoment. Er zijn maximaal drie meetmomenten, de *midline* (halverwege groep 7 en halverwege de interventieperiode), de *endline* (aan het eind van groep 7 en aan het eind van de interventieperiode) en de *followup* (halverwege groep 8 en ongeveer een half schooljaar na afloop van de interventie). Door weging wordt gecorrigeerd voor ongelijke kansen op toewijzing aan de interventiegroep en de controlegroep. Gewichten voor observaties in de interventiegroep zijn $1/p$. Gewichten voor observaties in de controlegroep zijn $1/(1-p)$. Hierbij is p de kans op toewijzing aan de interventiegroep.

In robuustheidsanalyses nemen we ook verschillende andere toetsmetingen uit de *baseline* periode mee. De *baseline* periode is de periode voorafgaand aan de start van de interventie. Het gaat hierbij om scores uit groep 5 of groep 6. Het is de verwachting dat dit voor de resultaten op basis van de rekentoetsen niet veel zal uitmaken, omdat de *baseline* scores al expliciet zijn meegenomen bij de *block* randomisatieprocedure. Met andere woorden, de voorspellende waarde van *baseline* scores is bij de rekentoetsen in verwachting minimaal, nadat ook voor de *sampling blocks* is gecontroleerd. Voor de resultaten op basis van de toetsen voor begrijpend lezen kan het controleren voor de *baseline* scores eventueel preciezere schattingen opleveren.

Standaardisatie van toetsscores. De vaardigheidsscores kunnen worden afgezet tegen landelijke spreiding van de relevante leeftijdsgroep. Met andere woorden, het is mogelijk om scores te beoordelen in vergelijking met scores van leeftijdsgenoten in de rest van Ne-

derland. In de analyses gebruiken we deze informatie bij het standaardiseren van de scores. Als eerste schatten we het populatiegemiddelde door het 20e en 80e percentiel te middelen. Vervolgens schatten we de populatiestandaarddeviatie door het verschil tussen het 20e en 80e percentiel te delen door 1.68. In de normale verdeling zijn het 20e en 80e percentiel ongeveer 1.68 standaarddeviaties van elkaar verwijderd. Deze benadering met betrekking tot het standaardiseren vereist dat de scores (ongeveer) normaal verdeeld zijn in de populatie. Deze aanname is op basis van de beschikbare documentatie aannemelijk [5].

B Opbrengsten uitgedrukt in Didactische Leeftijdsequivalenten (DLE's)

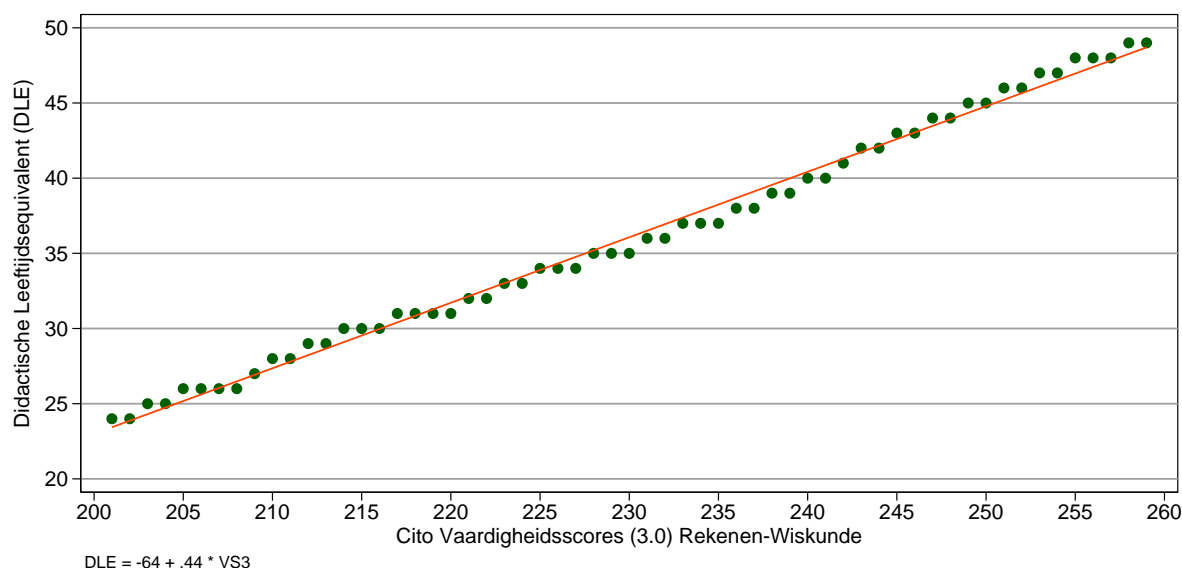
In Tabel 4 worden de opbrengsten van het HDT programma uitgedrukt in LVS vaardigheidsscores (generatie 3.0). Deze schaling/normalisatie heeft geen duidelijke betekenis. We zien dat deelname aan Bridge HDT bij deelnemers in het eerste en derde loopjaar ongeveer zeven vaardigheidspunten oplevert. Bij het tweede loopjaar, levert HDT ongeveer 14 extra vaardigheidspunten op.

Tabel 4: Resultaten [A] Rekenen & Wiskunden en [B] Begrijpend Lezen. De geschatte opbrengsten zijn uitgedrukt in Cito LVS Vaardigheidsscores (generatie 3.0).

	(1) <i>midline</i> [M7]	(2) <i>endline</i> [E7]	(3) <i>followup</i> [M8]
A: Rekenen & Wiskunde			
Loopjaar 1	7.38*** (2.69) [99]	6.09* (3.33) [95]	3.10 (3.47) [90]
Loopjaar 2	11.70*** (3.76) [97]	14.31*** (3.59) [96]	9.00** (3.86) [93]
Loopjaar 3	7.01** (2.88) [119]	5.68* (3.19) [116]	
B: Begrijpend Lezen			
Loopjaar 1	-1.15 (3.90) [98]	-3.16 (4.13) [93]	-0.85 (3.91) [90]
Loopjaar 2	0.16 (5.14) [92]	-0.37 (4.82) [95]	6.78 (4.59) [88]
Loopjaar 3	-0.08 (4.05) [119]	1.67 (4.22) [104]	

Noten. ***, **, * voor statische significantie op het 1, 5, en 10% niveau. Standaardfouten tussen haakjes en steekproefgrootte tussen rechte haken.

Figuur 3: Een lineair verband tussen de Didactische Leeftijdsequivalenten (DLE's) en de Cito-LVS Vaardigheidsscores (generatie 3.0)



Note: De figuur onderzoekt het verband tussen de DLE's en de vaardigheidsscores op een voor dit onderzoek relevant domein. Vaardigheidsscores die we in de data terug zien liggen met name in het interval 200-260.

Figuur 3 beschrijft de relatie tussen de LVS vaardigheidsscores op de horizontale as en de Didactische Leeftijdsequivalenten op de verticale as. De relatie tussen de beide eenheden is bij benadering lineair. Op dit (relevante) domein, komt elk vaardigheidspunt, ongeveer overeen met 0.44 DLE punten. Met andere woorden, een opbrengst van zeven vaardigheids-punten staat komen overeen met $0.44 \times 7 \approx 3$ DLE punten. De opbrengsten gemeten in het tweede loopjaar komen overeen met $0.44 \times 14 \approx 6$ DLE punten.

C Robuustheidsanalyses en validiteit

Gedurende de looptijd van het project, en met name in de laatste fase in de aanloop naar deze eindrapportage, zijn nieuwe toetsresultaten verzameld. Het gaat hierbij ook om scores op toetsen die door leerlingen in de *baseline* periode (voorafgaand aan de interventie) zijn gemaakt. In de Tabellen 5 en 6 onderzoeken we de robuustheid van de in de hoofdstekst

gerapporteerde effecten met betrekking tot het corrigeren voor deze *baseline* toetscores. In de kolommen (1-2) van Tabel 5A repliceren we de resultaten van Tabel 2A in de hoofdtekst. In de kolommen(1-2) van Tabel 6A repliceren we de resultaten van Tabel 2B in de hoofdtekst.

In de kolommen (3-10) corrigeren we in de regressieanalyses achtereenvolgens voor rekenscores (in Tabel 5) en voor scores op de toetsen voor begrijpend lezen (in Tabel 6) die zijn behaald aan het eind van groep 5 (E5), halverwege groep 6 (M6), aan het eind van groep 6 (E6) en voor de drie scores samen (E5, M6 en E6). In Tabel 5A gebruiken we de *sampling block fixed effects* regressies. De resultaten in Tabel 5B zijn op basis van regressieanalyses waarbij de *sampling block fixed effects* niet zijn meegenomen.

De resultaten van de verschillende modellen zijn redelijk vergelijkbaar, zoals vooraf kon worden verwacht. Voor resultaten op basis van de rekenscores zien we dat het controleren voor de *baseline*, voor loopjaar 1 soms tot licht sterkere effecten leidt (kolom 3-10). Voor loopjaar 2 zien we licht zwakkere effecten in sommige van de kolommen (3-10). Voor loopjaar 3 zien we in de kolommen (3-10) ook iets zwakkere effecten in sommige specificaties. De resultaten in Tabel 5A en 5B zijn vergelijkbaar.

Voor het Begrijpend Lezen in Tabel 6 zien we significant negatieve parameters in één specifiek geval. Dit ene resultaat is onvoldoende om er duidelijke conclusies aan te verbinden. Over het algemeen zien we dat de effecten op de prestaties in het begrijpend lezen niet significant verschillen van nul. Hierbij merken we op dat de parameters niet erg precies zijn geschat. Het corrigeren voor de *baseline* heeft maar beperkt invloed.

Tabel 5: Resultaten Rekenen-Wiskunde [A] op basis van *sampling block fixed effects* regressies en [B] op basis van reguliere regressies. De geschatte opbrengsten zijn uitgedrukt in populatie standaarddeviaties.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	no controls		E5 controls		M6 controls		E6 controls		E5,M6,E6 controls	
A: FE										
Loopjaar 1	0.30*** (0.11)	0.26* (0.14)	0.33*** (0.11)	0.28** (0.14)	0.31*** (0.11)	0.27* (0.14)	0.39*** (0.10)	0.35*** (0.12)	0.40*** (0.10)	0.35*** (0.13)
	[99]	[95]	[99]	[95]	[99]	[95]	[99]	[95]	[99]	[95]
Loopjaar 2	0.48*** (0.16)	0.62*** (0.15)	0.36*** (0.12)	0.48*** (0.15)	0.45*** (0.14)	0.57*** (0.15)	0.50*** (0.15)	0.62*** (0.14)	0.38*** (0.12)	0.50*** (0.14)
	[97]	[96]	[97]	[96]	[97]	[96]	[97]	[96]	[97]	[96]
Loopjaar 3	0.28** (0.12)	0.24* (0.14)	0.29** (0.12)	0.25* (0.14)	0.23** (0.11)	0.19 (0.13)	0.23** (0.10)	0.20* (0.12)	0.20** (0.10)	0.19 (0.12)
	[119]	[116]	[119]	[116]	[119]	[116]	[119]	[116]	[119]	[116]
B: OLS										
Loopjaar 1	0.31 (0.23)	0.29 (0.27)	0.37** (0.14)	0.32* (0.18)	0.30** (0.14)	0.27 (0.17)	0.39*** (0.10)	0.37** (0.15)	0.39*** (0.10)	0.36** (0.15)
	[99]	[95]	[99]	[95]	[99]	[95]	[99]	[95]	[99]	[95]
Loopjaar 2	0.48* (0.25)	0.64** (0.27)	0.33** (0.16)	0.45** (0.18)	0.45** (0.18)	0.57*** (0.19)	0.50*** (0.17)	0.61*** (0.18)	0.40*** (0.14)	0.53*** (0.17)
	[97]	[96]	[97]	[96]	[97]	[96]	[97]	[96]	[97]	[96]
Loopjaar 3	0.37 (0.23)	0.20 (0.27)	0.37*** (0.14)	0.34** (0.16)	0.21* (0.11)	0.19 (0.14)	0.22* (0.11)	0.17 (0.16)	0.22** (0.11)	0.23* (0.13)
	[119]	[116]	[119]	[116]	[119]	[116]	[119]	[116]	[119]	[116]

Noten. ***, **, * voor statistische significantie op het 1, 5, en 10% niveau.

Tabel 6: Resultaten Begrijpend Lezen [A] op basis van *sampling block fixed effects* regressies en [B] op basis van reguliere regressies. De geschatte opbrengsten zijn uitgedrukt in populatie standaarddeviaties.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	no controls		E5 controls		M6 controls		E6 controls		E5,M6,E6 controls	
A: FE										
Loopjaar 1	-0.05 (0.16)	-0.14 (0.18)	-0.01 (0.14)	-0.12 (0.16)	0.01 (0.15)	-0.11 (0.17)	0.03 (0.14)	-0.07 (0.15)	0.02 (0.13)	-0.08 (0.15)
	[98]	[93]	[97]	[91]	[97]	[91]	[97]	[91]	[97]	[91]
Loopjaar 2	0.01 (0.22)	-0.01 (0.21)	-0.31** (0.15)	-0.33* (0.18)	0.19 (0.16)	0.11 (0.16)	0.08 (0.15)	-0.00 (0.16)	-0.09 (0.15)	-0.10 (0.17)
	[92]	[95]	[92]	[95]	[92]	[95]	[92]	[95]	[92]	[95]
Loopjaar 3	-0.00 (0.17)	0.07 (0.18)	0.00 (0.14)	0.04 (0.14)	-0.12 (0.12)	-0.12 (0.14)	0.02 (0.12)	0.09 (0.14)	-0.03 (0.12)	-0.01 (0.13)
	[119]	[104]	[119]	[104]	[119]	[104]	[119]	[104]	[119]	[104]
B: OLS										
Loopjaar 1	-0.07 (0.23)	-0.11 (0.26)	-0.03 (0.15)	-0.15 (0.17)	0.00 (0.16)	-0.12 (0.19)	0.00 (0.15)	-0.11 (0.18)	-0.00 (0.14)	-0.12 (0.17)
	[98]	[93]	[97]	[91]	[97]	[91]	[97]	[91]	[97]	[91]
Loopjaar 2	0.04 (0.28)	0.05 (0.26)	-0.27 (0.19)	-0.27 (0.22)	0.22 (0.19)	0.14 (0.20)	0.09 (0.18)	0.02 (0.18)	-0.05 (0.17)	-0.04 (0.19)
	[92]	[95]	[92]	[95]	[92]	[95]	[92]	[95]	[92]	[95]
Loopjaar 3	0.00 (0.22)	0.07 (0.24)	-0.02 (0.16)	0.03 (0.18)	-0.15 (0.14)	-0.16 (0.16)	-0.02 (0.14)	0.05 (0.18)	-0.10 (0.13)	-0.08 (0.16)
	[119]	[104]	[119]	[104]	[119]	[104]	[119]	[104]	[119]	[104]

Noten. ***, **, * voor statistische significantie op het 1, 5, en 10% niveau.

In de Tabellen 7A en 8A onderzoeken we of leerlingen in voldoende mate de toetsen hebben gemaakt. Over het algemeen zien we een vrij hoge respons en weinig verschillen tussen de interventiegroep en de controlegroep. Op basis hiervan is er dus geen reden om aan de validiteit van het onderzoek te twijfelen. In de Tabellen 7B en 8B kijken we naar de verschillen in de scores bij de verschillende meetmomenten. In de *baseline* periode, vanaf halverwege groep 5 (M5) tot eind groep 6 (E6) zien we geen statistisch significante verschillen. Dit is zoals verwacht op basis van de willekeurige toewijzing van leerlingen. In de fase vanaf de start van de interventie in groep 7, zien we de verschillen in scores ontstaan.

Tabel 7: [A] Fractie geobserveerde toetsen Rekenen-Wiskunde, [B] Gestandaardiseerde Cito-LVS scores Rekenen-Wiskunde

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Loopjaar 1			Loopjaar 2 2			Loopjaar 3					
	T	C	diff	diff FE	T	C	diff	diff FE	T	C	diff	diff FE
A: Observed												
M5	0.78	0.76	0.01	0.01	0.92	0.90	0.02	0.02	0.95	0.94	0.01	0.01
E5	0.93	0.87	0.07	0.07	0.92	0.95	-0.04	-0.04	0.85	0.72	0.12	0.12***
M6	0.96	0.96	0.00	0.00	1.00	1.00	0.00	0.00	0.99	1.00	-0.01	-0.01
E6	0.98	0.96	0.02	0.02	0.84	0.84	0.00	0.00	0.96	0.99	-0.03	-0.03
M7	0.97	0.92	0.04	0.04	1.00	0.97	0.03	0.03	0.93	0.95	-0.02	-0.02
E7	0.93	0.89	0.04	0.04	0.98	0.97	0.01	0.01	0.92	0.92	0.01	0.01
M8	0.91	0.79	0.12*	0.12*	0.92	0.97	-0.05	-0.05	0.00	0.00	0.00	0.00
B: scores												
M5	-0.61	-0.77	0.16	0.06	-1.11	-1.24	0.13	0.18	-1.33	-1.30	-0.03	-0.02
E5	-0.95	-1.06	0.11	0.07	-1.21	-1.45	0.23	0.23	-1.53	-1.29	-0.23	-0.15
M6	-1.11	-1.13	0.03	0.01	-1.22	-1.24	0.03	0.03	-1.23	-1.30	0.07	0.07
E6	-1.29	-1.27	-0.02	-0.03	-1.48	-1.47	-0.01	-0.01	-1.22	-1.33	0.11	0.11
M7	-1.05	-1.36	0.31	0.29**	-0.83	-1.31	0.48*	0.49***	-0.89	-1.26	0.37	0.30**
E7	-1.12	-1.41	0.29	0.30*	-0.71	-1.35	0.64**	0.63***	-1.02	-1.22	0.20	0.25
M8	-1.05	-1.16	0.11	0.13	-1.20	-1.60	0.40	0.31*				

Noten. De tabel rapporteert voor de drie loopjaren, de (gewogen) gemiddeldes in de interventiegroep (T) en in de controlegroep (C), het verschil (diff) en het verschil waarbij voor de *sampling blocks* is gecorrigeerd (diff FE). ***, **, * voor statistische significantie op het 1, 5, en 10% niveau. De schattingen in deel B wijken soms iets af van de resultaten die in Tabel 2 zijn gerapporteerd. De reden hiervoor is dat voor de resultaten in de hoofdtekst niet voor elk meetmoment afzonderlijk is gecorrigeerd voor de *sampling blocks*. De resultaten in deze tabel kunnen dus ook worden gezien als robuustheidsanalyse.

Tabel 8: [A] Fractie geobserveerde toetsen Begrijpend Lezen, [B] Gestandaardiseerde Cito-IVS scores Begrijpend Lezen

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Loopjaar 1			Loopjaar 2 2			Loopjaar 3					
	T	C	diff	diff FE	T	C	diff	diff FE	T	C	diff	diff FE
A: Observed												
M5	0.73	0.69	0.05	0.05	0.94	0.92	0.02	0.02	0.96	0.93	0.03	0.03
E5	0.38	0.33	0.04	0.04	0.94	0.91	0.03	0.03	0.85	0.79	0.06	0.06
M6	0.94	0.96	-0.02	-0.02	1.00	1.00	0.00	0.00	1.00	0.98	0.02	0.02
E6	0.80	0.78	0.02	0.02	0.84	0.84	0.00	0.00	0.99	0.98	0.00	0.00
M7	0.97	0.90	0.06	0.06	0.98	0.87	0.12**	0.12**	0.95	0.94	0.01	0.01
E7	0.90	0.89	0.01	0.01	0.96	0.97	-0.01	-0.01	0.81	0.86	-0.05	-0.05
M8	0.90	0.82	0.08	0.08	0.87	0.91	-0.04	-0.04	0.00	0.00	0.00	0.00
B: Score												
M5	-0.49	-0.61	0.12	-0.03	-0.69	-0.99	0.30	0.32	-0.81	-0.65	-0.16	-0.14
E5	-0.74	-0.79	0.05	-0.04	-0.85	-1.20	0.35	0.28	-1.06	-1.03	-0.04	-0.02
M6	-1.10	-1.10	0.00	-0.07	-1.13	-1.03	-0.10	-0.10	-1.00	-1.11	0.11	0.09
E6	-1.19	-1.09	-0.10	-0.07	-1.16	-1.04	-0.13	-0.13	-1.00	-1.00	-0.00	-0.02
M7	-1.19	-1.12	-0.07	-0.09	-1.06	-1.10	0.04	-0.05	-1.23	-1.23	0.00	-0.02
E7	-1.18	-1.07	-0.11	-0.12	-0.89	-0.94	0.05	-0.01	-1.15	-1.22	0.07	0.10
M8	-1.07	-1.12	0.05	0.03	-0.98	-1.37	0.39*	0.24				

Noten. De tabel rapporteert voor de drie loopjaren, de (gewogen) gemiddeldes in de interventiegroep (T) en in de controlegroep (C), het verschil (diff) en het verschil waarbij voor de *sampling blocks* is gecorrigeerd (diff FE). ***, **, * voor statistische significantie op het 1, 5, en 10% niveau. De schattingen in deel B wijken soms iets af van de resultaten die in Tabel 2 zijn gerapporteerd. De reden hiervoor is dat voor de resultaten in de hoofdtekst niet voor elk meetmoment afzonderlijk is gecorrigeerd voor de *sampling blocks*. De resultaten in deze tabel kunnen dus ook worden gezien als robuustheidsanalyse.