

Leer je beter *Lezen met de BliksemBende?*

Resultaten van een effectonderzoek



ALEXANDER KREPEL

ILLA CARRIÓN BRAAKMAN

CIP-gegevens KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Krepel, A. & Carrión Braakman, I.
Beter leren Lezen met de BliksemBende
Resultaten van een effectonderzoek
Amstelveen: Kohnstamm Instituut.
(Rapport 1149, projectnummer 20896)

ISBN 978-94-6321-211-3

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher.

Dit onderzoek had niet tot stand kunnen komen zonder de inzet van het interventieteam van de Rijksuniversiteit Groningen met coördinatie van Mariëtte Hingstman en alle schoolbegeleiders. Wij danken daarnaast alle deelnemende scholen, leraren, tutors en leerlingen voor hun betrokkenheid, en de scriptiestudenten en studentassistenten voor hun bijdrage aan de dataverzameling. Het inhoudelijke advies van Femke Scheltinga (ITTA) en de hulp van Emiel Schoneveld (Universiteit van Amsterdam) bij de bootstrappinganalyses hebben dit onderzoek verder versterkt.

Uitgave en verspreiding:
Kohnstamm Instituut
Keizer Karelplein 1, Postbus 94208, 1090 GE Amsterdam
Tel.: 020-214 1400
www.kohnstammstituut.nl
Dataverwerking: Elion.nl
© Copyright Kohnstamm Instituut, 2025



Dit project is (mede) gefinancierd door het NRO (Projectnummer 405-00-860-156)

Inhoudsopgave

Publiekssamenvatting	1
1 Inleiding	2
1.1 Aanwijzingen voor effectiviteit	3
1.2 Voor wie werkt het?	4
1.3 Onderzoeksvragen	5
1.4 Leeswijzer	6
2 De interventie: Lezen met de BliksemBende	7
2.1 Doel/doelgroep van de interventie	7
2.2 Opzet van de interventie	7
2.3 Inhoud van de interventie	8
2.4 Begeleiding van scholen tijdens de interventie	10
2.5 Pilotonderzoek	11
3 Methode	12
3.1 Werving en selectie van scholen en leerlingen	12
3.2 Onderzoeksdesign	16
3.3 Meetinstrumenten	17
3.4 Analysestrategie	22
4 Resultaten	25
4.1 Implementatietrouw en verloop interventie	25
4.2 Effectevaluatie	29
4.3 Welke interventie- en oefenkenmerken zijn geassocieerd met de meeste vooruitgang?	37
4.4 Welke kind- en groepskenmerken zijn geassocieerd met de meeste vooruitgang?	40
4.5 Procesevaluatie interventieperiode	41
5 Discussie en conclusie	46
5.1 Wat levert de BliksemBende op?	46
5.2 Wanneer is de interventie het meest effectief?	48
5.3 Werkt de BliksemBende voor elke leerling even goed?	49
5.4 Procesevaluatie	50
5.5 Beperkingen	51
5.6 Implicaties voor verdere ontwikkeling en de praktijk	52
5.7 Conclusie	52
Referenties	54
Bijlage A Beschrijvende statistieken steekproef	57
Bijlage B Resultaten regressieanalyses	60
Bijlage C Behaalde power	65

Publiekssamenvatting

Internationaal onderzoek laat zien dat tutorprogramma's zoals *Tutoring with the Lightning Squad* veelbelovend zijn voor het verkleinen van leesachterstanden. De Nederlandse versie, *Lezen met de BliksemBende*, is in 2024–2025 in een grootschalig RCT-onderzoek getest onder zwakke lezers in groep 3 en 4.

Wat is Lezen met de BliksemBende?

De BliksemBende is een digitaal tutorprogramma, ontwikkeld door de Rijksuniversiteit Groningen. Leerlingen oefenen vier tot vijf keer per week ongeveer 25 minuten in duo's, begeleid door een tutor. Tijdens de sessies werken zij aan technisch lezen, spelling, woordenschat en begrijpend lezen.

Onderzoek

In totaal deden 23 basisscholen mee, met 77 klassen en 340 leerlingen met een leesachterstand. De leerlingen werkten 12 weken met de BliksemBende of kregen reguliere leesondersteuning van hun school. Naast scores uit het leerlingvolgsysteem zijn logboeken en programmadata bijgehouden om het gebruik van de interventie te monitoren.

Wat zijn de effecten?

De analyses laten zien dat bij een juiste implementatie de BliksemBende op de meeste uitkomstmaten geen significante effecten heeft. Alleen op begrijpend lezen bij leerlingen in groep 4 werd een klein positief effect gevonden. Dit effect verdween echter een half jaar na de interventie, mogelijk is blijvend oefenen nodig om leerwinst te behouden. Voor technisch lezen werd geen effect gevonden ten opzichte van de reguliere ondersteuning, wat mogelijk samenhangt met de gebruikte meetinstrumenten die leessnelheid meten in plaats van leesaccuratesse. Bij spelling zijn er aanwijzingen dat vooral de zwakkere spellers baat hebben, al waren deze resultaten niet robuust. Op rekenen-wiskunde trad een tijdelijk negatief effect op, waarschijnlijk doordat leerlingen lessen misten.

Wat betekent dit?

De BliksemBende werd goed ontvangen door scholen en uitgevoerd zoals beoogd, maar laat in deze studie beperkte effecten zien, met potentie voor begrijpend lezen in groep 4. De interventie lijkt meer geschikt voor leerlingen die al enige basis in technisch lezen hebben gelegd. Grotere en langdurigere studies zijn nodig om te bepalen onder welke condities de effecten duurzaam zijn en hoe scholen de interventie het beste kunnen inzetten.

1 Inleiding

In het kader van Nationaal Programma Onderwijs (NP Onderwijs) hebben scholen van 2021 tot en met schooljaar 2024-2025 extra financiële middelen ontvangen voor het wegwerken van leerachterstanden als gevolg van COVID-19 en voor het verbeteren van het onderwijs, door middel van kansrijke interventies. Om scholen te helpen om goed gefundeerde keuzes te maken en inzicht te krijgen in de effectiviteit van de verschillende aanpakken, worden meerdere als kansrijk getypeerde interventies onderzocht in het programma 'Effectmeting kansrijke interventies in het po en vo'. Eén van deze interventies is het leesprogramma *Lezen met de BliksemBende* (hierna: de BliksemBende). In het schooljaar 2024-2025 is de effectiviteit van deze interventie getest in een RCT-onderzoek.

Leesvaardigheid is een van de belangrijkste vaardigheden die leerlingen op school leren voor hun toekomst. Het onderwijs slaagt er nog onvoldoende in vroege verschillen tussen kinderen te verkleinen gedurende de schoolloopbaan (Inspectie van het onderwijs, 2022; Meelissen et al., 2023). Het inzetten van tutoring kan een effectieve manier zijn om vroegtijdig de leesvaardigheid van leerlingen een boost te geven (Nickow et al., 2024). Deze bevinding is vooral gedaan binnen onderzoeken in het buitenland. In de Nederlandse context is voorgaande jaren nog weinig onderzoek gedaan naar de effectiviteit van (lees)interventies.

Lezen met de BliksemBende is ontwikkeld door de Rijksuniversiteit Groningen en is gebaseerd op het Engelstalige programma *Tutoring with the Lightning Squad* (TWLS). De BliksemBende wordt onder die naam al gebruikt in de Verenigde Staten en in het Verenigd Koninkrijk. Ook Duitsland werkt met een eigen variant van het programma (*Lesen mit dem Turbo-Team*). In het programma oefenen duo's zwakke lezers 4-5 keer per week 25 minuten met lezen op de computer, onder begeleiding van een tutor. Deze tutor hoeft geen leerkracht te zijn, maar kan ook een volwassen vrijwilliger of onderwijsassistent zijn. Het idee is dat deze kleinschalige tutoring leidt tot verbetering van leesvaardigheid, zoals ook in verschillende Amerikaanse *Randomised Controlled Trials* (RCT's) is aangetoond (Madden & Slavin, 2017; Neitzel, 2024; Ross et al., 2017). Lezen met de BliksemBende kan worden getypeerd als een interventie op ondersteuningsniveau 2. De interventie richt zich op de 25% zwakste lezers die niet genoeg hebben aan het basisaanbod lees- en spellingonderwijs. Ondersteuning op ondersteuningsniveau 2 wordt ingezet om achterstanden te voorkomen die anders intensievere interventie op niveau 3 zouden vereisen, of als eerste stap om te kijken of meer intensieve interventie nodig is. Omdat de BliksemBende tijdens reguliere lestijd plaatsvindt, is het belangrijk te onderzoeken of deelname ten koste gaat van andere basisvaardigheden, zoals rekenen-wiskunde.

1.1 Aanwijzingen voor effectiviteit

Opzet en dosering

Leerlingen boeken het meeste vooruitgang wanneer een kwalitatief goede tutor de sessies begeleidt, wanneer de leerling minimaal drie tutorsessies per week ontvangt en wanneer de tutoring plaatsvindt in kleine groepjes leerlingen (Nickow et al., 2024). Hoewel verschillende personen de rol van tutor kunnen vervullen, is tutoring het meest effectief wanneer een onderwijsprofessional deze rol vervult. Het is bovendien belangrijk dat tutores vooraf een passende training volgen.

De dosering van de leesinterventie moet voldoende hoog zijn: oefening is een van de belangrijkste pijlers voor de leesontwikkeling. Door meer tijd aan lezen te besteden, kunnen leerlingen in staat worden gesteld hun leesachterstand in te halen. Tutorprogramma's met minimaal drie tutorsessies per week zijn het meest effectief (Nickow et al., 2024). Ook specifiek in Nederlandse context blijkt dat de hoeveelheid tutoring sessies in de praktijk samenhangt met de vooruitgang in lezen bij beginnende lezers (de Jong & Schreurs, 2023; Regtvoort et al., 2013). Bovendien laat een studie naar TWLS zien dat een hogere dosering samenhangt met betere leesuitkomsten (Wang et al., 2024). Voldoende dosering hangt samen met hoe goed de interventie wordt geïmplementeerd zoals bedoeld: *implementation fidelity* (Carroll et al., 2007). Tot slot zijn tutorprogramma's die tussen de twaalf en twintig weken voortduren, het meest effectief gebleken (Dietrichson et al., 2017).

In een kleinschalige setting kunnen leerlingen geconcentreerder werken dan in grotere groepen. Aangetoond is dat samenwerkend leren positieve effecten heeft op prestaties en betrokkenheid (Kyndt et al., 2013; Neitzel et al., 2022). Tutorinterventies kunnen inderdaad een positief effect hebben op leesmotivatie en taakgericht werken, ook buiten de tutorsessies om (Gest & Gest, 2005). Op zijn beurt is een hoge leesmotivatie gerelateerd aan betere leesvaardigheid (Hebbecker et al., 2019). Hoewel over het algemeen de beste resultaten worden geboekt bij één-op-één tutoring (Nickow et al., 2024), kwam in een studie naar de effectiviteit van TWLS naar voren dat leerlingen die aan samenwerkend leren deden, evenveel vooruitgingen als leerlingen die individuele tutoring ontvingen binnen het programma (Neitzel, 2024).

Taalinhoudelijke componenten

De BliksemBende bestaat uit zes spellen die zich richten op relevante deelaspecten van leesontwikkeling in de brede zin (van geletterdheid tot begrijpend lezen), namelijk **fonemisch bewustzijn, technisch lezen, spelling, woordenschat en begrijpend lezen**. Om goed te kunnen lezen en spellen, is specifieke kennis op woordniveau van belang. Hier gaan leerlingen binnen de BliksemBende dan ook onder andere mee aan de slag. Volgens de Lexical Quality Hypothesis, bestaat kennis van een woord uit kennis over de betekenis (semantische representatie), kennis over de uitspraak (fonologische representatie) en kennis over de woordspelling (orthografische representatie) (Perfetti, 2007). Deze representaties zijn nauw met elkaar verbonden wanneer een leerling hoogwaardige lexicale kennis van een woord bezit. Dan activeert bijvoorbeeld de confrontatie met de geschreven vorm van een woord (de spelling) bij een leerling onmiddellijk ook de uitspraak en betekenis ervan. Dit leidt er dus toe dat een leerling vloeiend kan lezen, en ook begrijpt wat er staat. Lexicale kennis van woorden vormt zo dus een scharnierpunt tussen technisch en begrijpend lezen.

De aandacht voor verschillende soorten woordkennis is terug te zien in de opzet van de oefeningen binnen de BliksemBende. Zo wordt er in *Superspeller* geoefend met een dictee waardoor de verbinding tussen de orthografische en fonologische representatie van woorden wordt verstevigd en is er in *Woordenschat* juist aandacht voor de betekenis van een woord. Omdat er binnen TWLS ook aandacht is voor begrijpend lezen, kan TWLS mogelijk worden gezien als een aanvulling op het Nederlandse tutorprogramma Bouw!. Dit is een effectief gebleken interventie die zich richt op technisch lezen bij jongere leerlingen op ondersteuningsniveau 3 (de Jong & Schreurs, 2023; Zijlstra, 2015).

Technologie

De inzet van **technologie** brengt ook verschillende voordelen met zich mee. Digitale hulpmiddelen zijn met name effectief als het gebruik ook wordt ondersteund door een volwassen tutor (McTigue et al., 2020). Digitale hulpmiddelen waarbij volledig zelfstandig gewerkt wordt, laten vaak juist kleine of geen effecten zien op lezen (Higgins et al., 2012; Neitzel et al., 2022). Juist deze combinatie van begeleiding en digitale hulpmiddelen kan effectief zijn, en minstens even goed werken als enkel het oefenen met een tutor (Chambers et al., 2011). Een ander voordeel is dat een tutor zo meerdere leerlingen tegelijkertijd kan begeleiden. Dit draagt dus bij aan de kostenefficiëntie van een tutorprogramma (Madden & Slavin, 2017). Digitale hulpmiddelen maken het ook mogelijk om directe auditieve en visuele feedback te geven. Feedback heeft een positief effect op leeruitkomsten, zeker als het gaat om het leren van een nieuwe vaardigheid (Wisniewski et al., 2020).

1.2 Voor wie werkt het?

Eerder onderzoek naar TWLS laat zien dat de effectiviteit van de interventie niet afhankelijk is van gender of etniciteit van de leerlingen (Nickow et al., 2020). Onderzoekresultaten geven geen uitsluitsel over de verschillen in effectiviteit per jaargroep. Zo vindt één studie dat TWLS even effectief is in de groepen 3 t/m 5 (Grade 1 t/m 3) (Nickow et al., 2020), terwijl in een andere studie naar een voorloper van TWLS grotere effecten worden gevonden in de groep 4 en 5 dan in groep 3 (Ross et al., 2017). Hieraan gerelateerd zou het beginniveau van leerlingen ook uit kunnen maken in de effectiviteit. Leerlingen met een grote leesachterstand hebben meer ondersteuning nodig dan leerlingen met een minder grote achterstand, mogelijk zijn de interventie-effecten dus kleiner voor leerlingen met een grotere leesachterstand (Stuebing et al., 2015).

Een ander belangrijk achtergrondkenmerk is thuistaal. Bij TWLS is nog niet specifiek onderzocht of het even goed werkt voor leerlingen die leren lezen in hun eerste of tweede taal. Eerdere meta-analyses laten zien dat een scala aan leesinterventies positieve effecten kan hebben op de leesvaardigheid van Engelse tweede taalleerders (Ludwig et al., 2019). Vooral programma's die gebruik maken van kleinschalige tutoring en samenwerkend leren zijn voor deze groep effectief (Cheung & Slavin, 2012). Specifiek voor tweede taalleerders met een risico op een leesachterstand is er nog weinig onderzoek, maar zijn verschillende interventies effectief gebleken, waaronder tutoring (Richards-Tutor et al., 2016). Het is daarom belangrijk om in kaart te brengen of de BliksemBende ook werkt voor NT2 leerlingen, want het is waarschijnlijk dat deze zijn oververtegenwoordigd in de groep van zwakke lezers (Inspectie van het Onderwijs, 2022). NT2 leerlingen die leesproblemen (lijken te) hebben, hebben niet altijd hetzelfde profiel als NT1 leerlingen. Achterblijvende scores op

begrijpend lezen kunnen bijvoorbeeld veroorzaakt worden door een kleinere woordenschat, in plaats van een achterstand in technisch lezen (Verhoeven & Van Leeuwe, 2012). Aangezien er in de BliksemBende ook expliciet aandacht is voor woordenschat, in tegenstelling tot andere leesinterventies, zou deze interventie mogelijk voor NT2 leerlingen goed geschikt kunnen zijn (Verhoeven & van Leeuwe, 2012).

Hoewel eerder gerapporteerde interventie-effecten van tutoring programma's dus over het algemeen weinig lijken af te hangen van achtergrondkenmerken van leerlingen (en de interventie dus even effectief is voor verschillende groepen), zijn leesachterstanden niet gelijk verdeeld over de leerlingpopulatie. Juist leerlingen uit meer kwetsbare groepen (zoals leerlingen waarvan de ouders een laag opleidingsniveau hebben of een laag inkomen, of leerlingen met een niet-westerse migratieachtergrond) laten de grootste vertraging in leergroei zien op begrijpend lezen en spelling (Inspectie van het Onderwijs, 2022). In het kader van kansengelijkheid is het daarom belangrijk om de werkzaamheid van de interventie voor verschillende leerlinggroepen systematisch in kaart te brengen.

1.3 Onderzoeksvragen

In deze RCT wordt onderzocht of Lezen met de BliksemBende een effectieve interventie is om leesachterstanden aan te pakken. Aangezien de BliksemBende onder leestijd plaatsvindt, missen leerlingen ook andere relevante stof. Daarom kijken wij ook of deelname aan de BliksemBende ten koste gaat van algemene rekenvaardigheden. Om het programma in de toekomst optimaal te kunnen implementeren, moet ook duidelijk worden wat de werkzame elementen zijn van de interventie, onder welke omstandigheden de meeste vooruitgang wordt geboekt, en voor welke leerlingen het programma het beste werkt. Ten slotte voeren we een procesevaluatie uit om te achterhalen hoe de interventie in de praktijk wordt ervaren en welke lessen dit oplevert voor scholen die met de BliksemBende aan de slag willen. Deze kwesties zijn verdeeld in de volgende hoofd- en deelvragen.

Effectevaluatie	
1.a	Heeft Lezen met de BliksemBende een positief effect op technisch lezen, spellen, woordenschat, en begrijpend lezen?
1.b	Heeft Lezen met de BliksemBende negatieve neveneffecten op rekenen-wiskunde?
1.c	Heeft Lezen met de BliksemBende een positief effect op leesmotivatie en heeft dit mogelijk een mediërend effect op vooruitgang in leesontwikkeling?
1.d	Heeft woordenschat een mediërend effect op de vooruitgang in leesontwikkeling?
Welke interventie- en oefenkenmerken zijn geassocieerd met de meeste vooruitgang?	
2.a	In welke mate is de interventie uitgevoerd zoals voorgenomen (treatment fidelity) als het gaat om het aantal sessies, minuten en vooruitgang binnen de oefenspellen?
2.b	Voorspelt het aantal sessies, minuten en vooruitgang binnen de spellen in de interventiegroep, de vooruitgang op lezen?
2.c	Zijn er verschillen in effectiviteit tussen verschillende type tutores (vrijwilligers, onderwijsassistenten, leerkrachten, etc.)?
Welke kind- en groepskenmerken zijn geassocieerd met de meeste vooruitgang? Is er verschil in effectiviteit op basis van...	
3.a	Gender
3.b	Thuis taal
3.c	Jaar laag (Groep 3 of 4)
3.d	Initiële leesachterstand
Procesevaluatie	
4.a	Welke succesfactoren en knelpunten zijn in de uitvoering zichtbaar geworden?
4.b	Welke ervaringen hebben leerlingen met het werken met de BliksemBende?
4.c	Welke praktische tips en aanbevelingen volgen hieruit voor scholen die met de BliksemBende aan de slag willen?

1.4 Leeswijzer

In dit rapport bespreken we eerst de opzet en inhoud van Lezen met de BliksemBende in hoofdstuk 2. In hoofdstuk 3 bespreken we de methode van het onderzoek, waaronder de werving/steekproef, onderzoeksdesign en de gebruikte meetinstrumenten. Vervolgens gaan we in op de resultaten van de analyses in Hoofdstuk 4. Ten slotte bespreken we de conclusies en implicaties in het discussiehoofdstuk 5.

2 De interventie: Lezen met de BliksemBende

2.1 Doel/doelgroep van de interventie

Het doel van Lezen met de BliksemBende is om leerlingen op een effectieve, efficiënte en leuke manier te helpen om betere lezers te worden. Het is een digitaal tutorprogramma om leerlingen met een (mogelijke) leesachterstand uit groep 3 en 4 te ondersteunen in hun leesontwikkeling. Het kan worden getypeerd als interventie op ondersteuningsniveau 2, hoewel leerlingen de tutorsessies buiten de klas ontvangen in plaats van in de klas. Ondersteuningsniveau 2 richt zich op de zwakste 25% van de lezers, die niet genoeg hebben aan het basisaanbod lees- en spellingonderwijs. Ondersteuning op ondersteuningsniveau 2 wordt ingezet om achterstanden te voorkomen die intensieve interventie op niveau 3 nodig maken, of als eerste stap om te kijken of meer intensieve interventie nodig is.

2.2 Opzet van de interventie

In het programma oefenen zwakke lezers in 4-5 sessies per week 25 minuten per sessie met lezen op de computer in kleine groepjes, onder begeleiding van een volwassen tutor. Deze tutor hoeft geen leerkracht te zijn, maar kan ook een vrijwilliger of onderwijsassistent zijn. Tijdens een tutorsessie werken vier tot zes leerlingen tegelijkertijd in duo's op een computer. Spelenderwijs werken zij aan verschillende oefeningen en geven zij feedback aan elkaar. Beurtelings nemen zij de rol van Lezer en van Maatje (coach) op zich. Deze oefeningen zijn steeds gebaseerd op een verhaal. De moeilijkheidsgraad van dit verhaal is afgestemd op het niveau van de leerlingen.

De taak van de tutor is om het leesproces en de samenwerking te observeren, begeleiden en monitoren. Daarnaast neemt de tutor beslissingen over het niveau, zoals het kiezen van een startniveau en het al dan niet behalen van de tussentijdse tutortoets (waarna een leerling doorgaat met een moeilijker verhaal). De tutor houdt de vorderingen bij en past niveaus aan waar nodig. Een tutorcoördinator (bijvoorbeeld een intern begeleider) coördineert het programma op school en ondersteunt de tutor.

De interventieperiode is flexibel, maar in deze RCT is de interventieperiode tussen scholen gelijkgetrokken op een periode van 12 weken. Aan het begin van de interventie wordt bij de leerlingen een programmagebonden woordleestaak afgenomen als instaptoets om tot een geadviseerd startniveau in de BliksemBende te komen. De instaptoets bestaat uit twee onderdelen: decoderen en vloeiend lezen. Binnen het onderdeel decoderen heeft de leerling

per woord 10 seconden om deze hardop voor te lezen. De moeilijkheidsgraad van de woorden neemt toe. De leerling gaat net zo lang door totdat er te veel woorden foutief gelezen zijn. Binnen het onderdeel vloeiend lezen leest de leerling 60 seconden lang hardop een tekstfragment. De tutor houdt de woorden bij die de leerling verkeerd leest. Vervolgens wordt de leerling gekoppeld aan een maatje met een vergelijkbaar leesniveau. Het systeem gaat altijd uit van het laagste verhaalnummer wanneer twee leerlingen aan elkaar gekoppeld worden. In het geval dat leerlingen naar verloop van tijd sterk uiteen gaan lopen qua niveau, kunnen nieuwe teams worden gevormd. Als het niet mogelijk is om een team te vormen, kunnen leerlingen eventueel ook individuele sessies volgen.

Opbouw tutorsessie

Een tutorsessie start met een check-in per team waarbij de tutor de aandachtspunten voor het leesproces en het samenwerken benoemt. De inschatting van de interventieaanbieder is dat leerlingen per verhaal drie tutorsessies nodig hebben, en dus ongeveer twee spellen per tutorsessie doen. Bij elk spel nemen de leerlingen om de beurt de rol in van lezer of van maatje. De lezer werkt aan de taak, en het maatje coacht en controleert. Na ieder spel geeft het programma ook standaard feedback: dan wordt het aantal goed gelezen letters of woorden per leerling weergegeven en gaat dit gepaard met bijvoorbeeld gejuich (meer dan 80% goed) of zien en horen de leerlingen 'Probeer het nog maar een keer' (minder dan 80% goed). Na ongeveer 25 minuten sluit de tutor de sessie af met een check-out per team en benoemt aandachtspunten voor het leesproces en het samenwerken.

2.3 Inhoud van de interventie

Het programma bevat 50 verhalen. Bij elk verhaal horen zes verschillende spellen op het gebied van vloeiend lezen, letters lezen, woorden lezen, spelling, woordenschat en begrijpend lezen. De vooruitgang van de leerlingen wordt bijgehouden door het programma. Na het afronden van een verhaal doen beide leerlingen individueel een tutortoets. De verschillende spellen en onderdelen worden hieronder toegelicht:



Leeshoek

Leerlingen lezen een eerder gelezen verhaal hardop en oefenen met vloeiend lezen. In de Leeshoek kiezen teams uit verhalen met een lager verhaalnummer dan het nieuwe verhaal waaraan zij werken. Dit spel wordt afgerond wanneer beide leerlingen het 'woorden per minuut'-doel behalen voor het verhaal.



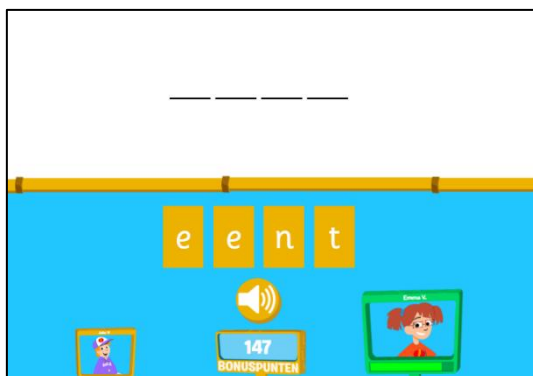
Letterlezer

Leerlingen lezen hardop letters of tweetekenklanken die in het nieuwe verhaal voorkomen. Indien nodig kunnen zij om hints vragen. De keuze voor letters die aan bod komen bij een verhaal, is gebaseerd op AVI-niveaus, de Leerlijn Estafette Nieuw en de Leerlijn technisch lezen van SLO. Vanaf verhaal 21 worden de letters en klanken bekend geacht en kan dit spel niet meer worden geoefend.



Woordlezer

Leerlingen lezen losse woorden uit het nieuwe verhaal hardop voor. Indien nodig kunnen zij om hints vragen. Bij de eerste hint wordt het beeld in beeld uitgerekt (bijvoorbeeld *m aa r*). Bij de tweede hint worden de losse klanken voorgelezen. Bij de derde hint wordt het hele verhaal voorgelezen.



Superspeller

Leerlingen horen een woord uit het nieuwe verhaal en moeten de correcte spelling typen. Er wordt automatisch een hint gegeven wanneer de leerling een fout maakt. Ook kunnen de leerlingen zelf om een hint vragen. Bij de eerste hint verschijnt een klankkaart/letterkaart in beeld. Bij de tweede hint wordt deze voorgelezen.



Woordenschat

Leerlingen oefenen met de woordenschat. De woorden horen bij het nieuwe verhaal waar zij op dat moment aan werken. Bij het opstarten van het spel worden de moeilijkste woorden uit het nieuwe verhaal uitgelegd in een animatievideo. Leerlingen krijgen daarna een zin te zien waarin één woord ontbreekt en selecteren uit een aantal woorden, het woord dat het beste in de zin past. Ook hier zijn hints beschikbaar in de vorm van

audio en afbeeldingen van de te kiezen woorden. Voor de selectie van woorden in dit spel is gebruik gemaakt van de BAK, Streefwoordenlijst en Digiwak.

Nieuw verhaal

Leerlingen lezen het verhaal waarvoor zij voorbereidende spellen hebben uitgevoerd. Ze beantwoorden ook begripend leesvragen bij dit verhaal. Dit spel wordt pas ontgrendeld wanneer de andere spellen binnen dit level goed zijn afgerond (minimaal 80% goed in elk spel).

Geen enkel verhaal is rechtstreeks vertaald vanuit het Engels. 32 verhalen zijn in een eerder RUG-project (Success for All) (Mullender-Wijnsma et al., 2020) geconstrueerd, inclusief illustraties. Achttien verhalen zijn nieuw geconstrueerd, maar de illustraties zijn wel hergebruikt van Engelse, door FFT ontwikkelde boekjes. Qua onderwerp zijn deze verhalen dus wel geïnspireerd op Engelse voorbeelden. De verhalen bouwen op in moeilijkheidsgraad (niveau B3 - E4) en in de context waarin het verhaal zich afspeelt (verhaal 1 t/m 11: thuis, verhaal 12 t/m 17: op school, verhaal 18 t/m 38: in de buurt, verhaal 39 t/m 50: verder weg in binnen- of buitenland).

Tutortoets

Nadat een team alle spellen bij een verhaal heeft afgerond (minimaal 80% goed), moeten zij beiden individueel de tutoertoets maken. Dit is een korte check over het verhaal waar de leerlingen mee bezig zijn geweest. De toets bestaat uit een programmegebonden woordleestaak waarin leerlingen lezen losse woorden voorlezen. Daarnaast lezen leerlingen een korte tekst uit het verhaal hardop voor, en beantwoorden een begripsvraag over het verhaal waar ze mee bezig zijn geweest. Tijdens de tutoertoets beoordeelt de tutor de voortgang van de leerlingen. Deze bepaalt of de leerlingen het verhaal beheersen en door mogen naar het volgende verhaal. De algemene richtlijn is dat leerlingen minimaal 80% van de woorden correct hebben gelezen, voldoende woorden lezen per minuut (per verhaal is hiervoor een drempelwaarde opgesteld) én de begripend leesvraag goed hebben geantwoord.

Beloning

Bij het afronden van elk spel komt er een felicitatiescherm in beeld en ontvangen leerlingen digitale punten. De punten kunnen worden uitgegeven in de digitale omgeving de Prijzenkast. Deze verschijnt na het behalen van de tutoertoets. In de Prijzenkast kunnen leerlingen ervoor kunnen kiezen om de punten uit te geven om verschillende educatieve filmpjes te bekijken. Er zitten in totaal 167 filmpjes in het programma, waarvan 147 afkomstig uit SchoolTV (gemiddelde duur 135 seconden). De overige filmpjes zijn door de interventieaanbieder ontwikkeld en bestaan voornamelijk uit moppen en spelletjes.

2.4 Begeleiding van scholen tijdens de interventie

In deze RCT oefenden leerlingen in een periode van 12 weken vier à vijf keer per met de BliksemBende. Voor enkele scholen was dit niet goed haalbaar en is ingezet op een frequentie van 4 keer per week, waarbij drie keer werd gezien als het minimum. Scholen werden gedurende het schooljaar intensief begeleid door begeleiders van het interventieteam van de Rijksuniversiteit Groningen. Hiervoor is gekozen om te kunnen waarborgen dat de implementatietrouw voldoende is, uitval te voorkomen en goed zicht te houden op het verloop van de interventie. Het doel was om scholen goed te ondersteunen en eventuele belemmeringen in de uitvoer vroeg op het spoor te komen en op te lossen. Het

implementatieteam heeft drie schoolbegeleiders (met elk 0,5 fte) aangesteld die elk zeven tot acht scholen onder hun hoede hadden.

De schoolbegeleiders waren een vast aanspreekpunt voor de scholen voor alles wat met de interventie te maken had. Ze bezochten de school voor een startgesprek, verzorgden de training, deden twee observatie-/coachingsbezoeken, monitorden de logfiles om te kijken of er opvallendheden waren, en waren beschikbaar voor extra ondersteuning indien nodig. Binnen elke school werd één persoon aangewezen als tutorcoördinator (meestal de ib'er). Deze stuurde de tutores aan. Tutores bestonden voor het overgrote deel uit onderwijsassistenten maar soms ook uit ander personeel van de school, zoals intern begeleiders of schoolleiders. Tutorcoördinatoren en tutores ontvingen in de periode oktober-januari een training van het implementatieteam van zes uur.

Halverwege de interventieperiode is een fysieke ontmoetingsmiddag georganiseerd met tutores en tutorcoördinatoren en het implementatieteam en onderzoeksteam. Vlak voor- en aan het einde van de interventieperiode vond een online meeting plaats voor een evaluatie met alle schoolleiders, waarbij ook meerdere tutorcoördinatoren aanwezig waren. Op deze momenten zijn ervaringen van de scholen opgehaald en onderling uitgewisseld en eventuele verbeterpunten in kaart gebracht rondom (de uitvoer van) het programma.

2.5 Pilotonderzoek

Voorafgaand aan de RCT is een pilot-onderzoek uitgevoerd op vier scholen. Het pilotonderzoek diende om de ervaringen met het programma als geheel op te halen (de training, samenwerkend leren, software etc.) en de moeilijkheidsgraad van de ontwikkelde materialen te testen bij de doelgroep in groep 3 en 4. In schooljaar 2022-2023 is een pilotonderzoek uitgevoerd voor de materialen voor groep 3. In schooljaar 2023-2024 is direct na de zomervakantie ook een pilot uitgevoerd voor de materialen in groep 4. De pilot-scholen waren geen onderdeel van de reguliere steekproef, maar bestonden uit scholen die al samenwerkten met de interventieaanbieder, of scholen die tijdens de reguliere wervingsprocedure aangaven niet in staat te zijn deel te nemen aan de volledige RCT, maar wel op deze manier aan het onderzoek wilden bijdragen. Naar aanleiding van de pilot zijn enkele kleine aanpassingen gedaan aan de interventie en de begeleiding, zo is de training iets uitgebreid in de informatievoorziening. Aan de inhoud van het programma zijn slechts marginale aanpassingen gedaan.

3 Methode

In dit hoofdstuk beschrijven we hoe het onderzoek naar de effectiviteit van Lezen met de BliksemBende is uitgevoerd. We bespreken achtereenvolgens de werving en selectie van scholen en leerlingen, de onderzoeksopzet en randomisatie, de meetinstrumenten en de analysestrategie.

3.1 Werving en selectie van scholen en leerlingen

Selectiecriteria voor scholen en klassen

Om in aanmerking te komen voor deelname aan de RCT, moesten scholen (bij voorkeur) aan enkele selectiecriteria voldoen. Alleen scholen in Noord-Nederland (Drenthe, Groningen, Friesland) en Noord-Holland zijn benaderd, zodat de interventie-aanbieder en het onderzoeksteam de scholen op locatie konden bezoeken. Scholen met combinatiegroepen waren uitgesloten van deelname¹. Omdat kleinere scholen vaker gebruik maken van combinatiegroepen, hebben we initieel alleen scholen met minimaal 200 leerlingen benaderd. Dit is tijdens het wervingsproces uitgebreid naar minimaal 150 leerlingen. Scholen moesten daarnaast bij gebruik maken van het Cito-leerlingvolgsysteem of bereid zijn om de vereiste toetsen af te nemen.

Omdat leerlingen uit meer kwetsbare groepen (zoals leerlingen waarvan de ouders een laag opleidingsniveau of inkomen hebben) de grootste vertraging in leergroei zien op begrijpend lezen en spelling², is er in eerste instantie gezocht naar scholen met een bovengemiddelde schoolweging (3^e of 4^e kwartiel). Gedurende het wervingsproces is dit uitgebreid naar het 2^e kwartiel om voldoende deelname te realiseren.

Wervingsproces

De werving vond plaats in de periode van maart 2023 tot en met november 2023. In alle communicatie stond de toegevoegde waarde voor scholen centraal (zie 'beloning scholen' hieronder) en is benadrukt dat het onderzoek zo is opgezet om scholen zoveel mogelijk te ontzorgen, doordat er bijvoorbeeld standaard LVS-toetsen worden afgenomen in plaats van dat er extra toetsmomenten moeten worden gecreëerd. De wervingsstrategie bestond uit een multi-channel strategie (digitale en geprinte flyers, website met ondersteunende informatie, persoonlijk belcontact en schoolbezoeken) met een nadruk op persoonlijk

¹ De reden dat scholen met combinatiegroepen waren uitgesloten heeft te maken met het onderzoeksdesign, zie bladzijde 15 voor een toelichting.

² Inspectie van het Onderwijs (2022). De Staat van het Onderwijs 2022. Utrecht: Inspectie van het Onderwijs.

contact. Deze wervingsstrategie is samen met de Rijksuniversiteit Groningen opgezet en uitgevoerd.



Figuur 3.1 Overzicht stappen in het wervingsproces van scholen

Gedurende het wervingsproces hebben we verschillende stappen gezet (zie Figuur 3.1). Voorafgaand aan de werving is een website met informatie en veelgestelde vragen opgezet. Ten eerste hebben we 297 scholen in Noord-Nederland per post een flyer gestuurd. Een overige 426 scholen in Noord-Holland hebben deze flyer digitaal ontvangen. Vervolgens hebben we 551 van deze scholen gebeld. In de eerste telefoongesprekken hebben we uitleg gegeven over het onderzoek en de eisen voor de school toegelicht. Veel scholen met interesse zeiden toch niet deel te kunnen of willen nemen, omdat zij niet met Cito-toetsen binnen de school werkten. Bij de 35 scholen die zeiden serieuze interesse te hebben om deel te willen nemen, zijn we op schoolbezoek gegaan. Hier hebben we een pitchpresentatie gegeven en openstaande vragen beantwoord over wat deelname aan de interventie en het onderzoek precies inhoudt. Na deze schoolbezoeken zijn 11 scholen uitgevallen om uiteenlopende redenen, zoals wisselingen van de directie of omdat de tijdsinvestering te groot bleek. Met de 24 overgebleven scholen hebben we samenwerkingsovereenkomsten gedeeld en met hen hebben we ook de (anonieme) screening van leerlingen in gang gezet. Ten tijde van de screening van de leerlingen heeft één laatste school zich toch afgemeld, vanwege de omvang van het onderzoek. De uiteindelijke sample bestond uit 23 scholen.

Beloning scholen

Er is op verschillende manieren ingezet om deelname aan dit onderzoek aantrekkelijk te maken voor scholen. Ten eerste ontvingen scholen kosteloze toegang tot het programma. Ten tweede ontvingen scholen vergoeding voor de extra te maken personeelskosten en eventueel extra aan te schaffen toetsen. Ten derde ontvingen de scholen intensieve begeleiding vanuit de interventie-aanbieder, zowel voorafgaand als tijdens de implementatie; zo kregen tutoren een trainingsdag, minimaal twee keer tussentijds een schoolbezoek van ondersteuners om de kwaliteit te borgen en om tussentijdse vragen te beantwoorden, en werd er twee keer een tussentijds overleg gepland met schooldirecteuren om de voortgang te bespreken. Bijkomende voordelen van deze begeleidingsactiviteiten waren dat deze ten goede kwamen aan het draagvlak van onderwijspersoneel om de gehele periode mee te werken en dat deze contactmomenten ook werden gebruikt voor de procesevaluatie. Ten vierde ontvingen de scholen na afloop van de interventieperiode een

schoolrapport met de resultaten, waarin de ontwikkeling van hun eigen leerlingen voor hen samengevat werd.

Selectiecriteria voor leerlingen

Om te kijken welke leerlingen in aanmerking komen voor deelname zijn 1377 leerlingen gescreend van de 23 deelnemende scholen uit groep 3 en 4 in oktober – november 2023. De screening bestond in groep 3 uit leestaken die horen bij de herfstsignalering van de leesmethode Veilig Leren Lezen en in groep 4 uit de gestandaardiseerde leestaken DMT en AVI (zie paragraaf 3.3 voor een nadere toelichting op deze instrumenten). Op basis van deze uitkomsten zijn leerlingen geselecteerd met een mogelijke leesachterstand, die baat zouden hebben bij een interventie op ondersteuningsniveau 2. In groep 3 was het selectie criterium leerlingen die binnen onze steekproef in de laagste 30% scoren op de Veilig Leren Lezen woordleestaak.

In groep 4 zijn leerlingen geselecteerd die een duidelijk risico op leesproblemen vertonen, zoals blijkt uit een niveau IV of V op de DMT of nog niet op beheersingsniveau lezen op AVI E3. Daarmee sluit de selectie nauw aan bij de doelgroep van ondersteuningsniveau 2. In sommige klassen kwamen minder dan 4 leerlingen in aanmerking. In deze gevallen zijn enkele leerlingen geselecteerd die vlak boven het selectie criterium scoorden, zodat in ieder geval groepjes van 4 konden worden gemaakt voor de tutoring.

Omdat de BlikseBende een interventie is op ondersteuningsniveau 2, werden leerlingen waarvoor definitief was vastgesteld dat ze meer baat hebben bij ondersteuning op ondersteuningsniveau 3 geëxcludeerd: Leerlingen die een E-score behaalden op de DMT én vanuit school al leesondersteuning ontvangen op ondersteuningsniveau 3 (zoals Bouw!), zijn niet geselecteerd. Niet elke leerling die op basis van de screening in aanmerking kwam is uiteindelijk geselecteerd voor het onderzoek. Dit geldt voor leerlingen waarvan de consentbrief niet is ondertekend en voor leerlingen waarvan de capaciteit van de school niet groot genoeg was om ze allemaal de interventie te laten volgen. In deze situatie is in overleg met de scholen een geschikte selectie van leerlingen gemaakt die in aanmerking komen voor een interventie op ondersteuningsniveau 2. Scholen regelden de oudertoestemming.

Steekproef

Aan deze RCT namen uiteindelijk 23 scholen deel (77 klassen, 340 leerlingen). De klassen zijn verdeeld in 37 klassen van groep 3 en 40 klassen van groep 4. De scholen liggen in overwegend matig tot weinig stedelijke gebieden, met daarnaast 3 scholen in een (zeer) stedelijk en 3 scholen in een niet – stedelijk gebied. De scholen hadden tussen de 170 en 500 leerlingen (gemiddeld 270). Voor 22 van de 23 scholen is de achterstandsscore beschikbaar³. Deze score wordt berekend op basis van kenmerken van de leerlingpopulatie, waaronder het opleidingsniveau van de ouders, herkomst van ouders en of ouders in de schuldsanering zitten. Hoe hoger de achterstandsscore, hoe groter de kans op leerachterstanden. De gemiddelde achterstandsscore in onze steekproef ligt met 155 boven de landelijke mediaan. De meeste scholen in onze studie (15) zitten in de middenmoot, met drie scholen in de laagste 25% en vier in de hoogste 25%. Dit betekent dat de interventie vooral is getest op scholen met een gemiddeld tot wat hoger achterstandrisico.

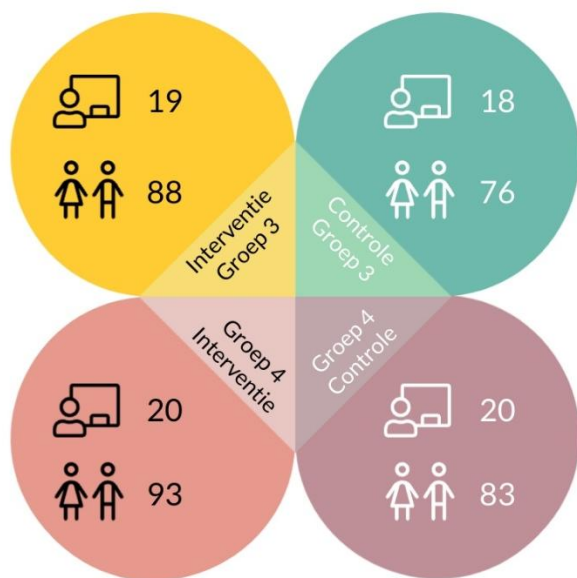
³ Gebaseerd op openbare DUO-data: achterstandsscores zonder drempel.

Op vier van de scholen wordt een specifiek onderwijsconcept gehanteerd in het kader van algemeen bijzonder onderwijs: dit betreffen een Daltonschool, een Jenaplanschool, een OGO-school en een begaafdheidsprofielschool. Op 17 van deze scholen wordt (veel) ingezet op samenwerkend leren door leerlingen en op zes scholen nauwelijks tot niet.

Randomisatie

21 scholen zijn random ingedeeld in groep A, waarin de interventie werd uitgevoerd in groep 3 en de controlegroep bestond uit groep 4, of groep B, waarin de interventie werd uitgevoerd in groep 4 en de controlegroep bestond uit groep 3. Leerlingen in de controlegroep ontvingen de leesondersteuning die de school normaal gesproken organiseerde (zolang dit niet De BliksemBende was). Om mogelijke *spill-over effecten* binnen de klas te voorkomen waren scholen met combinatiegroepen uitgesloten. Zo zitten groep 3-leerlingen uit de interventiegroep, niet in dezelfde klas als groep 4-leerlingen uit de controlegroep.

Voor twee scholen is een uitzondering gemaakt, omdat zij voorafgaand aan deelname aangaven de sterke voorkeur te hebben om de interventie in beide jaarlagen uit te voeren. Beide scholen deden mee met vier klassen. Bij deze scholen is binnen de jaarlaag gerandomiseerd welke klas met de interventie aan de slag ging en welke klas als controleklas fungeerde. Op deze scholen volgden groep 3a en 4a bijvoorbeeld de interventie, en waren groep 3b en 4b de controlegroep, of andersom. Een overzicht van het aantal klassen en leerlingen per groep is weergegeven in Figuur 3.2.



Figuur 3.2 Overzicht participanten op klas- en leerlingniveau

Achtergrondgegevens van de leerlingen zijn weergegeven in Tabel 3.1. De verdeling tussen jongens en meisjes is redelijk gelijk verdeeld. Ongeveer 20% van de leerlingen in de steekproef sprak thuis (ook) een thuistaal anders dan Nederlands. De meest gesproken thuistalen waren Arabisch, Pools en Engels.

Tabel 3.1 Achtergrondgegevens leerlingen

	Controlegroep		Interventiegroep	
	Groep 3	Groep 4	Groep 3	Groep 4
Jaarlaag				
Totaal	76	83	88	93
Meisjes	38	34	40	51
Jongens	37	49	48	42
Gemiddelde leeftijd* (SD)	6,65 (0,33)	8,01 (0,48)	6,79 (0,43)	7,86 (0,44)
(Deels) andere thuistaal	16	11	21	19

*leeftijd in jaren ten tijde van startdatum interventieperiode (februari 2024)

Noot: Bij één leerling in groep 3 is een andere antwoordoptie aangekruist dan jongen of meisje, waardoor de totalen van meisjes en jongens niet optellen tot 76.

3.2 Onderzoeksdesign

De scholen zijn in de periode februari–mei 2024 gedurende 12 schoolweken aan de slag gegaan met de interventie in groep 3 en groep 4. Wij hebben op vijf verschillende momenten data opgevraagd bij scholen. Deze momenten zijn onder te verdelen in een screeningsmoment, een voormeting, de interventieperiode, de nameting en de follow-up meting. Een overzicht van de verschillende meetmomenten en welke afnames hebben plaatsgevonden is weergegeven in Figuur 3.3.

Tijdens de screening eind 2023 is binnen groep 3 de herfstsignalering Veilig Leren Lezen afgenomen en binnen groep 4 de DMT en AVI. Leerlingen die voldeden aan de vooropgestelde selectiecriteria en waarvan oudertoestemming is verkregen namen vervolgens deel aan het onderzoek.

Bij geselecteerde leerlingen zijn de schoolprestaties in kaart gebracht in januari 2024 (voormeting), juni 2024 (nameting), en in januari 2025 (follow-up). We gebruiken de scores op de toetsen uit het leerlingvolgsysteem als primaire uitkomstmaat in plaats van toetsen binnen het computerprogramma zelf, omdat scholen deze ook door scholen zelf worden ingezet om de voortgang van leerlingen te volgen. Daarmee sluiten de metingen direct aan bij de praktijk en wordt de ecologische validiteit van de resultaten vergroot. Daarnaast zijn bij de voormeting en nameting de leesmotivatie van leerlingen in kaart gebracht. Tijdens de interventie zijn logboeken en logdata binnen het computerprogramma bijgehouden om te monitoren in welke mate er geoefend is.



Figuur 3.3. Overzicht van de tijdlijn en gebruikte meetinstrumenten

3.3 Meetinstrumenten

Wij hebben verschillende meetinstrumenten ingezet, verdeeld over de verschillende onderzoeksfases. Hieronder geven wij per meetinstrument een beschrijving over de meetpretentie, de opbouw van het instrument, de uitvoering, de scores/interpretatie, de betrouwbaarheid en validiteit, en een voorbeelditem.

Met scholen is afgesproken dat ze zelf de toestemmingsformulieren door ouders lieten ondertekenen, de data verzamelen bij de leerlingen en doorgeven aan de onderzoekers, aangezien het voor het grootste gedeelte om Cito-toetsen gaat die scholen zelf al afnemen bij leerlingen voor hun leerlingvolgsysteem. Bij enkele scholen zijn onderzoeksassistenten langsgekomen om deze toetsen af te nemen omdat de school deze toetsen niet standaard afnam in deze periode. Alle stappen van het onderzoek zijn beschreven in een handboek dat alle scholen aan het begin van schooljaar 2023-2024 opgestuurd kregen.

Leerlinguitkomsten

1. Cito – DMT

De Drie-Minuten-Toets (DMT) is een (gestandaardiseerde) Cito-toets. met als doel de technische leesvaardigheid van leerlingen in kaart te brengen. Leerlingen krijgen drie leeskaarten die oplopen in moeilijkheid. Op elke leeskaart staan 150 woorden. Per leeskaart krijgt de leerling een minuut om de woorden zo snel en accuraat mogelijk te lezen. De leerkracht neemt de toets af. De DMT wordt vanaf groep 3 afgenomen, met als eerste afnamemoment halverwege het schooljaar (jan). De betrouwbaarheid en validiteit van deze toets is in kaart gebracht en als voldoende beoordeeld (van Til et al., 2018a). Wij hebben van deze toets de ruwe score, vaardigheidsscore, het niveau (I-V en A-E), moment van afname en toetsversie opgevraagd (scholen zijn geïnstrueerd om zo mogelijk versie B af te nemen).

2. Cito – AVI

De AVI is een (gestandaardiseerde) Cito-toets, die ook meet hoe goed een leerling is in technisch lezen. In tegenstelling tot de DMT, wordt bij een AVI-toets het technisch lezen van een hele tekst getoetst. De leerling kan dus mogelijk een woord afleiden uit de context van de tekst. De toegekende score wordt bepaald op basis van het aantal gemaakte fouten en de tijd die het een leerling kost om de hele tekst te lezen. Aan deze score worden drie niveaus toegekend: beheersingsniveau, instructieniveau en frustratieniveau.

Beheersingsniveau betekent dat een leerling het leesniveau volledig beheerst, instructieniveau dat dit het leesniveau is waarop een leerling kan oefenen en frustratieniveau dat het leesniveau nog te hoog is. De AVI wordt vanaf groep 3 afgenomen, met als eerste afnamemoment halverwege het schooljaar (jan). De betrouwbaarheid en validiteit van deze toets is in kaart gebracht en als voldoende beoordeeld (van Til et al., 2018b). Wij hebben van deze toets het hoogst behaalde beheersingsniveau, moment van afname en toetsversie (de instructie was versie B) opgevraagd.

3. Cito – Begrijpend Lezen

Deze gestandaardiseerde toets heeft als meetpretentie om de begripsvaardigheid van geschreven teksten in kaart te brengen. Deze toets is geschikt voor leerlingen waarbij het technisch lezen voldoende geautomatiseerd is en kan vanaf eind groep 3 worden afgenomen. Voor leerlingen uit groep 3 is er dus geen voormeting voor Begrijpend Lezen in dit onderzoek.

De toets bestaat uit een tekst met daarbij meerkeuzevragen over (een deel van) deze tekst. Hierbij wordt getoetst in welke mate de leerling de tekst begrijpt. De betrouwbaarheid en validiteit van de papieren toets is in kaart gebracht voor groep 3 en 4 apart en als voldoende beoordeeld (Jolink et al., 2015a; Jolink et al., 2015b). Wij hebben van deze toets de ruwe score, vaardigheidsscore, het niveau (I-V), moment van afname en toetsversie (de instructie was versie B) opgevraagd.

4. Cito – Woordenschat

De meetpretentie van deze gestandaardiseerde toets is om een indicatie te bieden van de grootte van de algemene woordenschat. Deze toets wordt vanaf groep 3 afgenomen (zowel op papier als digitaal mogelijk), met als eerste afnamemoment halverwege het schooljaar (jan). In groep 3 wordt de woordenschat gemeten in opgaven met context. Deze opgaven worden voorgelezen door de leerkracht en de context bestaat uit tekeningen of foto's. Vanaf groep 4 bestaat de toets zowel uit opgaven met als zonder context. Wij hebben van deze toets de ruwe score, vaardigheidsscore, het niveau (I-V), moment van afname en toetsversie (de instructie was versie B) opgevraagd.

5. Cito – Taalverzorging

Deze gestandaardiseerde toets heeft als doel te meten hoe goed een leerling kan spellen. Dit wordt vastgesteld door aan de leerling te vragen om woorden op te schrijven. Deze toets wordt vanaf groep 3 afgenomen, met als eerste afnamemoment halverwege het schooljaar (jan). In groep 3 en 4 worden alleen niet-werkwoorden geïncorporeerd in de toets. Werkwoordspelling komt namelijk pas aan bod in hogere leerjaren. De betrouwbaarheid en validiteit van de papieren toets is in kaart gebracht voor groep 3 en 4 apart en als voldoende beoordeeld (Tomesen et al., 2015a; Tomesen et al., 2015b). Wij hebben van deze toets de ruwe score, vaardigheidsscore, het niveau (I-V), moment van afname en toetsversie (de instructie was versie B) opgevraagd.

6. Cito – Rekenen-Wiskunde

Deze gestandaardiseerde toets heeft als meetpretentie om de algemene rekenvaardigheid van de leerling in beeld te brengen. Deze toets wordt vanaf groep 3 afgenomen (zowel op papier als digitaal mogelijk), met als eerste afnamemoment halverwege het schooljaar (jan). De opgaven bestaan uit een plaatje. De leerkracht leest hierbij een tekst voor wanneer de

toets op papier wordt afgenomen, anders leest de computer de tekst voor. De meeste opgaven vragen om een kort antwoord, en enkele opgaven betreffen meerkeuzevragen. De betrouwbaarheid en validiteit van deze toets is in kaart gebracht voor groep 3 en 4 apart en als voldoende beoordeeld (Janssen et al., 2015a; Janssen et al., 2015b). Wij hebben van deze toets de ruwe score, vaardigheidsscore, het niveau (I-V), moment van afname en toetsversie (de instructie was versie B) opgevraagd.

7. Leerlingenvragenlijst leesmotivatie

Naast leerlingprestaties zijn de leesmotivatie van de leerling (voormeting en nameting) en de ervaringen van leerlingen met de interventie (nameting) in kaart gebracht. Dit is gemeten middels digitale vragenlijsten, die door een leerkracht of tutor samen met de leerling werden ingevuld. Leesmotivatie is gemeten in acht items. Zie Figuur 3.4 voor een voorbeelditem. Voor leesmotivatie zijn op basis van betrouwbaarheid- en factoranalyses drie items verwijderd die onvoldoende bijdroegen aan de interne consistentie. De uiteindelijke schaal bestaat uit vijf items, gericht op plezier in lezen ('Vind je lezen leuk?', 'Lees je graag als je thuis bent?'), intrinsieke leesinteresse ('Vind je het leuk om nieuwe dingen te leren door te lezen?', 'Vind je het leuk om een boek cadeau te krijgen?') en zelfperceptie van leesvaardigheid ('Vind je zelf dat je goed bent in lezen?'). De interne consistentie van de schaal was matig tot net acceptabel (Cronbach's α variërend van ,61 op de voormeting tot ,69 op de nameting). Voor de analyses is een gemiddelde schaalscore berekend, waarbij hogere waarden een sterkere leesmotivatie weerspiegelen.

	Nee, helemaal 	Nee, niet echt 	Soms 	Ja, best wel 	Ja, heel erg 	Weet ik niet/kind begrijpt vraag niet
Vind je lezen leuk?	☐	☐	☐	☐	☐	☐

Figuur 3.4 Voorbeelditem leesmotivatie leerlingen

Bij de nameting kregen interventieleerlingen daarnaast tien vragen voorgelegd over hoe zij het oefenen met de Bliksembende hebben ervaren, zoals of ze het leuk of leerzaam vonden en hoe de sessies verliepen. Een voorbeelditem is 'Hielp de tutor jou goed als jij vragen had?', met de antwoordopties *Nee helemaal niet* tot *Ja heel erg*.

8. Herfstsignalering Veilig Leren Lezen groep 3

De herfstsignalering van Veilig Leren Lezen is alleen tijdens de screening onder groep 3-leerlingen afgenomen. De meetpretentie van dit methodegebonden instrument, zoals gebruikt voor dit onderzoek, is inzicht krijgen in de letterkennis van de tot dan aangeboden letters en de leesvaardigheid van woorden. De herfstsignalering bestaat uit de gestandaardiseerde toetsen 'Letterkennis aangeboden letters', 'Woordleestoets Veilig & Vlot', 'Fonemendictee' en 'Spellingtoets'. De toetsen zijn op papier afgenomen door een leerkracht, student-assistent of de onderzoekers. Binnen de twee toetsen is zowel gescoord op leesaccuratesse als leessnelheid. Aangezien er geen normscores beschikbaar zijn voor deze taak, en de betrouwbaarheid en validiteit verder onbekend zijn, hebben we gekozen om

leerlingen te selecteren die in de laagste 30% scoorden op de woordleestaak (minder dan 12 woorden goed gelezen).

Interventieverloop en ervaringen

Tijdens de interventieperiode zijn meerdere instrumenten ingezet om de frequentie en kwaliteit van de interventie evenals de ervaringen in kaart te brengen.

1. Logboeken interventiegroep

De tutores kregen de opdracht om per tutorsessie een papieren logboek in te vullen, met als meetpretentie om inzicht te krijgen in hoe de tutorsessies over het algemeen zijn verlopen. Tutores werden gevraagd om per duo eventuele onregelmatigheden/bijzonderheden te noteren op het gebied van samenwerking en het leesproces, en daarnaast aan te geven in welke mate er geconcentreerd is gewerkt (zie Figuur 3.5). In dit logboek konden tutores ook aangeven indien er gewisseld werd in de samenstelling van duo's.

	Leerling 1 Aanwezig?	Leerling 2 Aanwezig?	Op tijd begonnen?	Is er geconcentreerd gewerkt? (helemaal niet - helemaal wel)	Duur tutorsessie (min)
Team 1	Ja/Nee	Ja/Nee	Ja/Nee	○ ○ ○ ○ ○	○

Figuur 3.5 Voorbeelditem logboek interventiegroep

2. Log-gegevens uit backend Bliksembende

Uit de backend van het programma hebben wij verschillende informatie opgehaald over de leerlingen die met de BliksemBende hebben gewerkt. Ten eerste konden wij hiermee nagaan of leerlingen uit de controlegroep niet toch met het programma hebben gewerkt (contaminatie), of juist dat interventieleerlingen in de praktijk nooit met het programma hebben geoefend. Daarnaast gaven de log-gegevens uit de backend informatie over het aantal tutorsessies per leerling, het startverhaal, het aantal voltooide verhalen, het aantal woorden dat de leerling tijdens de instaptoets kon lezen per minuut en het aantal woorden dat de leerling tijdens de laatste tutortoets kon lezen per minuut.

3. Logboeken controlegroep

De logboeken voor de controlegroep dienden om bij te houden hoeveel ondersteuning leerlingen in de controlegroep ontvingen gedurende de interventieperiode. Leerkrachten in de controlegroep zijn gevraagd om hiervoor een tweewekelijkse vragenlijst in te vullen. Per leerling in de controlegroep werden leerkrachten naar de type(s) leesondersteuning die de leerling ontving en de totale duur leesondersteuning in minuten in de afgelopen twee weken. Zie Figuur 3.6 voor een voorbeelditem. Aan leerkrachten in de controlegroep zijn gedurende de interventieperiode tweewekelijks logboeken gestuurd in de vorm van digitale vragenlijsten om bij te houden hoeveel leesondersteuning deze leerlingen hebben ontvangen. De respons op deze vragenlijsten werd tussendoor bijgehouden door het onderzoeksteam. Aan sommige leerkrachten waarbij de respons achterwege bleef, hebben wij extra reminders gestuurd.

	Type(s) leesondersteuning (meer antwoorden mogelijk)	Totale duur leesondersteuning in minuten
Leerling 1	○ Bouw! ○ Lezen met een tutor ○ Een ander digitaal leesprogramma, namelijk ○ Anders, namelijk	[open antwoord]

Figuur 3.6 Voorbeelditem observatieschema tutorsessies

Ondanks reminders via e-mail en telefoon vulden sommige leerkrachten de vragenlijsten weinig in. In totaal is 68% van deze vragenlijsten ingevuld, nadat leerkrachten per email en telefoon hiervoor reminders hebben ontvangen. Van de 33 leerkrachten van klassen binnen de controlegroep, hebben 31 leerkrachten minstens één sessie ingevuld. Van 151 leerlingen in de controlegroep is minstens één logboek bijgehouden. Meerdere scholen gaven hierbij aan dat hun ondersteuningsaanbod nauwelijks verschilt van week tot week.

Om ook in de controlegroep betrouwbare schattingen van leesondersteuning te verkrijgen, is bij ontbrekende logboeken daarom gebruikgemaakt van imputatie. Met een Monte Carlo-bootstrapanalyse is nagegaan of de gemiddelde duur van ondersteuning systematisch verschilt naar het aantal ingevulde logboeken (1 tot 6). Deze analyses laten zien dat de gemiddelde minuten ondersteuning duidelijk lager liggen bij slechts één of twee ingevulde logboeken (gemiddeld 62 respectievelijk 92 minuten). Vanaf drie logboeken stabiliseren de gemiddelden (99 minuten bij drie en 119 bij zes observaties). Dit wijst erop dat leerlingen met weinig ingevulde logboeken structureel lagere scores hebben en dus niet representatief zijn voor de groep als geheel.

Daarom is besloten alleen imputatie toe te passen voor leerlingen met minimaal drie ingevulde logboeken. Voor deze leerlingen zijn de ontbrekende waarden geïmputeerd op basis van het persoonlijke gemiddelde. In totaal zijn 97 datapunten geïmputeerd, verdeeld over 40 leerlingen.

4. Observaties

De drie schoolbegeleiders (vanuit het implementatieteam) hebben op elke school twee keer een sessie geobserveerd om de kwaliteit van de tutoring in kaart te brengen (en feedback te geven). Dit hebben zij gedaan aan de hand van een observatieschema. Dit schema bestond uit 15 items, onderverdeeld in drie categorieën: (1) organisatie van de tutoring, (2) kwaliteit van de taal-/leesinstructie, (3) betrokkenheid en samenwerking van de teams. Voor de categorie organisatie van de tutoring scoorden observators bijvoorbeeld 'De teams kunnen direct met de tutoring beginnen' als *niet zichtbaar*, *enigszins zichtbaar* of *duidelijk zichtbaar*. Ter beoordeling van de interbeoordelaarsbetrouwbaarheid tussen de drie schoolbegeleiders, hebben zij bij twee observaties samen gezeten en elk afzonderlijk het schema ingevuld. Uit de gewogen Cohen's Kappa's van de observaties samengevoegd, blijkt dat de interbeoordelaarsbetrouwbaarheid voldoende is. Deze maat corrigeert voor de kans dat beoordelaren toevallig hetzelfde scoren en houdt ook rekening met hoe ver de beoordelingen uit elkaar liggen. Over het algemeen wordt een score boven de ,60 als voldoende beschouwd. De gewogen Cohen's Kappa's van de twee observaties samengevoegd, lagen tussen de ,67 en ,92.

5. Tutorvragenlijst

Aan het einde van de interventieperiode zijn zelf ontwikkelde digitale vragenlijsten verspreid onder tutoren om hun perceptie van de interventie en het implementatieproces na te gaan. Deze vragenlijst bestond uit 16 items. Vier items gingen in op achtergrondinformatie van de tutor, zoals het aantal jaar werkervaring binnen het onderwijs en de functie binnen de school. De overige 12 items gingen over de ervaringen van tutoren met het werken met De BliksemBende en de ervaren opbrengsten van het programma. Tutoren gaven bijvoorbeeld aan ‘Kon u de leerlingen voldoende ondersteunen zodat zij succesvol met elkaar konden samenwerken?’ op een 5-punts Likertschaal (*Nee helemaal niet; Nee, niet echt; een beetje; Ja, best wel; Ja, heel erg*).

6. Leerkrachtvragenlijst

Aan het einde van de interventieperiode zijn zelf ontwikkelde digitale vragenlijsten verspreid onder leerkrachten van interventieklassen. Deze vragenlijst bestond uit 15 items. Twee items betroffen algemene vragen. De overige 13 items gingen in op hoe de leerkrachten terugblikten op de interventieperiode: hoe de organisatie volgens hen verliep rondom de tutorsessies en wat de ervaren opbrengsten waren van de leesinterventie. Zo gaven leraren bijvoorbeeld aan ‘Hebben de leerlingen volgens u teveel tijd gemist in het reguliere lesprogramma?’ op een 5-punt Likertschaal van *Nee, helemaal niet* tot *Ja, heel erg*.

7. Interviews tutorcoördinator

Alle scholen zijn uitgenodigd voor een semigestructureerd diepte-interview met de tutorcoördinator om te vertellen over de wijze waarop de interventie bij hen was georganiseerd. Hieraan hebben 4 scholen deelgenomen. De gesprekken duurden ongeveer een half uur. Deze interviews dienden als verdiepingsslag op de procesevaluatie van de interventieperiode.

3.4 Analysestrategie

Multilevel analyses (Snijders & Bosker, 2012) zijn gebruikt voor de analyses van de interventie-effecten. Analyses zijn uitgevoerd in R (R Core Team, 2018) met het nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018). Missings worden niet geïmputeerd, in plaats daarvan wordt gebruik gemaakt van Full Information Maximum Likelihood (FIML) om met ontbrekende waarden om te gaan. We hebben rekening gehouden met het gegeven dat leerlingen genest zijn in leerjaar per school, door middel van het toevoegen van random intercepts voor school-leerjaar. Het schoolniveau wordt niet meegenomen in deze modellen aangezien de randomisatie binnen scholen heeft plaatsgevonden en we maar kleine schooleffecten verwachten als er al is gecontroleerd voor het groepsniveau. Voor alle modellen wordt, naast random intercepts voor school-leerjaar, een Fixed effect van Interventie (controlegroep, interventiegroep) toegevoegd en wordt er gecorrigeerd voor Voormeting door deze toe te voegen als covariaat. We houden rekening met wijze van randomiseren (binnen jaarlaag of tussen jaarlagen) door een interactie tussen manier van randomisatie en Interventie toe te voegen. Een significante positieve estimate van Interventie zou betekenen dat leerlingen in de interventiegroep meer vooruitgaan dan leerlingen in de controlegroep.

Significantie van random intercepts wordt geschat via inspectie van de ICC en of toevoeging van intercept leidt tot een significant betere model fit aan de hand van een χ^2 -toets.

Significantie van Fixed effects worden getoetst door naar de p-waardes van het uiteindelijke model met alle Fixed effects te kijken en of het toevoegen van de Fixed effects leidt tot een betere model fit aan de hand van een χ^2 -toets. Binnen de interventiegroep wordt in Multilevel analyses gekeken of het aantal sessies en het aantal voltooide verhalen voorspellend is voor de vooruitgang.

Voor de analyses voor de follow-up meting gebruiken we ook Multilevel regressies met een random intercept voor school-leerjaar. Hier wordt Tijd als Within factor gebruikt (voormeting; nameting; follow-up) en een Fixed effect van Interventie. Ook hier wordt gecontroleerd voor randomisatie door een interactieterm toe te voegen. Door contrasten te inspecteren in de estimated marginal means, is onderzocht of leerlingen vooruit zijn gegaan tussen T1 en T2 en tussen T2 en T3, en of deze vooruitgang verschilde per groep. Daarnaast is onderzocht of de controlegroep en interventiegroep van elkaar verschillen op T2 en T3.

In de analyses met de follow-up meting wordt ook onderzocht of de vooruitgang in woordenschat een mediërend effect heeft op de leesontwikkeling, als mogelijk werkzaam mechanisme van de interventie. De mediatieanalyses zijn uitgevoerd met *lavaan* (versie 0.6-19) in R. De interventie (0 = controle, 1 = interventie) werd gemodelleerd als voorspeller van woordenschat op de nameting (T2), die op zijn beurt de uitkomsten op de follow-up (T3) voorspelde. Voor zowel de mediator als de uitkomst is, waar beschikbaar, gecontroleerd voor het bijbehorende baseline-niveau (T1). De modellen zijn geschat met Full Information Maximum Likelihood (FIML) om met ontbrekende waarden om te gaan, en de standaardfouten zijn gecorrigeerd voor clustering binnen school-leerjaar met de robuuste MLR-schatting. Indirecte, directe en totale effecten zijn in het model gedefinieerd en getoetst met 95%-betrouwbaarheidsintervallen.

A priori poweranalyse

Vooraf is een poweranalyse uitgevoerd die uitging van 48 clusters (24 deelnemende scholen met elk een controlegroep en interventiegroep), en een clustergrootte van 12. Dit komt neer op totaal 24 deelnemende leerlingen op een school, een realistisch schatting aangezien 10 scholen met meer dan 2 klassen meedoen. Uitgaande van een conservatieve effect-size van 0,35 wordt er een power bereikt van 0,80 om een effect van deze grootte te detecteren. In deze power-berekening is geen rekening gehouden met de verklaarde variantie van de voormeting op de nameting, waardoor de daadwerkelijke power hoger met deze aantallen hoger uit zal vallen. De uiteindelijke steekproef bestaat uit 46 clusters met een gemiddelde clustergrootte van 7,4.

Uitval en missende data

Er heeft nauwelijks uitval plaatsgevonden na de start van de interventie (één school was uitgevallen tijdens de screening van de leerlingen). Tijdens het project zijn één leerling in de controlegroep en één leerling in de interventiegroep uitgevallen wegens verhuizing. Van deze laatste leerling zijn ook geen gegevens voor de voormeting opgeleverd.

Doordat in dit project ingezet op een intensieve begeleiding vanuit het implementatieteam bleven de lijntjes met de scholen kort en zijn eventuele knelpunten tijdens de loop van het

project relatief snel opgespoord. Deze intensieve begeleiding heeft waarschijnlijk voor weinig uitval van scholen tijdens dit project gezorgd en voor een hoge implementatietrouw van de interventie.

Eén school heeft de Cito-woordenschat niet altijd afgenomen. Bij de nameting ontbreken relatief veel scores voor de nameting van Begrijpend Lezen in groep 3: deze taak is niet betrouwbaar in groep 3 als de technische leesvaardigheid niet op het gemiddelde niveau van eind groep 3 is. Daardoor hebben scholen er gezien de doelgroep van het onderzoek – zwakke lezers – relatief vaak voor gekozen om de toets niet af te nemen.

4 Resultaten

In dit hoofdstuk presenteren we de resultaten van het onderzoek. Eerst beschrijven we het verloop van de interventie en de interventietrouw. Vervolgens gaan we in op de effecten van de interventie op verschillende leeruitkomsten. Hierbij bespreken we mogelijke moderatoren en mediators die inzicht geven in voor wie en onder welke omstandigheden de interventie effectief is. Ten slotte bespreken we de implementatie en ervaringen van leerlingen, leerkrachten en tutores in de procesevaluatie

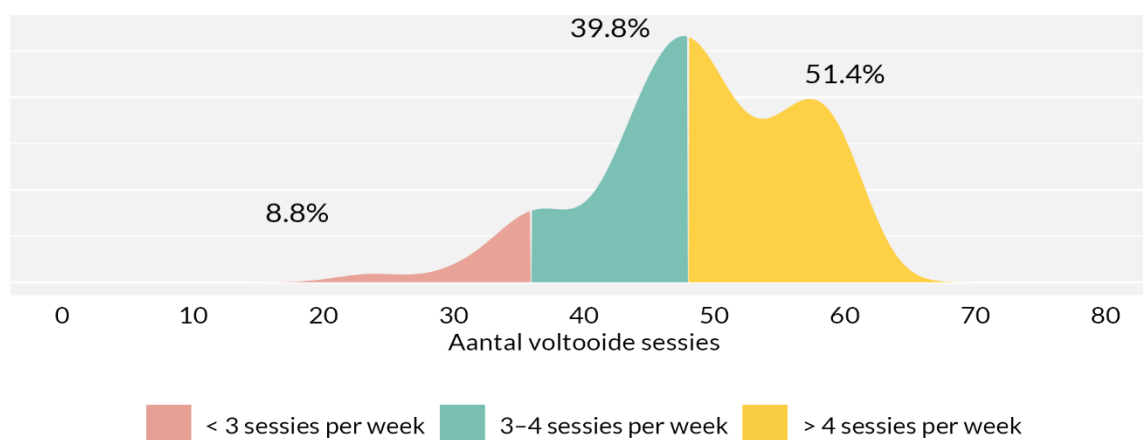
4.1 Implementatietrouw en verloop interventie

Interventiegroep

De implementatietrouw en verloop van de interventie zijn op meerdere manieren gemeten op basis van gegevens uit de back-end van BliksemBende, logboeken die ingevuld zijn door tutores en observaties van de schoolbegeleiders.

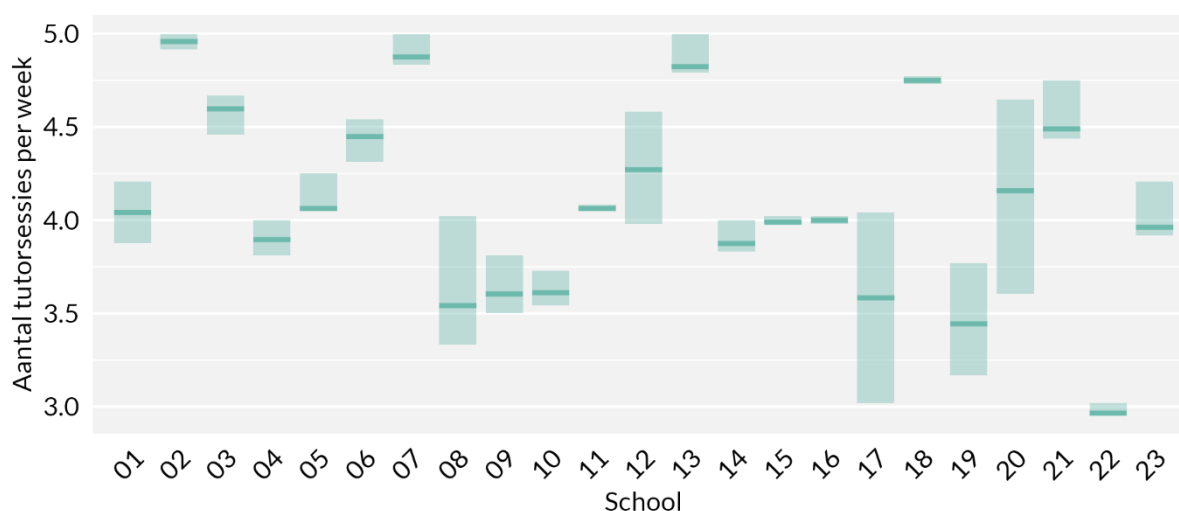
Oefenfrequentie: Log-gegevens uit backend BliksemBende

Uit de log-gegevens blijkt dat leerlingen gemiddeld 49 sessies voltooiden gedurende de interventieperiode (variërend van 23–68), wat neerkomt op een gemiddelde dosis van vier sessies per week. Slechts tien leerlingen hebben minder dan het minimum aantal sessies (36) behaald (Figuur 4.1). Dit komt doordat zij het moeilijkste verhaal in de BliksemBende gedurende de interventieperiode hadden voltooid en daarna minder sessies hebben gekregen.



Figuur 4.1 Het aantal voltooide tutorsessies per week

In Figuur 4.2 is af te lezen dat de meeste scholen rond dit gemiddelde van vier sessies zitten. Op 10 scholen zijn gemiddeld minder dan vier wekelijkse sessies voltooid, maar op elke school zijn minstens drie wekelijkse sessies gedaan – het minimum aantal benodigde sessies op basis van eerder onderzoek. Scholen verschillen onderling in de spreiding van het aantal sessies, maar over het algemeen kunnen we vaststellen dat scholen zich redelijk goed aan het doel hebben gehouden om vier á vijf sessies per week te organiseren voor elke deelnemende leerling.



Figuur 4.2 Gemiddeld aantal voltooide tutorssessies per school. De helft van de leerlingen valt binnen de gekleurde vlakken

In beide jaarlagen zijn leerlingen tijdens de interventie significant vooruitgegaan in de moeilijkheidsgraad van het verhaal en leessnelheid binnen de BliksemBende (Tabel 4.1). De leessnelheid aan de start wordt hierbij bepaald door het aantal gelezen woorden per minuut binnen de instaptoets voor de BliksemBende. De leessnelheid aan het eind wordt bepaald door het aantal gelezen woorden per minuut binnen de laatste tutortoets.

Tabel 4.1 Gemiddeld aantal sessies, voltooide verhalen en gelezen woorden per minuut binnen de BliksemBende

	Groep 3		Groep 4	
	Gem (SD)	Range	Gem (SD)	Range
Aantal sessies	53,32 (5,64)	40 - 60	44,71 (7,50)	23 - 60
Startverhaal	2,27 (1,59)	1 - 6	20,24 (11,23)	1 - 43
Eindverhaal	18,45 (5,61)	11 - 37	39,70 (13,23)	16 - 50
Programmagebonden woordleestaak T1	9,66 (11,74)	15 - 68	62,66 (17,27)	34 - 104
Programmagebonden woordleestaak T2	56,64 (20,67)	16 - 129	83,01 (18,44)	36 - 135

Planning tutorsessies

De tutorsessies werden voor nagenoeg alle leerlingen ingepland tijdens een vast tijdstip, of les, op de dag. Zo blijkt uit de vragenlijsten die leerkrachten na afloop van de interventieperiode invulden. Slechts één leerkracht gaf aan dat dit niet het geval was. De leerkrachten gaven aan dat de tutorsessies werden ingepland op een vast tijdstip en verschillende lessen (n = 14), tijdens rekenen (n = 1), tijdens taallessen (n = 1), tijdens creatieve vakken of zaakvakken (n = 3), of tijdens stillezen (n = 1).

Leerkrachten varieerden sterk in hoeveel overleg ze hadden gedurende de interventieperiode met de tutor(coördinator). Zeven leerkrachten hadden eens of meermaals per week overleg. Zeven andere leerkrachten hadden ongeveer eens in de twee weken tot eens per maand overleg. Drie leerkrachten hadden eens per zes weken tot eens in de twee maanden overleg. De overige vier leerkrachten hadden tweemaal, eenmaal, of nooit overleg gehad met de tutor(coördinator). Deze overleggen konden over verschillende dingen gaan:

1. de algemene voortgang van de leerlingen (n = 17);
2. de effectiviteit van lezen met de BliksemBende (n = 10);
3. in welke mate de kinderen goed samenwerkten (n = 13);
4. de mate van aansturing die nodig was bij de leerlingen (n = 10);
5. de planning (n = 1);
6. de (mogelijke) nadelen van de inzet van lezen met de BliksemBende (n = 1).

Verloop afzonderlijke tutorsessies: logboeken interventiegroep

Volgens de tutores verliepen de sessies overwegend vloeiend. Tutores hebben voorafgaand aan elke sessie in een logboek per duo bijgehouden of de sessie heeft plaatsgevonden en of er geconcentreerd is gewerkt. In totaal zijn er 4343 logboeken (deels) ingevuld. Tijdens de sessies hebben de leerlingen volgens de tutores overwegend geconcentreerd gewerkt en namen de tutorsessies, zoals bedoeld, ongeveer 25 minuten in beslag (Tabel 4.2). Daarnaast zijn nagenoeg alle tutorsessies op tijd volgens planning gestart en was het duo in de meeste gevallen compleet (Tabel 4.3).

Tabel 4.2 Beoordeling verloop tutores per tutorsessie

	Gem	SD	Range	Aantal sessies
Is er geconcentreerd gewerkt? (1 = helemaal niet – 5 = helemaal wel)	4,23	0,84	1 - 5	3453
Duur tutorsessie (min)	25,94	3,58	3 – 45	3449

Tabel 4.3 Procesevaluatie door tutores per tutorsessie

	Ja	Nee	Niet ingevuld
Op tijd begonnen met de tutorsessie?	3360	81	902
Beide leerlingen aanwezig?	3343	313	687

Kwaliteit van de sessies: Tussentijdse observaties door schoolbegeleiders

Volgens de schoolbegeleiders vanuit het implementatieteam verliep de organisatie van de tutoring goed op basis van de twee momenten waarop zij op elke school een tutorsessie hebben geobserveerd. De schoolbegeleider beoordeelde de tutor op aspecten organisatie van de tutoring, kwaliteit van de instructie, en betrokkenheid en samenwerking (1 = *niet zichtbaar*, 2 = *enigszins zichtbaar*, 3 = *duidelijk zichtbaar*). De scores op de meeste items waren hoog met beperkte variatie tussen scholen, wat wijst op een consistent hoge implementatiekwaliteit. Uit observaties bleek dat de instructie goed was als het ging om overzicht houden en uitleg en feedback geven (>80% van de scholen scoorde maximaal op deze items).. De betrokkenheid en samenwerking was redelijk zichtbaar, waarbij expliciete aandacht voor en feedback op samenwerken door de tutor hoog werd beoordeeld. Daarentegen werden check-ins en check-outs – zowel met betrekking tot het leesproces als het samenwerken – structureel weinig gedaan. Vanwege het plafondeffect op de meeste items is het niet mogelijk om variatie in implementatiekwaliteit te relateren aan verschillen in leerlinguitkomsten. De drie items met voldoende spreiding (check-out leesproces, teams helpen elkaar, expliciete aandacht voor samenwerken) betreffen randvoorwaarden van de implementatie en niet de kernkwaliteit van de instructie (zoals overzicht houden, uitleg en feedback geven), die op alle scholen als goed werd beoordeeld.

Tabel 4.4 Tussentijdse observaties door schoolbegeleiders

	Ronde 1			Ronde 2		
	Gem	SD	Range	Gem	SD	Range
Organisatie van de tutoring	2,87	0,28	2,00 – 3,00	2,90	0,21	2,20 – 3,00
Kwaliteit van de instructie	2,23	0,46	1,20 – 3,00	2,20	0,33	1,60 – 3,00
Betrokkenheid en samenwerking van de teams	1,95	0,48	1,20 – 3,00	2,03	0,44	1,20 – 3,00

Type tutoren

Tutoren vervulden, naast het tutorschap, verschillende functies binnen de school, zo blijkt uit vragenlijsten die tutoren invulden. Van de 26 tutoren, die op 15 verschillende scholen werkten, is het merendeel werkzaam als onderwijsassistent (n = 17). De overige tutoren zijn ib'er (n = 2), stagiair (n = 4), vrijwilliger, directie of ambulant begeleider. Drie van de tutoren hadden maximaal één jaar werkervaring binnen het onderwijs. De rest van de tutoren hadden allen minimaal 10 jaar werkervaring. Gezien het gebrek aan variatie in de achtergrondkenmerken van tutoren en het ontbreken van informatie over het type tutor op acht scholen, is het niet goed mogelijk om te toetsen of type tutoren verschillen in effectiviteit.

Controlegroep

Leerlingen in de controlegroep hebben tijdens de interventieperiode niet geoefend met de BliksemBende. Dit is geconstateerd uit de logdata van het programma, waarin we konden zien welke leerlingen precies oefenden. Leerkrachten in de controlegroep gaven in de tweewekelijkse logboeken aan dat de meest voorkomende leesondersteuning lezen met een tutor is (n = 423). Er is slechts vijfmaal aangegeven dat een leerling tijdens de leesondersteuningsmomenten alleen las met een medeleerling, en viermaal dat een leerling

alleen oefende met een computerprogramma. Vaker werd een combinatie van de drie bovenstaande typen leesondersteuning toegepast ($n = 83$). Er werd 109 keer aangegeven dat leerlingen in de afgelopen twee weken géén leesondersteuning ontvingen en 41 maal dat de leerkracht dit eigenlijk niet wist. Voor 22 leerlingen werd binnen alle ingevulde logboeken gerapporteerd dat deze géén leesondersteuning ontvingen.

De gemiddelde duur van gegeven leesondersteuning is 45 minuten per week (zie Tabel 4.5). Uit de hoge standaarddeviatie en range van het aantal minuten, kan worden afgeleid dat er een grote spreiding zit in de duur van de gegeven leesondersteuning.

Tabel 4.5 Duur leesondersteuning leerlingen in controlegroep per twee weken

	Gem	SD	Range	Mediaan
Duur leesondersteuning per 2 weken	89,45	73,96	0 – 306,67	103,00

Ondersteuning na de interventie

Na de interventieperiode en vóór de follow-up meting waren scholen vrij om leerlingen te laten oefenen met De BliksemBende. In de interventiegroep is daar door slechts 12 leerlingen gebruik van gemaakt. Zij ontvingen gemiddeld 26,4 sessies. Daarnaast hebben 14 leerlingen uit de controlegroep na de interventieperiode gebruik gemaakt van het programma (met een gemiddelde van 39,9 sessies).

4.2 Effectevaluatie

In deze paragraaf presenteren we de resultaten van de effectevaluatie. We gaan in op de vraag of deelname aan de BliksemBende heeft geleid tot verbetering van de leesvaardigheid van leerlingen en of er neveneffecten zijn op andere basisvaardigheden, zoals rekenen-wiskunde. Eerst beschrijven we de gemiddelde toetsscores van de leerlingen in de interventie- en controlegroep. Vervolgens analyseren we de effecten van de interventie op de verschillende toetsen en bekijken we in hoeverre deze effecten standhouden tot de follow-upmeting.

Gemiddelde scores

De effectiviteit van de BliksemBende wordt beoordeeld aan de hand van vooruitgang op de Cito-scores op de voormeting, nameting en follow-up. De vaardigheidsscores per groep zijn weergegeven in Tabel 4.6. Voor AVI zijn de behaalde niveaus omgezet naar een numerieke score (waarbij AVI-start = 1, M3 = 2, E3 = 3, etc.). Uit Tabel 4.6 blijkt dat leerlingen in zowel de interventie- als de controlegroep gemiddeld vooruitgingen op vrijwel alle uitkomstmaten tussen de voormeting, nameting en follow-up. Dit geldt met name voor technisch lezen (DMT, AVI) en rekenen-wiskunde, waar substantiële gemiddelde stijgingen zichtbaar zijn. Op begrijpend lezen en taalverzorging is eveneens vooruitgang zichtbaar, al verloopt de groei in beide groepen niet gelijkmatig. Voor woordenschat en leesmotivatie zijn de gemiddelde verschillen tussen meetmomenten klein. Deze beschrijvende resultaten laten zien dat alle leerlingen zich gedurende de onderzoeksperiode hebben ontwikkeld, maar dat de cruciale vraag is of de vooruitgang in de interventiegroep groter is dan in de controlegroep.

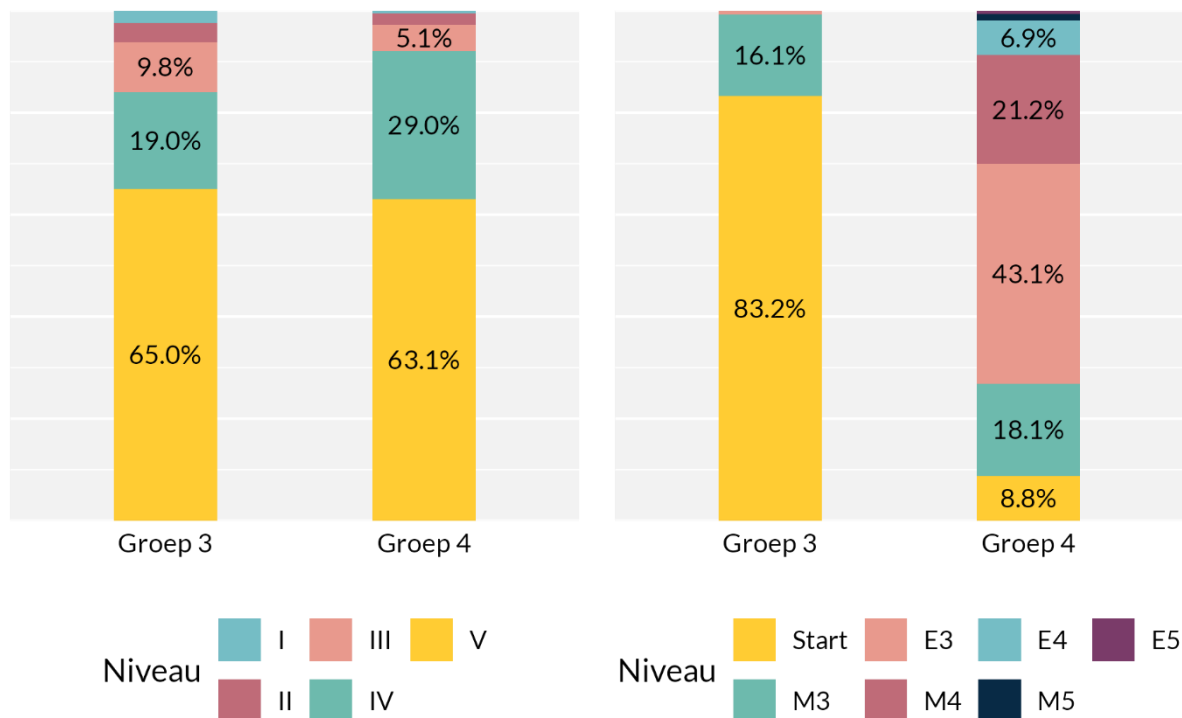
Tabel 4.6 Gemiddelde Cito vaardigheidsscores (standaarddeviatie) in de controlegroep en interventiegroep

Meetinstrument	Meetmoment	Controlegroep		Interventiegroep	
		groep 3	groep 4	groep 3	groep 4
DMT	Voormeting	10,53 (11,26)	24,08 (11,67)	6,75 (3,36)	24,42 (10,49)
	Nameting	19,31 (13,95)	32,08 (13,6)	15,26 (8,37)	32,41 (13,49)
	Follow-up	31,77 (17,55)	39,42 (15,41)	29,67 (19,33)	40,56 (16,77)
AVI	Voormeting	1,27 (0,48)	2,98 (1,19)	1,08 (0,27)	3,14 (1,02)
	Nameting	2,37 (1,03)	4,05 (1,55)	2,00 (1,11)	4,55 (1,50)
	Follow-up	3,87 (1,76)	5,42 (2,13)	3,44 (1,78)	5,40 (1,42)
Begrijpend Lezen	Voormeting	-	123,91 (20,53)	-	122,64 (22,33)
	Nameting	112,70 (33,18)	129,55 (23,67)	109,08 (32)	136,90 (22,41)
	Follow-up	128,24 (23,38)	152,75 (23,19)	122,19 (21,74)	151,41 (22,84)
Taalverzorging	Voormeting	122,37 (52,66)	226,08 (37,94)	117,05 (52,36)	236,31 (38,44)
	Nameting	184,17 (52,44)	247,94 (31,05)	190,38 (40,89)	258,27 (30,12)
	Follow-up	219,33 (47,39)	257,28 (30,04)	208,90 (46,97)	246,96 (76,39)
Woordenschat	Voormeting	452,04 (32,30)	465,43 (27,60)	453,86 (27,51)	470,81 (31,80)
	Nameting	451,59 (30,53)	472,75 (26,36)	452,59 (30,30)	478,44 (30,88)
	Follow-up	461,33 (27,34)	475,74 (21,44)	459,41 (30,74)	480,61 (25,27)
Rekenen-Wiskunde	Voormeting	101,76 (30,22)	161,84 (27,15)	107,38 (25,06)	156,03 (27,41)
	Nameting	132,75 (27,83)	188,23 (33,99)	128,80 (25,97)	178,98 (29,62)
	Follow-up	155,61 (33,48)	195,96 (27,33)	153,39 (31,03)	194,01 (29,54)
Leesmotivatie	Voormeting	4,04 (0,77)	3,70 (0,67)	3,70 (0,75)	3,70 (0,68)
	Nameting	3,95 (0,75)	3,57 (0,73)	3,72 (3,71)	3,71 (0,74)

De controlegroep en interventiegroep verschillen niet significant op de voormeting als groep 3 en 4 samen worden genomen. Binnen enkel groep 3 scoort de interventiegroep wel significant lager op de DMT, $t(85,248) = 2,806$, $p = ,006$. Op de andere uitkomstmaten verschillen de groepen niet op de voormeting. De correlaties tussen uitkomstmaten en voorspellers zijn weergegeven in Tabel A.2 van de bijlage.

Op de maten van technisch lezen is duidelijk af te lezen dat de leerlingen op het moment van de voormeting vallen onder de groep zwakke lezers, zie Figuur 4.3. Op de DMT scoort de overgrote meerderheid op niveau IV of V in beide jaarlagen. In groep 3 las meer dan 80% van

de leerlingen op AVI-niveau Start. Een klein deel had een AVI-beheersingsniveau van midden groep 3 (M3). In groep 4 las ongeveer één op de vijf leerlingen op het beheersingsniveau M4, en lazen de meeste andere leerlingen onder dit niveau.



Figuur 4.3 Niveau-indeling technisch lezen tijdens de voormeting

Interventie-effecten: BliksemBende heeft vooral een positief effect op begrijpend lezen

Voor elke uitkomst op de nameting is een aparte multilevel regressieanalyse uitgevoerd, met de predictor Interventie (controle/interventie) en de voormeting als covariaat. Een significant deel van de variantie op de Cito-uitkomsten wordt verklaard door verschillen tussen jaarlagen op scholen, de ICC op de voormeting varieert van 0,29 tot 0,49.

Voor begrijpend lezen geldt dat de analyse alleen is uitgevoerd in groep 4, omdat in groep 3 geen voormeting plaatsvond. In Tabel B.1 is een overzicht gegeven van de uitkomsten van de initiële regressieanalyses. Echter, aangezien de randomisatie op 2 andere scholen op een andere manier heeft plaatsgevonden, is het van belang om hier rekening mee te houden in de analyses, voor een accurate schatting van de effecten. Om deze reden baseren we onze interpretaties op de hoofdanalyse. In de hoofdanalyse modelleren we ook een eventueel effect van schoolrandomisatie (binnen een jaarlaag, tussen jaarlagen) en analyseren we of de effectiviteit van de interventie mogelijk afhangt van de wijze waarop is gerandomiseerd op scholen. Hierom is het interactie-effect schoolrandomisatie * Interventie (controle of interventie) opgenomen. De uitkomsten van deze analyses zijn opgenomen in

Tabel 4.7. We kijken hier voornamelijk naar de significantie van het hoofdeffect van Interventie. In Tabel C.1 wordt gerapporteerd over de behaalde power voor deze hoofdanalyses.

Tabel 4.7. Hoofdanalyse: Interventie-effecten controlerend voor randomisering

		Coëfficiënt	SE	T-waarde	p-waarde
DMT (n = 338)	Intercept	8,09	0,79	10,28	< ,001
	Voormeting	1,02	0,03	33,80	< ,001
	Schoolrandomisatie	-2,23	1,94	-1,15	,257
	Interventie	-0,38	0,80	-0,48	,630
	Schoolrandomisatie * Interventie	3,71	2,66	1,40	,164
AVI (n = 279)	Intercept	0,96	0,16	6,15	< ,001
	Voormeting	1,07	0,05	20,72	< ,001
	Schoolrandomisatie	-0,40	0,31	-1,28	,208
	Interventie	0,01	0,15	0,05	,962
	Schoolrandomisatie * Interventie	0,542	0,387	1,40	,163
Begrijpend Lezen (n = 149)	Intercept	55,65	9,63	5,78	< ,001
	Voormeting	0,60	0,07	7,98	< ,001
	Schoolrandomisatie	6,56	8,48	0,78	,449
	Interventie	8,17	4,07	2,01	,047
	Schoolrandomisatie * Interventie	-3,61	10,47	-0,35	,731
Taalverzorging (n = 338)	Intercept	113,91	4,93	23,09	< ,001
	Voormeting	0,59	0,02	25,86	< ,001
	Schoolrandomisatie	-3,48	8,35	-0,42	,679
	Interventie	4,77	3,93	1,21	,226
	Schoolrandomisatie * Interventie	23,02	10,01	2,30	,022
Woordenschat (n = 304)	Intercept	154,09	20,15	7,65	< ,001
	Voormeting	0,67	0,04	15,31	< ,001
	Schoolrandomisatie	16,20	8,13	1,99	,054
	Interventie	0,83	4,30	0,19	,847
	Schoolrandomisatie * Interventie	-12,52	8,81	-1,42	,156
Leesmotivatie (n = 296)	Intercept	1,32	0,20	6,43	< ,001
	Voormeting	0,64	0,05	12,95	< ,001
	Schoolrandomisatie	-0,21	0,19	-1,11	,274
	Interventie	0,02	0,09	0,22	,826
	Schoolrandomisatie * Interventie	0,20	0,24	0,82	,413
Rekenen-Wiskunde (n = 332)	Intercept	42,21	4,40	9,59	< ,001
	Voormeting	0,89	0,03	30,00	< ,001
	Schoolrandomisatie	-1,04	5,94	-0,18	,862
	Interventie	-6,33	2,58	-2,45	,015
	Schoolrandomisatie * Interventie	2,13	7,54	0,28	,777

Op het gebied van taalvaardigheid zien we duidelijk géén significante interventie-effecten op de uitkomstmaten van technisch lezen (DMT en AVI), met voor beide uitkomstmaten een hoge p-waarde. Ook voor woordenschat is er geen significant interventie-effect. Voor begrijpend lezen is er wel een significant effect van de interventie te zien. Leerlingen die de interventie hebben gevolgd, scoren bijna 8 punten hoger op de vaardigheidsscore Begrijpend Lezen tijdens de nameting.

Op Rekenen-Wiskunde is daarentegen duidelijk een *negatief* interventie-effect te zien. De controlegroep gaat méér vooruit op Rekenen-Wiskunde dan de interventiegroep, met een verschil van 6 punten op de vaardigheidsscore.

De resultaten voor Taalverzorging zijn opvallend, want uit de hoofdanalyse blijkt geen significant effect van de interventie, terwijl in de initiële analyses (waarin het effect van schoolrandomisatie *niet* is genomen, zie Tabel B.1), er *wel* een significant effect van de interventie was. In de hoofdanalyse verdwijnt interventie-effect. Tegelijkertijd zien we een significant interactie-effect tussen Schoolrandomisatie en Interventie (Tabel 4.7). Een eventueel interventie-effect lijkt groter te zijn op scholen waarbij de random toewijzing anders heeft plaatsgevonden. Dit betekent dat het eerder gevonden interventie-effect op taalverzorging grotendeels wordt gedreven door verschillen tussen deze scholen.

Voor de zekerheid hebben we ook de regressieanalyses uitgevoerd *zonder* de twee scholen mee te nemen waarop binnen jaarlaag is gerandomiseerd (gerapporteerd in Tabel B.1). De resultaten komen overeen met de hoofdanalyses, met een bijna significant effect voor Begrijpend Lezen ($p = 0,062$) en een significant effect op Rekenen-Wiskunde ($p = 0,016$) en geen effect op Taalverzorging ($p = 0,225$). Dit ondersteunt de interpretatie dat het effect op Taalverzorging grotendeels lijkt te worden gedreven door de twee scholen waarop afwijkend is gerandomiseerd; het effect is niet significant als deze scholen buiten beschouwing worden gelaten.

Naast de Cito-toetsen hebben we gekeken of de leesmotivatie is toegenomen door de interventie. Voor deze analyse geldt dat er geen significante verschillen in leesmotivatie worden gevonden tussen de controlegroep en de interventiegroep.

Hoewel de implementatietrouw goed te noemen is, heeft een aantal leerlingen minder geoefend dan het minimum aantal voorgeschreven sessies. Als leerlingen die minder dan 40 sessies hebben geoefend buiten beschouwing worden gelaten in de analyses, zijn er geen verschillen te zien in significantie van de interventie-effecten.

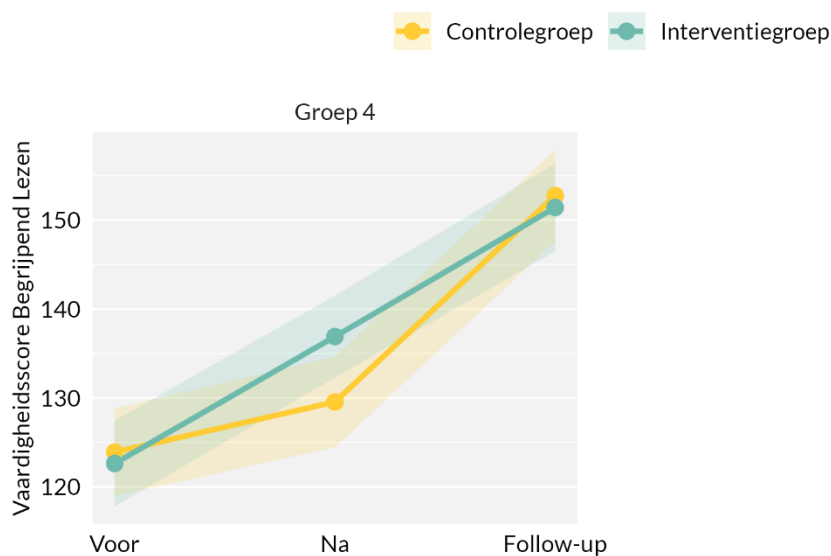
Positieve en negatieve effecten blijven niet hangen tot aan de follow-up meting

Voor de follow-up meting zijn multilevel analyses uitgevoerd waarin is gekeken of de ontwikkeling van voormeting naar nameting, tot aan de follow-up meting verschilt tussen de controlegroep en interventiegroep. Net als in de hoofdanalyses is gebruikgemaakt van een random intercept op school-leerjaar en is rekening gehouden met de wijze van randomisatie door dit als factor en in interactie met de interventiegroep op te nemen⁴. Voor deze en verdere aanvullende analyses in dit hoofdstuk rapporteren we alleen effecten op Begrijpend Lezen, Taalverzorging en Rekenen-Wiskunde, aangezien dit de enige uitkomsten zijn waarop de interventie een zeker effect lijkt te hebben. De analyses zijn ook uitgevoerd voor DMT, AVI, en Woordenschat, maar hieruit bleken geen significante relaties met relevante variabelen. Deze worden daarom niet gerapporteerd.

Zoals ook te zien is in Figuur 4.4, scoort de controlegroep lager dan de interventiegroep op de nameting op **Begrijpend Lezen**, maar wordt deze achterstand volledig ingehaald op het

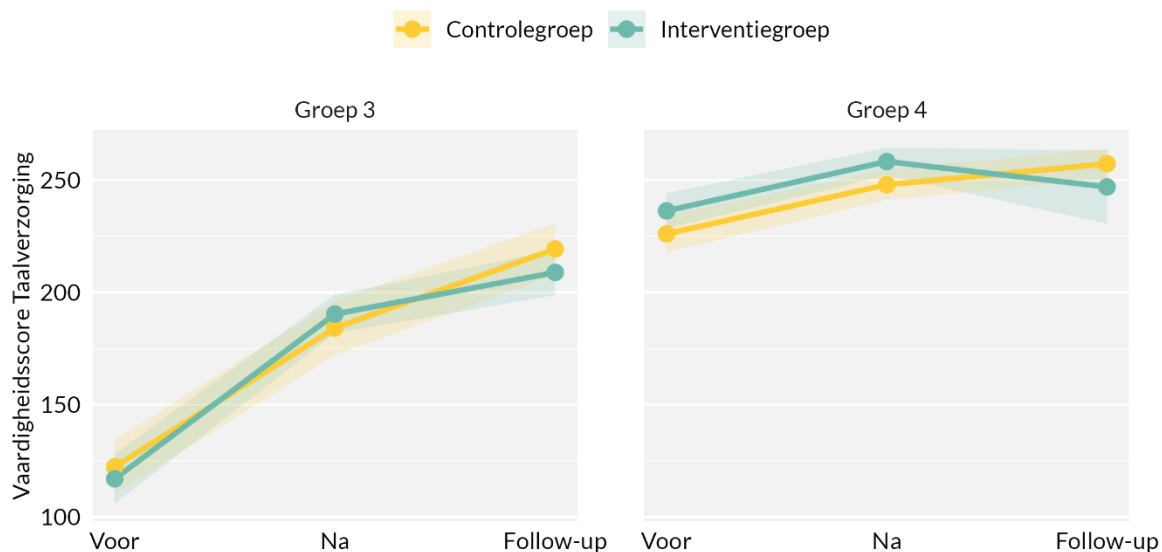
⁴ De significantie van interventie-effecten tussen T1 en T2 wijken af van de hoofdanalyse. Dit wordt waarschijnlijk veroorzaakt door een verminderde power in de huidige analyse doordat de voormeting niet als covariaat wordt meegenomen maar als een within-factor.

moment van de follow-up meting. De interventiegroep gaat significant vooruit tussen T1 en T2 ($p = ,03$), terwijl de controlegroep in diezelfde periode geen significante groei laat zien ($p = ,42$). Het verschil in groei tussen de groepen is echter niet significant ($p = ,41$) en de gemiddelden verschillen niet significant op T2 ($p = ,25$). Tussen T2 en T3 groeit de controlegroep significant verder ($p = ,004$), terwijl de interventiegroep een bijna-significante groei laat zien ($p = ,051$). Het verschil tussen de groepen in deze fase is opnieuw niet significant ($p = ,47$). Op T3 eindigen interventie- en controlegroep gemiddeld vrijwel gelijk ($p = ,86$).



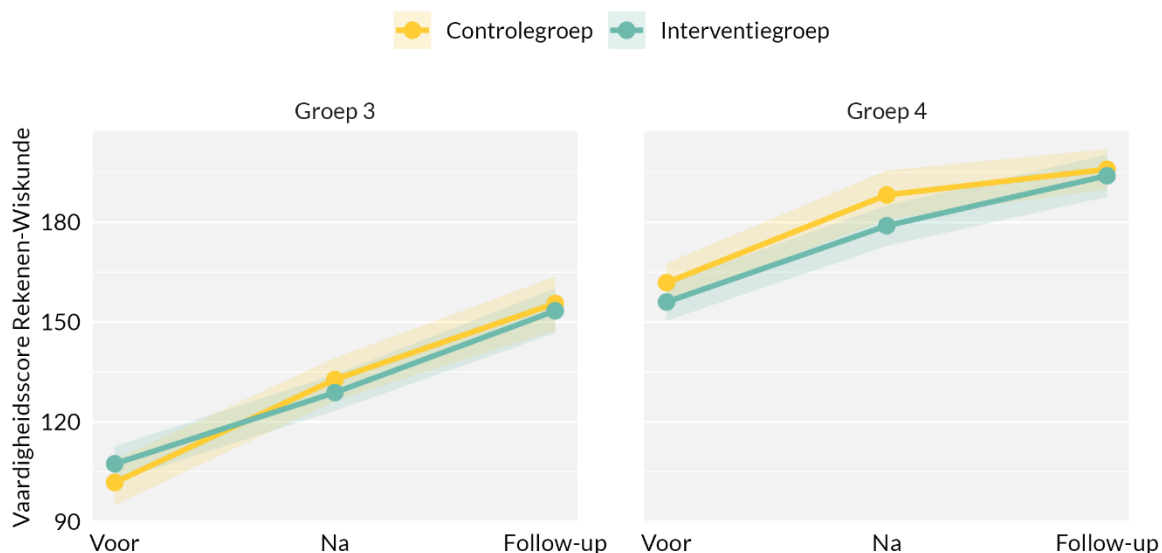
Figuur 4.4 Begrijpend Lezen (gemiddelde scores en 95% confidence-interval)

De resultaten voor **Taalverzorging** (zie Figuur 4.5) laten zien dat zowel interventie- als controlegroep van T1 naar T2 significant vooruitgaan ($p < ,001$). Op de nameting ligt het gemiddelde van de interventiegroep hoger dan dat van de controlegroep, maar dit groepsverschil is niet significant ($p = ,25$). Tussen T2 en T3 groeit de controlegroep significant verder ($p < ,001$), terwijl de interventiegroep in diezelfde periode geen significante vooruitgang meer laat zien ($p = ,75$). Het verschil in groei tussen groepen in deze fase is daarmee significant in het voordeel van de controlegroep ($p = ,01$). Hierdoor verdwijnt het aanvankelijke verschil: op T3 scoren beide groepen weer gelijk ($p = ,89$).



Figuur 4.5 Taalverzorging

Voor **Rekenen-Wiskunde** is in Figuur 4.6 te zien dat de controlegroep meer vooruit gaat tussen de voormeting en nameting. Beide groepen eindigen groepen ten tijde van de follow-up echter weer zeer dicht bij elkaar. Zowel interventie- als controlegroep gaan significant vooruit van T1 naar T2 ($p < ,001$). Op de nameting liggen de geschatte gemiddelden dicht bij elkaar; dit verschil is niet significant. Tussen T2 en T3 groeit de interventiegroep nog significant verder ($p < ,001$), terwijl de controlegroep in diezelfde periode slechts een kleine en niet-significante groei laat zien ($p = ,09$). Het verschil in groei tussen groepen is niet significant ($p = ,12$). Op T3 eindigen beide groepen rond hetzelfde niveau (niet significant).



Figuur 4.6 Rekenen-Wiskunde

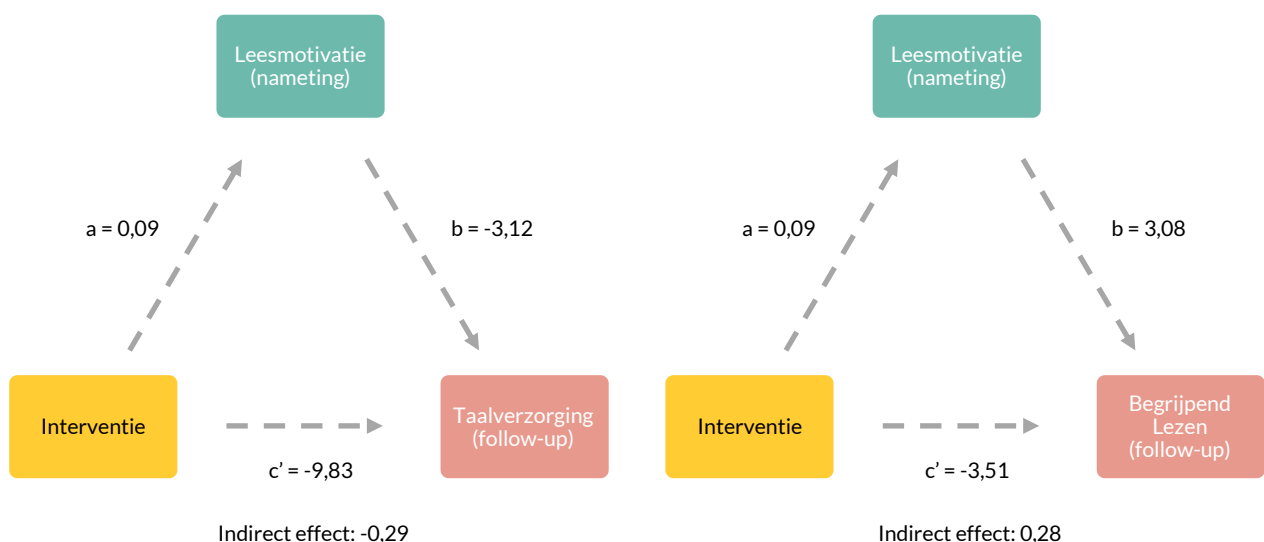
De analyses laten zien dat positieve effecten van de interventie (op Begrijpend Lezen en het mogelijke effect op Taalverzorging) niet blijven hangen tot aan de follow-up meting. Integendeel, tussen de nameting en de follow-up meting zien we dat de controlegroep méér

voortuitgaat dan de interventiegroep, en beide groepen op ongeveer hetzelfde niveau eindigen. Andersom zien we ook dat het negatieve effect van de interventie op Rekenen-Wiskunde niet blijvend is tot aan de follow-up meting. Hier halen de leerlingen uit de interventiegroep de achterstand weer in, en scoren beide groepen tijdens de follow-up meting gelijk.

Geen duidelijke rol van leesmotivatie en woordenschat

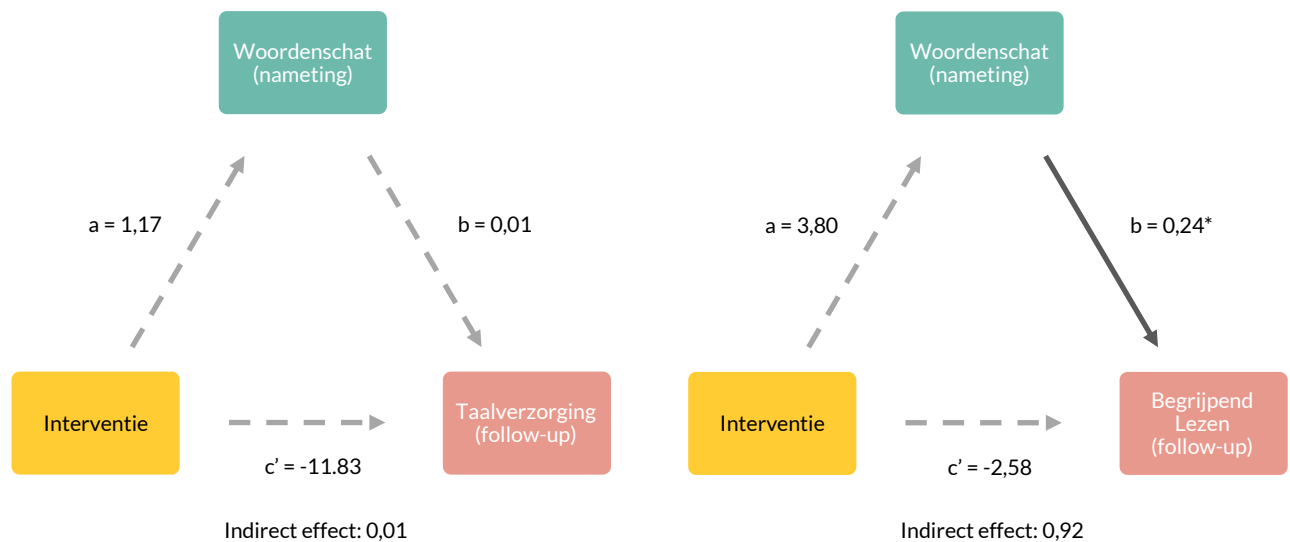
Om zicht te krijgen op mogelijke werkzame elementen van de interventie is aanvullend gekeken naar mediatie-effecten. We verwachtten dat de interventie zowel via een verbeterde woordenschat als via een hogere leesmotivatie zou kunnen bijdragen aan betere leesuitkomsten. Mediatie-analyses laten geen indirecte effecten zien van de interventie op Taalverzorging of Begrijpend Lezen via een verbeterde leesmotivatie. De uitkomsten van de mediatieanalyse voor Leesmotivatie zijn weergegeven in

Figuur 4.7. Hierin is af te lezen dat geen van de verbanden significant zijn. Dit is geen verassend resultaat, aangezien de interventie leesmotivatie niet lijkt te verbeteren en de cronbach's alpha van de vragenlijst matig tot net acceptabel was.



Figuur 4.7. Mediatie van Leesmotivatie op Taalverzorging en Begrijpend lezen

Als we kijken naar een mogelijk mediërend effect van de interventie via Woordenschat op Taalverzorging, blijken er geen significante verbanden te zijn en ook geen significant indirect effect (Figuur 4.8). Als het gaat om de uitkomsten op Begrijpend lezen blijkt dat Woordenschat wel voorspellend is voor Begrijpend Lezen op de follow-up. Er is echter geen significant indirect effect van de Interventie via Woordenschat. Deze resultaten zijn niet verbaasd, aangezien er in de hoofdanalyses ook geen interventie-effect zijn gevonden op woordenschat.



Figuur 4.8 Mediatie van Woordenschat op Taalverzorging en Begrijpend Lezen

4.3 Welke interventie- en oefenkenmerken zijn geassocieerd met de meeste vooruitgang?

Om inzicht te krijgen in *hoe* de BliksemBende leidt tot een mogelijke verbetering in taalvaardigheid, analyseren we de relatie tussen oefenfrequentie en de uitkomstmaten binnen de interventiegroep. Naast de oefenfrequentie kijken we hier naar de vooruitgang binnen de interventie (het aantal verhalen dat een leerling heeft voltooid). Van beide predictoren is in multilevel analyses getest of ze voorspellend zijn voor vooruitgang tussen de voormeting en nameting. Als additionele uitkomstmaat nemen we hierin ook de programmegebonden woordleestaak mee. Deze woordleestaak is als onderdeel van de interventie afgenomen binnen de omgeving van BliksemBende. De correlaties tussen kenmerken van interventietrouw en deze uitkomstmaten zijn weergegeven in Tabel A.3.

Tabel 4.8 Effectiviteit van het aantal tutorsessies en voltooide verhalen

		Estimate	SE	T-waarde	p-waarde
DMT (n = 181)	Intercept	10,75	4,40	2,44	,016
	Voormeting	1,05	0,05	20,63	< ,001
	Aantal tutorsessies	-0,26	0,09	-2,89	,004
	Aantal voltooide verhalen	0,52	0,12	4,48	< ,001
AVI (n = 138)	Intercept	2,44	0,73	3,35	,001
	Voormeting	1,07	0,07	14,70	< ,001
	Aantal tutorsessies	-0,05	0,01	-3,75	< ,001
	Aantal voltooide verhalen	0,06	0,02	3,26	,002
Begrijpend Lezen (n = 82)	Intercept	95,60	17,93	5,33	< ,001
	Voormeting	0,51	0,09	5,42	< ,001
	Aantal tutorsessies	-0,67	0,35	-1,92	,059
	Aantal voltooide verhalen	0,52	0,53	0,97	,335
Taalverzorging (n = 181)	Intercept	159,26	18,43	8,64	< ,001
	Voormeting	0,52	0,03	15,21	< ,001
	Aantal tutorsessies	-0,84	0,36	-2,36	,019
	Aantal voltooide verhalen	0,93	0,49	1,87	,063
Woordenschat (n = 157)	Intercept	148,32	36,06	4,11	< ,001
	Voormeting	0,70	0,06	11,30	< ,001
	Aantal tutorsessies	-0,40	0,35	-1,15	,253
	Aantal voltooide verhalen	0,565	0,45	1,27	,208
Rekenen-Wiskunde (n = 180)	Intercept	59,28	12,96	4,57	< ,001
	Voormeting	0,86	0,04	19,82	< ,001
	Aantal tutorsessies	-0,58	0,22	-2,62	,009
	Aantal voltooide verhalen	0,55	0,33	1,68	,095
Programmagebonden woordleestaak (n = 181)	Intercept	45,49	9,88	4,61	< ,001
	Voormeting	0,58	0,06	9,13	< ,001
	Aantal tutorsessies	-0,69	0,19	-3,63	< ,001
	Aantal voltooide verhalen	1,74	0,28	6,26	< ,001

De resultaten laten een opvallend patroon zien. Voor alle uitkomstmaten behalve woordenschat en begrijpend lezen is er inderdaad een significant effect van het aantal voltooide tutorssessies (voor begrijpend lezen nadert dit effect wel significantie, $p = ,059$). Het effect is echter negatief. Hoe meer sessies een leerling heeft voltooid, hoe *lager* zij scoren op de nameting. Dit verband lijkt op het eerste gezicht niet logisch. Echter, uit de correlaties is te zien dat leerlingen met een lagere leesvaardigheid zoals gemeten op de DMT over het algemeen meer hebben geoefend dan leerlingen met een relatief hogere leesvaardigheid ($r = -,470$). Een mogelijke verklaring is dat leerlingen met een grotere leesachterstand dus meer hebben geoefend met het programma.

Het aantal voltooide verhalen binnen de BliksemBende is daarentegen wel een significante voorspeller voor vooruitgang op toetsen van DMT, AVI en de programmagebonden woordleestaak (voor Taalverzorging neigt dit naar significantie, $p = 0,063$). Deze tegenstelde uitkomst is opvallend, aangezien leerlingen die vaker oefenen in het algemeen ook meer verhalen hebben voltooid (correlatie van ,353). Samen laten de uitkomsten van de analyses zien dat niet zozeer de oefenfrequentie van primair belang is, maar wél of leerlingen binnen die sessies vooruitgang maken in de omgeving van de BliksemBende.

De resultaten duiden er op dat er leerlingen zijn die veel oefenen en maar weinig verhalen voltooien, terwijl andere leerlingen met minder oefening juist meer verhalen voltooien. Dit is een indicatie dat de interventie mogelijk niet voor elke leerling even geschikt is. In aanvullende analyses zien we dat leerlingen die bij de start van de interventie meer moeite hebben met technisch lezen (op basis van de DMT) meer sessies nodig hebben om een verhaal af te ronden ($r = ,317$), met name leerlingen die in groep 3 op de voormeting bij technisch lezen in niveau V, voltooien per sessie significant minder verhalen dan andere leerlingen. In Tabel 4.9 is te zien dat een leerling uit groep 3 die leest op niveau II ongeveer twee sessies nodig hebben om een verhaal af te ronden, terwijl dit drie sessies zijn voor leerlingen op niveau V.

Tabel 4.9 Gemiddeld aantal voltooide verhalen per sessie (SD) per DMT-niveau op de voormeting

DMT-niveau	Groep 3	Groep 4
I	.	,31 (-)
II	,47 (-)	.
III	,41 (,13)	,39 (,07)
IV	,37 (,06)	,43 (,10)
V	,33 (,07)	,39 (,10)

4.4 Welke kind- en groepskenmerken zijn geassocieerd met de meeste vooruitgang?

Om onderzoeksvraag 3 te beantwoorden kijken we ook meer gedetailleerd of de interventie meer of minder effectief voor specifieke groepen leerlingen, zoals leerlingen in groep 3 of 4, leerlingen met een kleinere/grotere taalachterstand, en leerlingen die een andere thuistaal hebben dan Nederlands.

Thuistaal en gender lijken niet uit te maken voor effectiviteit interventie.

Eerst is gekeken of thuistaal een hoofdeffect heeft op de nameting, en of de mogelijke interventie-effecten mogelijk verschillen tussen leerlingen die alleen Nederlands als thuistaal hebben of thuis ook een andere taal spreken (zie Tabel B.5 in de bijlage). De interactie is niet significant, wat betekent dat de effectiviteit van de interventie niet verschilt voor leerlingen met een andere thuistaal dan Nederlands, of verschilt tussen jongens en meisjes.

Negatief effect op Rekenen-Wiskunde groter in groep 3

Om inzicht te krijgen of de interventie even effectief is geweest voor leerlingen in groep 3 en groep 4, hebben we de analyses ook apart uitgevoerd voor leerlingen in groep 3 en groep 4 (zie Tabel B.3 en Tabel B.4 in de bijlage). Uit deze analyses blijkt dat het negatieve effect op Rekenen-Wiskunde meer uitgesproken lijkt te zijn in groep 3 ($B = -8,18$, $p = 0,004$), aangezien er een kleiner en niet significant effect van de interventie is in groep 4 ($B = -4,72$, $p = 0,160$). Vooral de kleinere coëfficiënt (B) is hier relevant, aangezien de power om significante effecten te vinden in deze losse analyses lager is dan in de overkoepelende analyse. Zoals te verwachten blijft het interventie-effect op begrijpend lezen in groep 4 duidelijk zichtbaar.

Interventie mogelijk effectiever voor leerlingen met spellingsproblemen

Om in kaart te brengen of de effectiviteit van de interventie afhangt van de leesachterstand van leerlingen, hebben we onderzocht of de grootte van het interventie-effect samenhangt met de leesachterstand op de voormeting. Hiervoor hebben we de hoofdanalyses uitgebreid met een interactie-effect voor Voormeting * Interventie. Voor Taalverzorging geldt: als de interactie tussen voormeting en Interventie wordt gemodelleerd, is er een hoofdeffect van de interventie en een significante interactie. Een likelihood ratio test toont aan dat het model met deze interactie beter bij de data past dan het eenvoudigere model zonder dit effect. De interactie duidt erop dat leerlingen die lager scores op Taalverzorging meer vooruit gaan binnen de interventiegroep dan leerlingen die hoger scores op Taalverzorging. Voor Begrijpend lezen en Rekenen-Wiskunde is deze interactie niet significant.

Tabel 4.10 De interactie tussen beginniveau van leerlingen en de effectiviteit van de interventie

		Estimate	SE	T-waarde	p-waarde
Taalverzorging (n = 338)	Intercept	104,43	6,55	15,94	,000
	Voormeting	0,65	0,03	19,16	,000
	Schoolrandomisatie	-3,84	8,21	-0,47	,642
	Interventie	21,50	8,72	2,47	,014
	Schoolrandomisatie * Interventie	23,01	9,95	2,31	,021
	Voormeting * Interventie	-0,10	0,04	-2,14	,033
Begrijpend Lezen (n = 149)	Intercept	43,41	14,48	3,00	,003
	Voormeting	0,70	0,116	6,02	,000
	Schoolrandomisatie	5,18	8,53	0,61	,551
	Interventie	29,02	18,98	1,53	,128
	Schoolrandomisatie * Interventie	-1,13	10,68	-0,11	,915
	Voormeting * Interventie	-0,17	0,15	-1,13	,262
Rekenen-Wiskunde (n = 332)	Intercept	42,63	5,82	7,32	,000
	Voormeting	0,89	0,04	21,57	,000
	Schoolrandomisatie	-1,04	5,95	-0,17	,862
	Interventie	-7,23	8,28	-0,87	,383
	Schoolrandomisatie * Interventie	0,01	0,06	0,11	,909
	Voormeting * Interventie	2,12	7,56	0,28	,780

4.5 Procesevaluatie interventieperiode

Om inzicht te krijgen in de uitvoering en ervaren waarde van de BliksemBende, zijn na afloop leerlingen, tutores en leerkrachten bevraagd met vragenlijsten. Daarnaast werden tussentijdse signalen opgehaald tijdens overleggen en ontmoetingsdagen, en zijn verdiepende interviews gehouden met vier tutorcoördinatoren. Deze procesevaluatie laat zien hoe betrokkenen het programma hebben ervaren, welke knelpunten en succesfactoren zichtbaar werden, en welke lessen dit biedt voor verdere opschaling.

Ervaringen van leerlingen

Leerlingen keken over het algemeen positief terug op het werken met het programma (Tabel 4.11). Ze vonden het *best wel* leuk om met de BliksemBende te lezen en gaven aan goed geholpen te zijn door hun tutor. Ook het samenwerken met een maatje werd redelijk positief beoordeeld.

Tabel 4.11 Evaluatie vragenlijst leerlingen over de interventieperiode

	N	Gem	SD	Range
<i>1 = nee, helemaal niet – 5 = ja, heel erg*</i>				
Vond je het leuk om te lezen met de BliksemBende?	170	4,14	1,08	1-5
Hielp de tutor jou goed als jij vragen had?	166	4,38	0,74	1-5
Kon de tutor jou snel helpen als jij een spel lastig vond?	165	4,06	0,90	1-5
Kon je goed met je maatje samenwerken?	168	3,80	1,16	1-5

* Leerlingen konden ook kiezen voor 'weet ik niet'. Deze antwoorden zijn niet in dit overzicht meegenomen.

Wat betreft de spellen binnen het programma vonden leerlingen met name **Letterlezer/Woordlezer** en **Woordenschat** makkelijk, terwijl **Leeshoek** het minst eenvoudig werd gevonden. Dit suggereert dat de moeilijkheidsgraad en aantrekkelijkheid van de spellen per onderdeel varieert. Kortom, leerlingen waarderen het programma, voelen zich ondersteund en ervaren de inhoud grotendeels als toegankelijk, met kleine verschillen tussen spellen.

Tabel 4.12 Moeilijkheidsgraad van spellen volgens leerlingen

	N	Gem	SD	Range
<i>1 = nee, helemaal niet – 5 = ja, heel erg*</i>				
Vond jij Letterlezer en Woordlezer makkelijk?	162	4,31	0,90	1-5
Vond jij Superspeller makkelijk?	163	3,94	1,08	1-5
Vond jij Woordenschat makkelijk?	166	4,19	0,91	1-5
Vond jij Leeshoek makkelijk?	166	3,65	1,12	1-5

* Leerlingen konden ook kiezen voor 'weet ik niet'. Deze antwoorden zijn niet in dit overzicht meegenomen.

Procesevaluatie door tutoeren

Tutores bevestigen dit positieve beeld: zij denken dat hun leerlingen het leuk vonden om mee te doen (Tabel 4.13) en dat duo's doorgaans serieus samenwerkten. Tutores voelen zich voldoende in staat om leerlingen te ondersteunen en motiveren. Tegelijkertijd valt op dat tutoeren vaak verschillen in vooruitgang tussen maatjes signaleren.

Tabel 4.13 Evaluatie vragenlijst tutoren over de interventieperiode

	N	Gem	SD	Range
<i>1 = nee, helemaal niet – 5 = ja, heel erg*</i>				
Denkt u dat de leerlingen het leuk vonden om te lezen met de BliksemBende?	25	4,28	0,54	3-5
Kon u de leerlingen voldoende ondersteunen zodat zij succesvol met elkaar konden samenwerken?	25	3,92	0,91	1-5
Hebben de duo's gemiddeld genomen serieus samengewerkt tijdens de tutorsessies van lezen met de Bliksembende?	25	4,04	0,46	3-5
Lukte het u om de duo's gemiddeld genomen te motiveren om mee te doen met de tutorsessies?	24	4,08	0,50	3-5
Luisterden de duo's gemiddeld genomen met aandacht naar uw instructies tijdens de tutorsessies?	25	4,00	0,50	3-5
Is de leesvaardigheid van de duo's gemiddeld genomen volgens u vooruit gegaan tussen het begin en het einde, door te lezen met de Bliksembende?	23	4,26	0,62	3-5
Zat er verschil in vooruitgang tussen de leerlingen binnen een duo? Dat wil zeggen, gingen de leerlingen gelijk op, of maakte een van de leerlingen veel meer vooruitgang dan de ander en gingen de leerlingen hierdoor uit elkaar lopen in leesvaardigheid?	25	3,56	0,87	2-5

* Tutoren konden ook kiezen voor 'weet ik niet'. Deze antwoorden zijn niet in dit overzicht meegenomen.

Ervaringen van leerkrachten

Ook leerkrachten ervaren dat leerlingen met plezier deelnamen (Tabel 4.14). Tegelijkertijd signaleren zij knelpunten in de dagelijkse praktijk. Zo vonden ze het *een beetje* storend dat leerlingen elke dag uit de klas werden gehaald, en maken zij zich in enige mate zorgen over gemiste instructiemomenten of lestijd. Leerkrachten zien dat leerlingen enthousiast zijn, maar wijzen op de organisatorische belasting en de spanning met het reguliere lesprogramma.

Tabel 4.14 Evaluatie vragenlijst leerkrachten over de interventieperiode

	N	Gem	SD	Range
<i>1 = nee, helemaal niet – 5 = ja, heel erg*</i>				
Denkt u dat de leerlingen het leuk vonden om te lezen met de BliksemBende?	21	4,24	0,94	1-5
Past lezen met de BliksemBende goed bij de lessen en overige activiteiten op school?	21	3,57	1,08	1-5
Was het storend voor het verloop van de les dat de leerlingen elke dag even uit de klas werden gehaald?	21	3,05	1,28	1-5
Hebben de leerlingen standaard instructiemomenten tijdens de taallessen gemist wanneer zij aan het lezen waren met de BliksemBende?	21	2,62	1,40	1-5
Hebben de leerlingen volgens u teveel tijd gemist in het reguliere lesprogramma?	21	2,81	1,21	1-5

* Leerkrachten konden ook kiezen voor 'weet ik niet'. Deze antwoorden zijn niet in dit overzicht meegenomen.

Ervaringen van schoolleiders en tutorcoördinatoren

Uit ontmoetingsdagen, overleggen en interviews komt een genuanceerd beeld naar voren. De meeste scholen waren enthousiast en een deel wil graag doorgaan, maar anderen vonden de tijdsinvestering niet in verhouding tot het leerrendement. Sommige leerkrachten ervoeren de structurele onderbrekingen van de les als belastend en vreesden dat leerlingen hierdoor te veel lestijd misten. Zowel de interventie als het onderzoeksproject werden als arbeidsintensief ervaren. Voor een goede uitvoering was het dan ook cruciaal dat tutorcoördinatoren en leerkrachten nauw betrokken waren en dat deelname prioriteit had binnen de school. Begeleiding vanuit het interventieteam was hierbij essentieel: de doorlopende communicatie en het vaste aanspreekpunt werden zeer gewaardeerd en versterkten de betrokkenheid van de scholen. Enkele leerkrachten gaven aan dat zij graag meer tussentijdse feedback over leerlingvoortgang hadden ontvangen, zodat deze ook met ouders kon worden gedeeld.

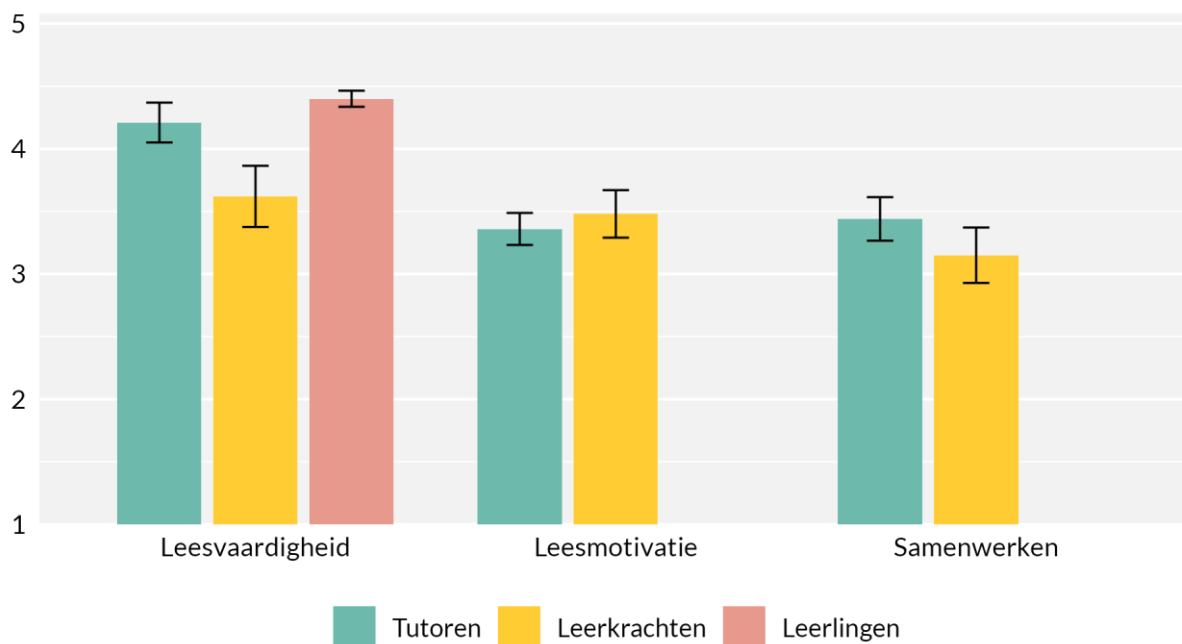
Wat betreft de inhoud van het programma gaven tutoren aan dat de interventie minder effectief leek voor de allerswakste lezers. BliksemBende werd daardoor eerder gezien als geschikt voor ondersteuningsniveau 2 dan voor niveau 3. Samenwerkend leren vormde soms een uitdaging, vooral bij leerlingen met gedragsproblemen of bij een te groot niveauverschil tussen maatjes. Daarnaast werd geobserveerd dat de motivatie van leerlingen tijdens de interventieperiode geleidelijk afnam. Tot slot was de periode tussen de tutortraining en de daadwerkelijke start van de interventie soms te lang, waardoor kennis bij tutoren deels was weggezakt.

Gepercipieerde effectiviteit

Leerkrachten, leerlingen en tutoren zijn gevraagd of zij vooruitgang zagen in de leesontwikkeling⁵ (1 = *Nee helemaal niet*; 5 = *Ja, heel erg*). Hoewel alle groepen overwegend positief zijn over de toegenomen leesvaardigheid, valt op dat met name leerlingen zelf denken dat ze *best wel* tot *heel erg* veel beter hebben leren lezen door te oefenen met de BliksemBende (Figuur 4.9)

Zowel leerkrachten als tutoren geven aan dat leerlingen *een beetje* tot *best wel* meer leesmotivatie hebben gekregen. Wat betreft samenwerken denken ze ook dat dit *een beetje* tot *best wel* is verbeterd, hoewel tutoren dit effect iets hoger inschatten.

⁵ Aan leerlingen is deze vraag gesteld als: 'Heb je beter leren lezen door te oefenen met de BliksemBende?'



Figuur 4.9 Gepercipieerde effectiviteit volgens leerkrachten, leerlingen en tutoeren (gemiddelde en standaarderror)

Hoewel tutoeren alle spellen in redelijke mate nuttig vonden, vonden ze sommige spellen nuttiger om leesvaardigheden te verbeteren dan andere spellen. Met name het lezen van een nieuw verhaal, het spel 'Woordlezer' en het spel 'Woordenschat' worden als nuttig bestempeld. Het spel 'Letterlezer' zagen zij in iets mindere mate als nuttig.

Tabel 4.15 Evaluatie vragenlijst tutoeren: Hoe nuttig vond u elk afzonderlijk spel om leesvaardigheden te verbeteren?

	N	Gem	SD	Range
1 = helemaal niet nuttig - 5 = heel erg nuttig*				
Leeshoek	25	4,12	0,83	2-5
Letterlezer	23	3,87	0,82	2-5
Woordlezer	24	4,29	0,55	3-5
Superspeller	25	4,20	0,71	2-5
Woordenschat	25	4,28	0,74	2-5
Nieuw verhaal	25	4,36	0,64	3-5

* Tutoren konden ook kiezen voor 'weet ik niet'. Deze antwoorden zijn niet in dit overzicht meegenomen.

5 Discussie en conclusie

Dit onderzoek had als doel om inzicht te krijgen in de effectiviteit van het leesprogramma Lezen met de BliksemBende. De effectiviteit van deze interventie is getest in een RCT-onderzoek in het schooljaar 2024-2025. Zwakke lezers in groep 3 en 4 deden mee aan de interventie, of kregen de reguliere leesondersteuning die de school bood. In dit onderzoek is gekeken of de BliksemBende positieve effecten heeft op taalontwikkeling en leesmotivatie van zwakke lezers, en welke elementen bepalen hoeveel vooruitgang een leerling boekt. Daarnaast is de uitvoer van de BliksemBende geëvalueerd om praktische tips en aanbevelingen voor mogelijk verdere implementatie te formuleren.

5.1 Wat levert de BliksemBende op?

Afwisselend effect op begrijpend lezen, spelling, technisch lezen en woordenschat

Er wordt duidelijk geen effect gevonden van Lezen met de BliksemBende op technisch lezen (onderzoeksvraag 1a). Dit duidt erop dat de interventie niet helpt in het verbeteren van de technische leesvaardigheid, óf dat de ondersteuning in de controlegroep van dusdanige kwaliteit is dat beide groepen significant vooruit gaan. Een alternatieve verklaring heeft te maken met de meetwijze van de DMT. Doordat er in de DMT op tijd wordt gelezen, zijn leessnelheid en accuratesse beiden van belang zijn voor de score. De meeste oefeningen in de BliksemBende richten zich op accuratesse (spelling en hardop lezen) van losse woorden. Het zou dus kunnen dat leerlingen meer op leesaccuratesse dan leessnelheid vooruit gaan. Amerikaanse studies van hetzelfde programma vinden inderdaad wel effecten op technisch lezen, wanneer dit gemeten wordt met taken die decodeeraccuratesse zonder tijdsdruk meten, zoals de Woodcock- Johnson Word Attack (Neitzel, 2024; Ross et al., 2017).

Voor begrijpend lezen is wél een positief effect gevonden. Dit effect hield echter geen stand bij de follow-upmeting een half jaar later, terwijl verwacht zou kunnen worden dat de interventie leerlingen tot een niveau brengt waarop zij 'normale' lezers blijven. Het zou kunnen dat blijvend oefenen nodig is (Bailey et al., 2017). Deze resultaten zijn in lijn met internationaal onderzoek naar tutoring, waarin uit een overzichtstudie ook blijkt dat positieve effecten van tutoring na verloop van tijd minder zichtbaar worden (Slavin et al., 2011). Slechts een klein aantal leerlingen heeft tussen de eindmeting en de follow-up zelfstandig doorgewerkt met de BliksemBende, waardoor het niet mogelijk was om na te gaan of de positieve effecten wél standhouden wanneer het programma structureel wordt voortgezet.

Onder de meest strikte onderzoekscondities, waarin rekening wordt gehouden met de specifieke manier van randomisatie in dit onderzoeksdesign, werd geen effect op spelling gevonden. Wanneer echter gecontroleerd wordt voor het beginniveau van leerlingen op het gebied van spelling, zien we dat leerlingen met een lagere startniveau meer vooruitgang boekten in de interventiegroep dan leerlingen met hogere startniveaus. Dit suggereert dat de interventie mogelijk een positief effect heeft op de spellingsvaardigheden, met name die van zwakkere spellers. Tegelijkertijd moet bij deze bevindingen een slag om de arm worden gehouden, gezien de complexiteit van de randomisatie en de beperkte robuustheid van de analyses.

We vonden geen effect van de BliksemBende op de ontwikkeling van woordenschat. Het is in dat licht dan ook niet verbazend dat we geen indirect effect van woordenschat op de vooruitgang in leesontwikkeling vonden (onderzoeksvraag 1d). We vinden dus geen ondersteuning dat vooruitgang in woordenschat een deel is van het werkzame mechanisme achter de interventie. Een verklaring voor het uitblijven van een (positief) effect van de interventie op woordenschat, is dat er andere woorden worden gebruikt in de interventie dan wat wordt gemeten in de Cito-toets. Deze toets is methode-onafhankelijk. Bij methode gebonden toetsen wordt direct gemeten hoe goed specifieke lesstof is geleerd, terwijl methode-onafhankelijke toetsen erop gericht zijn om te laten zien of leerlingen in allerlei contexten woorden hebben verworven (Hilte & Verhallen, 2014). Woordenschatverwerving generaliseert niet goed naar onbekende woorden en interventie-effecten op woordenschat worden daarom doorgaans sneller gevonden wanneer deze aansluiten op de interventie (Marulis & Neuman, 2010).

Kortdurende negatieve neveneffecten op rekenvaardigheid

Lezen met de BliksemBende heeft op korte termijn negatieve neveneffecten op rekenvaardigheid, maar deze effecten zijn een half jaar later verdwenen (onderzoeksvraag 1b). Leerlingen die *niet* met de BliksemBende oefenden, gingen tijdens de interventie meer vooruit op rekenen-wiskunde dan de leerlingen die wél meededen aan de BliksemBende. Dit negatieve neveneffect lijkt meer uitgesproken te zijn in groep 3 dan in groep 4. Het effect is enigszins opmerkelijk, omdat leerlingen die deelnamen aan de BliksemBende niet structureel rekenlessen misten. Zij werden vooral op vaste momenten uit verschillende lessen gehaald (waaronder ook rekenen). Uiteindelijk eindigden beide groepen rond hetzelfde niveau. Een mogelijke verklaring voor deze 'inhaalslag' is dat leerlingen tijdens de interventie in de praktijk toch voornamelijk rekenlessen gemist en in het half jaar tot aan de follow-up meting niet meer.

Effect van interventie op leesmotivatie onduidelijk

Het is onduidelijk of de BliksemBende een positief effect heeft op leesmotivatie en ook of dit een indirect effect heeft op vooruitgang in leesontwikkeling (onderzoeksvraag 1c). Een belangrijke kanttekening is echter dat in dit onderzoek de betrouwbaarheid van de leerlingvragenlijst over leesmotivatie slechts matig tot net acceptabel was. Bovendien waren de correlaties tussen leesmotivatie en de leesuitkomsten klein of niet significant in tegenstelling tot andere studies (Malanchini et al., 2017). Het is daardoor aannemelijk dat leesmotivatie in deze groep niet betrouwbaar genoeg is gemeten om uitspraken te kunnen doen over de rol van leesmotivatie.

Mogelijk speelt leesmotivatie op jonge leeftijd bovendien nog geen grote rol in de ontwikkeling van leesvaardigheid, aangezien dit verband vooral is aangetoond bij oudere leerlingen (Smith et al., 2012; Malanchini et al., 2017). Om hier meer inzicht in te krijgen is het voor vervolgonderzoek van belang om meetinstrumenten te ontwikkelen die leesmotivatie betrouwbaar kunnen meten voor beginnende lezers.

Een alternatieve verklaring is dat de interventie daadwerkelijk geen positief effect heeft op de leesmotivatie van leerlingen, in tegenstelling tot sommige andere tutorinventies (Gest & Gest, 2005; Hebbecker et al., 2019). Mogelijk richt de BliksemBende zich onvoldoende expliciet op het stimuleren van leesmotivatie. Uit een recente meta-analyse blijkt dat juist interventies die expliciet de interesse van leerlingen proberen te wekken de grootste effecten hebben op leesmotivatie en leesbegrip (Van der Sande et al., 2023). Het zou kunnen dat de BliksemBende zich in de uitvoer sterker richt op het ontwikkelen van leesvaardigheid dan op het aanwakken van motivatie.

5.2 Wanneer is de interventie het meest effectief?

Scholen hebben de interventie in de meeste gevallen uitgevoerd zoals gepland, met minimaal drie tot vier sessies per week per leerling (onderzoeksvraag 2a), waarbij op elke school minstens drie wekelijkse sessies zijn gedaan. Uit eerder onderzoek weten we dat leerlingen minstens drie tutorsessies per week nodig hebben om vooruitgang te boeken, een dosering die in het huidige onderzoek ruimschoots is gehaald (Nickow et al., 2024). Het is daardoor niet aannemelijk dat het uitblijven van grote of significante effecten op technisch lezen te wijten is aan een gebrek aan interventietrouw als het gaat om voldoende frequentie. Hoewel de wekelijkse frequentie dus adequaat was, moet wel worden opgemerkt dat de totale interventieduur in deze studie (12 weken) aanzienlijk korter was dan in eerdere implementaties van Lezen met de BliksemBende, waar wél effecten werden gevonden op technisch lezen. In Amerikaanse studies van hetzelfde programma duurde de interventie 24-28 weken, met 3-4 sessies per week (Neitzel, 2024; Madden & Slavin, 2017). Aangezien de overgang van decodeeraccuratesse naar geautomatiseerd vloeiend lezen een gradueel proces is dat extensieve oefening vereist (Ehri, 2005), zou de kortere interventieduur in deze studie het uitblijven van effecten op technisch lezen (deels) kunnen verklaren.

Daarnaast namen de tutorsessies, zoals bedoeld, ongeveer 25 minuten in beslag en meestal waren beide leerlingen aanwezig. Binnen het programma maakten leerlingen consequent vooruitgang. In zowel groep 3 als groep 4 zijn leerlingen vooruitgegaan in de moeilijkheidsgraad van het verhaal en kregen ze dus steeds moeilijkere oefeningen.

Oefenfrequentie niet altijd gerelateerd aan vooruitgang

Het is opvallend dat hoe meer tutorsessies een leerling heeft voltooid binnen Lezen met de BliksemBende, hoe lager deze leerling scoort op de Cito-toetsen op de nameting, met uitzondering van woordenschat (onderzoeksvraag 2b). Deze opvallende uitkomst kan mogelijk worden verklaard doordat leerlingen met een grotere initiële technische leesachterstand (op basis van de DMT) meer hebben meer geoefend met het programma. Leerlingen die het meeste ondersteuning nodig hebben, kregen dus de hoogste dosering. Aangezien leerlingen met een lagere leesvaardigheid in het algemeen minder vooruit gaan

in een bepaalde periode dan leerlingen met een kleinere leesachterstand (Stuebing et al., 2015), zou dit kunnen verklaren waarom in deze studie oefenfrequentie een negatief effect lijkt te hebben op leesontwikkeling.

Vooruitgang binnen de spellen belangrijker dan oefenfrequentie

De vooruitgang binnen de spellen, gemeten door het aantal voltooide verhalen binnen Lezen met de BliksemBende, voorspelt de vooruitgang op technisch lezen evenals de woordleestaak binnen het programma. Ook deze uitkomst is opvallend, omdat leerlingen die vaker oefenen in het algemeen ook meer verhalen hebben voltooid. Terwijl over oefenfrequentie juist werd gevonden dat dit negatief samenhangt met vooruitgang. Samen laten deze uitkomsten zien dat niet zozeer de oefenfrequentie van primair belang lijkt te zijn, maar wél of leerlingen binnen die sessies vooruitgang maken in de omgeving van de BliksemBende. De kwaliteit of vooruitgang binnen een sessie is bepalender voor leesontwikkeling dan het aantal voltooide sessies. Verkennende analyses suggereren dat sommige leerlingen weinig oefenen en veel vooruitgaan binnen de interventie, terwijl andere leerlingen veel oefenen en weinig vooruitgaan boeken. Mogelijk is deze laatste groep minder responsief voor de interventie. Met name de zwakste lezers binnen groep 3 (die scoren op niveau V van de DMT) hebben meer oefening nodig om vooruit te gaan. Dit is een mogelijke aanwijzing dat de interventie voor deze groep leerling wellicht nog te moeilijk of minder geschikt is.

Weinig variatie in type tutor en kwaliteit

Het bleek niet mogelijk om te toetsen of er verschillen in effectiviteit zijn tussen verschillende type tutores (onderzoeksvraag 2c); er was een gebrek aan variatie in de achtergrondkenmerken van tutores en van een deel van de scholen ontbrak informatie over het type tutor. Het merendeel van de tutores is werkzaam als onderwijsassistent en heeft bovendien minimaal 10 jaar werkervaring in het onderwijs. In tussentijdse observaties door schoolbegeleider werden tutores op elke school als kwalitatief goed beoordeeld op het gebied van organisatie van de tutoring, kwaliteit van de instructie, en betrokkenheid en samenwerking. Ook hierin zat weinig variatie tussen tutores.

5.3 Werkt de BliksemBende voor elke leerling even goed?

Geen verschil in effectiviteit op basis van gender of thuistaal

Gender en thuistaal lijken niet uit te maken voor de effectiviteit van de interventie (onderzoeksvraag 3a en 3b). Dit onderzoek vond geen verschil in effectiviteit van de interventie tussen jongens en meisjes. Er werd ook geen verschil gevonden tussen leerlingen die alleen Nederlands als thuistaal hebben en leerlingen die thuis (óók) een andere thuistaal spreken: de interventie werkt dus in gelijke mate voor leerlingen met een andere thuistaal. Beide bevindingen zijn in lijn met eerdere bevindingen van de Engelstalige versie van de BliksemBende (*Tutoring with the Lightning Squad*, Nickow et al., 2020). In dit onderzoek waren de aantallen te klein om specifiek leerlingen die alléén thuis Nederlands spreken te vergelijken met leerlingen die thuis (óók) een andere taal spreken. Daarom kunnen geen gerichte uitspraken worden gedaan over de mate waarin de BliksemBende werkt voor NT2-leerlingen.

Weinig verschil tussen groep 3 en groep 4

We zien geen verschil in effectiviteit van de interventie tussen leerlingen in groep 3 en groep 4 (onderzoeksvraag 3c). Dit is in lijn met eerdere onderzoeksresultaten (Nickow et al., 2020). Alleen voor rekenen-wiskunde blijkt het negatieve neveneffect meer uitgesproken voor groep 3 dan groep 4. Het positieve effect op begrijpend lezen is bovendien alleen getoetst voor leerlingen in groep 4 (aangezien voor deze toets geen groep 3 voormeting beschikbaar is).

Meer verkennende analyses wijzen er echter op dat de interventie mogelijk geschikter is voor leerlingen in groep 4 dan in groep 3. Met name zwakke lezers in groep 3 maken weinig vooruitgang binnen de BliksemBende terwijl ze juist veel oefenen. Mogelijk waren de oefeningen in de interventie nog wat te moeilijk voor hen. Deze bevinding komt overeen met eerder onderzoek naar de Engelstalige versie van de BliksemBende, dat grotere effecten vond voor leerlingen in groep 4 dan groep 3 (Ross et al., 2017).

Effecten voor leerlingen met verschillende startniveaus

De achterstand van de leerling zoals gemeten met de Cito-toetsen leek op de meeste vlakken niet uit te maken voor de effectiviteit van de interventie. Alleen voor spelling is gevonden dat de interventie effectiever was voor leerlingen met een lager beginniveau (zie ook het antwoord op onderzoeksvraag 1). Dit suggereert dat de BliksemBende mogelijk meer geschikt is voor zwakke spellers.

5.4 Procesevaluatie

Uit vragenlijsten, tussentijdse overleggen en verdiepende interviews zijn verschillende succesfactoren en aandachtspunten zichtbaar geworden (onderzoeksvraag 4a). Over het algemeen wordt Lezen met de BliksemBende breed gewaardeerd door leerkrachten, tutores en schoolleiders. Leerkrachten en tutores bevestigen dat leerlingen met plezier deelnamen en dat er tijdens de tutorsessies doorgaans serieus werd samengewerkt.

Tegelijkertijd komt naar voren dat de uitvoer in de schoolpraktijk arbeidsintensief kan zijn. Het dagelijkse uit de klas halen van leerlingen werd door sommige leerkrachten als storend ervaren en sommigen vreesden dat de leerlingen te veel lestijd hebben gemist. Deelname vraagt om een goede planning en duidelijke prioriteitstelling binnen de school. Naarmate die organisatie meer op orde is en er een betrokken tutorcoördinator aanwezig is, verloopt de implementatie soepeler.

De communicatie en intensieve begeleiding vanuit het interventieteam worden positief gewaardeerd en zijn cruciaal geweest voor de hoge implementatietrouw. Deze begeleiders waren een vast aanspreekpunt voor de scholen en wisten eventuele belemmeringen in de uitvoer vroeg op het spoor te komen en op te lossen. Scholen gaven aan dat het wenselijk zou zijn om meer tussentijdse terugkoppeling te kunnen geven over de voortgang van leerlingen, zowel naar leerkrachten als naar ouders toe. Op deze manier zouden ze draagvlak van de interventie hoger kunnen houden onder deze groep. Wat betreft de gepercipieerde opbrengsten zijn de leerkrachten en tutores positief: zij zien vooruitgang in leesvaardigheid, en in mindere mate ook in motivatie en samenwerking.

Aandachtspunt: Voor welke leerlingen is de BliksemBende het meest geschikt?

Een belangrijk aandachtspunt betreft de doelgroep. Scholen en tutores signaleren dat de interventie minder geschikt lijkt voor de allerzwakste lezers, terwijl zij wel intensieve ondersteuning vragen. Sommige tutorcoördinatoren gaven aan dat ook de samenwerking voor sommige leerlingen in groep 3 nog echt een uitdaging kan zijn, terwijl dat een essentieel onderdeel is van de BliksemBende. Daarnaast bleek het soms lastig om maatjes met uiteenlopende niveaus goed te laten samenwerken, zo laten de tutores weten. Tutores gaven aan dat de écht zwakke lezers minder baat bij de interventie leken te hebben. Hieruit kan worden beargumenteerd dat De BliksemBende mogelijk meer geschikt is als interventie op ondersteuningsniveau 2 dan op ondersteuningsniveau 3. Dit past ook bij de opzet van de interventie, gezien oefening in kleine groepjes wordt gezien als passend bij ondersteuningsniveau 2.

Waar de statistische analyses lieten zien dat zwakke spellers relatief meer lijken te profiteren van de BliksemBende dan betere spellers, ervaren leerkrachten en tutores dus dat de interventie niet altijd aansluit bij de allerzwakste lezers. Deze ogenschijnlijke spanning vraagt om een genuanceerde interpretatie: de BliksemBende lijkt minder passend voor leerlingen die nog nauwelijks een basis in technisch lezen hebben, maar zou een mogelijk meerwaarde kunnen hebben voor leerlingen met specifieke spellingsproblemen.

Leerlingen hebben positieve ervaringen met Lezen met de BliksemBende

Leerlingen gaven aan dat ze het leuk vonden om te werken met de BliksemBende (onderzoeksvraag 4b). Ze voelden zich bovendien goed ondersteund door hun tutor en ervaren de inhoud grotendeels als toegankelijk, met kleine verschillen tussen spellen. Ook het samenwerken met een maatje werd als redelijk positief ervaren en hadden ze het gevoel dat ze vooruit waren gegaan door het programma. De motivatie van leerlingen leek einde van de interventieperiode wel wat af te nemen, aldus leerkrachten. Meer variatie in de oefeningen kan dit mogelijk ondervangen.

5.5 Beperkingen

Een belangrijke beperking van dit onderzoek betreft de **statistische power**. Hoewel bij de start een zo groot mogelijke steekproef is nagestreefd, was de gerealiseerde power lager dan beoogd. Dit betekent dat kleine of subtiele effecten mogelijk niet betrouwbaar konden worden vastgesteld. Dit punt is extra relevant omdat de gevonden effecten in dit onderzoek over het algemeen klein van omvang waren. Een grotere steekproef zou nodig zijn geweest om deze effecten met voldoende zekerheid te kunnen detecteren. Daarnaast moet worden opgemerkt dat er in dit onderzoek meerdere toetsen zijn uitgevoerd, telkens met een alfa van 0,05. Hierdoor neemt de kans op het vinden van toevallige significante resultaten toe (type I-fout).

Een tweede beperking betreft de **generalisering** en **ecologische validiteit**. De steekproef bestond voornamelijk uit reguliere basisscholen met een gemiddeld tot verhoogd achterstandsrisico, verspreid over matig tot weinig stedelijke gebieden. Daarmee zijn de resultaten enigszins representatief voor een brede groep scholen. Deze scholen zijn intensief begeleid door het interventieteam, wat heeft bijgedragen aan een hoge mate van implementatietrouw. Dat is waardevol en zelfs noodzakelijk in dit eerste onderzoek, maar

roept ook de vraag op in hoeverre de resultaten generaliseerbaar zijn naar de dagelijkse schoolpraktijk. In een bredere implementatie zal de begeleiding onvermijdelijk minder intensief zijn, waardoor de mate van naleving en de effecten mogelijk lager uitvallen (De Jong et al., 2023). Dit benadrukt het belang om in vervolgonderzoek ook te kijken naar de effectiviteit van de interventie in een setting waarin scholen zelf verantwoordelijk zijn voor de implementatie.

5.6 Implicaties voor verdere ontwikkeling en de praktijk

De resultaten van dit onderzoek geven aanknopingspunten voor zowel de verdere ontwikkeling van de BliksemBende als voor scholen die het programma in de praktijk willen inzetten. Hieronder benoemen we verbeter suggesties voor de interventie zelf en praktische aanbevelingen voor een succesvolle implementatie in de schoolpraktijk.

Verbetersuggesties voor de interventie

De procesevaluatie wijst erop dat de BliksemBende potentie heeft als interventie op ondersteuningsniveau 2. Voor een succesvolle opschaling is het belangrijk om aandacht te besteden aan het bereiken van de juiste doelgroep (niet de allerzwakste lezers), voldoende variatie in het oefenmateriaal te bieden en te zorgen dat tutores regelmatig met leerkrachten (en eventueel) ouders kunnen communiceren over de voortgang van de leerlingen.

Tips en aanbevelingen voor scholen

Uit dit onderzoek volgen verschillende praktische tips en aanbevelingen voor scholen die met de BliksemBende aan de slag willen (onderzoeksvraag 4c):

1. Investeer in een duidelijke organisatiestructuur en planning binnen de school. Het helpt wanneer een tutorcoördinator wordt aangewezen en leerkrachten nauw betrokken zijn.
2. Wissel indien mogelijk tussen maatjes wanneer het niveauverschil te groot wordt, zodat leerlingen elkaar niet in hun leerproces belemmeren.
3. Zorg ervoor dat de tutores de gehele training doorlopen. Na het volgen van deze training voelden de tutores in dit onderzoek zich voldoende voorbereid om de sessies te leiden.
4. Haal leerlingen niet standaard tijdens de rekenles uit de klas, maar wissel af met andere momenten.
5. Doe als tutor regelmatig check-ins en check-outs als het gaat om het leesproces en de samenwerking van leerlingen. Dit werd soms vergeten door de tutores in dit onderzoek, terwijl dit leidt tot een betere naleving van gedrag en meer focus.

5.7 Conclusie

Samenvattend laat dit onderzoek zien dat *Lezen met de BliksemBende* bij een goede implementatie op korte termijn een bescheiden positief effect heeft op begrijpend lezen, al verdwijnt dit effect na een half jaar zonder verdere oefening. Voor technisch lezen en woordenschat lijkt het programma niet effectiever te zijn dan reguliere ondersteuning, en voor spelling zijn er alleen aanwijzingen dat zwakkere spellers meer profiteren. Het feit dat

leerlingen lestijd missen kan er mogelijk voor zorgen dat er minder vooruitgang wordt gemaakt op het gebied van rekenen, maar dit verdween na verloop van tijd. Hoewel de uitkomsten niet verschillen voor verschillende groepen, geven tutores aan dat de interventie meer geschikt lijkt voor leerlingen die al enige basis in technisch lezen hebben gelegd, terwijl de interventie minder aansluit bij de allermakste lezers en soms te uitdagend is voor groep 3. Alles bij elkaar genomen laten de bevindingen zien dat de BliksemBende op onderdelen potentie heeft, maar dat de effecten in dit onderzoek klein waren.

Referenties

- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 1-9.
- Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P., Logan, M. K., & Gifford, R. (2011). Small-group, computer-assisted tutoring to improve reading outcomes for struggling first and second graders. *The Elementary School Journal*, 111(4), 625-640.
- Cheung, A. C., & Slavin, R. E. (2012). Effective reading programs for Spanish-dominant English language learners (ELLs) in the elementary grades: A synthesis of research. *Review of Educational Research*, 82(4), 351-395.
- Dietrichson, J., Bøg, M., Filges, T. & Jørgenson, A.M.K. (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.
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of reading*, 9(2), 167-188.
- Gest, S. D., & Gest, J. M. (2005). Reading tutoring for students at academic and behavioral risk: Effects on time-on-task in the classroom. *Education and Treatment of Children*, 25-47.
- Hebbecke, K., Förster, N., & Souvignier, E. (2019). Reciprocal effects between reading achievement and intrinsic and extrinsic reading motivation. *Scientific Studies of Reading*, 23(5), 419-436.
- Higgins, S., Xiao, Z., & Katsipataki, M. (2012). The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation. Full Report. *Education Endowment Foundation*.
- Hilte, M., & Verhallen, M. (2014). Woordkennis toetsen: Weet wat je meet. *Jeugd in School en Wereld*, 6-9.
- Inspectie van het Onderwijs (2022). *De Staat van het Onderwijs 2022*. Utrecht: Inspectie van het Onderwijs.
- Janssen, J., Hop, M., & Wouda, J. (2015a). Wetenschappelijke verantwoording Rekenen - Wiskunde 3.0 voor groep 4. *Arnhem: Cito*.
- Janssen, J., Hop, M., Wouda, J., & Hollenberg, J. (2015b). Wetenschappelijke verantwoording Rekenen - Wiskunde 3.0 voor groep 3. *Arnhem: Cito*.
- Jolink, A., Tomesen, M., Hilte, M., Weekers, A., Engelen, R., & van Boxtel, H. (2015a). Wetenschappelijke verantwoording Begrijpend lezen 3.0 voor groep 3. *Arnhem: Cito*.
- Jolink, A., Tomesen, M., Hilte, M., Weekers, A., & Engelen, R. (2015b). Wetenschappelijke verantwoording Begrijpend lezen 3.0 voor groep 4. *Arnhem: Cito*.

- Jong, P.F. de & Schreurs, B.G.M. (2023). *Wat maakt de grootschalige implementatie van een evidence-based programma voor de preventie van leesproblemen tot een succes?* Universiteit van Amsterdam & NRO.
- Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review*, 10(2), 133-149.
- Ludwig, C., Guo, K., & Georgiou, G. K. (2019). Are reading interventions for English language learners effective? A meta-analysis. *Journal of Learning Disabilities*, 52(3), 220-231.
- Madden, N.A., & Slavin, R.E. (2017). Evaluations of technology-assisted small-group tutoring for struggling readers. *Reading & Writing Quarterly*, 33(4), 327-334.
- Malanchini, M., Wang, Z., Voronin, I., Schenker, V. J., Plomin, R., Petrill, S. A., & Kovas, Y. (2017). Reading self-perceived ability, enjoyment and achievement: A genetically informative study of their reciprocal links over time. *Developmental Psychology*, 53(4), 698.
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, 80(3), 300-335.
- McTigue, E. M., Solheim, O. J., Zimmer, W. K., & Uppstad, P. H. (2020). Critically reviewing GraphoGame across the world: Recommendations and cautions for research and implementation of computer-assisted instruction for word-reading acquisition. *Reading Research Quarterly*, 55(1), 45-73.
- Meelissen, M. R. M., Maassen, N. A. M., Gubbels, J., van Langen, A. M. L., Valk, J., Dood, C., Derks, I., In 't Zandt, M., & Wolbers, M. (2023). Resultaten PISA-2022 in vogelvlucht. *Universiteit Twente*.
- Mullender-Wijnsma, M. J., Veldman, M. A., de Boer, H., van Kuijk, M., & Bosker, R. J. (2020). *Implementatie en effecten van Success for All in Nederland*. GION onderwijs/onderzoek.
- Neitzel, A. J. (2024). An Evaluation of Tutoring with the Lightning Squad on the Reading Achievement of Struggling Readers in Grades 2 and 3: Technical Report.
- Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly*, 57(1), 149-179.
- 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. Working Paper 27476. *National Bureau of Economic Research*.
- Nickow, A., Oreopoulos, P., & Quan, V. (2024). The Promise of Tutoring for PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence. *American Educational Research Journal*, 61(1), 74-107.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383.
- Regtvoort, A., Zijlstra, H., & van der Leij, A. (2013). The effectiveness of a 2-year supplementary tutor-assisted computerized intervention on the reading development of beginning readers at risk for reading difficulties: A randomized controlled trial. *Dyslexia*, 19(4), 256-280.
- Richards-Tutor, C., Baker, D. L., Gersten, R., Baker, S. K., & Smith, J. M. (2016). The effectiveness of reading interventions for English learners: A research synthesis. *Exceptional Children*, 82(2), 144-169.
- Ross, S.M., Laurenzano, M. & Madden, N.A. (2017). An Evaluation of the Lightning Squad Computer-Assisted Small Group Tutoring Program on the Reading Achievement of

- Disadvantaged Students in Grades 1–3. *Center for Research and Reform in Education, Johns Hopkins University*.
- Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). *Effective programs for struggling readers: A best-evidence synthesis*. *Educational Research Review*, 6(1), 1–26.
- Smith, J. K., Smith, L. F., Gilmore, A., & Jameson, M. (2012). Students' self-perception of reading ability, enjoyment of reading and reading achievement. *Learning and individual differences*, 22(2), 202–206.
- Snijders, T., & Bosker, R. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publishers.
- Stuebing, K. K., Barth, A. E., Trahan, L. H., Reddy, R. R., Miciak, J., & Fletcher, J. M. (2015). Are child cognitive characteristics strong predictors of responses to intervention? A meta-analysis. *Review of Educational Research*, 85(3), 395–429.
- Tomesen, M., Wouda, J., Mols, A., & Horsels, L. (2015a). Wetenschappelijke verantwoording van de LVS-toetsen: Spelling 3.0 groep 3. *Arnhem: Cito*.
- Tomesen, M., Wouda, J., Mols, A., & Horsels, L. (2015b). Wetenschappelijke verantwoording van de LVS-toetsen: Spelling 3.0 groep 4. *Arnhem: Cito*.
- van Til, A., Kamphuis, F., Keuning, J., Gijssels, M., Vloedgraven, J., & de Wijs, A. (2018a). Wetenschappelijke verantwoording LVS-toetsen DMT. *Arnhem: Cito*.
- van Til, A., Kamphuis, F., Keuning, J., Gijssels, M., & de Wijs, A. (2018b). Wetenschappelijke verantwoording LVS-toetsen AVI. *Arnhem: Cito*.
- van der Sande, L., van Steensel, R., Fikrat-Wevers, S., & Arends, L. (2023). Effectiveness of interventions that foster reading motivation: A meta-analysis. *Educational Psychology Review*, 35(1), 21.
- Verhoeven, L., & Van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, 25(8), 1805–1818.
- Wang, X., Neitzel, A., & Madden, N. (2024). Lightning Squad: Assessing the Dosage Effect of Computer-Assisted Tutoring with Cooperative Learning for Struggling Readers. *Journal of Education for Students Placed at Risk (JESPAR)*, 1–19.
- Wisniewski, B., Zierer, K. & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10.
- Zijlstra, A. H. (2015). *Early grade learning: The role of teacher-child interaction and tutor-assisted intervention (Doctoral dissertation)*. Universiteit van Amsterdam.

Bijlage A Beschrijvende statistieken steekproef

Tabel A.1 Aantal missende waarden per uitkomstmaat

Jaarlaag	Groep	Meting	DMT	AVI	Begrijpend Lezen	Taal- verzorging	Woorden- schat	Rekenen- Wiskunde
3	Controle	Voor	1	5	-	1	6	1
3	Controle	Na	2	9	29	1	6	5
3	Controle	Follow-up	10	13	22	12	34	10
3	Interventie	Voor	0	22	-	0	4	0
3	Interventie	Na	0	9	16	0	0	0
3	Interventie	Follow-up	3	27	15	4	20	5
4	Controle	Voor	0	0	16	0	2	0
4	Controle	Na	0	5	0	0	0	1
4	Controle	Follow-up	6	23	4	5	22	5
4	Interventie	Voor	0	16	10	0	20	0
4	Interventie	Na	0	5	1	0	4	1
4	Interventie	Follow-up	12	9	10	9	37	9

Tabel A.2 Correlaties tussen uitkomstmaten apart voor groep 3 (boven diagonaal) en groep 4 (onder diagonaal)

	DMT T1	DMT T2	DMT T3	AVI T1	AVI T2	AVI T3	TV T1	TV T2	TV T3	WS T1	WS T2	WS T3	BL T1	BL T2	BL T3	RW T1	TW T2	TW T3	LM T1	LM T2
DMT T1	-	,779**	,526**	,496**	,362**	,577**	,291**	,297**	,382**	,030	-,023	,043	-	,085	,149	,254**	,280**	,249**	,212**	,139
DMT T2	,871**	-	,808**	,549**	,680**	,763**	,321**	,407**	,484**	-,034	-,062	-,002	-	,208*	,255**	,227**	,216**	,275**	,196*	,228**
DMT T3	,774**	,802**	-	,443**	,727**	,815**	,275**	,427**	,478**	,016	-,013	,049	-	,285**	,333**	,277**	,294**	,360**	,132	,247**
AVI T1	,807**	,779**	,719**	-	,542**	,475**	,266**	,200*	,328**	-,059	-,059	-,050	-	,222*	,252**	,216*	,237**	,187*	,203*	,242**
AVI T2	,703**	,749**	,652**	,734**	-	,721**	,303**	,429**	,400**	,099	,004	,033	-	,470**	,327**	,232**	,251**	,222**	,239**	,320**
AVI T3	,689**	,756**	,694**	,695**	,746**	-	,384**	,419**	,413**	,125	,073	,079	-	,419**	,424**	,277**	,308**	,290**	,176	,269**
Taalverzorging T1	,315**	,315**	,152	,254**	,355**	,356**	-	,682**	,555**	,349**	,337**	,322**	-	,438**	,430**	,506**	,420**	,280**	,159*	,120
Taalverzorging T2	,372**	,399**	,207**	,300**	,428**	,387**	,811**	-	,648**	,322**	,261**	,312**	-	,494**	,323**	,441**	,373**	,292**	,157*	,143
Taalverzorging T3	,313**	,350**	,200*	,165*	,179*	,216**	,260**	,265**	-	,266**	,149	,242*	-	,309**	,453**	,280**	,317**	,477**	,143	,187*
Woordenschat T1	,047	,108	-,017	,140	,167*	,160	,229**	,140	,254**	-	,600**	,546**	-	,314**	,404**	,234**	,188*	,099	,010	,031
Woordenschat T2	,094	,128	,046	,167*	,224**	,181*	,276**	,191*	,259**	,740**	-	,586**	-	,418**	,399**	,309**	,251**	,140	,002	-,026
Woordenschat T3	,051	,105	,117	,156	,164	,151	,155	,135	,102	,565**	,571**	-	-	,470**	,537**	,321**	,252**	,253**	-,084	-,035
Begrijpend Lezen T1	,357**	,366**	,199*	,419**	,449**	,378**	,260**	,276**	,289**	,538**	,540**	,493**	-							
Begrijpend Lezen T2	,104	,222**	,162*	,207**	,336**	,356**	,266**	,253**	,130	,524**	,561**	,510**	,574**	-	,529**	,448**	,422**	,361**	,047	,020
Begrijpend Lezen T3	,013	,120	,079	,100	,285**	,296**	,155*	,206**	,130	,400**	,460**	,382**	,439**	,568**	-	,248**	,229**	,312**	,032	,142
Rekenen-Wiskunde T1	,118	,144	,084	,165*	,240**	,272**	,283**	,310**	,162*	,341**	,357**	,229*	,379**	,378**	,256**	-	,776**	,618**	-,007	-,058
Rekenen-Wiskunde T2	,161*	,242**	,157	,204*	,277**	,286**	,272**	,329**	,162*	,311**	,353**	,208*	,378**	,407**	,237**	,754**	-	,758**	-,015	,001
Rekenen-Wiskunde T3	,183*	,282**	,217**	,238**	,316**	,287**	,273**	,353**	,195*	,247**	,283**	,284**	,315**	,377**	,302**	,671**	,726**	-	-,116	-,035
Leesmotivatie T1	,161*	,165*	,153	,135	,225**	,299**	,215**	,262**	,134	,108	,111	,084	,152	,165*	,251**	,043	,086	-,014	-	,541**
Leesmotivatie T2	,128	,076	,122	,124	,232**	,207*	,086	,104	-,073	-,091	-,072	-,019	,100	,034	,200*	-,112	-,090	-,102	,657**	-

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

Tabel A.3 Correlaties uitkomstmaten en interventietrouw binnen de interventiegroep

	Groep 3				Groep 4			
	Aantal tutorsessies	Voltooide verhalen	Programmagebonden Woordleestaak T1	Programmagebonden Woordleestaak T2	Aantal tutorsessies	Voltooide verhalen	Programmagebonden Woordleestaak T1	Programmagebonden Woordleestaak T2
Aantal tutorsessies	-	,216*	-,224*	-,161	-	,527**	-,178	-,1
Voltooide verhalen	,216*	-	,409**	,548**	,527**	-	,106	,302**
Programmagebonden Woordleestaak T1	-,224*	,409**	-	,458**	-,178	,106	-	,615**
Programmagebonden Woordleestaak T2	-,161	,548**	,458**	-	-,1	,302**	,615**	-
DMT T1	-,135	,409**	,574**	,457**	-,115	-,023	,615**	,525**
DMT T2	-,104	,488**	,602**	,708**	-,177	,03	,655**	,590**
DMT T3	-,157	,413**	,542**	,688**	-,147	-,115	,631**	,471**
AVI T1	-,15	,235	,378**	,428**	-,083	-,151	,645**	,519**
AVI T2	-,162	,448**	,413**	,632**	-,268*	-,113	,601**	,503**
AVI T3	-,343**	,452**	,375**	,667**	-,330**	-,131	,527**	,495**
Taalverzorging T1	,303**	,374**	,267*	,242*	-,147	-,015	,115	,280**
Taalverzorging T2	-,026	,352**	,377**	,388**	-,133	-,055	,161	,242*
Taalverzorging T3	,02	,388**	,432**	,413**	,332**	,564**	,233*	,309**
Woordenschat T1	-,259*	-,098	,011	,094	-,116	,347**	,134	,350**
Woordenschat T2	-,012	,038	,07	,119	,075	,415**	,119	,302**
Woordenschat T3	-,207	-,075	,086	,168	-,058	,232	,282*	,343**
Begrijpend Lezen T1	-	-	-	-	-,02	,197	,520**	,481**
Begrijpend Lezen T2	,123	,385**	,21	,369**	-,230*	,062	,296**	,451**
Begrijpend Lezen T3	-,17	,258*	,402**	,395**	,016	,153	,144	,213
Rekenen-Wiskunde T1	,145	,141	,095	,073	-,074	,099	,087	,260*
Rekenen-Wiskunde T2	,102	,212*	,136	,104	-,188	,031	,155	,371**
Rekenen-Wiskunde T3	,059	,296**	,264*	,260*	-,007	,15	,175	,358**
Leesmotivatie T1	-,111	,145	,175	,229*	-,097	,01	,151	,118
Leesmotivatie T2	-,164	,234*	,208	,397**	,019	-,118	,122	,073

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

Bijlage B Resultaten regressieanalyses

Tabel B.1 Uitkomsten initiële regressieanalyses zonder schoolrandomisatie

		Estimate	SE	T-waarde	p-waarde
DMT (n = 338)	Intercept	7,86	0,77	10,23	< ,001
	Voormeting	1,03	0,03	34,08	< ,001
	Interventie	-0,05	0,76	-0,07	,948
AVI (n = 279)	Intercept	0,90	0,15	5,91	< ,001
	Voormeting	1,08	0,05	20,75	< ,001
	Interventie	0,09	0,14	0,65	,519
Begrijpend Lezen (n = 149)	Intercept	55,30	9,50	5,82	< ,001
	Voormeting	0,61	0,07	8,27	< ,001
	Interventie	7,64	3,75	2,04	,043
Taalverzorging (n = 338)	Intercept	113,19	4,98	22,72	< ,001
	Voormeting	0,59	0,02	25,19	< ,001
	Interventie	8,60	3,74	2,30	,022
Woordenschat (n = 304)	Intercept	156,38	20,17	7,75	< ,001
	Voormeting	0,67	0,04	15,29	< ,001
	Interventie	-2,18	3,83	-0,57	,570
	Intercept	42,08	4,35	9,68	< ,001
	Voormeting	0,89	0,03	30,09	< ,001
	Interventie	-6,07	2,42	-2,51	,013
Leesmotivatie (n = 296)	Intercept	1,29	0,20	6,35	,000
	Voormeting	0,64	0,05	12,96	,000
	Interventie	0,05	0,09	0,60	,547

Tabel B.2 Uitkomsten regressieanalyses met enkel de 21 scholen waar is gerandomiseerd per jaarlaag

		Estimate	SE	T-waarde	p-waarde
DMT (n = 308)	Intercept	7,78	0,83	9,34	< ,001
	Voormeting	1,04	0,03	31,54	< ,001
	Interventie	-0,34	0,83	-0,41	,685
AVI (n = 249)	Intercept	0,99	0,16	6,11	< ,001
	Voormeting	1,06	0,06	18,99	< ,001
	Interventie	0,01	0,15	0,03	,973
Begrijpend Lezen (n = 134)	Intercept	51,66	9,82	5,26	< ,001
	Voormeting	0,63	0,08	8,24	< ,001
	Interventie	8,26	4,12	2,01	,062
Taalverzorging (n = 308)	Intercept	114,48	5,00	22,88	< ,001
	Voormeting	0,59	0,02	24,92	< ,001
	Interventie	4,73	3,83	1,23	,225
Woordenschat (n = 274)	Intercept	127,66	20,84	6,13	< ,001
	Voormeting	0,72	0,05	16,06	< ,001
	Interventie	0,60	4,05	0,15	,883
Rekenen-Wiskunde (n = 303)	Intercept	43,78	4,48	9,78	< ,001
	Voormeting	0,88	0,03	28,92	< ,001
	Interventie	-6,40	2,53	-2,53	,016

Tabel B.3 Uitkomsten regressieanalyses in groep 3

		Estimate	SE	T-waarde	p-waarde
Taalverzorging (n = 162)	Intercept	113,70	8,35	13,61	< ,001
	Voormeting	0,60	0,05	11,57	< ,001
	Schoolrandomisatie	-5,62	15,22	-,37	,715
	Interventie	6,70	7,42	,90	,369
	Schoolrandomisatie * Interventie	26,04	17,85	1,46	,147
Rekenen-Wiskunde (n = 158)	Intercept	54,35	5,42	1,03	< ,001
	Voormeting	0,77	0,05	15,83	< ,001
	Schoolrandomisatie	-12,42	7,08	-1,75	,094
	Interventie	-8,18	2,79	-2,93	,004
	Schoolrandomisatie * Interventie	2,13	10,28	9,40	1,09

Tabel B.4 Uitkomsten regressieanalyses in groep 4

		Estimate	SE	T-waarde	p-waarde
Begrijpend Lezen (n = 149)	Intercept	55,65	9,63	5,78	< ,001
	Voormeting	0,60	0,07	7,98	< ,001
	Schoolrandomisatie	6,56	8,48	,78	,449
	Interventie	8,17	4,07	2,01	,047
	Schoolrandomisatie * Interventie	-3,61	10,47	-0,35	,731
Taalverzorging (n = 176)	Intercept	101,17	8,24	12,28	< ,001
	Voormeting	0,65	0,04	18,21	< ,001
	Schoolrandomisatie	-2,29	7,10	-,32	,750
	Interventie	1,91	2,85	,67	,503
	Schoolrandomisatie * Interventie	20,62	9,70	2,21	,035
Rekenen-Wiskunde (n = 174)	Intercept	44,45	9,88	4,50	< ,001
	Voormeting	0,89	0,06	14,90	< ,001
	Schoolrandomisatie	8,41	8,30	1,01	,323
	Interventie	-4,72	3,35	-1,41	,160
	Schoolrandomisatie * Interventie	-2,09	11,38	-0,18	,855

Tabel B.5 Uitkomsten regressieanalyses naar achtergrondkenmerk thuistaal

		Estimate	SE	T-waarde	p-waarde
Begrijpend Lezen (n = 140)	Intercept	71,20	14,14	5,04	< ,001
	Voormeting	0,55	0,08	6,78	< ,001
	Thuistaal	-8,00	7,15	-1,12	,266
	Schoolrandomisatie	6,14	9,09	,68	,509
	Interventie	7,10	11,43	,62	,536
	Schoolrandomisatie * Interventie	-2,15	10,99	-0,20	,845
	Thuistaal * Interventie	1,10	8,99	0,12	,903
Taalverzorging (n = 318)	Intercept	112,76	9,01	12,52	< ,001
	Voormeting	0,59	0,02	24,50	< ,001
	Thuistaal	1,21	5,86	,21	,837
	Schoolrandomisatie	-2,77	8,66	-,32	,751
	Interventie	6,15	10,12	,61	,544
	Schoolrandomisatie * Interventie	24,01	10,52	2,28	,023
	Thuistaal * Interventie	-0,89	7,66	-0,12	,908
Rekenen-Wiskunde (n = 312)	Intercept	44,84	7,09	6,32	< ,001
	Voormeting	0,88	0,03	28,65	< ,001
	Thuistaal	-1,63	4,34	-,38	,708
	Schoolrandomisatie	-0,23	5,99	-,04	,970
	Interventie	-6,87	7,27	-,94	,346
	Schoolrandomisatie * Interventie	2,05	7,73	0,27	,791
	Schoolrandomisatie * Thuistaal	1,14	5,60	0,20	,839

Tabel B.6 Uitkomsten regressieanalyses naar achtergrondkenmerk gender

		Estimate	SE	T-waarde	p-waarde
Begrijpend Lezen (n = 140)	Intercept	53,26	9,94	5,36	< ,001
	Voormeting	0,60	0,08	7,99	< ,001
	Gender	3,08	4,59	,67	,503
	Schoolrandomisatie	6,65	8,90	,75	,464
	Interventie	10,30	5,40	1,91	,059
	Schoolrandomisatie * Interventie	-2,16	10,57	-0,20	,838
	Gender * Interventie	-6,37	6,26	-1,02	,311
Taalverzorging (n = 318)	Intercept	112,49	5,18	21,70	< ,001
	Voormeting	0,60	0,02	25,86	< ,001
	Gender	-1,57	4,15	-,38	,706
	Schoolrandomisatie	-2,43	8,28	-,29	,770
	Interventie	6,75	4,77	1,41	,159
	Schoolrandomisatie * Interventie	21,97	9,95	2,21	,028
	Gender * Interventie	-1,51	5,63	-0,27	,789
Rekenen-Wiskunde (n = 312)	Intercept	46,78	4,98	9,40	< ,001
	Voormeting	0,88	0,03	28,28	< ,001
	Gender	-5,60	3,14	-1,78	,076
	Schoolrandomisatie	-0,63	6,06	-,10	,917
	Interventie	-6,99	3,41	-2,05	,041
	Schoolrandomisatie * Interventie	2,25	7,48	0,30	,764
	Gender * Thuis taal	1,54	4,22	0,36	,716

Bijlage C Behaalde power

Om na te gaan in hoeverre ons onderzoeksdesign in staat was de gevonden effecten te detecteren, hebben we een post-hoc poweranalyse uitgevoerd met het pakket *simr* in R. In deze analyse wordt het geschatte multilevelmodel van de hoofdanalyse gebruikt, waarna herhaalde simulaties worden gedaan om te bepalen hoe vaak een effect opnieuw significant zou worden bevonden. Dit levert de behaalde power op.

Een veelvoorkomende kritiek is dat deze benadering sterk samenhangt met de gevonden p-waarde⁶. Als een effect in de oorspronkelijke analyse niet significant was, zal de berekende power in de regel laag zijn, en omgekeerd. De behaalde power is daarmee deels te zien als een alternatieve manier om vergelijkbare informatie weer te geven.

De post-hoc poweranalyses bevestigen dat de kans om effecten te detecteren in ons onderzoeksdesign over het algemeen beperkt was. Voor de meeste uitkomsten (DMT, AVI, Woordenschat en Taalverzorging) lag de behaalde power rond de 5%. Gezien de geschatte effecten hier bovendien zeer klein of vrijwel nul waren, is het aannemelijk dat er in werkelijkheid geen effect van de interventie aanwezig was. Dat betekent dat de kans om een bestaand effect in dit onderzoeksdesign aan te tonen zeer klein was.

Voor Begrijpend lezen en Rekenen-Wiskunde was de behaalde power duidelijk hoger (respectievelijk 55% en 64%). Dat betekent dat de kans om een bestaand effect te detecteren substantieel groter was, maar nog steeds onvoldoende om stevige conclusies te trekken. Dit verklaart waarom het interventie-effect op deze domeinen in sommige modellen wel en in andere niet significant naar voren kwam.

Tabel C.1 Post hoc power voor de hoofdanalyse

Uitkomst	Geobserveerd effect Interventie	Behaalde power (95% CI)
DMT	$b \approx -0,37$	5,6% (4,3–7,2%)
AVI	$b \approx 0,009$	0% (0–0,4%)
Begrijpend lezen	$b \approx 8,2$	55% (52–58%)
Taalverzorging	$b \approx 2,0$	5% (3,8–6,7%)
Woordenschat	$b \approx 0,83$	5% (3,7–6,5%)
Rekenen-Wiskunde	$b \approx -6,3$	64% (61–67%)

6 Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician*, 55(1), 19–24.