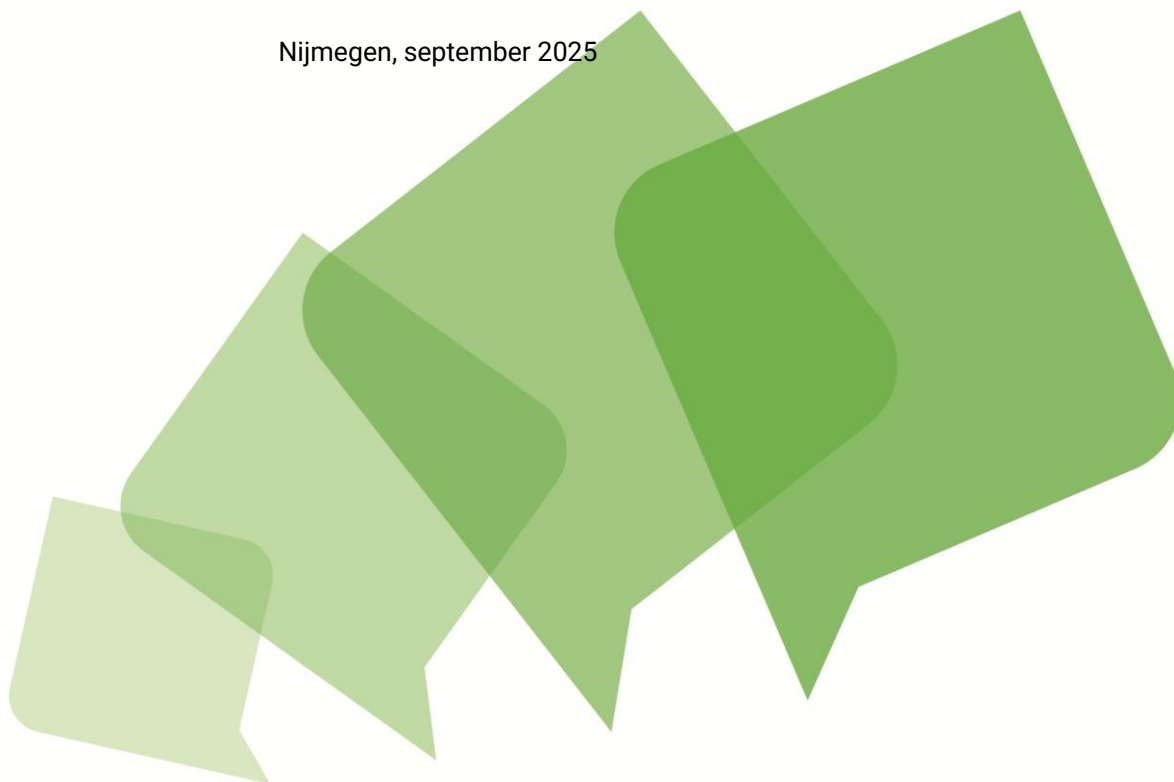


Het effect van de F&L-methode (opnieuw) onderzocht

*Onderzoek naar de effectiviteit van de F&L-methode
voor leerlingen met dyslexie*

Melissa in 't Zandt en Martine Gijsel

Nijmegen, september 2025



Project: E2022511

© 2025 Expertisecentrum Nederlands

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook, en evenmin in een retrieval systeem worden opgeslagen, zonder de voorafgaande schriftelijke toestemming van het Expertisecentrum Nederlandste Nijmegen.

No part of this book/publication may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.

Inhoud

1	Inleiding	4
2	Onderzoeksvragen	5
3	Methode	6
	3.1 Deelnemers	6
	3.2 Materiaal	6
	3.3 Procedure	7
	3.4 Bewerkingen dataset	7
	3.5 Analyses	8
4	Resultaten	10
	4.1 Effectiviteit F&L-methode	10
	4.2 Verschil effectiviteit naar doublure	12
	4.3 Verschil effectiviteit naar bouw	13
	4.4 Verschil effectiviteit naar geslacht	16
	4.5 Verschil effectiviteit naar comorbiditeit	17
	4.6 Verschil effectiviteit naar meertaligheid	19
	4.7 Samenhang tijdsinterval tussen doorverwijzing en start behandeling en effectiviteit van de behandeling	20
5	Samenvatting en conclusie	21
6	Referentielijst	24

1 Inleiding

Onderzoekers van het Expertisecentrum Nederlands (EN) zijn gevraagd om de Fonologische & Leerpsychologische Methode (hierna: de F&L-methode), op effectiviteit te onderzoeken. Het onderzoek moet antwoord geven op de vraag wat het effect is van de F&L-methode in de behandeling van leerlingen met dyslexie. Het gaat hierbij om leerlingen bij wie dyslexie is gediagnosticeerd en behandeling met de F&L-methode heeft plaatsgevonden binnen de kaders van het Protocol Dyslexie Diagnostiek en Behandeling (PDDDB) 2.0. Het onderzoek is een vervolg op de onderzoeken naar de behandelresultaten van de F&L-methode in 2008-2010 (Van Geffen et al., 2008; Gijsel, 2009; Gijsel & Bosman, 2010).

De opdrachtgever wilde graag weten of de behandelresultaten met de F&L-methode bij de leerlingen met dyslexie binnen de kaders van het PDDDB 2.0 vergelijkbaar zijn met de resultaten van de leerlingen uit de eerdere effectstudies. De onderzoeksgroepen verschillen namelijk op een aantal aspecten. Zo waren in de eerdere effectstudie leerlingen opgenomen met een diagnose dyslexie of met een vermoeden van (ernstige) dyslexie. De leerlingen uit de huidige onderzoeksgroep hebben allemaal de diagnose dyslexie gekregen alvorens met de behandeling te starten. De leerlingen uit de eerdere effectstudies hadden bovendien niet altijd gestructureerde interventie op het gebied van lezen en spellen van minimaal 20 weken gehad op school. Soms werd er enige begeleiding op school gegeven, maar soms ook niet. De hardnekkigheid van de problematiek van de leerlingen kon in de eerdere studies dan ook niet vooraf worden vastgesteld.

Ten behoeve van het onderzoek zijn gegevens aangeleverd van leerlingen die in de periode 2009 t/m 2024 in behandeling zijn geweest bij Stichting Taalhulp of bij Krebbers Logopedie en dyslexie. Het betreft ruim 1000 kinderen die een volledig behandeltraject hebben doorlopen in de vergoede dyslexiezorg. Alle leerlingen hebben vóór aanmelding op school een gestructureerde lees- en/of spellinginterventie gehad op ondersteuningsniveau 3 (ON3), volgens de richtlijnen van het Protocol Dyslexie Diagnostiek en Behandeling (PDDDB) 2.0. Omdat de problematiek zeer hardnekkig was en de interventie onvoldoende resultaat had, werden de leerlingen aangemeld voor het vergoede dyslexietraject (diagnostiek en behandeling volgens het Protocol Dyslexie Diagnostiek en Behandeling 2.0). Toestemming voor (geanonimiseerde) verwerking van de gegevens was geregeld door de behandelaars die werken met de F&L-methode. Middels een gegevensuitwisselingsovereenkomst is ervoor gezorgd dat de onderzoekers van EN deze data konden gebruiken voor de effectstudie. Het gereed maken van het databestand is in overleg tussen de opdrachtgever en de onderzoekers gebeurd; in juli 2024 is de definitieve versie van het databestand gedeeld met de onderzoekers.

Met de opdrachtgever is afgesproken om net als in de eerdere effectstudie (Gijsel, 2009; Gijsel & Bosman, 2010) gebruik te maken van de didactische leeftijdsequivalenten (dle's) als maat voor de lees- en spellingvaardigheid van leerlingen. We zijn ons bewust van de kritiek op het gebruik van dle's, maar gezien het veelvuldige gebruik hiervan in de Nederlandse behandelpraktijk en de mogelijkheid om de vooruitgang van leerlingen tijdens de behandeling te kunnen bepalen zonder een controlegroep, is het gebruik van dle's in dit onderzoek goed te onderbouwen.

In deze rapportage beschrijven we de bevindingen van het onderzoek. In paragraaf 2 is het overzicht van de onderzoeksvragen te vinden. Paragraaf 3 beschrijft de methode van het onderzoek. In paragraaf 4 worden de toetsresultaten gepresenteerd en wordt antwoord gegeven op de onderzoeksvragen. In paragraaf 5 ten slotte, is de samenvatting en conclusie weergegeven.

2 Onderzoeksvragen

In overleg met de opdrachtgever is de volgende hoofdonderzoeksvraag opgesteld:

1. In hoeverre is de F&L-methode effectief voor het bevorderen van de lees- en spellingvaardigheid op woord- en zinsniveau?

Op basis van eerdere onderzoeken (Van Geffen et al., 2008; Gijsel, 2009; Gijsel & Bosman, 2010) is onze verwachting dat de F&L-methode een positief effect heeft op zowel de lees- als spellingvaardigheid van leerlingen met dyslexie.

Naast het onderzoeken van de effectiviteit voor de totale groep leerlingen, wordt onderzocht of de effectiviteit anders is voor groepen leerlingen met verschillende leerlingkenmerken. De hierbij horende onderzoeksvragen zijn als volgt:

2. Is er een verschil in de effectiviteit van de F&L-methode tussen
 - a. leerlingen die gedoubleerd hebben en leerlingen die niet gedoubleerd hebben?
 - b. leerlingen in de onder- (groep 3 en 4), midden- (groep 5 en 6) en bovenbouw (groep 7 en 8) van de basisschool?
 - c. jongens en meisjes?
 - d. leerlingen met comorbide problemen en leerlingen zonder comorbide problemen?
 - e. meertalige leerlingen en niet-meertalige leerlingen?

We verwachten dat de effectiviteit van de F&L-methode groter is in de onderbouw dan in de midden- en bovenbouw. In de studie van Gijsel & Bosman (2010) bleek een interactie-effect tussen bouw en testmoment bij het PI-dictee en DMT kaart 3: het verschil in leerrendement tussen het eerste en tweede moment was groter bij de onderbouw (groep 3 en 4) dan bij de middenbouw (groep 5 en 6) en bovenbouw (groep 7-8). Ten aanzien van de andere leerlingkenmerken (doubleure, geslacht, comorbiditeit en meertaligheid) hebben we vooraf geen specifieke hypothesen.

De laatste onderzoeksvraag die wordt beantwoord, is:

3. Hangt het tijdsinterval tussen de derde hoofdmeting op school en de start van de behandeling samen met de effectiviteit van de behandeling op de lees- en spellingresultaten?

Ook hier hebben we vooraf geen specifieke verwachting over de resultaten.

3 Methode

3.1 Deelnemers

De opdrachtgever heeft een dataset aangeleverd, met daarin gegevens van in totaal 1056 leerlingen. Voor het uitvoeren van de analyses zijn een aantal exclusiecriteria gehanteerd. Zo zijn leerlingen met meer dan zeven testmomenten uit de dataset verwijderd ($n = 4$) en er is een leerling zonder resultaten van de diagnostiek verwijderd ($n=1$). Ook leerlingen bij wie geen enkele toets op zowel het eerste (t_1) als het laatste testmoment (t_2) was afgenomen, zijn verwijderd uit de dataset ($n = 48$), omdat voor hen geen vooruitgang te bepalen was van het begin tot het eind van de behandeling.

Dit resulteerde in 1003 leerlingen die we hebben meegenomen in de analyses, waarvan minimaal gegevens waren van diagnostiek en één aanvullend testmoment (evaluatie of eindmeting). De onderzoeksgroep bestond uit 552 jongens (55.0%) en 451 meisjes (45.0%), waarvan 911 leerlingen bij Stichting Taalhulp in behandeling waren en 92 leerlingen bij praktijk Krebbers Logopedie en dyslexie. Op het moment van aanmelding bij de dyslexiepraktijk, zat 32.7% ($n = 328$) van de leerlingen in de onderbouw, 57.9% ($n = 581$) in de middenbouw en 9.4% ($n = 94$) in de bovenbouw van het basisonderwijs. In totaal hebben 255 leerlingen (25.4%) een keer gedoubleerd in groep 3 t/m 8, op het moment van aanmelding of op een ander moment tijdens de looptijd van de behandeling. Andere kenmerken van de totale onderzoeksgroep zijn weergegeven in Tabel 1.

Tabel 1. Gemiddelde (en standaardafwijking) leeftijd, aantal testmomenten en aantal behandelssessies en bereik ($n = 1003$).

Start van de behandeling					
		Onderbouw	Middenbouw	Bovenbouw	Totaal
Leeftijd bij aanmelding (in maanden)					
	gemiddelde	101.3 (6.6)	115.3 (9.5)	136.4 (8.6)	112.7 (13.1)
	bereik	85.9 – 124.8	74.9 – 152.6	101.4 – 156.9	74.9 – 156.9
Aantal testmomenten (excl. diagnostiek)					
	gemiddelde	3.3 (1.0)	3.0 (0.8)	2.4 (0.7)	3.0 (0.9)
	bereik	1 – 6	1 – 6	1 – 5	1 – 6
Aantal behandelssessies					
	gemiddelde	47.0 (9.0)	44.4 (9.4)	36.9 (10.8)	44.5 (9.8)
	bereik	12 – 94	10 - 100	11 – 59	10 – 100

Van de leerlingen die in behandeling waren bij Stichting Taalhulp ($n = 911$), zijn aanvullende gegevens bekend over: het voorkomen van lees- en/of spellingproblemen bij vader, moeder, broer of zus, meertaligheid en comorbiditeit (bijv. AD(H)D, autisme, dyscalculie). Bij 69.0% van deze leerlingen ($n = 625$) was sprake van een indicatie van erfelijkheid van de lees- en/of spellingproblemen, bij 7.4% ($n = 67$) van meertaligheid en bij 6.7% ($n = 61$) van (gediagnosticeerd) comorbiditeit.

3.2 Materiaal

Om de effectiviteit van de F&L-methode in kaart te brengen, is gebruikgemaakt van zes verschillende toetsen voor verschillende lees- en spellingvaardigheden. Voor het vaststellen van het leesniveau van losse woorden zijn de Drie-Minuten-Toets (alle kaarten) (DMT; Cito, 2009; Cito, 2017b) en de Een-Minuut

Test (EMT; Brus & Voeten, 1973) gebruikt. De leesvaardigheid van pseudowoorden is vastgesteld aan de hand van de Klepel (Van den Bos, et al., 1994). Het meten van de leesvaardigheid van teksten is gedaan met behulp van de AVI-niveaukaarten (Cito, 2009; Cito 2017a; Visser et al., 1994). De spellingvaardigheid van leerlingen is vastgesteld met twee verschillende toetsen, namelijk het PI-dictee (woorddictee) (Geelhoed & Reitsma, 1999) en het ZISB-dictee (zinnendictee) afkomstig van het Orthopedagogisch Instituut te Amsterdam.

In de dataset waren scores opgenomen van verschillende versies van eenzelfde toets: van de DMT waren scores aanwezig van de versies uit 2009 en 2017, van de EMT, Klepel en het PI-dictee van de versies uit 1999 en 2018 en van de AVI waren scores beschikbaar van de versies uit 1994, 2009 en 2017. Voor het ZISB-dictee zijn alle verschillende lengtes afgenomen, oftewel 4, 8, 12, 16 en 20 zinnen. Deze verschillende versies van de toetsen kwamen voor op verschillende testmomenten in de dataset.

3.3 Procedure

Het effectonderzoek had een quasi-experimenteel design. Er was geen controlegroep mogelijk, omdat alle leerlingen behandeling volgens de F&L-methode hadden ontvangen en het onderzoek is uitgevoerd op een bestaande dataset. Vanwege het ontbreken van de controlegroep is op een andere manier de effectiviteit van de behandeling vastgesteld (zie hieronder).

Om de vooruitgang van de leerlingen te bepalen, zijn de ruwe scores op de zes verschillende lees- en spellingtoetsen op het eerste testmoment (t1, diagnostiek, pre-test) en het laatste testmoment (t2, eindmeting, post-test) van een leerling omgezet naar didactische leeftijdsequivalenten (dle) en leerrendementen (lr). Het eerste testmoment was voor elke leerling het diagnostisch onderzoek. Het laatste testmoment was de eindmeting van de betreffende leerling, na afloop van de behandeling. Het verschil tussen leerlingen om welk testmoment het precies ging: het laatste testmoment kan voor leerlingen het tweede t/m het zevende testmoment zijn. De vooruitgang van leerlingen is dus bepaald aan de hand van de scores voor en na de behandeling (pre- en post).

Het didactische leeftijdsequivalent (dle) geeft de didactische leeftijd aan waarop de bijbehorende score gemiddeld wordt behaald. De didactische leeftijd (dl) verwijst naar het aantal maanden dat een leerling formeel lees- en spellingonderwijs heeft genoten. Met deze twee gegevens is het leerrendement (lr) op het eerste (t1) en laatste testmoment (t2) bepaald ($lr = dle/dl$). De gemiddelde (niet-dyslectische) leerling heeft een leerrendement van 1, het niveau (dle) komt dan dus overeen met de didactische leeftijd (dl). Daarnaast is ook het leerrendement tussen de twee testmomenten berekend, om de vooruitgang van een leerling te bepalen ($lr \text{ van behandeling} = \frac{dle_2 - dle_1}{dl_2 - dl_1}$).

3.4 Bewerkingen dataset

De dataset die de onderzoekers hebben ontvangen, moest gecontroleerd en deels bewerkt worden om de analyses goed uit te kunnen voeren. Ten eerste was er een aanpassing nodig in de weergave van de data. Daarnaast is voor elke leerling de eindmeting bepaald en op de juiste manier in de dataset aangegeven. Ook moesten er extra variabelen aangemaakt worden (zoals leerrendement per testmoment). Bij het controleren van de data bleek dat er verschillende onjuistheden in de data zaten, zoals ontbrekende waarden waar dat niet zou moeten, een andere dl bij verschillende toetsen op één testmoment en een ruwe score en dle die niet overeenkwamen. Deze incongruenties zijn voorgelegd aan de opdrachtgever en zo goed als kon opgelost door deze gegevens nogmaals op te zoeken in de dossiers. Hoewel er nog een kans bestaat dat er nog steeds kleine onjuistheden in de dataset zitten,

gaan we er gezien de omvang van de dataset niet vanuit dat dit de uitkomsten van de analyses zal beïnvloeden.

Naast bovenstaande bewerkingen, moest ook een oplossing gevonden worden voor het feit dat sommige leerlingen op één testmoment een score voor meerdere versies van eenzelfde toets hadden (bijv. een score voor zowel DMT2009 als DMT2018 op het moment van diagnostiek). Dit kwam voor bij 45 leerlingen, op het eerste en/of het laatste testmoment, bij een of meerdere van deze toetsen: DMT, EMT, Klepel, AVI en het PI-dictee. Bij deze leerlingen is gekeken of dit op beide testmomenten het geval was. Indien een leerling op zowel het eerste als het laatste testmoment voor twee versies van een toets een score had, is de score op de meest recente versie van de toets behouden (bijv. de score op DMT2018 i.p.v. DMT2009). Indien een leerling enkel op een van de twee testmomenten een score op meerdere versies van een toets had, is gekeken welke versie behouden moest blijven, om een verschilscore tussen het eerste en het laatste testmoment te kunnen berekenen. Als een leerling bijvoorbeeld een score op AVI1994 en AVI2009 had op het eerste testmoment, maar alleen een score op AVI1994 op het laatste testmoment, dan is de score op AVI1994 behouden en die op AVI2009 verwijderd bij het eerste testmoment.

Bij het ZISB-dictee zagen we bij 18 leerlingen dat op het eerste en/of het laatste testmoment scores voor verschillende lengteversies van het dictee voorkwamen. Bij dit dictee is ervoor gekozen om de score van de versie waarbij de meeste zinnen zijn afgenomen, te behouden¹. Wanneer bij een leerling bijvoorbeeld zowel een score voor vier als voor acht zinnen bekend was, is alleen de score van acht zinnen behouden voor het onderzoek.

Daarnaast is op verzoek van de opdrachtgever gekeken hoe leerlingen die te laag scoorden om een dle toegekend te krijgen, toch meegenomen konden worden in de analyse. De dle-normering begint past bij dle=5, dus deze leerlingen hadden een ruwe score die overeenkwam met een dle < 5. In overleg met de opdrachtgever is afgesproken om deze leerlingen allemaal op dle = 3 te zetten voor de analyse. Op die manier konden alle leerlingen, ook die leerlingen die net onder de dle-normering scoorden, meegenomen worden in de analyses. Op de pre-test varieerde het aantal leerlingen met dle=3 van 99 bij de AVI tot 232 bij de Klepel en op de post-test van één leerling bij het PI-dictee tot 18 leerlingen bij de Klepel.

3.5 Analyses

Voor het beantwoorden van onderzoeksvragen 1 (effectiviteit) en 2a, 2b en 2c (effect doublure, bouw en geslacht) is voor elke toets een variantieanalyse uitgevoerd, met leerrendement als afhankelijke variabele, moment (pre- versus post-test) als binnenproefpersonenfactor en doublure (wel vs. niet doublure), bouw (onder- vs. midden- vs. bovenbouw) en geslacht (jongen vs. meisje) als tussenproefpersonenfactoren.

Voor het beantwoorden van onderzoeksvragen 2d (comorbiditeit) en 2e (meertaligheid) zijn aparte onderzoeksgroepen gecreëerd, om de leerlingen met comorbide problemen en meertalige leerlingen mee te kunnen vergelijken. Er zat namelijk een te groot verschil tussen de groepsgroottes met en zonder deze kenmerken wanneer we met alle leerlingen uit de dataset zouden werken. Bij comorbiditeit zou het dan gaan om een vergelijking van n = 845 (geen comorbide problemen) met n = 61 leerlingen (wel comorbide problemen) en bij meertaligheid om een vergelijking tussen n = 844 (niet-meertalig) en n =

¹ In de afnameprocedure is vastgesteld dat op basis van het aantal fouten dat een leerling maakt, wordt doorgetoetst, dat wil zeggen meer zinnen worden gedictreed.

67 (meertalig). Bij deze aantallen gaat het alleen om leerlingen van Stichting Taalhulp, omdat deze achtergrondgegevens bij de leerlingen van praktijk Krebbers niet bekend waren. Door de grote verschillen in groepsgroottes hebben we ervoor gekozen om twee groepen van vergelijkbare groottes te creëren voor beide factoren (comorbiditeit en meertaligheid) en daarmee de analyses uit te voeren. Het creëren van de vergelijkingsgroepen is gedaan via matching (Propensity Score Matching, PMS) op basis van de volgende kenmerken: geslacht, leeftijd (in maanden), aantal ontvangen behandelsessies en een gemiddelde score van de drie op de zes lees- en spellingtoetsen² op het eerste testmoment (diagnostiek). Op basis van deze kenmerken kon voor bijna alle meertalige leerlingen en leerlingen met comorbide problemen een match gevonden worden. Voor comorbiditeit bestond de onderzoeksgroep uit 112 leerlingen (wel comorbide problemen = 56, geen comorbide problemen = 56) en voor meertaligheid uit 122 leerlingen (meertalig = 61, niet-meertalig = 61).

Voor onderzoeksvraag 2d is voor elke toets een 2 (comorbiditeit: wel vs. niet) x 2 (moment: pre- versus post-test) variantieanalyse uitgevoerd, met leerrendement als afhankelijke variabele, moment als binnenproefpersonenfactor en comorbiditeit als tussenproefpersonenfactoren.

Voor onderzoeksvraag 2e is voor elke toets een 2 (meertalig: wel vs. niet) x 2 (moment: pre- versus post-test) variantieanalyse uitgevoerd, met leerrendement als afhankelijke variabele, moment als binnenproefpersonenfactor en meertaligheid als tussenproefpersonenfactoren.

Voor het beantwoorden van onderzoeksvraag 3 zijn drie regressieanalyses uitgevoerd, met het tijdsinterval tussen hoofdmeting 3 op school en de start van de behandeling als onafhankelijke variabele en het verschil in leerrendement tussen de pre- en post-test (delta leerrendement) voor de DMT, AVI en het PI-dictee als afhankelijke variabelen.

² Sommige leerlingen misten scores voor een of meerdere van deze zes toetsen. Voor hen is het gemiddelde berekend van de scores op de toetsen die ze wel hadden (bijv. vier of vijf van de toetsen).

4 Resultaten

In Tabel 2 zijn de ruwe scores op de pre- en post-test en het aantal waarnemingen bij de verschillende versies van de zes toetsen weergegeven. Bij de DMT, EMT en Klepel is dit het aantal goed gelezen woorden per minuut, bij de AVI het hoogste beheersingsniveau, bij het PI-dictee het aantal goed geschreven woorden en bij het ZISB-dictee het aantal gemaakte fouten. Het aantal leerlingen in de analyses per toets verschilt, omdat niet alle leerlingen alle lees- en spellingtoetsen aangeboden hebben gekregen op alle testmomenten.

Tabel 2. Gemiddelde ruwe score (en standaarddeviatie) en aantal observaties op de pre- en post-test, bij de verschillende toetsen.

	Pre-test		Post-test	
	m (sd)	n	m (sd)	n
Woorden lezen				
DMT2009	106.1 (52.4)	655	186.4 (56.9)	643
DMT2018	92.0 (47.7)	301	160.4 (54.9)	313
EMT1999	28.0 (12.0)	797	46.9 (12.8)	793
EMT2018	27.5 (11.4)	95 ³	44.5 (12.9)	99
Pseudowoorden lezen				
Klepel1999	17.8 (8.8)	779	34.0 (11.9)	775
Klepel2018	17.4 (9.1)	94	31.5 (11.0)	98
Tekst lezen				
AVI1994 ⁴	AVI-1 instructieniveau	23	AVI 8	19
AVI2009	AVI-E3	497	AVI-E5	465
AVI2018	AVI-E3	264	AVI-M6	300
Woorden spellen				
PI-dictee1999	43.8 (22.5)	808	96.9 (21.1)	802
PI-dictee2018	45.1 (27.1)	85	96.9 (21.8)	91
Zinnen spellen				
ZISB – 4 zinnen	9.8 (5.8)	80	5.5 (3.5)	2
ZISB – 8 zinnen	16.3 (6.8)	259	8.4 (4.7)	27
ZISB – 12 zinnen	18.6 (8.6)	282	9.5 (5.1)	186
ZISB – 16 zinnen	20.0 (9.3)	71	10.3 (5.6)	186
ZISB – 20 zinnen	21.7 (15.6)	32	8.9 (5.2)	323

4.1 Effectiviteit F&L-methode

De eerste onderzoeksvraag was: In hoeverre is de F&L-methode effectief voor het bevorderen van de lees- en spellingvaardigheid op woord- en zinsniveau? We verwachtten een positief effect van de behandeling op zowel lezen als spellen. Voor het berekenen van het leerrendement zijn de verschillende

³ Het totaal aantal leerlingen bij EMT2009 en EMT2018 (n=892) en bij PI-dictee1999 en PI-dictee2018 (n=893) wijkt af van het aantal leerlingen in de analyses, doordat de ruwe score van een aantal leerlingen ontbrak, maar de dle en dl wel bekend waren.

⁴ Bij de scores op AVI1994, AVI2009 en AVI2018 gaat het om de modus i.p.v. het gemiddelde.

versies van een toets samengenomen⁵. In Tabel 3 t/m 8 komt elke toets daarom maar een keer voor. De hoofdeffecten van testmoment zijn weergegeven in Tabel 3.

De resultaten zijn in lijn met onze verwachting: bij alle zes de toetsen steeg het leerrendement significant tussen de pre-test en post-test. Het gaat hier om grote effecten (zie Cohen, 1988 voor een interpretatie van de effectgroottes). Voor spelling waren de effecten groter dan voor lezen en voor tekst lezen groter dan voor woorden lezen.

Tabel 3. Gemiddelde dl, dle en lr op de pre- en post-test, verschil in lr tussen pre- en post-test, en de hoofdeffecten van testmoment bij de verschillende toetsen.

		m (sd)		Delta lr (pre-test – post-test)		Hoofdeffect testmoment
		n	pre-test	post-test		
DMT		956			0.83	$F(1,951) = 408.67; p < 0.001$; partiële $\eta^2 = 0.30$
	dl		30.1	44.5		
	dle		10.0	21.6		
	lr		0.32	0.48		
EMT		933			0.73	$F(1,928) = 227.47; p < 0.001$; partiële $\eta^2 = 0.20$
	dl		30.1	44.4		
	dle		11.9	22.0		
	lr		0.39	0.50		
Klepel		873			0.79	$F(1,868) = 244.61; p < 0.001$; partiële $\eta^2 = 0.22$
	dl		30.3	44.5		
	dle		9.3	20.1		
	lr		0.30	0.46		
AVI		784			1.16	$F(1,779) = 459.20; p < 0.001$; partiële $\eta^2 = 0.37$
	dl		30.3	44.6		
	dle		14.7	30.8		
	lr		0.46	0.68		
PI-dictee		933			1.69	$F(1,928) = 1282.01; p < 0.001$; partiële $\eta^2 = 0.58$
	dl		30.1	44.4		
	dle		12.6	35.9		
	lr		0.41	0.81		
ZISB-dictee		724			1.60	$F(1,719) = 647.44; p < 0.001$; partiële $\eta^2 = 0.47$
	dl		30.0	44.1		
	dle		13.2	35.1		
	lr		0.43	0.80		

⁵ Bij Gerard Melis, mede-ontwikkelaar van de dle-schaal, is nagevraagd of dle's van meerdere versies van eenzelfde toets samengevoegd mogen worden. Dit bleek inderdaad het geval te zijn.

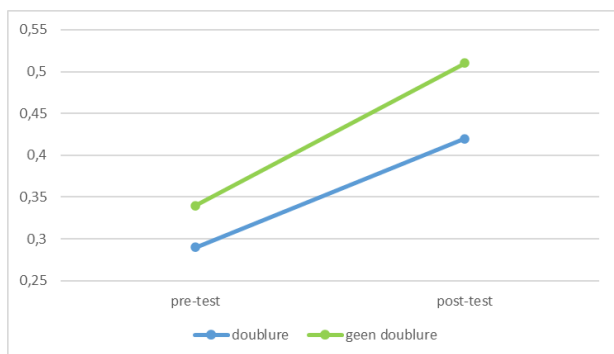
4.2 Verschil effectiviteit naar doublure

Vraag 2a betrof het effect van doublure: is er een verschil in de effectiviteit van de F&L-methode tussen leerlingen die gedoubleerd hebben en leerlingen die niet gedoubleerd hebben? We hadden hier vooraf geen specifieke verwachting over. De interactie-effecten tussen doublure en testmoment zijn weergegeven in Tabel 4.

Uit de resultaten bleek een significant interactie-effect bij alle zes de toetsen. De stijging in leerrendement tussen pre-test en post-test was significant groter voor leerlingen die niet gedoubleerd hadden, dan voor leerlingen die wel gedoubleerd hadden. Dit betrof kleine effecten, bij alle zes de toetsen. De interacties bij de zes verschillende toetsen zijn geïllustreerd in Figuur 1 t/m 6.

Tabel 4. Gemiddelde leerrendement (lr) op de pre- en post-test, uitgesplitst voor doublure, en de interactie-effecten bij de verschillende toetsen.

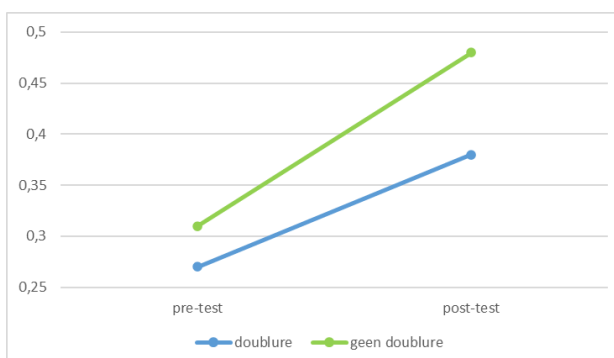
		m (pre-test)	m (post-test)	Interactie testmoment * doublure
DMT				$F(1,951) = 10.29$; $p = 0.001$; partiële $\eta^2 = 0.01$
doublure	ja	0.29	0.42	
	nee	0.34	0.51	
EMT				$F(1,928) = 6.45$; $p = 0.011$; partiële $\eta^2 = 0.01$
doublure	ja	0.33	0.42	
	nee	0.41	0.52	
Klepel				$F(1,868) = 15.32$; $p < 0.001$; partiële $\eta^2 = 0.02$
doublure	ja	0.27	0.38	
	nee	0.31	0.48	
AVI				$F(1,779) = 6.75$; $p = 0.010$; partiële $\eta^2 = 0.01$
doublure	ja	0.41	0.58	
	nee	0.50	0.71	
PI-dictee				$F(1,928) = 30.12$; $p < 0.001$; partiële $\eta^2 = 0.03$
doublure	ja	0.36	0.67	
	nee	0.46	0.85	
ZISB-dictee				$F(1,719) = 11.61$; $p < 0.001$; partiële $\eta^2 = 0.02$
doublure	ja	0.36	0.66	
	nee	0.47	0.84	



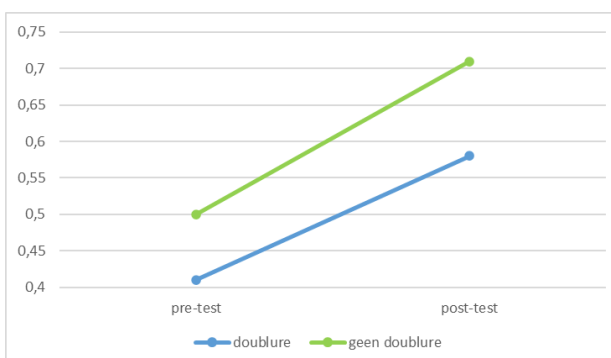
Figuur 1. Leerrendement op de DMT op pre- en post-test, uitgesplitst voor doublure.



Figuur 2. Leerrendement op de EMT op pre- en post-test, uitgesplitst voor doublure.



Figuur 3. Leerrendement op de Klepel op pre- en post-test, uitgesplitst voor doublure.



Figuur 4. Leerrendement op de AVI op pre- en post-test, uitgesplitst voor doublure.



Figuur 5. Leerrendement op het PI-dictee op pre- en post-test, uitgesplitst voor doublure.



Figuur 6. Leerrendement op het ZISB-dictee op pre- en post-test, uitgesplitst voor doublure.

4.3 Verschil effectiviteit naar bouw

Vraag 2b betrof het effect van bouw: is er een verschil in de effectiviteit van de F&L-methode tussen leerlingen in de onder- (groep 3 en 4), midden- (groep 5 en 6) en bovenbouw (groep 7 en 8) van de basisschool? Onze verwachting was dat de effectiviteit groter zou zijn in de onderbouw dan in de midden- en bovenbouw. De interactie-effecten tussen bouw en testmoment zijn weergegeven in Tabel 5.

De resultaten lieten een significant interactie-effect bij alle toetsen zien, behalve bij de DMT. Dit betekent dat de stijging in leerrendement tussen pre-test en post-test niet hetzelfde was voor leerlingen in de onder-, midden- en bovenbouw. Het ging hierbij om kleine tot middelgrote effecten.

Tabel 5. Gemiddelde leerrendement (lr) op de pre- en post-test, uitgesplitst voor bouw, en de interactie-effecten bij de verschillende toetsen.

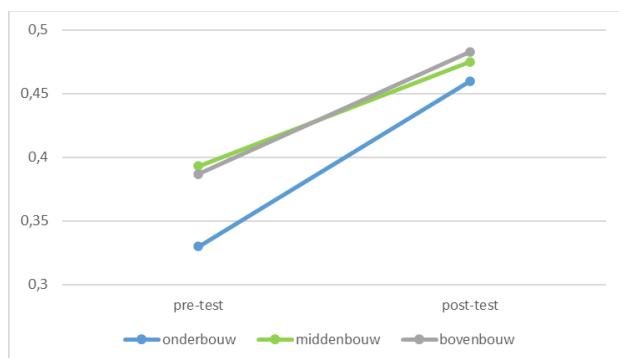
		m (pre-test)	m (post-test)	Interactie testmoment * bouw
DMT				$F(2,951) = 0.32; p = 0.724$; partiële $\eta^2 = 0.001$
	onderbouw	0.25	0.40	
	middenbouw	0.33	0.48	
	bovenbouw	0.38	0.51	
EMT				$F(2,928) = 11.21; p < 0.001$; partiële $\eta^2 = 0.02$
	onderbouw	0.33	0.46	
	middenbouw	0.39	0.48	
	bovenbouw	0.39	0.48	
Klepel				$F(2,868) = 8.63; p < 0.001$; partiële $\eta^2 = 0.02$
	onderbouw	0.26	0.43	
	middenbouw	0.30	0.43	
	bovenbouw	0.32	0.42	
AVI				$F(2,779) = 16.57; p < 0.001$; partiële $\eta^2 = 0.04$
	onderbouw	0.34	0.58	
	middenbouw	0.47	0.69	
	bovenbouw	0.56	0.67	
PI-dictee				$F(2,928) = 16.64; p < 0.001$; partiële $\eta^2 = 0.04$
	onderbouw	0.33	0.74	
	middenbouw	0.40	0.79	
	bovenbouw	0.49	0.75	
ZISB-dictee				$F(2,719) = 17.26; p < 0.001$; partiële $\eta^2 = 0.05$
	onderbouw	0.34	0.76	
	middenbouw	0.43	0.77	
	bovenbouw	0.48	0.71	

Om meer zicht te krijgen op het verschil in de stijging tussen pre- en post-test tussen de bouwen, zijn afzonderlijke analyses voor onderbouw, middenbouw en bovenbouw uitgevoerd. Die lieten zien dat er bij alle toetsen een significant hoofdeffect van moment was bij alle drie de bouwen, oftewel: voor onderbouw, middenbouw en bovenbouw is er een significante stijging in leerrendement tussen pre-test en post-test. De resultaten van deze afzonderlijke toetsen zijn te vinden in Tabel 6.

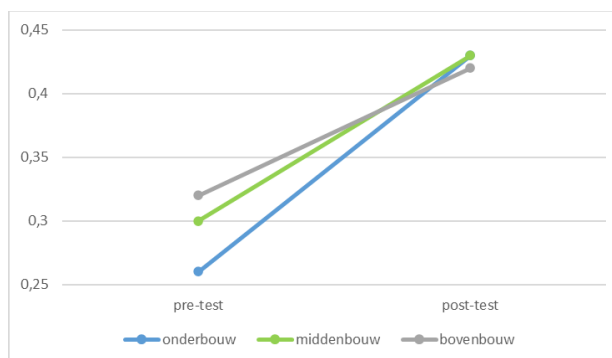
Tabel 6. Resultaten afzonderlijke analyses voor onderbouw, middenbouw en bovenbouw bij de verschillende toetsen (behalve de DMT).

Resultaten		
EMT	onderbouw	$F(1, 928) = 247.57, p < 0.001$, partiële $\eta^2 = .21$
	middenbouw	$F(1, 928) = 139.86, p < 0.001$, partiële $\eta^2 = .13$
	bovenbouw	$F(1, 928) = 36.14, p < 0.001$, partiële $\eta^2 = .04$
Klepel	onderbouw	$F(1, 868) = 275.30, p < 0.001$, partiële $\eta^2 = .24$
	middenbouw	$F(1, 868) = 231.59, p < 0.001$, partiële $\eta^2 = .21$
	bovenbouw	$F(1, 868) = 23.03, p < 0.001$, partiële $\eta^2 = .03$
AVI	onderbouw	$F(1, 779) = 493.70, p < 0.001$, partiële $\eta^2 = .39$
	middenbouw	$F(1, 779) = 544.46, p < 0.001$, partiële $\eta^2 = .41$
	bovenbouw	$F(1, 779) = 30.66, p < 0.001$, partiële $\eta^2 = .04$
PI-dictee	onderbouw	$F(1, 928) = 1149.78, p < 0.001$, partiële $\eta^2 = .55$
	middenbouw	$F(1, 928) = 1542.52, p < 0.001$, partiële $\eta^2 = .62$
	bovenbouw	$F(1, 928) = 126.66, p < 0.001$, partiële $\eta^2 = .12$
ZISB-dictee	onderbouw	$F(1, 719) = 725.78, p < 0.001$, partiële $\eta^2 = .50$
	middenbouw	$F(1, 719) = 713.97, p < 0.001$, partiële $\eta^2 = .50$
	bovenbouw	$F(1, 719) = 54.74, p < 0.001$, partiële $\eta^2 = .07$

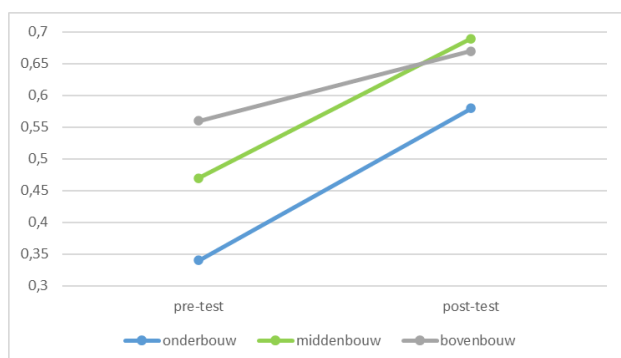
De resultaten in Tabel 6 laten zien dat de stijging in leerrendement het grootst was in de onderbouw bij de EMT, het grootst in de middenbouw bij het PI-dictee en het grootst in de onderbouw en middenbouw bij de Klepel, AVI en het ZISB-dictee. Deze uitkomsten zijn deels in lijn met onze verwachting, dat de effectiviteit groter zou zijn in de onderbouw. De interacties bij de vijf verschillende toetsen zijn geïllustreerd in Figuur 7 t/m 11 hieronder.



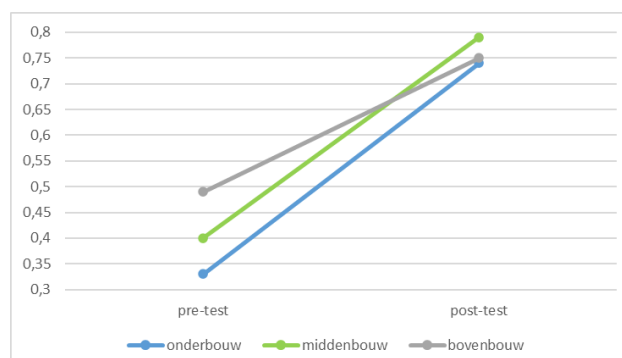
Figuur 7. Leerrendement op de EMT op pre- en post-test, uitgesplitst voor bouw.



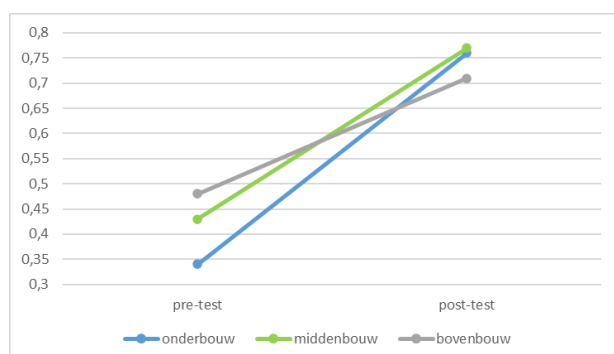
Figuur 8. Leerrendement op de Klepel op pre- en post-test, uitgesplitst voor bouw.



Figuur 9. Leerrendement op de AVI op pre- en post-test, uitgesplitst voor bouw.



Figuur 10. Leerrendement op het PI-dictee op pre- en post-test, uitgesplitst voor bouw.



Figuur 11. Leerrendement op het ZISB-dictee op pre- en post-test, uitgesplitst voor bouw.

4.4 Verschil effectiviteit naar geslacht

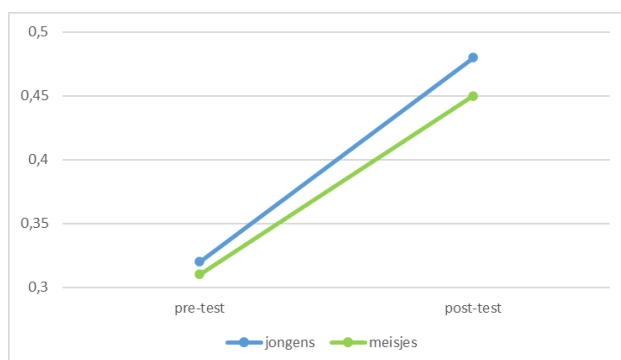
Vraag 2c betrof het effect van geslacht: is er een verschil in de effectiviteit van de F&L-methode tussen jongens en meisjes? We hadden hier vooraf geen specifieke verwachting over. De interactie-effecten tussen geslacht en testmoment zijn weergegeven in Tabel 7.

Tabel 7. Gemiddelde leerrendement (lr) op de pre- en post-test, uitgesplitst voor geslacht, en de interactie-effecten bij de verschillende toetsen.

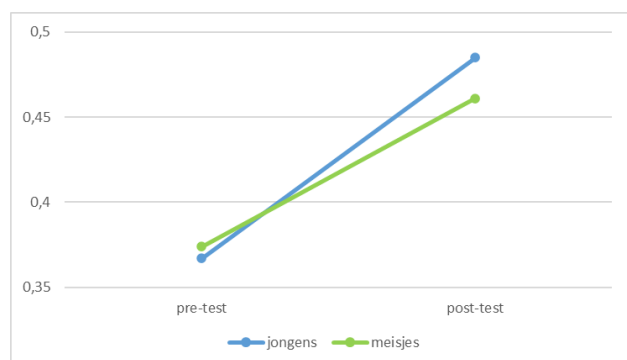
		m (pre-test)	m (post-test)	Interactie testmoment * geslacht
DMT				$F(1,951) = 4.10$; $p = 0.043$; partiële $\eta^2 = 0.004$
	Jongens	0.32	0.48	
	Meisjes	0.31	0.45	
EMT				$F(1,928) = 11.33$; $p < 0.001$; partiële $\eta^2 = 0.01$
	Jongens	0.37	0.49	
	Meisjes	0.37	0.46	
Klepel				$F(1,868) = 0.51$; $p = 0.475$; partiële $\eta^2 = 0.001$
	Jongens	0.31	0.45	
	Meisjes	0.28	0.41	
AVI				$F(1,779) = 1.87$; $p = 0.172$; partiële $\eta^2 = 0.002$

	Jongens	0.45	0.65	
	Meisjes	0.47	0.65	
PI-dictee				$F(1,928) = 3.47; p = 0.063$; partiële $\eta^2 = 0.004$
	Jongens	0.38	0.74	
	Meisjes	0.45	0.78	
ZISB-dictee				$F(1,719) = 0.12; p = 0.732$; partiële $\eta^2 < 0.001$
	Jongens	0.38	0.72	
	Meisjes	0.45	0.78	

Uit de resultaten bleek een significant interactie-effect bij de DMT en EMT. De stijging in leerrendement tussen pre-test en post-test was significant groter voor jongens dan voor meisjes. Dit betrof (zeer) kleine effecten. De interacties bij de twee toetsen zijn geïllustreerd in Figuur 12 en 13 hieronder.



Figuur 12. Leerrendement op de DMT op pre- en post-test, uitgesplitst voor geslacht.



Figuur 13. Leerrendement op de EMT op pre- en post-test, uitgesplitst voor geslacht.

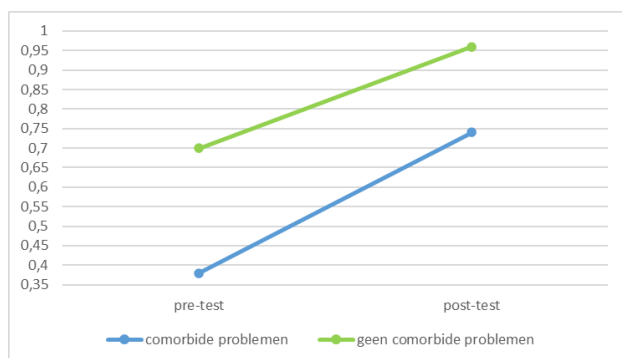
4.5 Verschil effectiviteit naar comorbiditeit

Vraag 2d betrof het effect van comorbiditeit: is er een verschil in de effectiviteit van de F&L-methode tussen leerlingen met comorbide problemen en leerlingen zonder comorbide problemen? We hadden hier vooraf geen specifieke verwachting over. Voor de analyse voor deze onderzoeksvraag is gebruikgemaakt van een subgroep van de dataset, bestaande uit 112 leerlingen (56 met comorbiditeit, 56 als 'controlegroep'). De interactie-effecten tussen comorbiditeit en testmoment zijn weergegeven in Tabel 8.

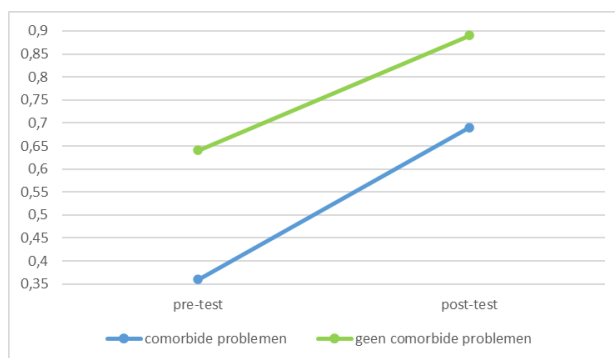
Tabel 8. Gemiddelde leerrendement (lr) op de pre- en post-test, uitgesplitst voor comorbiditeit, en de interactie-effecten bij de verschillende toetsen.

		m (pre-test)	m (post-test)	Interactie testmoment * comorbiditeit
DMT				$F(1,105) = 0.59$; $p = 0.446$; partiële $\eta^2 = 0.01$
Comorbiditeit	Ja	0.30	0.43	
	Nee	0.39	0.54	
EMT				$F(1,109) = 0.35$; $p = 0.556$; partiële $\eta^2 = 0.003$
Comorbiditeit	Ja	0.36	0.46	
	Nee	0.42	0.50	
Klepel				$F(1,107) = 0.87$; $p = 0.354$; partiële $\eta^2 = 0.01$
Comorbiditeit	Ja	0.28	0.40	
	Nee	0.36	0.51	
AVI				$F(1,91) = 1.92$; $p = 0.169$; partiële $\eta^2 = 0.02$
Comorbiditeit	Ja	0.43	0.66	
	Nee	0.57	0.74	
PI-dictee				$F(1,110) = 4.85$; $p = 0.030$; partiële $\eta^2 = 0.04$
Comorbiditeit	Ja	0.38	0.74	
	Nee	0.70	0.96	
ZISB-dictee				$F(1,86) = 4.25$; $p = 0.042$; partiële $\eta^2 = 0.05$
Comorbiditeit	Ja	0.36	0.69	
	Nee	0.64	0.89	

Uit de resultaten bleek een significant interactie-effect bij het PI-dictee en ZISB-dictee. De stijging in leerrendement tussen pre-test en post-test was significant groter voor leerlingen met comorbide problemen, dan voor leerlingen zonder comorbide problemen. Dit betrof kleine tot middelgrote effecten. De interacties bij de twee toetsen zijn geïllustreerd in Figuur 14 en 15 hieronder.



Figuur 14. Leerrendement op het PI-dictee op pre- en post-test, uitgesplitst voor comorbiditeit.



Figuur 15. Leerrendement op het ZISB-dictee op pre- en post-test, uitgesplitst voor comorbiditeit.

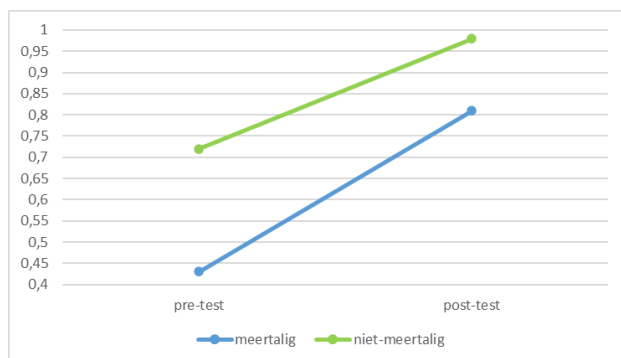
4.6 Verschil effectiviteit naar meertaligheid

Vraag 2e betrof het effect van meertaligheid: is er een verschil in de effectiviteit van de F&L-methode tussen meertalige leerlingen en niet-meertalige leerlingen? We hadden hier vooraf geen specifieke verwachting over. Voor deze onderzoeksvraag is gebruikgemaakt van een subgroep van de dataset, bestaande uit 122 leerlingen (61 met comorbiditeit, 61 controle). De interactie-effecten tussen meertaligheid en testmoment zijn weergegeven in Tabel 9.

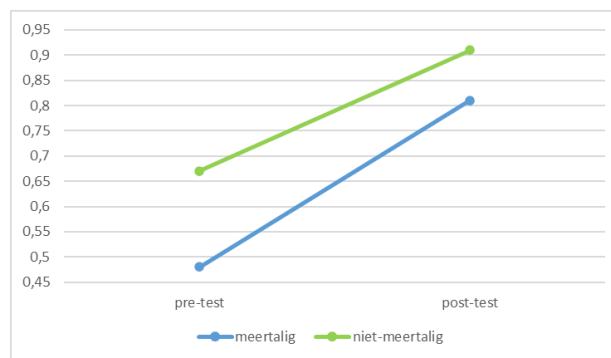
Tabel 9. Gemiddelde leerrendement (lr) op de pre- en post-test, uitgesplitst voor meertaligheid, en de interactie-effecten bij de verschillende toetsen.

		m (pre-test)	m (post-test)	Interactie testmoment * meertaligheid
DMT				$F(1,113) = 0.01$; $p = 0.932$; partiële $\eta^2 < 0.001$
Meertalig	Ja	0.32	0.46	
	Nee	0.39	0.54	
EMT				$F(1,118) = 0.04$; $p = 0.837$; partiële $\eta^2 < 0.001$
Meertalig	Ja	0.40	0.47	
	Nee	0.42	0.51	
Klepel				$F(1,117) = 0.64$; $p = 0.426$; partiële $\eta^2 = 0.01$
Meertalig	Ja	0.29	0.42	
	Nee	0.37	0.48	
AVI				$F(1,93) = 0.38$; $p = 0.540$; partiële $\eta^2 = 0.004$
Meertalig	Ja	0.44	0.65	
	Nee	0.54	0.74	
PI-dictee				$F(1,118) = 10.24$; $p = 0.002$; partiële $\eta^2 = 0.08$
Meertalig	Ja	0.43	0.81	
	Nee	0.72	0.98	
ZISB-dictee				$F(1,92) = 4.82$; $p = 0.031$; partiële $\eta^2 = 0.05$
Meertalig	Ja	0.48	0.81	
	Nee	0.67	0.91	

Uit de resultaten bleek een significant interactie-effect bij het PI-dictee en ZISB-dictee. De stijging in leerrendement tussen pre-test en post-test was significant groter voor meertalige leerlingen, dan voor niet-meertalige leerlingen. Dit betrof middelgrote effecten. De interacties bij de twee toetsen zijn geïllustreerd in Figuur 16 en 17 hieronder.



Figuur 16. Leerrendement op het PI-dictee op pre- en post-test, uitgesplitst voor meertaligheid.



Figuur 17. Leerrendement op het ZISB-dictee op pre- en post-test, uitgesplitst voor meertaligheid.

4.7 Samenhang tijdsinterval tussen doorverwijzing en start behandeling en effectiviteit van de behandeling

De derde onderzoeksvraag was of het tijdsinterval tussen de derde hoofdmeting op school en de start van de behandeling samenhangt met de effectiviteit van de behandeling op de lees- en spellingresultaten. Hierover hadden we op voorhand geen verwachtingen.

De resultaten lieten zien dat het aantal dagen tussen de derde hoofdmeting op school en de start van de behandeling nogal varieerde tussen de leerlingen. Het minimale aantal dagen dat hiertussen zat was 7 en het maximale was 657 dagen (bijna twee jaar⁶). Gemiddeld zat er zo'n 180 dagen tussen de derde hoofdmeting op school en de start van de behandeling (sd = 89).

Voor zowel de leesresultaten op de DMT en de AVI en de spellingresultaten op het PI-dictee is gekeken of er een samenhang was tussen dit tijdsinterval en de effectiviteit van de F&L-methode. Dit bleek voor alle drie de toetsen niet het geval te zijn. Het tijdsinterval tussen de derde hoofdmeting op school en de start van de behandeling hing niet samen met het verschil in leerrendement tussen de pre-test en post-test van de lees- en spellingresultaten (delta leerrendement) (DMT; $b < 0.001$, $SE < 0.001$, $\beta = -0.05$, $t(884) = -1.37$, $p = 0.170$, AVI; $b < 0.001$, $SE < 0.001$, $\beta = -0.07$, $t(308) = -1.15$, $p = 0.252$, PI-dictee; $b < 0.001$, $SE = 0.001$, $\beta = -0.06$, $t(104) = -0.59$, $p = 0.555$).

⁶ Deze grote tijdsintervallen kwam incidenteel voor, bijvoorbeeld omdat leerlingen eerst op een wachtlijst van een andere behandelaar stonden.

5 Samenvatting en conclusie

De eerste onderzoeksvraag was: In hoeverre is de F&L-methode effectief voor het bevorderen van de lees- en spellingvaardigheid op woord- en zinsniveau? De resultaten lieten voor alle lees- en spellingtoetsen een significant verschil zien in leerrendement tussen de begin- en eindmeting: het leerrendement van leerlingen bleek na afloop van de behandeling significant hoger op alle toetsen dan op de beginmeting. Dat impliceert dat de behandeling met de F&L-methode effectief is geweest, voor het verbeteren van de lees- en spellingresultaten van leerlingen met dyslexie. De effecten bleken voor zowel lezen als spellen groot en waren het grootst voor spellingvaardigheid, daarna voor de leesvaardigheid van teksten en het kleinst voor de leesvaardigheid van losse woorden. Dit betekent dat leerlingen met dyslexie, met een grote achterstand en hardnekkigheid van de lees- en/of spellingproblemen ondanks extra structurele hulp op school wel degelijk vooruitgang boeken na een specialistische behandeling met de F&L-methode in de gezondheidszorg. Deze resultaten bevestigen de eerdere onderzoeksuitkomsten bij een iets bredere doelgroep (Gijsel & Bosman, 2010). Ook in de voorgaande studie lieten leerlingen een significante verbetering zien van hun lees- en spellingprestaties, waarbij de effecten voor het lezen van teksten groter waren dan voor het lezen van losse woorden en de effecten voor het spellen, net als in de huidige studie, beduidend groter waren dan de effecten voor het lezen.

De effectgroottes zijn in beide studies ongeveer vergelijkbaar voor de DMT (vooral kaart 3), AVI en het PI-dictee. Bij de EMT en de Klepel is het effect in de huidige studie iets groter dan in de eerdere studie (Gijsel & Bosman, 2010). In de huidige studie is de onderzoeksgroep groter en ook de context is anders. Sinds 2015 zijn de gemeenten verantwoordelijk voor gespecialiseerde dyslexiezorg voor basisschoolleerlingen (zowel voor onderzoek als behandeling). Gemeenten betalen het diagnostische onderzoek en de behandeling, wanneer er sprake is van Ernstige enkelvoudige Dyslexie (EED; tot 1 januari 2022), vastgesteld op basis van het Protocol Diagnostiek en Behandeling 2.0 (2013). Leerlingen in deze studie zijn gediagnosticeerd en behandeld binnen de kaders van het Protocol Dyslexie Diagnostiek en Behandeling 2.0⁷. De lees- en/of spellingproblemen in de huidige doelgroep zouden dus wat hardnekkiger kunnen zijn dan in de vorige studie.

De tweede onderzoeksvraag was of effectiviteit anders is voor groepen leerlingen met verschillende leerlingkenmerken. Net als in de vorige studie (Gijsel & Bosman, 2010) is daarbij het effect van bouw onderzocht. In zowel de onder-, midden- als bovenbouw maakten de leerlingen een significante groei door van het begin tot het eind van de behandeling. Het bleek dat deze groei, bij alle toetsen met uitzondering van de DMT, verschilde tussen de bouwen. Zo bleek dat de stijging in scores op de EMT het grootst was in de onderbouw. Een mogelijke verklaring hiervoor is dat het leerrendement op de pre-test in de onderbouw bij de EMT een stuk lager was dan in de andere bouwen; dit laat mogelijk ruimte voor een sterkere groei. De groei in scores bij het tekst lezen (AVI), lezen van pseudowoorden (Klepel) en spellen van zinnen (ZISB-dictee) bleek het grootst in de onder- en middenbouw ten opzichte van de bovenbouw. Bij de spelling van woorden (gemeten met het PI-dictee) bleek de groei in scores in de middenbouw het grootst.

In aanvulling op de studie van Gijsel & Bosman (201) is op verzoek van de opdrachtgever ook het verschil in effectiviteit onderzocht voor jongens en meisjes en leerlingen die wel en niet gedoubleerd hebben. Uit de resultaten bleek dat de jongens significant meer vooruitgang boekten op de woordleestoetsen

⁷ Sinds 2022 wordt onderscheid gemaakt tussen dyslexie en Ernstige Dyslexie (ED) i.p.v. EED en wordt het Protocol Diagnostiek en Behandeling 3.0 (2022) toegepast.

(DMT en EMT) dan de meisjes. Op de overige toetsen werd geen verschil gevonden. Daarnaast lieten de resultaten zien dat de groei in lees- en spellingprestaties tussen de beginmeting en de eindmeting verschilde, afhankelijk van de schoolloopbaan van de leerlingen. Leerlingen die niet gedoubleerd hadden bleken een sterkere groei door te maken op lezen en spellen dan leerlingen die wel gedoubleerd hadden. Oftewel: de behandeling bleek effectiever voor leerlingen die niet gedoubleerd hadden. Het is aannemelijk dat de didactische resistentie bij leerlingen die al eens zijn blijven zitten groter is dan bij leerlingen zonder doublure(s). Bovendien lag hun lees- en spellingvaardigheid bij de start van de behandeling lager dan die van leerlingen die niet waren blijven zitten.

Ook is op verzoek van de opdrachtgever gekeken naar de invloed van comorbiditeit en meertaligheid op de effectiviteit van de behandeling. Uit de resultaten op een kleine subset van de data bleek een significant interactie-effect bij het PI-dictee en ZISB-dictee. Alle leerlingen maakten een groei door in leerrendement van het begin tot het eind van de behandeling. De stijging in leerrendement tussen pre-test en post-test was significant groter voor leerlingen met comorbide problemen, dan voor leerlingen zonder comorbide problemen. Dit betrof kleine tot middelgrote effecten. Dit is een belangrijk resultaat in het licht van de wijzigingen in de criteria van de vergoede dyslexiezorg. Sinds 2022 komen leerlingen in aanmerking voor vergoede dyslexiezorg wanneer sprake is van ernstige Dyslexie (ED) in plaats van Ernstige Enkelvoudig Dyslexie (EED). Dat betekent dat ook leerlingen waarbij sprake is van comorbide problematiek in aanmerking kunnen komen voor vergoede zorg. De huidige resultaten laten zien dat juist ook voor deze leerlingen een dyslexiebehandeling in de zorg voor vooruitgang op lezen en spellen kan zorgen en voor spellen zelfs voor een grotere vooruitgang dan leerlingen zonder comorbiditeit. Ditzelfde resultaat geldt voor het verschil tussen meertalige en eentalige kinderen. Alle leerlingen maakten een groei door in leerrendement van het begin tot het eind van de behandeling. De stijging in leerrendement tussen pre-test en post-test bij spelling (PI-dictee en ZISB-dictee) was significant groter voor meertalige leerlingen, dan voor niet-meertalige leerlingen. Dit betrof middelgrote effecten. Een mogelijke verklaring voor deze resultaten is dat de verschillen tussen de groepen (wel versus geen comorbide problemen en wel versus geen meertaligheid) groot waren bij aanvang van de behandeling: de scores van de groep met comorbiditeit en de groep met meertaligheid lagen bij de pre-test een stuk lager dan die van de groep zonder comorbiditeit en zonder meertaligheid. Er was dus meer ruimte voor groei.

De derde onderzoeksvraag was of het tijdsinterval tussen de derde hoofdmeting op school en de start van de behandeling samenhangt met de effectiviteit van de behandeling op de lees- en spellingresultaten. Uit de resultaten bleek dat het tijdsinterval tussen de derde hoofdmeting op school en de start van de behandeling niet samenhangt met het verschil in leerrendement tussen de pre-test en post-test van de lees- en spellingresultaten. Dit betekent dat de duur tussen de doorverwijzing en de start van de behandeling niet samenhangt met de effectiviteit van de F&L-methode.

Uit onze studie kan dus met enige voorzichtigheid worden geconcludeerd dat behandeling volgens de F&L-methode voor zowel lezen als spellen effectief is voor leerlingen met dyslexie. De resultaten lieten zien dat leerlingen aan het eind van de behandeling significant vooruit waren gegaan in hun lees- en spellingresultaten. Deze positieve effecten komen overeen met resultaten uit eerdere effectstudies naar de effectiviteit van de F&L methode (Gijssel, 2009; Van Geffen et al., 2008). Het lijkt dus verantwoord om leerlingen een behandeling volgens de F&L-methode aan te bieden. De voorzichtigheid waar we naar verwijzen heeft te maken met het quasi-experimentele design en retrospectieve opzet van het effectonderzoek. Doordat er geen controlegroep opgenomen kon worden, kunnen we niet met volle zekerheid zeggen dat de vooruitgang van leerlingen tussen de pre-

test en post-test volledig toe te schrijven is aan de F&L-behandelmethod. Er kunnen naast de inzet van de behandeling andere factoren zijn, die een invloed hebben gehad op de groei in leerrendement tijdens de behandeling (bijv. de begeleiding op school). Dit achten wij echter niet waarschijnlijk. Alle leerlingen in deze studie hebben juist vóóraf aan de diagnostiek al gestructureerde begeleiding gehad op school (interventie van minimaal 20 weken op ON3 volgens PDDb), welke onvoldoende resultaat had. Het ligt niet voor de hand dat als een leerling wordt doorverwezen naar een extern behandelinstituut, school een andere en meer intensieve interventie zal inzetten bij deze leerling dan voorheen.

6 Referentielijst

- Bos, K. P., van den, Lutje Spelberg, H. C., Scheepstra, A. J. M., & de Vries, J. (1994). De Klepel. Vorm A en B. Een test voor de leesvaardigheid van pseudowoorden. Nijmegen, Nederland: Berkhout.
- Brus, B.T. en Voeten, M.J.M. (1973). *Eén-Minuut-Test vorm A en B. Verantwoording en handleiding*. Nijmegen: Berkhout Testmateriaal BV.
- Cito (2009). *Cito Leerling- en onderwijsvolgsysteem Primair onderwijs. DMT en AVI. Groep 3 tot en met 8*. Arnhem: Cito.
- Cito (2017a). *Cito Volgsysteem Primair en speciaal onderwijs. AVI. Groep 3 tot en met 8*. Arnhem: Cito.
- Cito (2017b). *Cito Volgsysteem Primair en speciaal onderwijs. DMT. Groep 3 tot en met 8*. Arnhem: Cito.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Geelhoed, J. W., & Reitsma, P. (1999). *PI-dictee*. Lisse, Nederland: Swets & Zeitlinger.
- Geffen, E.C., van, Berends, M., & Franssens, J. (2008). Effectonderzoek naar de Fonologische en Leerpsychologische methode voor behandeling van dyslexie. *Tijdschrift voor Orthopedagogiek*, 47, 365-375.
- Gijssel, M. A.R. (2009). Lees- en spellingprestaties met de F&L-methode en de voorspellende factoren voor succes. *Tijdschrift voor Orthopedagogiek*, 48, 307-320.
- Gijssel, M. A. R., & Bosman, A. M. T. (2010). Het effect van de Fonologische en Leerpsychologische methode bij leerlingen met dyslexie. *Pedagogische Studiën*, 87, 118-133.
- Visser, J., Laarhoven, A., van, & Beek, A. ter. (1994). *AVI-toetspakket*. 's-Hertogenbosch, Nederland: Katholiek Pedagogisch Centrum.