Het effect van de Nederlandse versie van de *Nuffield Early Language Intervention* - NELI - op de taalvaardigheid van jonge kinderen



Dr. Sharisse van Driel
Prof.dr. Martijn Meeter
Dr. Femke van der Wilt

COLOFON

December 2025

Eindrapport van het onderzoeksproject ‘Het effect van de Nederlandse versie van de Nuffield Early Language Intervention - NELI - op de taalvaardigheid van jonge kinderen’. Dit project is gefinancierd door het Nationaal Regieorgaan Onderwijsonderzoek (NRO). Dossiernummer: 405-00-860-154

Dr. Sharisse van Driel, Vrije Universiteit Amsterdam
Prof.dr. Martijn Meeter, Vrije Universiteit Amsterdam
Dr. Femke van der Wilt, Vrije Universiteit Amsterdam

Met medewerking van:
Een groot aantal betrokken basisscholen, leerkrachten, tutors, kinderen en hun ouders, en studenten van de Vrije Universiteit Amsterdam. Om anonimiteit van deelnemers aan dit onderzoeksproject te waarborgen, zullen deze betrokkenen niet met naam worden genoemd.

INHOUDSOPGAVE

Samenvatting	4
1. Inleiding	5
2. Methode	6
3. Resultaten	14
4. Discussie	19
Referenties	22

SAMENVATTING

Het ontwikkelen van een goede taalvaardigheid is belangrijk voor de verdere ontwikkeling van kinderen. Kinderen die moeite hebben met het leren van taal lopen een groter risico op latere achterstanden. Het tutorprogramma *Nuffield Early Language Intervention* (NELI) bleek in het Verenigd Koninkrijk effectief te zijn in het verbeteren van de mondelinge taalvaardigheden van kleuters met een taalachterstand. In het huidige onderzoek is onderzocht of de Nederlandse versie van NELI ook effectief is in het verbeteren van de taalvaardigheid van kleuters met een relatief zwakke taalvaardigheid (ten opzichte van hun klasgenoten). Aan het onderzoek deden 44 basisscholen mee, verdeeld over 116 klassen, met in totaal 692 kinderen van 4 tot 7 jaar. De klassen werden willekeurig toegewezen aan een interventiegroep of een controlegroep. Uit elke klas werden vijf of zes kinderen geselecteerd die relatief zwak scoorden op de Nederlandse taal. De kinderen in de interventiegroep volgden gedurende twintig weken het NELI-programma en kregen dagelijks extra ondersteuning van een tutor; individueel of in een klein groepje. De kinderen in de controlegroep namen niet deel aan NELI en volgden het gewone onderwijsprogramma. Voor en na de interventie werden de taalvaardigheden van de kinderen gemeten met verschillende toetsen die onder andere gericht waren op woordenschat en vertelvaardigheden. De resultaten laten zien dat kinderen die deelnamen aan NELI meer vooruitgang boekten in hun taalvaardigheid dan de kinderen in de controlegroep. Het effect was klein, maar significant. Er werden geen aanwijzingen gevonden dat het effect van de interventie verschilde naar gelang het beginniveau van taalvaardigheid, de thuistaal of het geslacht van de kinderen.

INLEIDING

De taalvaardigheid van jonge kinderen speelt een belangrijke rol in hun verdere ontwikkeling (Bleses et al., 2016; Petersen & LeBeau, 2021). Meerdere studies hebben aangetoond dat de mondelinge taalvaardigheid op jonge leeftijd een sterke voorspeller is voor latere leesvaardigheid (Hjetland et al., 2020; Hulme et al., 2015). Bovendien is taalvaardigheid van belang voor de sociale ontwikkeling van kinderen (Bulotsky-Shearer, et al., 2012; Petersen & LeBeau, 2021). Zo heeft onderzoek laten zien dat kleuters met een zwakke mondelinge taalvaardigheid vaker worden afgewezen door hun klasgenoten (e.g., Van der Wilt et al., 2018a; 2018b). Het is daarom belangrijk dat de taalvaardigheid van kinderen die moeite hebben met taal vanaf jonge leeftijd wordt bevorderd (Conti-Ramsden & Durkin, 2012).

Een effectieve manier om de taalvaardigheid van kinderen te verbeteren is tutoring (Hulme et al., 2015). Tutoring is gerichte en intensieve ondersteuning, en is een aanvulling op het reguliere onderwijsaanbod. Meestal wordt tutoring onder schooltijd gegeven aan kinderen die in bepaalde domeinen achterblijven (Nickow et al., 2020). Er kan onderscheid worden gemaakt tussen individuele tutoring en tutoring in kleine groepjes. Individuele tutoring maakt het mogelijk om de ondersteuning volledig af te stemmen op de specifieke behoeften van het kind (Nickow et al., 2020). De tutor kan hierbij het niveau, tempo en de inhoud van de instructie aanpassen, terwijl er ook een persoonlijke band met het kind wordt opgebouwd. Dit maatwerk is cruciaal voor kinderen met taalachterstanden, omdat zij vaak specifieke hiaten hebben in hun woordenschat, zinsstructuur of vertelvaardigheden (Baye et al., 2019; Neitzel et al., 2021). Tutoring in kleine groepjes biedt kinderen de mogelijkheid om van elkaar te leren. In een dergelijke setting kan de tutor gerichte feedback geven, wat bijdraagt aan een efficiënter leerproces (Baye et al., 2019; Neitzel et al, 2021). Onderzoek heeft aangetoond dat zowel individuele tutoring als tutoring in kleine groepjes effectief kan zijn in het verbeteren van leerresultaten, mits de intensiteit en kwaliteit van de ondersteuning voldoende zijn (Baye et al., 2019; Neitzel et al, 2021).

Een veelbelovende vorm van tutoring die specifiek gericht is op het ondersteunen van de taalontwikkeling van jonge kinderen, is de *Nuffield Early Language Intervention* (NELI; Fricke et al., 2018). Dit programma is oorspronkelijk ontwikkeld in het Verenigd Koninkrijk en wordt daar op grote schaal toegepast. NELI is een twintig weken durend tutorprogramma, ontworpen om de woordenschat, vertelvaardigheden en het luisterbegrip van jonge kinderen te verbeteren. Het programma wordt uitgevoerd door getraind schoolpersoneel, vaak onderwijsassistenten, die dagelijks individuele tutorsessies of sessies met een kleine groep verzorgen. Meerdere grootschalige, cluster-gerandomiseerde onderzoeken in het Verenigd Koninkrijk hebben de effectiviteit van NELI aangetoond (Bowyer et al., 2008; Fricke et al., 2013; Fricke et al., 2017; West et al., 2021). Evaluaties van de twintig weken durende versie van NELI lieten zien dat de mondelinge taalvaardigheid van kinderen die deelnamen aan NELI significant sterker was

verbeterd dan die van kinderen in de controlegroep (*d* = 0.21 in Fricke et al., 2017; *d* = 0.26 in West el al., 2021).

Op basis van de veelbelovende resultaten uit eerdere studies is NELI vertaald naar het Nederlands en aangepast om beter aan te sluiten bij de Nederlandse onderwijscontext. Hoewel NELI in het Verenigd Koninkrijk positieve effecten heeft laten zien, is het belangrijk om te onderzoeken of deze effecten in Nederland kunnen worden gerepliceerd. Er zijn namelijk verschillen in taalstructuur, lespraktijken, onderwijsbeleid en sociale context die van invloed kunnen zijn op de effectiviteit van een interventie als NELI. Het primaire doel van het huidige onderzoek was daarom het evalueren van de effecten van de Nederlandse aanpassing van NELI op de mondelinge taalvaardigheid van kinderen met een relatief zwakke taalvaardigheid in de Nederlandse taal. Op basis van eerder onderzoek in het Verenigd Koninkrijk (Fricke et al., 2013, 2017; West et al., 2021) werd verwacht dat kinderen die deelnamen aan NELI sterkere verbeteringen zouden laten zien in hun mondelinge taalvaardigheid dan kinderen in de controlegroep. Daarnaast is onderzocht of de effecten van NELI variëren afhankelijk van de taalvaardigheid van kinderen voorafgaand aan de interventie, de thuistaal (of er thuis Nederlands gesproken wordt) en het geslacht van de kinderen.

METHODE

Design

Het design, de meetinstrumenten en de analysemethoden zijn in juni 2023 vooraf geregistreerd in het ISRCTN Register (https://doi.org/10.1186/ISRCTN10927572). Het onderzoek betrof een cluster-gerandomiseerde gecontroleerde trial (RCT) met een voor- en nameting en werd uitgevoerd in kleuterklassen. De randomisatie vond plaats op klasniveau met behulp van een online randomisatietool (https://www.randomizer.org/). Klassen binnen scholen werden willekeurig toegewezen aan de interventie- of controlegroep. Scholen namen deel met twee klassen (n = 30), drie klassen (n = 13) of vier klassen (n = 1).

Bij deelname met drie klassen werden de klassen willekeurig toegewezen aan conditie 1, 2 of 3. Op deze scholen verliep het onderzoek in twee fasen: Fase 1 (sept. 2023 – feb. 2024): kinderen in conditie 1 namen deel aan NELI (interventie), terwijl kinderen in conditie 2 het reguliere programma volgden (business-as-usual controle). Fase 2 (jan. 2024 – juni 2024): kinderen in conditie 2 namen deel aan NELI (interventie), terwijl conditie 3 het reguliere programma volgde (controle). Conditie 2 fungeerde dus eerst als controlegroep (fase 1) en daarna als interventiegroep (fase 2). Bij scholen met twee of vier klassen vond het onderzoek plaats in één fase; in de school met vier klassen namen twee klassen gelijktijdig deel aan NELI.

De startdatum varieerde van september 2023 tot januari 2024. In totaal waren er 58 interventieklassen en 58 controleklassen. De mondelinge taalvaardigheid van de deelnemende kinderen werd vóór en na de interventieperiode gemeten. Voor conditie 2 (in scholen met drie klassen) gold de nameting van fase 1 als voormeting van fase 2. Omdat een schooljaar te kort was voor twee volledige fasen, werd fase 1 voor conditie 2 met ongeveer een maand ingekort. Hierdoor vond de nameting in fase 1 voor deze groep een maand eerder plaats. Hiervoor is gecontroleerd in de analyse (zie ‘ Data-analyse’).

Met behulp van Optimal Design (Raudenbush et al., 2011) is een a priori poweranalyse uitgevoerd. Voor het detecteren van een gemiddeld effect (d = 0.26; West et al., 2021) was een minimale steekproef van n = 118 klassen nodig (power = .80, α = .05). Omdat conditie 2 dubbel meetelde (eerst controle, daarna interventie), kwam de uiteindelijke steekproef uit op 116 klassen, wat betekent dat de power net iets minder dan .80 was.

De Nederlandse aanbieder van NELI (Stichting Vinci Toponderwijs, https://www.vinci-toponderwijs.nl/) was verantwoordelijk voor de vertaling en aanpassing van het programma, de eerste screening van de kinderen, de implementatie van NELI en de training en ondersteuning van de tutors. Het onderzoeksteam van de Vrije Universiteit Amsterdam was verantwoordelijk voor het onderzoeksproces.

Ethiek

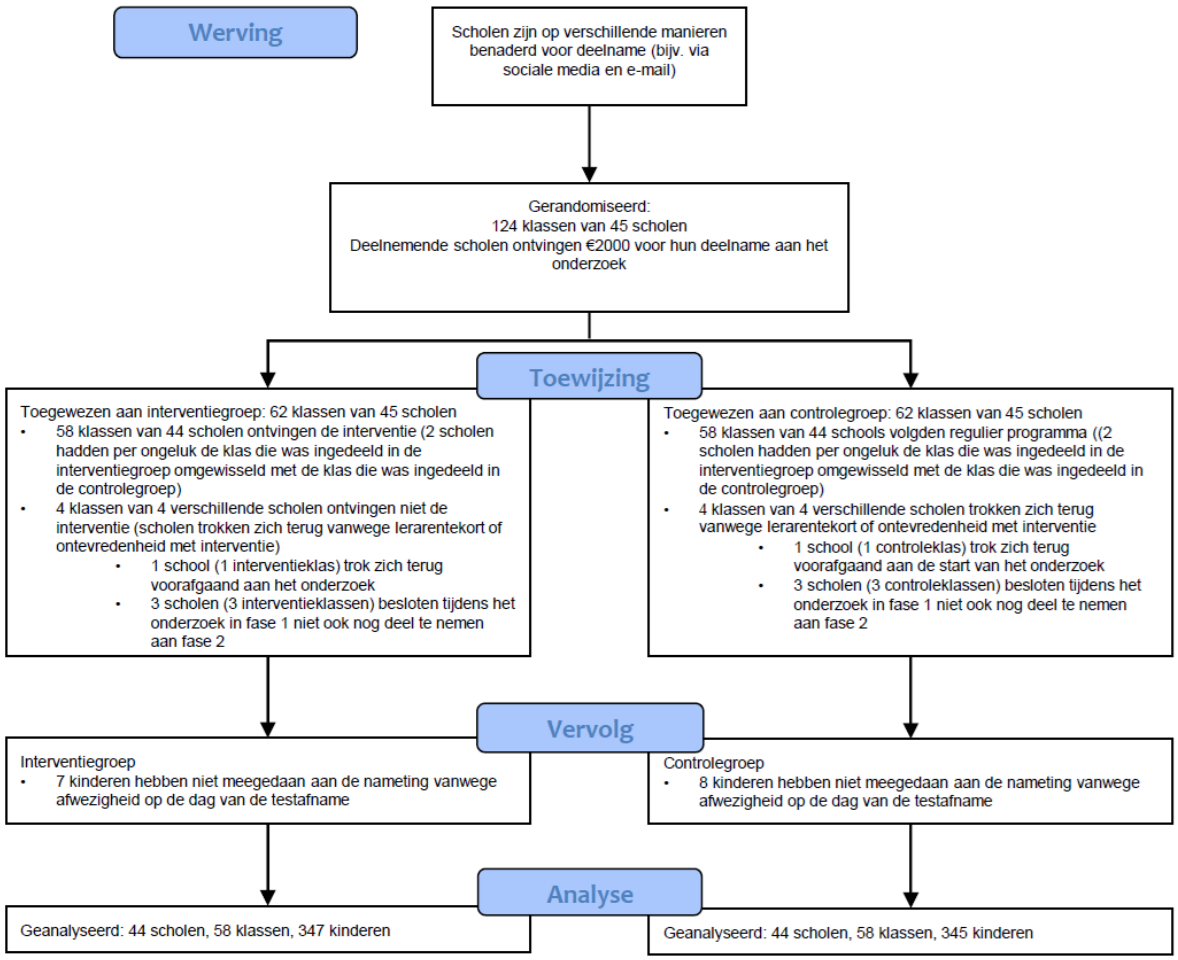
Dit onderzoek heeft voorafgaand aan de uitvoering een positief advies ontvangen van de Vaste Commissie voor Wetenschap en Ethiek van de Vrije Universiteit Amsterdam (VCWE-2022-150R1). Schooldirecteuren gaven toestemming voor deelname van hun school en leraren en ouders werden uitgebreid geïnformeerd via bijeenkomsten en brieven. Deelname van kinderen was volledig vrijwillig. Ouders werden gevraagd om toestemming te verlenen voor het onderzoek door het invullen van een toestemmingsformulier. Kinderen zonder ouderlijke toestemming zijn niet meegenomen in het onderzoek. Om de privacy te waarborgen werden alle data geanonimiseerd en veilig opgeslagen op de beschermde servers van de universiteit, alleen toegankelijk voor de primaire onderzoekers. Er is geen identificeerbare informatie over scholen, leerkrachten of leerlingen gedeeld in rapportages of publicaties.

Participanten

In totaal namen *N* = 692 kinderen deel aan dit onderzoek, afkomstig uit 116 klassen van 44 basisscholen in Nederland (zie Figuur 1). De leeftijd van de kinderen varieerde tussen de 3.44 en 7.12 jaar (*M* = 5.21, *SD* = 0.57) en er waren 317 meisjes (45.8%) en 375 jongens (54.2%). De meerderheid van de kinderen (79.3%) was in Nederland geboren, een hoger percentage dan het landelijke gemiddelde van 74.0% (CBS, 2024). Andere geboortelanden waren in Europa (6.6%), Azië (5.9%), Afrika (1.4%), Zuid-Amerika (1.3%) en Noord-Amerika (0.1%; 5.2% ontbrekende waarden). Wat betreft thuistaal sprak 28.0% uitsluitend Nederlands, 41.8% zowel Nederlands als een andere taal, en 29.9% alleen een andere taal (0.3% ontbrekende waarden). Opleidingsniveau van ouders varieerde: 26.7% laag, 33.1% gemiddeld en 15.2% hoog (25.0% ontbrekende waarden). Het aandeel laag- en gemiddeld opgeleide ouders kwam overeen met het landelijk gemiddelde, maar hoogopgeleide ouders waren ondervertegenwoordigd (CBS, 2024).

Er zaten 333 kinderen (48.1%) in groep 1 en 336 kinderen (48.6%) in groep 2 (3.3% ontbrekende waarden). Het aantal leerlingen per klas varieerde van 13 tot 37 leerlingen (*M* = 23.07, *SD* = 4.53; 3.4% ontbrekende waarden). Leraren beoordeelden de gemiddelde mondelinge taalvaardigheid van de leerlingen in hun klas op een 3-puntsschaal (1 = beneden gemiddeld, 2 = gemiddeld, 3 = boven gemiddeld). Het gemiddelde was 1.50 (*SD* = 0.60; 5.2% ontbrekende waarden). Het totaal aantal leerlingen per school varieerde van 82 tot 680 leerlingen (*M* = 262.56, *SD* = 113.10; 2.3% ontbrekende waarden).

Het NELI-programma werd uitgevoerd door getraind schoolpersoneel. In 41 klassen gaf één tutor de sessies, in 14 klassen waren dit twee tutors, en in drie klassen drie tot vijf tutors. Tutors waren 20 tot 70 jaar oud (*M* = 41.53, *SD* = 12.58) en voornamelijk vrouwen (76 vrouwen, 1 man; 6.1% ontbrekende waarden). Hun opleidingsniveau varieerde van middelhoog (60.5%) tot hoog (30.9%; 8.6% ontbrekende waarden). Functies liepen uiteen van leerkracht of onderwijsassistent (in opleiding) tot intern begeleider, pedagogisch medewerker, remedial teacher, schooldirecteur of vrijwilliger.



Figuur 1. Weergaven van onderzoeksdeelname in CONSORT diagram. Op twee scholen werden per ongeluk interventie- en controleklassen omgewisseld. Na overleg met de scholen is bevestigd dat dit geen opzet betrof. Voor deze incidentele fouten is daarom niet gecontroleerd in de analyses.

Screening en selectie

Het huidige onderzoek richtte zich op reguliere basisscholen in Nederland. Alleen scholen met ten minste twee deelnemende kleuterklassen kwamen in aanmerking, omdat randomisatie op klasniveau binnen scholen werd uitgevoerd (zie 'Design'). Hoewel er geen aanvullende selectiecriteria voor scholen werden gehanteerd, richtte de werving zich voornamelijk op scholen met een Onderwijsachterstandenbeleid (OAB-scholen). Leerlingen werden geselecteerd op basis van hun taalvaardigheid. Voorafgaand aan het onderzoek werden alle kinderen uit de deelnemende klassen gescreend met de Nederlandse versie van de LanguageScreen, een digitale taaltoets op een tablet (https://www.vinci-toponderwijs.nl/onderwijsconcepten/languagescreen). Eerder onderzoek naar de Engelse versie toonde de validiteit van dit instrument aan (West et al., 2021). De Nederlandse versie werd vertaald en aangepast door de aanbieder van NELI. Een kleinschalige pilot (niet uitgevoerd door huidig onderzoeksteam) leidde tot verdere verfijning van de tool. Schoolpersoneel voerde de screening uit bij alle kinderen in groep 1/2 van de deelnemende scholen. De LanguageScreen bevat subtests voor expressieve en receptieve woordenschat, zinsherhaling en luisterbegrip. Scores werden automatisch berekend en veilig opgeslagen. Schoolpersoneel (bv. leraar, Ib-er)

selecteerden vervolgens uit elke deelnemende klas vijf of zes kinderen voor deelname aan het onderzoek op basis van een aantal selectiecriteria. De zes kinderen uit een klas die het laagst scoorden op de LanguageScreen kwamen als eerst in aanmerking voor deelname aan het onderzoek. Kinderen met ernstige visuele, auditieve of aandachtsproblemen, evenals kinderen die nog helemaal geen Nederlands spraken, selectief mutisme vertoonden of extreem verlegen kinderen werden van deelname uitgesloten. Leraren beoordeelden bovendien of de geselecteerde kinderen geschikt waren voor deelname; soms werd een kind dat al intensieve taalondersteuning ontving niet geselecteerd. De verantwoordelijkheid voor het screeningsproces en de opslag van de bijbehorende data lagen bij de aanbieder van NELI. Het onderzoeksteam kreeg enkel toegang tot de LanguageScreen-scores van kinderen die ouderlijke toestemming hadden en daadwerkelijk zouden deelnemen aan het onderzoek.

Meetinstrumenten

Voor en na de interventieperiode werd de taalvaardigheid van de kinderen getoetst door getrainde testassistenten, die niet wisten of kinderen in de controle- of interventiegroep zaten. Hierbij werden de Nederlandse versies gebruikt van vier gestandaardiseerde tests: twee subtests van de *Clinical Evaluation of Language Fundamentals Preschool* (CELF Preschool-2NL, Wiig et al., 2012), de *Renfrew Action Picture Test* (APT; Manders et al., 2015) en het *Multilingual Assessment Instrument for Narratives* (MAIN, Gagarina et al., 2019). De tests werden individueel afgenomen in een rustige ruimte binnen de eigen school van de kinderen en testafname duurde in totaal ongeveer 15 minuten per kind. Er werd een audio-opname gemaakt van de testafnames, zodat die naderhand gescoord konden worden.

Expressieve woordenschat

De expressieve woordenschat van de kinderen werd gemeten met de subtest Expressieve Woordenschat van de CELF (Wiig et al., 2012). Het doel van deze subtest is het beoordelen van het vermogen van het kind om verschillende afbeeldingen te benoemen, zoals voorwerpen, handelingen en zelfstandige naamwoorden, en daarmee specifieke woorden op te halen en uit te spreken. Tijdens de testafname wordt een afbeelding aan het kind getoond en wordt het kind gevraagd deze te benoemen. Bij een afbeelding van een wortel moet het kind bijvoorbeeld het juiste woord, ‘wortel’, zeggen. De subtest bestaat uit 20 items. Correcte antwoorden krijgen twee punten, gedeeltelijk correcte antwoorden één punt en foutieve of geen antwoorden nul punten. De totaalscore wordt berekend door de punten van alle items op te tellen. De betrouwbaarheid van de CELF Expressieve Woordenschat is goed bevonden, met een gecorrigeerde betrouwbaarheidscoëfficiënt van 0.67 (Wiig et al., 2012). Wat betreft constructvaliditeit is aangetoond dat de correlatie met een andere taaltest (CELF-4NL) sterk is (r = 0.70; Wiig et al., 2012).

Zinnen herhalen

De subtest Zinnen Herhalen van de CELF werd gebruikt om het vermogen van kinderen om zinnen te reproduceren te meten (Wiig et al., 2012). Deze subtest beoordeelt syntactische en

morfologische ontwikkeling door te kijken hoe goed een kind zinnen kan herhalen die mondeling door de testassistent worden voorgelezen. De procedure is als volgt: de testassistent leest een zin hardop voor, bijvoorbeeld 'De jongen is blij'. Vervolgens wordt het kind gevraagd de zin precies te herhalen, zonder enige wijziging. De moeilijkheid van de zinnen neemt toe naarmate de test vordert, beginnend met korte, eenvoudige zinnen en oplopend naar langere, complexere zinnen. De subtest bestaat uit 13 items, die nul tot drie punten krijgen afhankelijk van het aantal fouten dat het kind maakt (bijvoorbeeld door het weglaten, wijzigen of herhalen van woorden). De punten per item worden opgeteld tot een totaalscore. De betrouwbaarheid van de CELF Zinnen Herhalen is goed, met een gecorrigeerde betrouwbaarheidscoëfficiënt van 0.84 (Wiig et al., 2012). De constructvaliditeit van de subtest is aangetoond door een sterke correlatie met de CELF-4NL (r = 0.65; Wiig et al., 2012).

Informatie en grammatica

De APT is ontworpen om de gesproken taal van kinderen te meten (Manders et al., 2015). De test richt zich op de inhoud van informatie en grammatica in de antwoorden van kinderen op vragen over afbeeldingen die tijdens de test worden getoond. De APT bestaat uit 10 afbeeldingen waarop een handeling wordt getoond, met elk een korte begeleidende vraag, zoals 'Wat doet dit meisje?'. De antwoorden worden gescoord op informatiewaarde en grammaticale complexiteit. Onderzoek naar de constructvaliditeit van de APT toont aan dat de informatiewaarden van de APT matig tot sterk correleren met scores van andere taaltests (r = 0.61 tussen APT Informatie en de *Word Finding Vocabulary Test* totaalscore en r = 0.55 tussen APT Information score en *Bus Story Test*; Manders et al., 2015). We hebben een scoringsprotocol ontwikkeld om de APT Grammar score te verkrijgen. De interbeoordelaarbetrouwbaarheid is vastgesteld door het berekenen van de intraclass correlation coefficients (ICCs) met behulp van de R/shiny application ICC4IRR (Psychogyiopoulos, Koopman & Ten Hove, 2025). Dit resulteerde in een hoge interbeoordelaarbetrouwbaarheid (ICC(Q, $\hat{k}$) = 0.94).

Vertelvaardigheden

De MAIN is speciaal ontwikkeld om de vertelvaardigheden van kinderen van 3 tot 10 jaar te beoordelen (Gagarina et al., 2019). De MAIN richt zich op het vermogen om samenhangende verhalen te vertellen. In dit onderzoek werd vertelvaardigheid gemeten via een naverteltaak op basis van afbeeldingen. Voor de voor- en nametingen gebruikten we de *'Cat story'*-sequentie, bestaande uit zes kleurenafbeeldingen die samen een verhaal vormen. Tijdens de testafname leest de testassistent het verhaal eenmaal hardop voor, terwijl het kind naar de afbeeldingen kijkt. Daarna wordt het kind gevraagd het verhaal na te vertellen met behulp van de afbeeldingen. De MAIN beoordeelt verschillende aspecten van verhalen, waaronder de verhaalcomponenten (bijvoorbeeld de setting en doelen van de personages) en interne toestandstermen (zoals emoties van de personages). Daarnaast werd het totale aantal woorden dat het kind gebruikte, het aantal verschillende woorden en de gemiddelde lengte van de uitingen van kinderen geregistreerd. De kwaliteit van de Duitse versie van MAIN is

geëvalueerd door Lautenschläger et al. (2021). De resultaten toonden een zeer hoge interbeoordelaarsbetrouwbaarheid (98.1%) en een bijna perfecte test-hertest betrouwbaarheid over twee weken (r = 0.98). Wat betreft constructvaliditeit liet de studie een sterke correlatie zien tussen de MAIN-scores en scores van een test voor expressieve woordenschat (r = 0.65). We hebben scoringsprotocollen ontwikkeld om de scores voor de MAIN Verhaalelementen en MAIN Interne Toestand Termen te bepalen. De interbeoordelaarbetrouwbaarheid is vastgesteld door het berekenen van de intraclass correlation coefficients (ICCs) met behulp van de R/shiny application ICC4IRR (Psychogyiopoulos, Koopman & Ten Hove, 2025). Dit resulteerde in hoge interbeoordelaarbetrouwbaarheid voor de schalen MAIN Verhaalelementen en MAIN Interne Toestand Termen (respectievelijk, ICC(Q, $\hat{k}$) = 0.91 en 0.84).

NELI interventie

De *Nuffield Early Language Intervention* (NELI) is een gestructureerd en doelgericht tutoringprogramma van twintig weken, dat zich richt op het verbeteren van de mondelinge taalvaardigheid van kleuters met een zwakke taalontwikkeling (zie https://www.teachneli.org/). In 2022 werd het programma door Stichting Vinci Toponderwijs vertaald en aangepast aan de Nederlandse onderwijscontext. Het programma is ontwikkeld in samenwerking met leerkrachten en logopedisten en is gebaseerd op onderwijskundige kaders. De doelstellingen zijn onder meer het vergroten van de woordenschat, het versterken van narratieve vaardigheden, het bevorderen van actief luisteren en het vergroten van het zelfvertrouwen in zelfstandig spreken. Kinderen die deelnemen aan NELI krijgen wekelijks drie tutorsessies in een klein groepje en twee individuele tutorsessies, waarin woordenschat, narratieve vaardigheden en taalproductie worden gestimuleerd via een herhalende aanpak. In totaal omvat het programma 57 groepssessies van elk 30 minuten en 36 individuele sessies van elk 15 minuten. Alle sessies volgen een vaste opbouw en introduceren nieuwe woorden binnen een specifiek thema. Het programma wordt uitgevoerd door schoolmedewerkers, zoals getrainde onderwijsassistenten, pedagogisch coaches of leerkrachten.

Treatment fidelity

De aanbieder van de Nederlandse NELI-versie verzorgde training en begeleiding aan tutoren op de deelnemende scholen om een effectieve uitvoering en hoge *treatment fidelity* te waarborgen. Iedere tutor die aan het onderzoek deelnam volgde eerst een online training van circa tien tot twaalf uur, ontwikkeld door ervaren logopedisten. De training richtte zich op het vergroten van kennis over mondelinge taalontwikkeling en bood een gedetailleerde uitleg van het NELI-programma. Deelnemers konden de training in eigen tempo doorlopen en kregen videovoorbeelden van daadwerkelijke sessies, voorzien van expertcommentaar. Aan het eind van elk onderdeel was de mogelijkheid tot online reflectie en discussie met collega's en mentoren binnen de online leeromgeving.. Korte toetsen aan het einde van elk onderdeel hielpen deelnemers hun voortgang te monitoren. Het studiemateriaal bleef beschikbaar tijdens en na afloop van de training. Na afronding van minimaal tachtig procent van de modules en het

behalen van een eindtoets met minimaal zeventig procent correcte antwoorden ontvingen deelnemers een praktijkcertificaat. Tutors dienden eerst de training succesvol af te ronden voordat de zij met de uitvoering van NELI konden starten. De basistraining werd aangevuld met verdere ondersteuning, waaronder twee coachingsessies op school (aan het begin en halverwege het programma), e-mailondersteuning en een ondersteunend *webinar*, dat onder andere gericht was op het delen van uitdagingen in de uitvoering.

Het Nederlandse NELI-programma omvat 57 groepssessies en 36 individuele sessies. Van de 58 interventieklassen voltooiden 33 klassen het volledige programma tijdens de onderzoeksperiode. In 25 klassen is het niet gelukt om het programma helemaal af te ronden. De reden hiervoor was vaak dat tutoren door het lerarentekort klassen moesten overnemen waardoor tutorsessies niet door konden gaan. Tutors registreerden per sessie de aanwezigheid van kinderen via registratieformulieren, die na afloop van het onderzoek door het onderzoeksteam werden verzameld. Op basis van deze gegevens ontvingen kinderen gemiddeld 42.27 (*SD* = 12.98) van de 57 groepssessies (74%) en 25.97 (*SD* = 10.84) van de 36 individuele sessies (72%; 0.9% missende waarden). Dit was vergelijkbaar met West et al. (2021) waarin kinderen gemiddeld 83% groepssessies en 65% individuele sessies hadden gevolgd. Het merendeel van de kinderen hebben dus niet het volledige NELI-programma ontvangen. Aangezien er twijfels zijn over de betrouwbaarheid van de ingevulde registratieformulieren, hebben we het aantal NELI-sessies niet als variabele meegenomen in de data-analyse.

Data-analyse

De analysemethode in dit onderzoek bouwt voort op het werk van West et al. (2021), die de effecten van het oorspronkelijke NELI-programma in het Verenigd Koninkrijk onderzochten. De data bestonden uit scores van kinderen op de taaltesten op de voor- en nameting. Voor het opschonen van de data en het berekenen van de beschrijvende statistieken werd gebruikgemaakt van SPSS, terwijl de meer complexe analyses in R werden uitgevoerd. Het percentage ontbrekende data bedroeg 1.7% op de voormeting en 2.3% op de nameting. Ontbrekende data waren voornamelijk het gevolg van afwezigheid van kinderen door ziekte of verhuizing, of doordat kinderen niet wilden of durfden te spreken. Volgens de *Little's MCAR-test* waren de data volledig willekeurig ontbrekend (voormeting: $\chi^2(2161) = 1671.43$, $p = 1.000$; nameting: $\chi^2(665) = 657.15$, $p = 0.58$).

Voor de analyses is *Structural Equation Modeling* (SEM) toegepast met het Lavaan-pakket in R. Er zijn modellen geschat op de voormeting en de nameting voor twee latente variabelen. Er is gekozen voor twee aparte SEM modellen, omdat het combineren van de twee latente variabelen in één model leidde tot slechte modelfit. Bovendien sluit de eerste latente variabele op deze manier aan bij die uit het onderzoek van West et al. (2021), wat vergelijking tussen bevindingen vergemakkelijkt. De eerste latente variabele, *Lexicale en Syntactische Vaardigheid*, werd opgebouwd uit de scores op de subtests van de CELF (Expressieve

Woordenschat en Zinnen Herhalen) en de subscores van de APT (Informatie, Grammatica). De tweede latente variabele, *Narratieve Vaardigheid*, werd opgebouwd uit drie subscores van de MAIN (Verhaalelementen, Aantal Verschillende Woorden, en Gemiddelde Lengte van Uitingen (*Mean Length of Utterance*; MLU). De subscores Interne Toestand Termen (lage betrouwbaarheid, zie Tabel 1) en Aantal Woorden (sterk gecorreleerd met Aantal Verschillende Woorden; $r = 0.90$ op de voormeting) zijn niet meegenomen in de analyses.
Voor het omgaan met ontbrekende data is gebruikgemaakt van *Full Information Maximum Likelihood* (FIML). Om clusteringseffecten op klasniveau te corrigeren, zijn robuuste (Huber-White) standaardfouten toegepast. In beide modellen werd de variabele conditie (interventie versus controle) opgenomen als dummy (hypothese 1). Daarnaast is een dummyvariabele toegevoegd voor kinderen bij wie de nameting vroegtijdig werd afgenomen (in conditie 2; zie 'Design'). Om te onderzoeken of het effect van NELI afhankelijk was van het beginniveau van de taalvaardigheid van het kind, is een interactieterm opgenomen tussen conditie en het initiële niveau van taalvaardigheid (hypothese 2). Tot slot is met behulp van multi-group modellen gekeken naar de modererende rol van thuistaal en geslacht (hypothese 2).

Bij de toetsing van de modelaannames bleek dat de observaties kleine afwijkingen vertoonden van univariate normaliteit en dat multivariate normaliteit in beide modellen geschonden werd volgens de test van Mardia. Daarom is gebruikgemaakt van de *robuuste maximum likelihood* schatter (MLR), die standaardfouten en toetsingsstatistieken corrigeert voor niet normaal verdeelde data. Er bleek geen sprake te zijn van multicollineariteit: de correlatie tussen de latente scores op de voormeting en conditie was in beide modellen laag (Model 1: $r = 0.03$; Model 2: $r = 0.05$). De lineariteitsaanname werd getoetst met *scatterplots*, die een duidelijke positieve lineaire relatie lieten zien tussen de scores op de voor- en nameting op de latente variabelen. De aanname van onafhankelijkheid van observaties werd echter gedeeltelijk geschonden doordat een kleine groep kinderen (conditie 2 in scholen met drie klassen; $n = 65$) tweemaal aan het onderzoek deelnam, eenmaal in de controlegroep tijdens fase 1 en eenmaal in de interventiegroep tijdens fase 2. Om rekening te houden met de geneste datastructuur en de schending van de onafhankelijkheidsaanname zijn de modellen geclusterd op klasniveau.

RESULTATEN

Beschrijvende Statistieken

De beschrijvende statistieken zijn weergegeven in Tabel 1, met gemiddeldes en standaarddeviaties afzonderlijk gerapporteerd voor de interventie- en controlegroep. Onafhankelijke t-toetsen zijn uitgevoerd om te bepalen of kinderen in de interventie- en controlegroep bij aanvang van het onderzoek van elkaar verschilden in hun scores op de geteste taalvaardigheden. Daarbij bleek dat de groepen significant van elkaar verschilden op de MAIN-subscore *Gemiddeld Aantal Woorden per Uiting* bij de voormeting ($t(675) = -2.11$, $p = 0.04$). Voor de overige taalvaardigheden werden bij de voormeting geen significante verschillen tussen de condities gevonden ($|t| < 1.42$, $p > 0.16$). Daarnaast is onderzocht of de interventie- en controlegroep significant verschilden op een aantal achtergrondvariabelen, namelijk de scores op de LanguageScreen, leeftijd, geslacht en thuistaal. Onafhankelijke t-toetsen lieten zien dat er geen significante verschillen waren tussen de condities wat betreft scores op de LanguageScreen ($|t| < 1.63$, $p > 0.10$) en leeftijd bij de voormeting ($t = -1.12$, $p = 0.26$). Chi-kwadraattoetsen wezen bovendien uit dat de groepen niet significant van elkaar verschilden in geslacht ($\chi^2 = 1.87$, $p = 0.17$) en thuistaal ($\chi^2 = 0.04$, $p = 0.84$). Op basis van deze resultaten kan geconcludeerd worden dat de condities bij de voormeting vergelijkbaar waren wat betreft taalvaardigheden en achtergrondkenmerken.

Met covariantieanalyses (ANCOVA's) is het effect van de interventie op elke gemeten taalvaardigheid afzonderlijk getoetst. De Cohen's *d's* wijzen op kleine positieve effecten voor alle gemeten taalvaardigheden, variërend van 0.06 tot 0.19, met uitzondering van de MAIN Interne Toestand Termen. Statistisch significante interventie-effecten werden uitsluitend gevonden voor de CELF-subtest Expressieve Woordenschat en de MAIN-subtest Aantal Verschillende Woorden.

Tabel 1. *Beschrijvende Statistieken op Voor (t1)- en Nameting (t2) per Conditie en het Effect van NELI*

		Interventie *n* = 347			Controle *n* = 345			
	Betrouwbaarheid[b]	*n*	*M*	*SD*	*n*	*M*	*SD*	Cohen's d [95% CI][c]
Leeftijd (jaren)								
t1		344	5.18	0.55	344	5.23	0.59	
LanguageScreen (t0)								
Expressieve Woordenschat								
t1	-	347	76.54	11.12	345	77.96	11.72	-
Receptieve Woordenschat								
t1	-	347	89.29	14.70	345	89.74	15.63	-
Zinnen Herhalen								
t1	-	347	86.35	9.86	345	86.17	10.04	-
Luisterbegrip								
t1	-	347	83.23	11.14	345	84.21	11.30	-
Gemeten taalvaardigheden[a]								
CELF – Expressieve Woordenschat								
t1	0.88	339	11.26	8.63	341	11.30	8.20	
t2	0.88	338	15.68	8.80	334	14.31	8.72	0.13 [0.05, 0.22]
CELF – Zinnen Herhalen								
t1	0.84	327	10.78	5.83	317	10.44	5.40	
t2	0.85	339	13.44	6.43	337	12.36	5.78	0.14 [-0.01, 0.29]
APT – Informatie								
t1	0.86	339	32.83	13.30	344	32.15	13.29	
t2	0.83	339	40.79	11.42	334	39.13	12.42	0.10 [-0.04, 0.24]
APT – Grammatica								
t1	0.76	341	16.73	10.00	344	15.93	9.46	
t2	0.73	340	22.26	9.07	334	21.07	9.74	0.06 [-0.08, 0.20]
MAIN – Verhaalelementen								
t1	0.67	335	4.52	2.71	342	4.33	2.90	
t2	0.54	340	6.59	2.45	336	6.02	2.76	0.11 [-0.04, 0.25]
MAIN – Interne Toestand Termen								
t1	0.26	335	3.63	2.98	342	3.32	2.79	
t2	0.24	340	4.91	2.97	336	4.77	3.24	-0.04 [-0.23, 0.15]
MAIN – Aantal Woorden								
t1	-	335	68.37	44.67	342	64.48	39.66	
t2	-	340	89.86	46.40	336	80.42	40.68	0.12 [-0.05, 0.29]
MAIN – Aantal Verschillende Woorden								
t1	-	335	29.96	14.35	342	29.37	15.10	
t2	-	340	38.50	12.04	336	35.02	14.19	0.19 [0.04, 0.34]
MAIN - Gemiddeld Aantal Woorden per Uiting								
t1	-	335	5.37	1.78	342	5.07	1.95	
t2	-	340	6.31	1.47	336	5.90	1.68	0.13 [-0.01, 0.26]

[a] Gebaseerd op ruwe scores, ongecorrigeerd voor te vroege nameting (zie 'Design').
[b] Cronbach's alpha, ontbreekt voor LanguageScreen omdat wij geen toegang hebben tot de itemscores
[c] Uitkomsten ANCOVA met scores op de nameting als uitkomstvariabele en conditie als voorspeller, controlerend voor scores op de voormeting en te vroege nameting (zie 'Design'), berekend met pooled voormeting *SD* en robuuste standaardfouten (Morris, 2008).

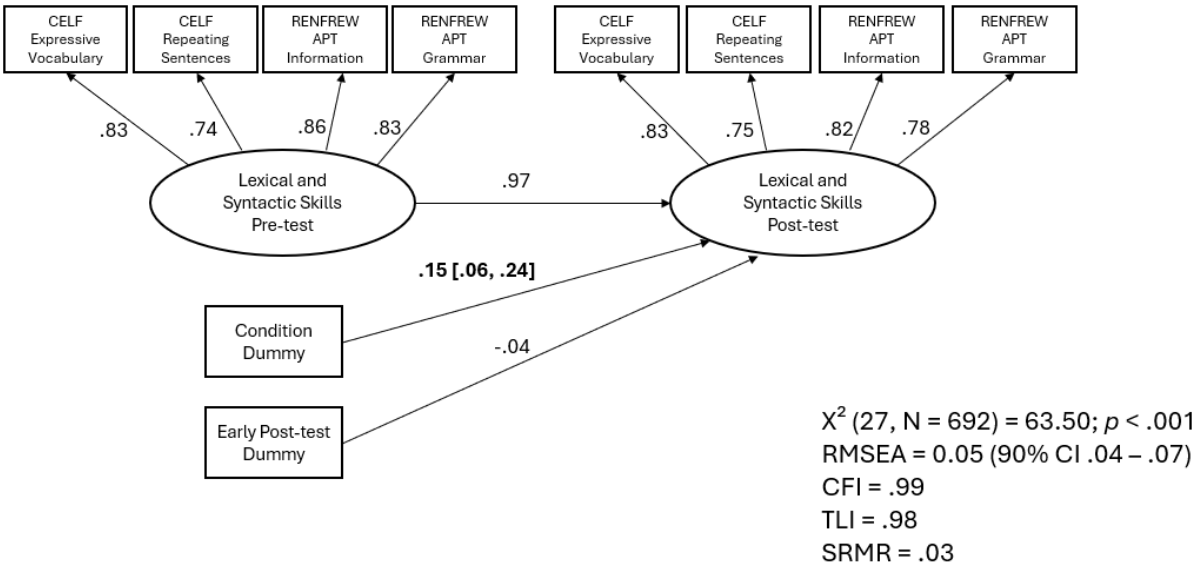
Lexicale en Syntactische Vaardigheid

Voor de latente variabele *Lexicale en Syntactische Vaardigheid* toont Figuur 2 de modelstructuur. De fitindices geven aan dat dit een goed passend model is (X^2 (27, N = 692) = 63.50; $p < 0.001$; RMSEA = 0.05 (90% CI 0.04–0.07); CFI = 0.99; TLI = 0.98; SRMR = 0.03). De

ladingen van de gemeten taalvaardigheden (> 0.74) wijzen op een sterke relatie met de latente variabele *Lexicale en Syntactische Vaardigheid*. De latente variabele op de voormeting is een sterke voorspeller van de latente variabele op de nameting, met een gestandaardiseerde effectgrootte van 0.97. Dit duidt erop dat het grootste deel van de variatie in de scores op de latente variabele op de nameting verklaard wordt door die op de voormeting, hetgeen wijst op een hoge mate van stabiliteit in deze latente variabele. De interventiegroep laat bij de nameting een significant sterkere vooruitgang zien in *Lexicale en Syntactische Vaardigheid* dan de controlegroep (d = 0.15, 95% CI [0.06, 0.24]). De effectgrootte is klein. Het toevoegen van een interactieterm tussen beginniveau van taalvaardigheid en conditie toonde geen significante verschillen tussen condities (β = -0.03; p = 0.17), hetgeen suggereert dat de interventie voor alle kinderen even effectief kan zijn, ongeacht hun beginniveau van taalvaardigheid.

Verder is onderzocht of de interventie-effecten op *Lexicale en Syntactische Vaardigheid* verschilden tussen kinderen die thuis wel (NL) of geen Nederlands spraken (GNL). Omdat het *multigroup model* geen metrische invariantie liet zien, zijn *multigroup modellen* uitgevoerd voor de afzonderlijke taalvaardigheden en niet voor de latente variabele. Deze analyses, gebaseerd op een vrij (niet-beperkt) model, toonden geen significante groepsverschillen, wat suggereert dat de interventie even effectief was voor beide groepen (CELF Expressieve Woordenschat d = 0.41, 95% CI [0.11, 0.72] voor NL en d = 0.25, 95% CI [0.05, 0.44] voor GNL (Wald test: χ^2 = 0.55, df = 1, p = 0.46); CELF Zinnen Herhalen d = 0.31, 95% CI [-0.04, 0.66] voor NL en d = 0.18, 95% CI [-0.07, 0.43] voor GNL (Wald test: χ^2 = 0.08, df = 1, p = 0.78); APT Informatie d = 0.38, 95% CI [-0.03, 0.78] voor NL en d = 0.09 , 95% CI [-0.16, 0.34] voor GNL (Wald test: χ^2 = 1.24, df = 1, p = 0.26); APT Grammatica d = 0.18, 95% CI [-0.12, 0.48] voor NL en d = 0.06, 95% CI [-0.17, 0.29] voor GNL (Wald test: χ^2 = 0.34, df = 1, p = .56). Het valt echter op dat de betrouwbaarheidsintervallen breed zijn, waardoor de power om verschillen te detecteren tussen de groepen beperkt is.

Ook is nagegaan of de interventie-effecten verschilden op basis van geslacht. Hoewel het *multigroup model* metrische invariantie aantoonde voor de latente variabele, werd er geen scalare invariantie gevonden. Daarom zijn opnieuw afzonderlijke *multigroup modellen* uitgevoerd, per geteste taalvaardigheid. Geen van deze modellen liet significante verschillen zien tussen jongens en meisjes, hetgeen consistent is met de aanname dat de interventie voor beide groepen even effectief was (CELF Expressieve Woordenschat (meisjes d = 0.19, 95% CI [-0.06, 0.44]; jongens d = 0.40, 95% CI [0.17, 0.62]; Wald test: χ^2 = 1.52, df = 1, p = 0.22); CELF Zinnen Herhalen (meisjes d = 0.13, 95% CI [-0.14, 0.40]; jongens d = 0.32, 95% CI [0.05, 0.59]; Wald test: χ^2 = 0.84, df = 1, p = 0.36); APT Informatie (meisjes d = 0.17, 95% CI [-0.13, 0.46]; jongens d = 0.17, 95% CI [-0.10, 0.44]; Wald test: χ^2 = 0.001, df = 1, p = 0.98); APT Grammatica: (meisjes d = 0.13, 95% CI [-0.13, 0.39]; jongens d = 0.07, 95% CI [-0.18, 0.32]; Wald test: χ^2 = 0.08, df = 1, p = 0.78). Het valt echter op dat een aantal betrouwbaarheidsintervallen breed zijn, waardoor de power om verschillen te detecteren tussen de groepen beperkt is.



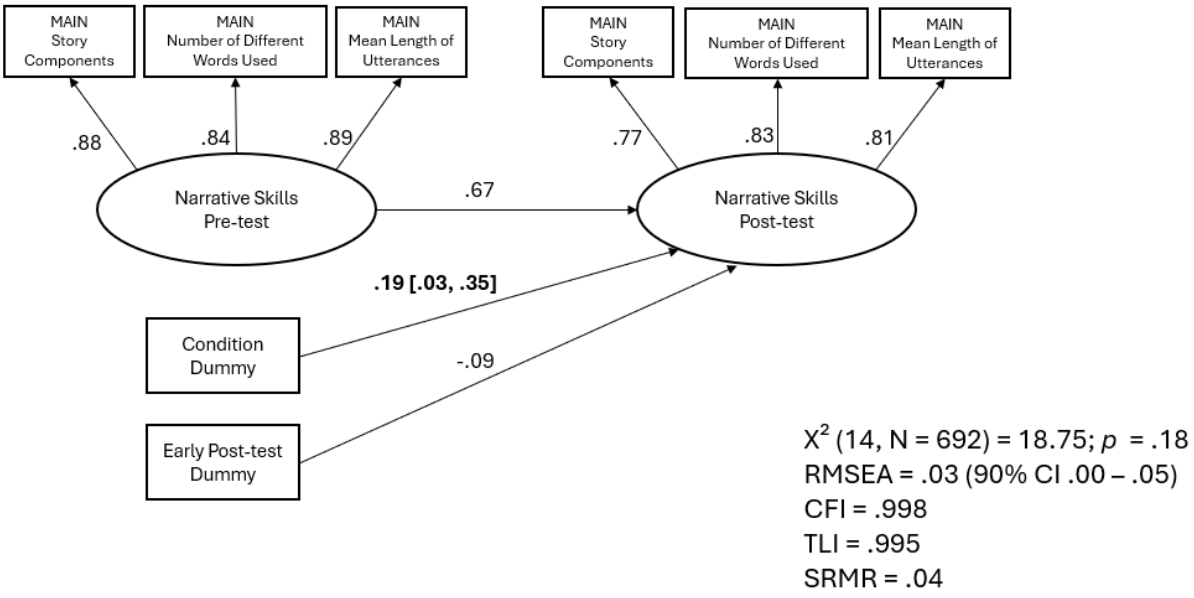
Figuur 2. Paddiagram met lading van CELF- en APT-testen op de latente variabele Lexicale en Syntactische Vaardigheid en het gestandaardiseerde interventie-effect. De dummy 'vroege nameting' is meegenomen als covariaat. De coëfficiënt voor conditie is y-gestandaardiseerd (vergelijkbaar met Cohen's d), gevolgd door een 95% robuust Huber-White betrouwbaarheidsinterval. Covarianties die zijn meegenomen maar niet zijn opgenomen in de figuur: voor- en nameting CELF Expressieve Woordenschat (0.70), voor- en nameting CELF Zinnen Herhalen (0.54), voor- en nameting APT Informatie (0.16), voor- en nameting APT Grammatica (0.11), voormeting APT Informatie en APT Grammatica (0.48), nameting APT Informatie en APT Grammatica (0.49).

Narratieve Vaardigheid

Voor de latente variabele *Narratieve Vaardigheid* laat Figuur 3 de modelstructuur zien. De fitindices wijzen op een goede modelpassing (X^2 (14, N = 692) = 18.75; p = 0.18; RMSEA = 0.03 (90% CI 0.00–0.05); CFI = 0.998; TLI = 0.995; SRMR = 0.04). De ladingen van de gemeten taalvaardigheden (> 0.77) tonen een sterke relatie met de latente variabele. Scores op de latente variabele op de voormeting voorspellen de scores op de latente variabele op de nameting gemiddeld, met een gestandaardiseerde effectgrootte van 0.67, hetgeen duidt op een redelijke stabiliteit van het construct over de tijd. De interventiegroep laat bij de nameting een significant sterkere vooruitgang zien in *Narratieve Vaardigheid* dan de controlegroep (d = 0.15, 95% CI [0.06, 0.24]). Ook hier is de effectgrootte klein. Het opnemen van een interactieterm tussen het beginniveau van taalvaardigheid en conditie liet geen significante verschillen zien (β = -0.04; p = 0.21), wat consistent is met de aanname dat de interventie even effectief was, ongeacht de initiële taalvaardigheid van de kinderen.

Bij de analyse naar verschillen in effectiviteit tussen kinderen die thuis wel of geen Nederlands spraken, liet het *multigroup model* voor de latente variabele *Narratieve Vaardigheid* metrische en scalare invariantie zien. In het latente *multigroup model* werden de regressie coëfficiënten vrij geschat voor de twee groepen. Het model liet geen significante verschillen zien in de effectiviteit van de interventie tussen de twee groepen, wat suggereert dat de interventie voor beide groepen vergelijkbaar effectief kan zijn (NL d = 0.05, 95% CI [-0.41, 0.81]; GNL d = 0.14, 95% CI [0.14, 0.89]; Wald test: χ^2 = 0.75, df = 1, p = 0.39). Opvallend is dat enkele betrouwbaarheidsintervallen breed zijn, wat wijst op beperkte power om verschillen tussen de groepen te detecteren.

Ten slotte is nagegaan of de interventie-effecten op *Narratieve Vaardigheid* verschilden op basis van geslacht. Het multigroup model liet geen metrische invariantie zien. Daarom zijn ook hier afzonderlijke *multigroup modellen* uitgevoerd. Geen van deze modellen liet significante verschillen tussen jongens en meisjes zien, wat suggereert dat de interventie voor beide groepen even effectief kan zijn (MAIN Verhaalelementen (meisjes d = 0.04, 95% CI [-0.20, 0.28]; jongens d = 0.24, 95% CI [-0.02, 0.50]; Wald test: χ^2 = 1.31, df = 1, p = 0.25); MAIN Aantal Verschillende Woorden (meisjes d = 0.26, 95% CI [-0.01, 0.53]; jongens d = 0.26, 95% CI [0.01, 0.51]; Wald test: χ^2 = 0.01, df = 1, p = 0.94); MAIN Gemiddeld Aantal Woorden per Uiting (meisjes d = 0.21, 95% CI [-0.06, 0.48]; jongens d = 0.19, 95% CI [-0.07, 0.44]; Wald test: χ^2 = 0.03, df = 1, p = 0.86). Ook hier valt op dat een aantal betrouwbaarheidsintervallen breed zijn, waardoor de power om verschillen te detecteren tussen de groepen beperkt is.



Figuur 3. Paddiagram met lading van CELF- en APT-testen op de latente variabele Lexicale en Syntactische Vaardigheid en het gestandaardiseerde interventie-effect. De dummy 'vroege nameting' is meegenomen als covariaat. De coëfficiënt voor conditie is y-gestandaardiseerd (vergelijkbaar met Cohen's d), gevolgd door een 95% robuust Huber-White betrouwbaarheidsinterval. Covarianties die zijn meegenomen maar niet zijn opgenomen in de figuur: voor- en nameting MAIN Verhaalelementen (0.24), voor- en nameting MAIN Aantal Verschillende Woorden (0.20), voor- en nameting MAIN Mean Length of Utterances (0.14), voormeting MAIN Verhaalelementen en voormeting MAIN Mean Length of Utterances (-0.59).

DISCUSSIE

Het primaire doel van dit onderzoek was het evalueren van de effecten van de Nederlandse versie van de twintig weken durende *Nuffield Early Language Intervention* (NELI) op de taalvaardigheid van kinderen met relatief zwakke taalvaardigheden in de instructietaal (Nederlands). Naast het vaststellen van de algemene effectiviteit van het programma is ook onderzocht of de effecten verschilden afhankelijk van het beginniveau van taalvaardigheid, de thuistaal (of er thuis Nederlands werd gesproken) en het geslacht van de kinderen.

Uitkomsten van de analyses lieten zien dat kinderen in de interventiegroep, vergeleken met de controlegroep, significant meer vooruitgang boekten op de latente variabelen *Lexicale en Syntactische Vaardigheid* (*d* = .15) en *Narratieve Vaardigheid* (*d* = .19). Wanneer afzonderlijk werd gekeken naar de gemeten taalvaardigheden leverde de interventie alleen significante effecten op bij de CELF Expressieve Woordenschat en Aantal Verschillende Woorden van de MAIN. Belangrijk is dat de effecten niet werden gemodereerd door het beginniveau van taalvaardigheid, de thuistaal of het geslacht van de kinderen. Er was dus geen evidentie dat groepen kinderen meer of minder profiteerden van de interventie afhankelijk van hun beginniveau van taalvaardigheid, of zij thuis Nederlands spraken of niet, of hun geslacht.

De positieve effecten op de latente taalvariabelen sluiten grotendeels aan bij eerdere bevindingen uit het Verenigd Koninkrijk, al zijn ze kleiner van omvang (Fricke et al., 2013; 2017; West et al., 2021). Voor de latente variabele *Lexicale en Syntactische Vaardigheid*, die grotendeels gebaseerd is op dezelfde taaltesten als in de Britse studies (CELF Expressieve Woordenschat, CELF Zinnen Herhalen, APT Informatie en APT Grammatica), werden eerder kleine tot gemiddelde effecten gevonden (*d* = .21 bij Fricke et al., 2017; *d* = .26 bij West et al., 2021). Voor de latente variabele *Narratieve Vaardigheid* is slechts één vergelijkbare Britse studie beschikbaar (Fricke et al., 2013), waarin een NELI-programma van dertig weken werd onderzocht en een groter effect werd gerapporteerd (*d* = .39) dan in het huidige onderzoek. Omdat in het huidige onderzoek een kortere interventieperiode van twintig weken werd gehanteerd, zijn de resultaten niet direct vergelijkbaar. Dat de effecten in het huidige onderzoek niet werden gemodereerd door beginscores, thuistaal of geslacht komt overeen met de bevindingen van West et al. (2021).

De positieve effecten van NELI op de taalvaardigheid van jonge kinderen bevestigen dat doelgerichte en intensieve ondersteuning, zoals NELI biedt, krachtig kan zijn in het ondersteunen van kinderen met relatief zwakke mondelinge taalvaardigheden (Nickow et al., 2020). Deelname aan de NELI-sessies lijkt bovendien de natuurlijke snelle taalontwikkeling die kenmerkend is voor de vroege kindertijd (Nærland, 2011) te versterken. Toch bleken de gevonden effectgroottes klein. In de literatuur wordt gesteld dat effectgroottes boven de 0.25 standaarddeviatie-eenheden pas onderwijskundig relevant zijn (West et al., 2021). Hoewel een

kleinere effectgrootte een bekend fenomeen is bij replicatiestudies (Pietschnig et al., 2019), zijn er ook andere mogelijke verklaringen voor de relatief beperkte effecten in dit onderzoek.

Ten eerste werd NELI in Nederland voor het eerst op grote schaal geïmplementeerd. Schoolbesturen en tutors hadden geen eerdere ervaring met het programma. De initiële implementatie van een interventie is doorgaans uitdagend, omdat uitvoerders nieuwe vaardigheden moeten ontwikkelen, vertrouwd moeten raken met materialen en tegelijkertijd geconfronteerd worden met organisatorische barrières, zoals beperkte tijd, ruimte of ondersteuning (Century & Cassata, 2016; Fixsen et al., 2005). Om die reden leveren eerste implementaties vaak minder sterke effecten op dan latere edities, waarin de uitvoeringskwaliteit doorgaans toeneemt (Fixsen et al., 2005).

Ten tweede is NELI oorspronkelijk ontwikkeld voor het Britse kleuter- en basisonderwijs. Dit kan verschillen van de Nederlandse context, onder meer in pedagogische aanpak, interactie tussen tutor en kind, en de leefwereld en interesses van kinderen. Het is denkbaar dat de Nederlandse versie van NELI verdere aanpassingen behoeft om beter aan te sluiten bij het Nederlandse onderwijs.

Ten derde stond de uitvoering van het programma onder druk door het huidige lerarentekort in het Nederlandse basisonderwijs. Dit leidde ertoe dat de planning en consistentie van de sessies niet altijd optimaal waren. Hoewel het gemiddelde percentage afgeronde sessies vergelijkbaar was met dat in het Britse onderzoek van West et al. (2021), kreeg een groot deel van de kinderen niet het volledige programma aangeboden, wat waarschijnlijk de effectiviteit ervan heeft verminderd.

Naast deze contextuele factoren kent het huidige onderzoek ook enkele methodologische beperkingen. Een kleine groep kinderen nam deel aan twee onderzoeksfasen: eerst in de controlegroep en vervolgens in de interventiegroep. Omdat een schooljaar te kort was voor twee volledige fasen, werd de nameting van fase 1 voor deze kinderen een maand eerder afgenomen dan gepland. Hoewel in de analyses is gecontroleerd voor het mogelijke effect hiervan, kan dit de resultaten hebben beïnvloed. Daarnaast was de power net iets minder dan .80: er namen 116 klassen deel aan huidig onderzoek, terwijl een poweranalyse had uitgewezen dat 118 nodig waren voor een power van .80.

Verder is, net als in het onderzoek van West et al. (2021), de selectie van kinderen gebaseerd op relatieve prestaties binnen de klas: de zes kinderen met de laagste scores op een screeningsinstrument kwamen als eerste in aanmerking voor deelname (als ook aan de andere criteria was voldaan). Omdat geen vaste afkapwaarde werd gehanteerd, kan de onderzoeksgroep heterogeen zijn geweest, met mogelijk ook kinderen die relatief hoge scores hadden op de screening.

Ten slotte was inzicht in de implementatiebeperkt tot het aantal uitgevoerde sessies, waarvan de betrouwbaarheid bovendien onvoldoende gewaarborgd bleek. Deze informatie kon daarom niet worden meegenomen in de effectanalyses, waardoor het mogelijke effect van de interventiedosering niet nader kon worden onderzocht. Aanvullende gegevens over interventietrouw en uitvoeringskwaliteit, zoals kwalitatieve informatie over de inhoudelijke uitvoering van de NELI-sessies, waren ten tijde van deze rapportage nog niet beschikbaar. Hierdoor blijft voor nu onduidelijk in hoeverre NELI is uitgevoerd zoals bedoeld en hoe variatie in uitvoering samenhangt met de gevonden effecten.

Voor toekomstig onderzoek is het van belang om de effecten van NELI in Nederland te onderzoeken bij scholen die het programma over een langere periode implementeren en vervolgmetingen toe te voegen om te kunnen nagaan of de interventie-effecten op de langere termijn behouden blijven (Hulme et al., 2025). Daarnaast zou verdere verfijning van het programma, afgestemd op de Nederlandse context, kunnen bijdragen aan sterkere effecten. Tot slot is diepgaander onderzoek naar interventietrouw en uitvoeringskwaliteit noodzakelijk. Analyse van video-opnames van sessies, aangevuld met andere vormen van procesevaluatie, wordt op dit moment uitgevoerd. De resultaten hiervan zullen naar verwachting inzichtbieden in onderliggende mechanismen en verklarende factoren, en daarmee richting geven aan optimalisatie van NELI in de Nederlandse praktijk.

Samenvattend laat dit onderzoek zien dat de Nederlandse aanpassing van NELI effectief is in het ondersteunen van jonge kinderen met relatief zwakke mondelinge taalvaardigheden in de taal van instructie (Nederlands). De gevonden effecten zijn klein en de gemiddelde taalvooruitgang van kinderen in de interventiegroep was beperkt in vergelijking met die in de controlegroep. Dit betekent dat NELI, ondanks de forse investering in onderwijstijd, expertise en personele inzet, een bescheiden bijdrage levert aan de taalontwikkeling van kinderen. De resultaten laten zien dat het programma in de huidige vorm meetbare effecten heeft en daarmee waardevol kan zijn voor kinderen met een relatief zwakke taalvaardigheid, maar ook dat optimalisatie en verdere afstemming op de praktijk nodig lijken zijn om grotere effecten te realiseren.

REFERENTIES

1. Baye, A., Inns, A., Lake, C., & Slavin, R.E. (2019). A Synthesis of Quantitative Research on Reading Programs for Secondary Students. *Reading Research Quarterly, 54*(2), 133-166. https://doi.org/10.1002/rrq.229
2. Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics, 37*(6), 1461-1476. https://doi.org/10.1017/S0142716416000060
3. Bowyer-Crane, C., Snowling, M.J., Duff, F.J., Fieldsend, E., Carroll, J.M., Miles, J., . . . & Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry, 49,* 422–432. https://doi.org/10.1111/j.14697610.2007.01849.
4. Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what, how, why, where, and who. *Review of Research in Education, 40*(1), 169–215. https://doi.org/10.3102/0091732X16665332
5. Clegg, J., Law, J., Rush, R., Peters, T. J., & Roulstone, S. (2015). The contribution of early language development to children's emotional and behavioural functioning at 6 years: An analysis of data from the Children in Focus sample from the ALSPAC birth cohort. Journal of Child Psychology and Psychiatry, 56(1), 67-75. https://doi.org/10.1111/jcpp.12281
6. Conti-Ramsden, G., & Durkin, K. (2012). Language development and assessment in the preschool period. *Neuropsychology Review, 22*(4), 384-401.https://doi.org/10.1007/s11065-012-9208-z
7. Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
8. Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54*(3), 280–290. https://doi.org/10.1111/jcpp.12010
9. Fricke, S., Bowyer-Crane, C., Snowling, M.J., & Hulme, C. (2018). *The Nuffield early language intervention.* Oxford: Oxford University Press.
10. Fricke, S., Burgoyne, K., Bowyer-Crane, C., Kyriacou, M., Zosimidou, A., Maxwell, L., Lervåg, A., Snowling, M. J., & Hulme, C. (2017). The efficacy of early language intervention in mainstream school settings: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 58*(10), 1141–1151. https://doi.org/10.1111/jcpp.12737
11. Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U. & Walters, J. (2019). *MAIN: Multilingual Assessment Instrument for Narratives – Revised*. ZAS Papers in Linguistics, 63.
12. Hjetland, H.N., Brinchmann, E.I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review, 30,* 100323.

https://doi.org/10.1016/j.edurev.2020.100323
13. Hulme, C., Snowling, M. J., West, G., Lervåg, A., & Melby-Lervåg, M. (2020). Children's Language Skills Can Be Improved: Lessons From Psychological Science for Educational Policy. *Current Directions in Psychological Science, 29*(4), 372–377. https://doi.org/10.1177/0963721420923684
14. Hulme, C., West, G., Rios Diaz, M., Hearne, S., Korell, C., Duta, M., & Snowling, M. J. (2025). The Nuffield Early Language Intervention (NELI) programme is associated with lasting improvements in children's language and reading skills. *Journal Of Child Psychology and Psychiatry, and Allied Disciplines, 66*(9), 1357–1365. https://doi.org/10.1111/jcpp.14157
15. Law, J., Charlton, J., and Asmussen, K. (2017). *Language as a child wellbeing indicator*. London: Early Intervention Foundation.
16. Neitzel, A., Lake, C., Pellegrini, M. & Slavin, R.E. (2021). A Synthesis of Quantitative Research on Programs for Struggling Readers in Elementary Schools. *Reading Research Quarterly. 57*(1), 149-179. https://doi.org/10.1002/rrq.379
17. Nickow, A.J., Oreopoulos, P. & Quan, V. (2020). *The Impressive Effects of Tutoring on PreK-12 Learing: A Systematic Review and Meta-Analysis of the Experimental Evidence*. NBER Working Paper, No. 27476. https://doi.org/10.3386/w27476
18. Petersen, I. T., & LeBeau, B. (2021). Language ability in the development of externalizing behavior problems in childhood. *Journal of Educational Psychology, 113*(1), 68–85. https://doi.org/10.1037/edu0000461
19. Pietschnig, J., Siegel, M., Eder, J. S. N., & Gittler, G. (2019). Effect declines are systematic, strong, and ubiquitous: A meta-meta-analysis of the decline effect in intelligence research. *Frontiers in Psychology, 10,* 2874. https://doi.org/10.3389/fpsyg.2019.02874
20. Psychogyiopoulos, A., Koopman L. & Ten Hove, D. (2025). ICC4IRR: A shiny application to estimate interrater reliability using intraclass correlation coefficient. https://tasospsy.shinyapps.io/icc4irr_app/
21. Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X., & Martinez, A. (2011). *Optimal Design for longitudinal and multilevel research: Documentation for the "Optimal Design" software (Version 3.01).* University of Michigan. Retrieved from https://hlmsoft.net/od/
22. Van der Wilt, F., Van der Veen, C., Van Kruistum, C., & Van Oers, B. (2018a). Why can't I join? Peer rejection in early childhood education and the role of oral communicative competence. *Contemporary Educational Psychology, 54,* 247-254. https://doi.org/10.1016/j.cedpsych.2018.06.007
23. Van der Wilt, F., Van der Veen, C., Van Kruistum, C., & Van Oers, B. (2018b). Popular, rejected, neglected, controversial or average: Do young children of different sociometric groups differ in their level of oral communicative competence? *Social Development, 27*(4), 793-807. https://doi.org/10.1111/sode.12316
24. West, G., Snowling, M. J., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., McLachlan, H., & Hulme, C. (2021). Early language screening and intervention can be delivered successfully at scale: evidence from a cluster randomized controlled trial. *Journal of Child*

Psychology and Psychiatry, 62(12), 1425–1434. https://doi.org/10.1111/jcpp.13415
25. Wiefferink, K., van Beugen, C., Wegener Sleeswijk, B., & Gerrits, E. (2020). Children with language delay referred to Dutch speech and hearing centres: caseload characteristics. *International Journal of Language & Communication Disorders, 55*(4), 573-582. https://doi.org/10.1111/1460-6984.12540
26. Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*(1), 165–200.