



Universiteit
Utrecht

EVENING

*Een evaluatie van 960 uur voorschoolse educatie en de
urennorm voor de pedagogisch beleidsmedewerker*

September 2025

Thomas van Huizen, Paul Leseman, Ora Oudgenoeg-Paz, Ryanne Francot,
Ekin Yurdakul & Vaso Tiliopoulou



EVENING: Een evaluatie van 960 uur voorschoolse educatie en de urennorm voor de pedagogisch beleidsmedewerker

September 2025

Thomas van Huizen, Paul Leseman, Ora Oudgenoeg-Paz, Ryanne Francot, Ekin Yurdakul
& Vaso Tiliopoulou

Universiteit Utrecht



Deze rapportage is de eindrapportage van het NRO project EVENING (Effectstudie Voorschoolse Educatie: een natuurlijk experiment in Nederlandse gemeenten; dossiernummer 40.5.19938.004). EVENING maakt deel uit van het bredere onderzoeksprogramma Onderwijsachterstandenbeleid, opgesteld door het ministerie van OCW. Dit onderzoek is (mede) gefinancierd door het Nationaal Regieorgaan Onderwijsonderzoek.

Inhoud

Voorwoord	3
Samenvatting	4
Onderzoeksopzet	5
Effecten van 960 uur voorschoolse educatie	6
Effecten van de van pedagogisch beleidsmedewerker	7
Conclusies en beleidsaanbevelingen	8
1. Inleiding	10
1.1. Het EVENING-project	10
1.2. Overzicht van dit rapport	12
2. Steekproef en procedures	13
2.1. Procedure en response	13
2.2. Gemeenten	15
2.3. Aanbieders	17
2.4. Kinderen en gezinnen	17
3. Kindcentra	19
3.1. Kenmerken	19
3.2. Proceskwaliteit	21
3.3. Activiteitsaanbod	27
4. Gezinnen	30
4.1. Steekproef	30
4.2. Achtergrondkenmerken	31
4.3. Taalomgeving	33
5. Kinderen	34
5.1. Kenmerken van de kinderen	34
5.2. Metingen van de ontwikkeling	35
5.3. Meetinvariantie en berekenen van geschaalde scores	40
5.4. Kinduitkomsten	42
6. Effecten 960 uur voorschoolse educatie	52
6.1. Inleiding	52
6.2. Implementatie van 960 uur voorschoolse educatie	53
6.3. Identificatiestrategie	56
6.4. Resultaten	61
6.5. Conclusie en discussie	65
7. Effecten van de pedagogisch beleidsmedewerker	67
7.1. Inleiding	67

7.2.	Implementatie van de pedagogisch beleidsmedewerker	68
7.3.	Identificatiestrategie	74
7.4.	Resultaten	81
7.5.	Conclusie en discussie	88
8.	Conclusies	89
8.1.	Effecten van meer uren voorschoolse educatie	89
8.2.	Effecten van de pedagogisch beleidsmedewerker	89
8.3.	Aanbevelingen en vervolgonderzoek	89
	Literatuur	91
	Appendix A. Kernconcepten en kerndomeinen	94
	Appendix B. Non-respons	97
	Appendix C. Databeschrijving per meetronde	98
	Appendix D. Activiteitsaanbod	104
	Appendix E. Meetinvariantie en schalen van scores	112
	Appendix F. Domeinscores	122
	Appendix G. Additionele resultaten hoofdstuk 6	124
	Appendix H. Additionele resultaten hoofdstuk 7	129

Voorwoord

In dit vierde rapport van het EVENING project, het eindrapport, presenteren we de belangrijkste uitkomsten van het onderzoek dat is uitgevoerd naar de implementatie en effecten van 960 uur voorschoolse educatie en de nieuwe urennorm voor de inzet van de pedagogisch beleidsmedewerker. Het rapport geeft inzicht in de uitvoering en kwaliteit van voorschoolse educatie in Nederland en hoe het aanbod met extra uren en de grotere inzet van pedagogisch beleidsmedewerkers in praktijk vorm heeft gekregen en wat deze maatregelen betekenen voor de ontwikkeling van kinderen. Hoewel met dit rapport een belangrijk deel van het EVENING project wordt afgesloten, zijn er vervolgrapportages gepland op twee specifieke thema's: 1) de effecten van eerder starten met voorschoolse educatie (op tweejarige leeftijd); en 2) effecten van extra uren voorschoolse educatie op begrijpend lezen en rekenen in groep 3.

Zonder de welwillende medewerking van de kinderopvangorganisaties was de uitvoering van het EVENING-project onmogelijk geweest. Daarom willen wij allereerst onze grote dank uitspreken aan alle pedagogisch professionals en leidinggevendenden die onze observatoren gastvrij hebben ontvangen en vragenlijsten hebben ingevuld. Verder willen we graag alle deelnemende gezinnen hartelijk bedanken voor hun tijd en inzet om onze onderzoeksassistenten thuis te ontvangen en hun welwillendheid om mee te werken aan de interviews. Ook de beleidsmedewerkers van de deelnemende gemeenten willen we bedanken voor hun steun aan het onderzoek en hun bereidwilligheid om informatie over het gemeentelijk beleid omtrent voorschoolse educatie te delen en toe te lichten.

Daarnaast zijn wij dankbaar voor de inzet en het enthousiasme waarmee het observatoren- en assistententeam door het hele land locatiebezoeken en huisbezoeken heeft afgelegd. In alfabetische volgorde: Ariana Ahmady, Annemiek Balt, Anne Roos Bergwerff, Paulien Brink, Sania Chaudhry, Renske Dekker, Eline van Diest, Alyza Dikilitas, Kyra El Atiki, Isabelle Evers, Sophie Fransen, Zeynep Gökdoğan, Justine Griffioen, Anouk Hermens, Loes Hogenkamp, Jessie Jochems, Angela Jong, Nikki Koot, Marjolein Looijen, Melanie Megens, Pleun Meister, Marit van de Mheen, Safae Moukhouh, Hanna Nelissen, Claudia Otter, Daniëlle vd Pavert, Limke Schopman, Anna Schouws, Ceyda Sekerci, Yasemin Sinir, Marlies Smit, Julia Swiebel, Marieke Verhoef, Rivka van der Vlugt, Lotte Werner, Laura Wielemaker, Lara van der Wielen, Margaux Zanati en Safaa Zkik.

De uitvoering van het onderzoek werd belemmerd door diverse uitdagingen, waaronder de COVID-pandemie en gerelateerde maatregelen. We waren daarom genooddaakt de onderzoeksopzet en de planning te herzien. We bedanken de NRO Programmacommissie Onderwijsachterstandenbeleid, de betrokken beleidsmedewerkers bij het NRO en het Ministerie van OCW voor de prettige samenwerking en de constructieve houding ten aanzien van de voorgestelde aanpassingen van het onderzoek.

Samenvatting

Sociale ongelijkheid in de ontwikkeling van kinderen manifesteert zich al op jonge leeftijd, ruim voor de deelname aan het primair onderwijs. Door peuters met een risico op een onderwijsachterstand beter op de basisschool voor te bereiden, kunnen voorschoolse voorzieningen een belangrijke rol spelen bij het creëren van meer kansengelijkheid. De internationale literatuur laat zien dat voorschoolse opvang en educatie een positief effect kan hebben op de ontwikkeling van kinderen uit kansarme gezinnen. In lijn met de internationale literatuur heeft Nederlands onderzoek ook aangetoond dat extra investeringen in voorschoolse educatie hebben geleid tot minder kleuterbouwverlenging. Dit duidt op een positief effect van voorschoolse educatie op de schoolrijpheid. De nieuwste resultaten uit het EVENING project tonen aan dat doelgroeppeuters baat hebben gehad van een ruimer uuraanbod en een hogere kwaliteit voorschoolse educatie.

Het EVENING onderzoek evalueert twee landelijke beleidsmaatregelen in de voorschoolse educatie:

- 1) Sinds 1 augustus 2020 zijn gemeenten verplicht om kinderen met een VVE-indicatie 960 uur voorschoolse educatie te bieden tussen de leeftijd van 2.5 en 4 jaar. In de praktijk komt dit vaak neer op 16 uur per week tijdens de schoolweken. Dit is een toename van 60 procent: voor augustus 2020 was een aanbod van 600 uur (10 uur per week) verplicht.
- 2) Sinds 1 januari 2022 geldt een landelijke norm voor de inzet van pedagogisch beleidsmedewerkers in de voorschoolse educatie. De norm betreft een minimum aantal uren per jaar, dat bepaald wordt door het aantal doelgroeppeuters per kindcentrum te vermenigvuldigen met 10.

Voor de uitvoering van de twee beleidsmaatregelen is structureel € 170 miljoen beschikbaar vanuit het Rijk. Met de invoering van de beleidsmaatregelen is vanaf 2020 het budget voor gemeenten structureel met ruim 50 procent toegenomen tot ongeveer een half miljard op jaarbasis. Het uiteindelijke doel van deze maatregelen is om de ontwikkeling van kinderen met een risico op taal- of onderwijsachterstand te versterken en daarmee de kansengelijkheid te vergroten.

EVENING richt zich op de effectiviteit en doelmatigheid van deze twee beleidsmaatregelen en probeert de volgende onderzoeksvragen te beantwoorden:

- 1) Wat zijn de effecten van extra uren voorschoolse educatie op a) de kwaliteit en het activiteitenaanbod; b) het feitelijk gebruik van voorschoolse educatie; c) de ontwikkeling van doelgroepkinderen?

- 2) Wat zijn de effecten van de nieuwe urennorm voor de pedagogisch beleidsmedewerker op a) de kwaliteit en het activiteitenaanbod; b) de ontwikkeling van doelgroepkinderen?

De twee beleidshervormingen worden gebruikt als natuurlijke experimenten om de effecten van meer uren voorschoolse educatie en de extra inzet van de pedagogisch beleidsmedewerker te schatten. Het EVENING onderzoek is opgezet als quasi-experimentele evaluatie van voorschoolse educatie waarbij econometrische technieken worden toegepast. Naast het hoofdonderzoek is door consortiumpartner Sardes verdiepend onderzoek verricht naar de wijzen waarop pedagogisch beleidsmedewerkers worden ingezet. Dat deelonderzoek is beschreven in een apart rapport.

De resultaten van het EVENING project tonen aan dat de twee beleidsmaatregelen de ontwikkeling van doelgroepeuters positief hebben beïnvloed en daarmee de kansengelijkheid hebben vergroot. De bevindingen laten zien dat de uitbreiding van het aantal uren voorschoolse educatie en de versterking van de kwaliteit door inzet van pedagogisch beleidsmedewerkers effectieve beleidsmaatregelen zijn.

Onderzoeksopzet

Het EVENING onderzoek is in april 2019 gestart. De laatste gegevens zijn in 2024 verzameld. EVENING is een uniek project waarin de ontwikkeling van de hele keten van voorschoolse educatie over een langere periode in beeld is gebracht: het gemeentelijk beleid, kindcentra en daarbinnen de groepen, en gezinnen en kinderen die deelnemen aan voorschoolse educatie. Door middel van documentanalyses, interviews en vragenlijsten voor gemeentelijke beleidsmedewerkers is het gemeentelijk beleid over onder andere de doelgroepdefinitie, de startleeftijd van voorschoolse educatie, het uren aanbod, de invoering van 960 uur voorschoolse educatie en de inzet van pedagogisch beleidsmedewerkers in gemeenten in beeld gebracht. In kindcentra zijn observaties op de groepen uitgevoerd; de observaties met het Classroom Assessment Scoring System (CLASS, peutersversie) geven een beeld van de emotionele en educatieve proceskwaliteit en de veranderingen daarin in de loop van de jaren. Met vragenlijsten voor leidinggevend en pedagogisch medewerkers werden onder andere het activiteiten aanbod en de frequentie en groepeeringsvormen van specifieke VVE-activiteiten vastgesteld. Ten slotte werd door middel van persoonlijke interviews met ouders het gebruik (beginleeftijd, uren, verandering in uren, cumulatieve aantal uren) van voorschoolse opvang en educatie in kaart gebracht, naast structurele en pedagogische gezinskenmerken. Tijdens de huisbezoeken werden ook de vaardigheden van de kinderen bepaald met een psychometrisch gevalideerde testbatterij voor (Nederlandse) fonologische kennis, woordenschat, selectieve aandacht, verbaal en ruimtelijk geheugen, en zelfcontrole (uitstel van beloning). De sociaalemotionele ontwikkeling werd gemeten aan de hand van gedragsbeoordelingen door pedagogisch medewerkers en ouders met verkorte versies van instrumenten als de Strengths and Difficulties Questionnaire (externaliserend, 'druk' gedrag en internaliserend, 'teruggetrokken' gedrag).

De steekproeftrekking volgde een gestratificeerd steekproefmodel om dekking naar regio en urbanisatiegraad te waarborgen. Kleinere gemeenten (minder dan 20.000 inwoners) en gemeenten met een kleine doelgroepopulatie volgens CBS-gegevens werden om logistieke redenen uitgesloten. De positieve respons op gemeente, centrum en medewerkerniveau was in het algemeen goed (tussen 67% en 77%). De respons van ouders kon vanwege de getrapte wervingsprocedure waarbij ouders via de kindcentra werden benaderd niet bepaald worden, maar is waarschijnlijk aanzienlijk lager. In totaal namen 314 kinderen en hun ouders deel aan EVENING. Ongeveer twee op de drie kinderen behoorden tot de doelgroep van het gemeentelijk onderwijsachterstandenbeleid. In bijna 50% van de gezinnen waren de ouders eerste generatie migranten; 53% van de kinderen groeiden op in een meertalig gezin. In 10% van de gezinnen hadden beide ouders maximaal vmbo als hoogst bereikte onderwijsniveau. In 25% was tenminste één ouder laagopgeleid en de andere ouder hooguit gemiddeld (mbo). In 35% van de gezinnen waren beide ouders hoogopgeleid (hbo en wo), waaronder veel ouders die een opleiding in het buitenland hadden gevolgd. In voorscholen is in totaal, over alle meetmomenten, in 231 unieke groepen geobserveerd, waarbij 429 (herhaalde) observaties werden verricht gedurende het onderzoek. EVENING is daarmee het meest grootschalige Nederlandse onderzoek naar de kwaliteit van voorschoolse educatie. Daarnaast werden voor 264 groepen vragenlijstdata verkregen van 176 leidinggevend en 413 pedagogisch medewerkers (deels herhaald). De kindcentra waren gevestigd in 30 gemeenten, zowel grote, middelgrote als kleine gemeenten verspreid over het land.

De uitvoering van het onderzoek werd belemmerd door de COVID-19 pandemie en de herhaalde sluiting van de kinderopvang. In 2020 kon om die reden het veldwerk slechts deels worden uitgevoerd. In de nasleep van de pandemie was er vanuit de kinderopvang lange tijd minder contact met ouders, was er verhoogd ziekteverzuim en nam het personeelstekort sterk toe. De sector kreeg te maken met ongekende personeelstekorten, die nog altijd impact hebben. Door al deze omstandigheden is de beoogde omvang van de onderzoeksgroep uiteindelijk niet bereikt. Hoewel dit het statistisch onderscheidingsvermogen van de hoofdanalyses heeft beperkt, konden de centrale onderzoeksvragen van het onderzoek adequaat worden beantwoord.

Effecten van 960 uur voorschoolse educatie

De effecten van een ruimer ureaanbod zijn geschat met behulp van een verschillen-in-verschillen benadering. Voor de analyses reconstrueren we voor elk kind het cumulatief ureaanbod vanaf leeftijd 2.5 tot het testmoment. Het cumulatief ureaanbod hangt af van de gemeente, de geboortedatum van kind en de datum van het testmoment. We houden daarbij ook rekening met sluitingen vanwege de COVID-19 pandemie en regionale verschillen in vakanties. Er is veel variatie in het ureaanbod omdat er grote verschillen zijn tussen gemeenten in zowel het ureaanbod voor de hervorming als in de timing van de invoering van 960 uur (16 uur per week). In sommige gemeenten was er een aanbod van 10 uur per week, terwijl in andere gemeenten het aanbod ruimer was, bijvoorbeeld 12 of 15 uur per week. En in sommige gemeenten werd 960 uur ingevoerd op de wettelijke invoeringsdatum (augustus 2020), terwijl in andere gemeenten de maatregel bijvoorbeeld al in januari 2020 of augustus 2019 werd

ingevoerd. Dat betekent dat ook bij dezelfde testleeftijd er grote verschillen kunnen zijn in het cumulatief urenaanbod, afhankelijk van het geboortecohort van het kind en de gemeente waar het gezin woont. Naast het cumulatief urenaanbod (recht) reconstrueren we ook voor elk kind het cumulatief urengebruik vanaf leeftijd 2.5 tot het testmoment. Hierbij maken we gebruik van een uitgebreide oudervragenlijst. Het cumulatieve urengebruik en urenaanbod komen dus via onafhankelijke bronnen tot stand: het urenaanbod wordt berekend op basis van gemeentelijk beleid, het urengebruik wordt berekend op basis van het ouderinterview.

De resultaten laten een positief effect zien van een ruimer voorschools urenaanbod op het feitelijk gebruik van uren voorschoolse educatie door doelgroepkinderen. Ook hebben extra uren voorschoolse educatie geleid tot hogere scores op de tests voor taal, executieve functies en fijne motoriek, en tot een positievere beoordeling van het sociaal-emotionele gedrag. Er worden geen significante effecten gevonden op zelfcontrole. Ook worden er voor niet-doelgroepkinderen geen effecten gevonden. Eerdere resultaten uit EVENING lieten zien dat de extra uren voorschoolse educatie doelmatig werden ingezet voor gerichte educatieve activiteiten, meestal in een kleine groepjes. Dit geeft een beeld van het causale pad van de maatregel naar uitkomsten: de kern van het werkingsmechanisme lijkt een hoger gebruik in combinatie met meer blootstelling aan gestructureerde en intensief begeleide educatieve activiteiten in kleine groepen.

Effecten van de van pedagogisch beleidsmedewerker

EVENING laat zien dat de educatieve proceskwaliteit van voorschoolse educatie tussen 2021 en 2023 substantieel is toegenomen. Dit kan duiden op een positief effect van de pedagogisch beleidsmedewerker. Het is echter ook mogelijk dat deze toename (deels) veroorzaakt is door andere ontwikkelingen. Om het effect van de pedagogisch beleidsmedewerker te isoleren is gebruik gemaakt van een verschillen-in-verschillen benadering, waarmee onder andere gecontroleerd wordt voor algemene (niet-geobserveerde) ontwikkelingen over de tijd en stabiele, niet-geobserveerde gemeentekennmerken. Omdat de wettelijk norm een inzet van minimaal 10 uur per doelgroepeuter per jaar voorschrijft, is de verwachting dat centra met meer (een hoger aandeel) doelgroepeuters sterker zijn beïnvloed door de maatregel dan centra met minder (een lager aandeel) doelgroepeuters.

De resultaten laten een significant positief effect zien op de geobserveerde educatieve proceskwaliteit, met name op de wijze waarop pedagogisch medewerkers kinderen feedback en ondersteuning geven en aan rijk en gevorderd taalgebruik blootstellen. De kwaliteitsverbeteringen worden geconstateerd tijdens (begeleid) spelmomenten en gestructureerde educatieve activiteiten, maar worden niet gevonden bij metingen tijdens eet- en drinkmomenten. Er was geen verbetering zichtbaar in de emotionele proceskwaliteit, die voor de invoering van de maatregel over de hele linie al hoog was. Ook rapporteerden pedagogisch medewerkers vaker dat ze VVE-activiteiten individueel of in kleine groepjes met alleen doelgroepkinderen aanboden (tutoring). Uit analyses blijkt verder dat pedagogisch medewerkers vaker worden gecoacht, maar er is geen significante verandering op andere dimensies van professionalisering. Dit wijst er op dat

de positieve effecten op de educatieve proceskwaliteit en op de wijze waarop VVE-activiteiten worden uitgevoerd door intensievere coaching van medewerkers tot stand is gekomen.

Tot slot zijn ook de effecten van de maatregel op de ontwikkeling van peuters getoetst. Eerder Nederlands onderzoek op basis van pre-COOL liet al zien dat een hogere educatieve proceskwaliteit een positieve bijdrage levert aan de (taal)ontwikkeling van kinderen. Omdat de pedagogisch beleidsmedewerker de educatieve proceskwaliteit in de voorschoolse educatie significant heeft verhoogd kan daarom verwacht worden dat kinderen baat hebben gehad bij deze beleidsmaatregel. In EVENING is met behulp van een verschillen-in-verschillen benadering onderzocht of de inzet van de pedagogisch beleidsmedewerker effect heeft gehad op de ontwikkeling van kinderen. Het model is gebaseerd op dezelfde variatie als de analyses op kwaliteit: het is de verwachting dat de effecten sterker zijn in centra met een hoger aandeel doelgroepeuters. Daarnaast wordt ook gebruik gemaakt van de duur van “exposure”: er is een groter effect te verwachten voor peuters die meer tijd in de voorschoolse educatie hebben doorgebracht na de implementatie van de norm. Voor kinderen die korter konden profiteren van hogere kwaliteit voorschoolse educatie omdat zij al eerder na de invoering uitstroonden naar de basisschool zijn kleinere effecten te verwachten.

De analyses laten significant positieve effecten zien op taal bij doelgroepeuters, maar geen significante effecten op hun executieve functies, motorische vaardigheden of hun sociaal-emotioneel functioneren. De effecten zijn kleiner en marginaal significant voor de hele steekproef (inclusief niet-doelgroepeuters). Dat de positieve effecten geconcentreerd zijn binnen de groep doelgroepeuters is in lijn met de bevinding dat de maatregel ook heeft geleid tot een andere werkwijze, namelijk dat VVE-activiteiten vaker in kleine groepen met alleen doelgroepeuters (of individueel) worden uitgevoerd. De steekproef is relatief klein en de verwachting is dat de effecten pas volledig zichtbaar zijn na een langere periode na de implementatie. Het is daarom mogelijk dat effecten op domeinen anders dan taal niet konden worden aangetoond in het onderzoek. Desalniettemin laten de resultaten zien dat doelgroepeuters baat hebben gehad bij deze hervorming.

Conclusies en beleidsaanbevelingen

In het EVENING onderzoek zijn twee landelijke beleidsmaatregelen in de voorschoolse educatie geëvalueerd: een uitbreiding van het urenaanbod (augustus 2020) en de invoering van een norm voor de inzet van de pedagogisch beleidsmedewerker in de voorschoolse educatie (januari 2022). De resultaten suggereren dat de uitbreiding van het aantal uren voorschoolse educatie tot 960 uur tussen de leeftijd 2.5 en 4 jaar een positief effect heeft gehad op de ontwikkeling van doelgroepkinderen. Een ruimer urenaanbod leidt tot meer uren gebruik van voorschoolse educatie en de extra uren worden ingezet voor extra aanbod van educatieve activiteiten. De bevindingen laten zien dat doelgroepeuters baat hebben bij een ruimer urenaanbod: taalvaardigheden, executieve functies en het sociaal-emotioneel functioneren zijn verbeterd met een ruimer urenaanbod. Dit beleid dient daarom gecontinueerd te worden. Het verder uitbreiden van het urenaanbod, via meer uren per week dan wel een start voor de

leeftijd van 2.5 jaar, lijkt een kansrijke strategie en vraagt nader onderzoek. In een vervolgrapportage zal EVENING aandacht besteden aan het starten met voorschoolse educatie vanaf leeftijd 2 jaar.

De tweede maatregel, de norm voor de inzet van pedagogisch beleidsmedewerker, heeft eveneens het gewenste effect. De educatieve proceskwaliteit in de voorschoolse educatie is in de afgelopen jaren fors toegenomen. De analyses laten zien dat de nieuwe norm voor de inzet van de pedagogisch beleidsmedewerker heeft geleid tot hogere educatieve proceskwaliteit tijdens belangrijke, in evenwicht aangeboden, curriculumactiviteiten, namelijk gestructureerde educatieve activiteiten en begeleid spel. In lijn met die bevindingen heeft de extra inzet van pedagogisch beleidsmedewerkers een significant positief effect op de taalvaardigheid van doelgroepkinderen. Ook deze beleidsmaatregel verdient daarom continuering.

De resultaten laten zien dat de maatregel sterkere effecten heeft gehad op centra met een hoger aandeel doelgroepeuters. Dit kan wenselijk zijn omdat daarmee, gegeven de beschikbare middelen, een groter aantal doelgroepeuters wordt bereikt. Maar deze uitkomst kan ook als onwenselijk worden beschouwd omdat centra met een lager aandeel doelgroepeuters achterblijven in termen van kwaliteit; kinderen in deze centra kunnen daarom niet optimaal profiteren van voorschoolse educatie. Om dit effect te beperken is het aan te bevelen om naast een norm van 10 uur per doelgroepeuter per jaar een minimumnorm per kindcentrum (of per voorschoolse educatie groep) te overwegen. Bij een minimumnorm van bijvoorbeeld 80 uur per VE-groep per jaar is het waarschijnlijker dat de inzet van de pedagogisch beleidsmedewerker impact heeft op de kwaliteit, ook wanneer er minder dan 8 kinderen met een VVE-indicatie deelnemen. In een aantal gemeenten wordt al gewerkt met een minimumnorm.

Tenslotte laat kwalitatief onderzoek zien dat er veel variatie is in de wijze waarop pedagogisch beleidsmedewerkers worden ingezet. Het is echter niet duidelijk hoe de pedagogisch beleidsmedewerker het meest effectief kan worden ingezet. Het is daarom aan te bevelen om via een gerandomiseerd experiment te onderzoeken hoe – gegeven een kwantitatieve norm – de pedagogisch beleidsmedewerker effectiever kan worden ingezet. Het is denkbaar dat de middelen doelmatiger kunnen worden ingezet.

Het EVENING project sluit aan op het pre-COOL project dat in 2022 werd afgerond. Uit het pre-COOL onderzoek bleek onder andere dat doelgroepkinderen al op tweejarige leeftijd een grote achterstand hadden. Met EVENING is er nieuw en vooral methodologisch sterker bewijs geleverd dat investeringen in voorschoolse educatie een belangrijke bijdrage kunnen leveren aan kansengelijkheid. Hoewel EVENING de effectiviteit van twee specifieke landelijke beleidsmaatregelen heeft geëvalueerd, kunnen de uitkomsten algemener gezien worden als bewijs dat voorschoolse educatie de ontwikkelingskansen van kinderen met een risico op taal- of onderwijsachterstand versterkt.

1. Inleiding

1.1. Het EVENING-project

Sociale ongelijkheid in de ontwikkeling van kinderen manifesteert zich al op jonge leeftijd, ruim voor de deelname aan het primair onderwijs (Leseman et al., 2017; Leseman & Veen, 2022; van Huizen, 2018; CPB, 2020). Door peuters met een risico op een onderwijsachterstand beter op de basisschool voor te bereiden, kunnen voorschoolse voorzieningen een belangrijke rol spelen bij het creëren van een meer gelijk speelveld. De internationale literatuur laat zien dat voorschoolse voorzieningen een positief effect kunnen hebben op kinderen uit kansarme gezinnen (Bailey et al., 2021; Leseman & Veen, 2022; van Huizen & Plantenga, 2018). Ook Nederlands onderzoek naar de effecten van de bestuursafspraken met de G37 en de gerelateerde financiële impuls voor voorschoolse educatie (VE) laat zien dat doelgroepkinderen positieve effecten ondervinden van extra investeringen in voorschoolse educatie (Akgündüz & Heijnen, 2018).

Het EVENING onderzoek evalueert de effecten van een verhoging van de kwantiteit (urenaanbod) en de kwaliteit van voorschoolse educatie. Dit onderzoek sluit daarmee aan op het pre-COOL project dat in 2022 werd afgerond (Leseman & Veen, 2022). In pre-COOL werden aanwijzingen gevonden dat kinderen van praktisch opgeleide ouders, met een migratieachtergrond of een andere thuistaal dan het Nederlands baat kunnen hebben bij deelname aan voorschoolse opvang en educatie. Al vroeg bestaande verschillen in taalvaardigheid en executieve functies tussen kinderen naar gezinsachtergrond halveren in de periode van 2 tot 6 jaar. Hoewel het mede op grond van bewijs uit verschillende landen 'theoretisch' plausibel is dat de verkleining van de vroege kloof (mede) het gevolg is van deelname aan voor- en vroegschoolse educatie, laat de onderzoeksopzet van pre-COOL met beperkte controle van (niet-geobserveerde) selectiemechanismen, geen causale conclusie toe over het effect van voor- en vroegschoolse educatie. Het is daarom niet uit te sluiten dat er alternatieve verklaringen zijn voor de patronen die gevonden worden in het pre-COOL onderzoek. Al met al is er voor de Nederlandse situatie behoefte aan meer en vooral methodologisch sterker bewijs. Het doel van EVENING is om hier een belangrijke bijdrage aan te leveren.

EVENING is opgezet als een quasi-experimentele evaluatie van twee beleidsmaatregelen in de voorschoolse educatie: 1) de uitbreiding van het minimum urenaanbod van 600 naar 960 uur tussen de leeftijd van 2.5 en 4 jaar in augustus 2020; en 2) de invoering van een landelijke urennorm voor de inzet van een pedagogisch beleidsmedewerker in 2022. Het doel van de beleidsmaatregelen is het vergroten van kansengelijkheid door de ontwikkeling van kinderen met een risico op taal- of onderwijsachterstand te versterken. Voor de uitvoering van deze twee beleidsmaatregelen is voor gemeenten structureel € 170 miljoen beschikbaar vanuit het Rijk. Daarmee is vanaf 2020 het budget

voor gemeenten structureel met ruim 50 procent toegenomen tot ongeveer een half miljard op jaarbasis.¹

EVENING evalueert de effecten van de twee hervormingen en probeert de volgende onderzoeksvragen te beantwoorden:

- 1 *Wat zijn de effecten van het extra urenaanbod van voorschoolse educatie op a) de kwaliteit en het activiteitenaanbod; b) het urengebruik van voorschoolse educatie; c) de ontwikkeling van (doelgroep)kinderen?*
- 2 *Wat zijn de effecten van de nieuwe urennorm voor de pedagogisch beleidsmedewerker op a) de kwaliteit en het activiteitenaanbod; b) de ontwikkeling van (doelgroep)kinderen?*

Het EVENING onderzoek is uitgevoerd door een multidisciplinair team van de Universiteit Utrecht en Sardes en maakt deel uit van een breed monitorings- en beleidsevaluatieprogramma opgesteld door het ministerie van OCW om het onderwijsachterstandenbeleid te evalueren.

Het project is in april 2019 gestart. De laatste gegevens werden in de eerste helft van 2024 verzameld. De uitvoering van het onderzoek is ernstig belemmerd door de COVID-pandemie en de herhaalde sluiting van kinderopvang. Dit maakte aanpassingen van de oorspronkelijke onderzoeksaanpak noodzakelijk, zoals die in de voorgaande rapporten zijn beschreven. In de nasleep van de pandemie was er vanuit de kinderopvang lange tijd minder contact met ouders, was er verhoogd ziekteverzuim en nam het personeelstekort sterk toe. Door al deze omstandigheden is de beoogde omvang van de steekproef uiteindelijk niet bereikt. Hoewel dit het statistisch onderscheidingsvermogen van de hoofdanalyses heeft beperkt, konden de hoofdvragen van het onderzoek adequaat worden beantwoord.

Een uniek aspect van het EVENING project is dat data verzameld is in de hele keten van voorschoolse educatie: gemeenten, kindcentra (aanbieders van voorschoolse educatie) en gezinnen en kinderen. Om ontwikkelingen in gemeentelijk beleid in kaart te brengen is er op meerdere momenten data verzameld bij gemeenten via vragenlijsten en gesprekken en zijn relevante beleidsdocumenten bestudeerd. In drie meetronden werden data verzameld bij kindcentra over onder andere de kwaliteit en het activiteitenaanbod door middel van vragenlijsten en observaties. Tenslotte werden met behulp van tests en interviews doorlopend data verzameld over kinderen en gezinnen. Een innovatieve onderzoeksopzet was mogelijk door gegevens over (veranderingen in) gemeentelijk beleid te koppelen aan data over de kwaliteit van voorschoolse educatie en de ontwikkeling van doelgroepeuters. De gegevens zijn verzameld rond de twee

¹ Staatsblad van het Koninkrijk der Nederlanden (2019), Besluit van 20 september 2019 tot wijziging van het Besluit basisvoorwaarden kwaliteit voorschoolse educatie in verband met de verhoging van het minimaal aantal uren aanbod voorschoolse educatie en de inzet van een pedagogisch beleidsmedewerker, jaargang 315, 1-24.

beleidshervormingen, waardoor de effecten van de maatregelen konden worden geëvalueerd met behulp van “natuurlijk experiment” methoden.

1.2. Overzicht van dit rapport

Dit rapport bestaat uit twee delen. Het eerste deel (hoofdstukken 2 t/m 5) beschrijft de onderzoeksopzet, de werving van deelnemende gemeenten, kindercentra, ouders en kinderen, en de gebruikte meetinstrumenten. In hoofdstuk 2 geven we een overzicht van de EVENING steekproef over alle meetronden heen. Dit hoofdstuk bevat verder een beschrijving van de data die in de periode 2020 tot en met 2024 op de groepen en bij de gezinnen zijn verzameld. In hoofdstuk 3 beschrijven we de resultaten van de kwaliteitsobservaties met de CLASS en analyseren we de veranderingen in de geobserveerde kwaliteit over de drie meetronden. Hoofdstuk 4 gaat in op de deelnemende gezinnen. We beschrijven de sociaaleconomische, sociolinguïstische en etnisch-culturele kenmerken van de gezinnen en laten zien hoe het gebruik van VE-uren is veranderd sinds begin 2020. Hoofdstuk 5 staat in het teken van de metingen bij kinderen. We verantwoorden de keuze van de meetinstrumenten en rapporteren de psychometrische analyses die zijn uitgevoerd om tot de uiteindelijke constructen te komen. Het tweede deel van dit rapport is gewijd aan de effectevaluaties van de twee beleidsmaatregelen: de invoering van 960 uur voorschoolse educatie (hoofdstuk 6) en de norm voor de inzet van de pedagogisch beleidsmedewerker (hoofdstuk 7). De methodologische benadering en identificatiestrategie van het quasi-experimentele onderzoeksdesign worden verder toegelicht in deze hoofdstukken. Het laatste hoofdstuk geeft een beknopt overzicht van de belangrijkste bevindingen, geeft antwoord op de centrale onderzoeksvragen en beschrijft de beleidsimplicaties.

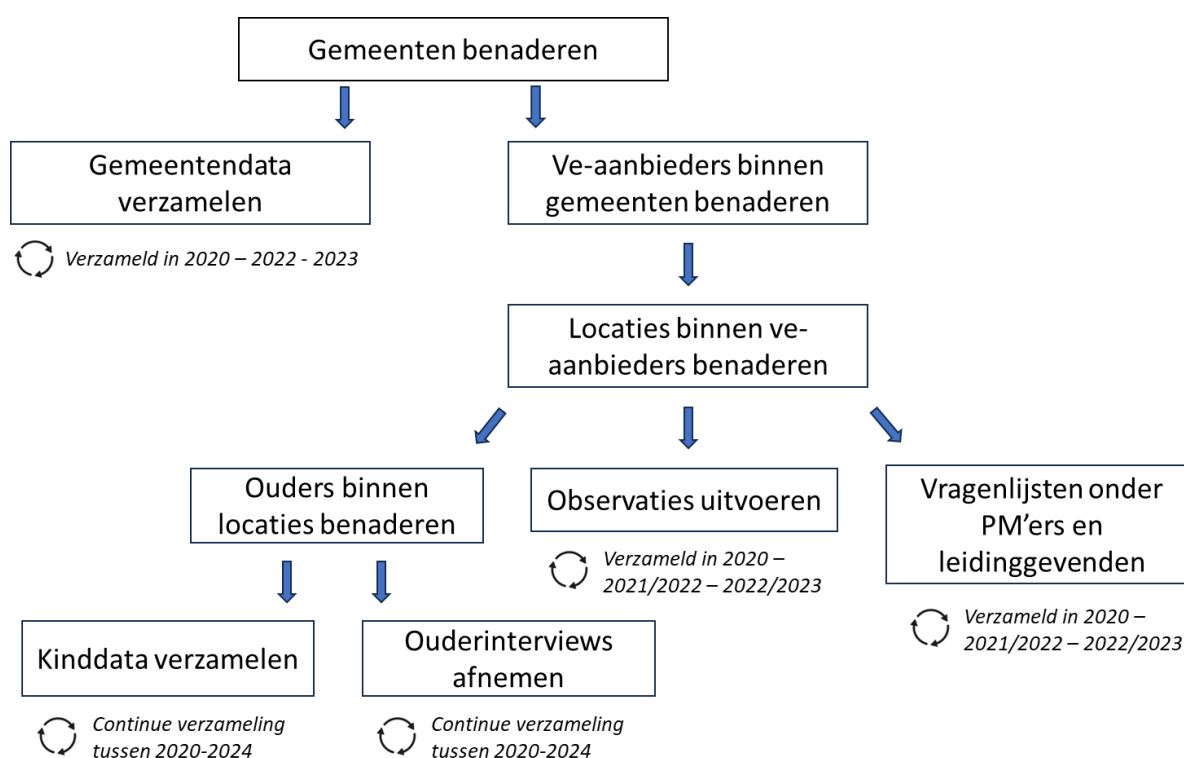
2. Steekproef en procedures

In dit hoofdstuk wordt de steekproef en de gehanteerde procedures rondom de dataverzameling toegelicht. Er wordt beschreven hoe de selectie van gemeenten, VE-aanbieders en kinderen heeft plaatsgevonden. Daarnaast wordt ingegaan op de respons- en non-respons.

2.1. Procedure en response

De steekproeftrekking voor EVENING bestond uit verschillende stappen. Eerst is een steekproef van gemeenten getrokken. Vervolgens zijn, na overleg met de organisaties die VE aanbieden in deze gemeenten, VE-locaties binnen de gemeente benaderd voor deelname. Via de deelnemende kindcentra zijn als volgende stap ouders en hun kinderen geworven. Figuur 2-1 geeft een overzicht van de verschillende stappen binnen de steekproeftrekking en dataverzameling. Tabel 2-1 geeft een overzicht van de verschillende deelnemers. In Appendix A wordt een overzicht gegeven van de domeinen die worden uitgevraagd of getest. De steekproef uit 156 kindercentra met minimaal één kwaliteitsmeting (429 observaties in totaal) en 314 kinderen en ouders die van begin 2020 tot juni 2024 aan het onderzoek hebben deelgenomen.

Figuur 2-1 Flowchart van de steekproeftrekking en dataverzameling



Tabel 2-1 Overzicht van de deelnemers

	N 2020- 2021	N 2021- 2022	N 2022- 2023	N 2023- 2024	Opmerking
Gemeenten	21	25	25	7	Sinds de start van het onderzoek zijn 30 gemeenten benaderd. De aangegeven cijfers betreffen het aantal gemeenten waar observaties en/of huisbezoeken zijn uitgevoerd, naast het verzamelen van gemeentedata.
Deelnemende groepen	200	178	186	-	Aantal deelnemende groepen, locaties en aanbieders per meetronde.
Locaties	135	125	128	-	
Aanbieders/ houders	40	37	41	-	
Uitgevoerde observaties	68	177	184	-	Aantal uitgevoerde observaties op de deelnemende groepen. In 2020 minder vanwege COVID.
Vragenlijst – leidinggevenden	107	95	83	-	Aantal ingevulde vragenlijsten door leidinggevenden, volledig of minimaal 30% ingevuld. Deze vragenlijsten worden per deelnemende locatie ingevuld. Sommige leidinggevenden coördineren meer dan één locatie en vullen daarom meerdere vragenlijsten in.
Volledig	94	86	74		
Gedeeltelijk	13	9	9		
Vragenlijst - pedagogisch medewerkers	127	141	136	-	Aantal ingevulde vragenlijsten door pedagogisch medewerkers, volledig of minimaal 30% ingevuld. Deze vragenlijsten worden per deelnemende groep ingevuld.
Volledig	122	137	127		
Gedeeltelijk	5	4	9		
Gezinnen en kinderen (aantal eerste afspraken)	113	171	29	1	In totaal hebben tot en met juni 2024 314 gezinnen meegedaan. Bij 110 kinderen/gezinnen zijn twee huisbezoeken gedaan.
Aantal tweede afspraken	-	33	60	17	

Tabel 2-2 geeft de respons van de VE-groepen en de redenen van non-respons weer. In de eerste meetronde (2020-2021) waren er 196 deelnemende groepen. Het aantal

groepen waar geobserveerd kon worden was vanwege de COVID-maatregelen laag (68 groepen), maar er is wel een relatief hoog percentage van online vragenlijsten ingevuld (60,2% van de leidinggevend en 63,5% van de pedagogisch medewerkers). In 2021-2022 en 2022-2023 zijn alle groepen waar we een positieve respons kregen geobserveerd en is een hoog percentage vragenlijsten door zowel leidinggevend en pedagogisch medewerkers ingevuld. In de eerste meetronde waren COVID-gerelateerde problemen de meest genoemde reden om deel te nemen aan het onderzoek. Bij de latere meetronden had de non-respons vaker te maken met redenen op organisatie- en beleidsniveau (zoals fusies) of werd een te hoge werkdruk als reden genoemd. In Appendix B worden de verschillende inhoudelijke redenen toegelicht.

Tabel 2-2 Respons en non-respons van VE groepen (2020-2023)

VE groepen	2020-2021	2021-2022	2022-2023
Totaal benaderd	325 (100%)	308 (100%)	227 (100%)
Respons	200 (61.5%)	178 (57.8%)	186 (81.9%)
Geobserveerde groepen	68 (34%)	177 (99.4%)	184 (98.9%)
Ingevulde vragenlijsten – leidinggevend en	131 (65.5%)	135 (75.8%)	124 (66.7%)
Ingevulde vragenlijsten – pm'ers	127 (63.5%)	141 (79.2%)	136 (73.1%)

Inhoudelijke redenen non-respons

Onderzoek	0 (0%)	2 (1.5%)	6 (14.6%)
Opvanglocatie	1 (0.8%)	28 (21.5 %)	9 (22%)
Beroepskrachten	0 (0%)	0 (0%)	1 (2.4%)
Organisatie en beleid	2 (1.6%)	6 (4.6%)	9 (22%)
Corona-maatregelen	116 (92.8%)	13 (10%)	0 (0%)
Ouders	1 (0.8%)	6 (4.6%)	0 (0%)
Kinderen	1 (0.8%)	0 (0%)	3 (7.3%)
Overig	4 (3.2%)	75 (57.7%)	13 (31.7%)

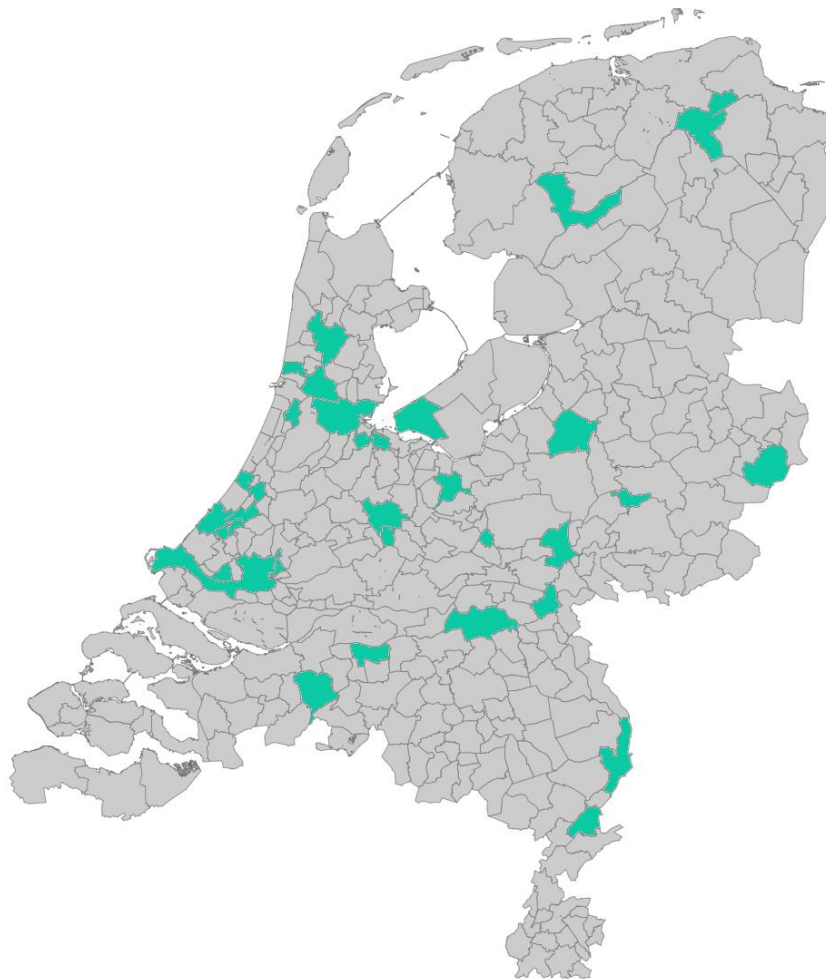
Noot. Het gerapporteerde aantal ingevulde vragenlijsten door leidinggevend en pedagogisch medewerkers (pm'ers) betreft het aantal groepen.

2.2. Gemeenten

Bij het trekken van de steekproef van gemeenten is een aantal criteria toegepast. Om redenen van efficiëntie zijn gemeenten met minder dan 20.000 inwoners en/of minder dan 50 doelgroepkinderen volgens de CBS-gegevens niet meegenomen. De kleinste Nederlandse gemeenten en gemeenten met een zeer klein aantal doelgroepkinderen zijn daarom niet vertegenwoordigd in het onderzoek. Voor het onderzoek is het ook van belang dat er gedurende de relevante onderzoeksperiode al een redelijke VE-infrastructuur (d.w.z. voldoende VE-groepen) bestond in de gemeente. Om dat te garanderen hebben we alleen gemeenten meegenomen in de trekking met een voldoende hoog niveau van GOAB-middelen (minimaal €250.000 in 2018 en minimaal €400.000 in 2020). Ook hebben we gemeenten die ontstaan zijn door fusies per 1 januari 2019 niet meegenomen, aangezien in deze gemeenten de vergelijking met voorgaand beleid moeilijk of niet mogelijk is.

Binnen de overgebleven groep gemeenten hebben we gemeenten aselekt getrokken, gestratificeerd naar urbanisatiegraad en regio in het land. Hierbij hebben we de verhouding van aantal doelgroepkinderen in de verschillende type gemeenten en regio's meegewogen. Volgens de CBS schatting woont meer dan de helft van alle doelgroepkinderen bijvoorbeeld in één van de 37 grootste gemeenten. Daarom is ervoor gekozen ook een relatief groot aandeel van de G37 gemeenten te trekken. Wanneer gemeenten besloten niet deel te nemen aan het onderzoek zijn er uit dezelfde cel van het steekproefontwerp (gedefinieerd door urbanisatiegraad en regio) aselekt alternatieve gemeenten getrokken.

Figuur 2-2 Geografische spreiding van deelnemende EVENING-groepen in 2020-2023



Toelichting: De blauwgroene gemeenten hebben deelgenomen aan het EVENING project.

In totaal hebben we data verzameld in 30 gemeenten. In 2020 kon het onderzoek vanwege de COVID-maatregelen slechts in beperkte mate in 21 gemeenten worden uitgevoerd. In de tweede en derde meetronde beschikken we over data van 25 gemeenten. Het onderzoeksteam heeft kwantitatief en kwalitatief onderzoek uitgevoerd in de gemeenten naar het gemeentelijk VE-beleid. In dit onderzoek zijn de beleidsontwikkelingen gemeten via vragenlijsten, door relevante beleidsdocumenten te bestuderen en met interviews met verantwoordelijke beleidsmedewerkers. Als

aanvulling op de vragenlijsten van de Monitor Implementatie en Besteding van het GOAB (2019-2021), hebben we in juli 2020, april 2022 en juli 2023 een aanvullende vragenlijst uitgezet onder de deelnemende gemeenten waarin gevraagd wordt naar het beleid ten aanzien van voorschoolse educatie. Deze vragenlijst richt zich vooral op het aanbod van voorschoolse educatie, de inzet van pedagogisch beleidsmedewerkers en de gehanteerde doelgroepdefinitie. In 2020 hadden we complete vragenlijsten voor 25 gemeenten, 18 in 2022 en 22 in 2023. In de laatste vragenlijsten worden ook recente beleidsveranderingen uitgevraagd, zodat in kaart gebracht kan worden hoe het beleid zich in de afgelopen periode heeft ontwikkeld. Figuur 2-2 geeft aan welke gemeenten (op enig moment) deelnamen in 2020-2023.

2.3. Aanbieders

Binnen gemeenten zijn aanbieders van voorschoolse educatie geworven voor deelname aan het onderzoek. Hierbij is rekening gehouden met evenwichtige verdeling naar type aanbieders (o.a. kinderopvang en peuteropvang; lokaal, regionaal, nationaal). In totaal namen 49 verschillende aanbieders actief deel met verschillende locaties en groepen. In totaal zijn er 429 observaties uitgevoerd op 264 groepen.

Binnen de kindcentra is gedurende de onderzoeksperiode een aantal keer een dagdeel geobserveerd op de groepen om de educatieve en emotionele proceskwaliteit te meten. In het volgende hoofdstuk geven we meer achtergrondinformatie en beschrijvende resultaten van de observaties uitgevoerd middels de CLASS Toddler, het gebruikte observatie-instrument om de proceskwaliteit van voorschoolse voorzieningen te meten.

Bij de evaluatie van de maatregelen wordt gekeken naar de manier waarop het VE-urenaanbod in praktijk wordt vormgegeven en hoe de pedagogisch beleidsmedewerker wordt ingezet. Hiervoor zijn tijdens iedere meetronde vragenlijsten uitgezet bij locatiemanagers en pedagogisch medewerkers. In Appendix A staat een overzicht van de kernconcepten uit alle vragenlijsten.

2.4. Kinderen en gezinnen

De gezinnen zijn geworven via de deelnemende groepen. Pedagogisch medewerkers brachten het onderzoek onder de aandacht bij de ouders middels (meertalige) folders. Op meerdere momenten werd de folder uitgedeeld, bijvoorbeeld tijdens het intakegesprek met ouders, of rondom ouderbijeenkomsten of inloopochtenden. Ook werd er in nieuwsbrieven of in het digitale ouderportaal gewezen op het EVENING project en de mogelijkheid om deel te nemen.

Tijdens huisbezoeken werd een gestructureerd persoonlijk interview afgenomen met de ouders waarin onder andere vragen werden gesteld over ervaringen met de voorschool, het urengebruik van voorschoolse educatie, activiteiten die plaatsvinden in de thuissituatie, en een aantal achtergrondgegevens van de gezinnen. Bij de kinderen werd in de thuisomgeving met speelse taakjes het ontwikkelingsniveau aan het begin (rond leeftijd 2,5 jaar) en aan het eind (bijna of net 4 jaar) van hun voorschoolse periode vastgesteld, gericht op de brede ontwikkeling van het kind op het gebied van

Nederlandse taalvaardigheid, executieve functies, sociaal-emotioneel functioneren en motoriek. Om kindcentra zoveel mogelijk te ontlasten, werden de taakjes voor kinderen en de interviews met ouders in de thuissituatie uitgevoerd.

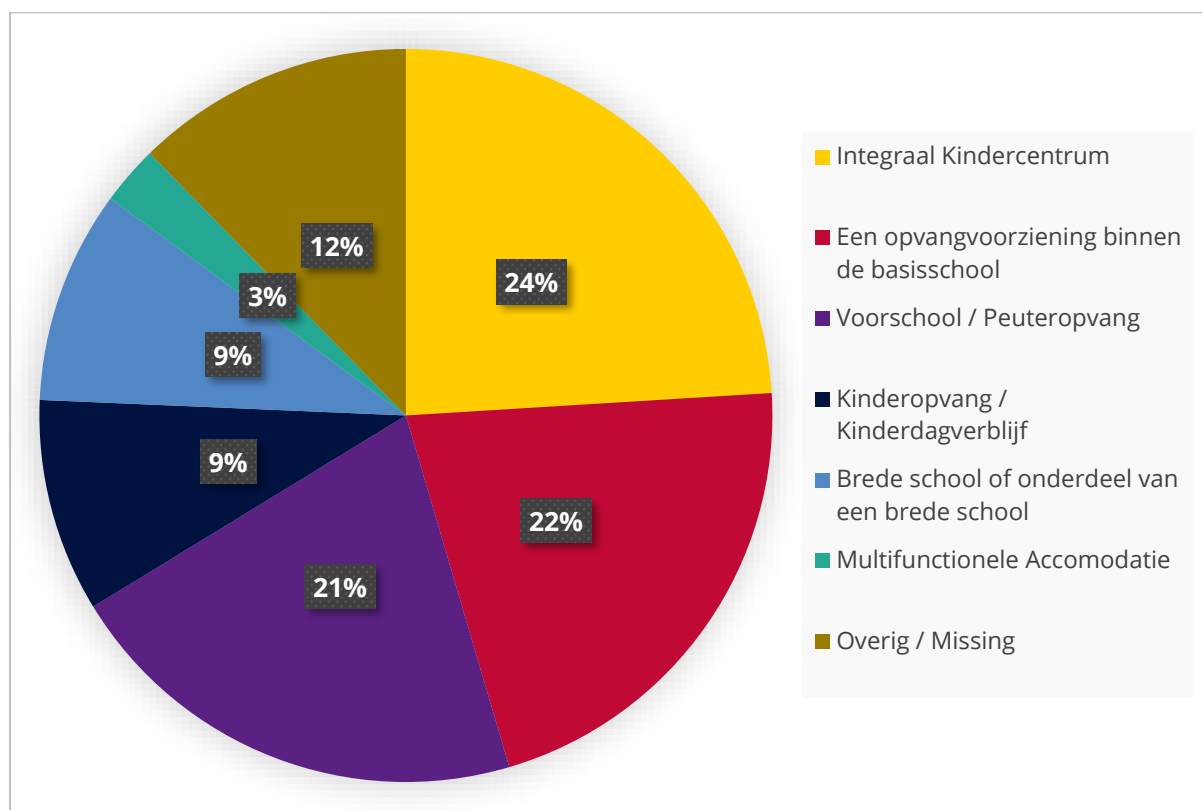
3. Kindcentra

In dit hoofdstuk worden de kenmerken van de kindercentra beschreven, evenals achtergrondkenmerken van de pedagogisch medewerkers en leidinggevenden. We gaan onder andere in op de respons per meetronde en de verdeling van de deelnemende locaties naar opvangtype. Daarnaast lichten we het instrument toe dat is gebruikt om de emotionele en educatieve proceskwaliteit binnen VE te meten. De scores op proceskwaliteit worden gepresenteerd voor de volledige steekproef, per meetronde en naar type activiteit. Ook worden deze resultaten vergeleken met bevindingen uit andere (inter)nationale studies.

3.1. Kenmerken

In verband met de coronacrisis is in 2020 de beslissing genomen om de persoonlijke interviews met de pedagogisch medewerkers en leidinggevenden om te zetten in online vragenlijsten. Gezien de positieve ervaringen in het veld met de online versie en het belang van continuïteit van de dataverzamelingsprocedure is voor de vervolgmetingen eveneens gebruik gemaakt van online vragenlijsten. Dit rapport beschrijft resultaten van volledig en gedeeltelijk (minimaal 30%) ingevulde vragenlijsten.

Figuur 3-1 Verdeling naar type opvanglocatie (2020-2023)



Respons

In 2020 hebben we vanwege de COVID-maatregelen niet alle deelnemende groepen kunnen observeren. Om deze reden hebben we besloten om de vragenlijst voor de pedagogisch medewerkers en de leidinggevenden naar alle deelnemende groepen te sturen, ongeacht of we een observatie konden plannen. In 2020 zijn de vragenlijsten uitgezet van april tot oktober 2020. In totaal zijn pedagogisch medewerkers van 200 groepen benaderd voor het invullen van de vragenlijst: 127 (63.5%) pedagogisch medewerkers hebben de vragenlijst volledig of voor minimaal 30% ingevuld. Verder hebben 135 verschillende locaties vragenlijsten voor de leidinggevenden gekregen en zijn de gegevens van de leidinggevenden van 107 locaties (79.3%), met vaak meerdere EVENING-groepen, meegenomen in deze rapportage.

De vragenlijsten voor meetmoment 2 en 3 zijn respectievelijk in de perioden van september 2021 tot en met juni 2022 en van oktober 2022 tot met augustus 2023 uitgezet; ze beslaan daarom schooljaren 2021/2022 en 2022/2023. De vragenlijst voor pedagogisch medewerkers is in 2021-2022 uitgezet bij 178 groepen en in 2022-2023 bij 186 groepen. In 2021-2022 vulden pedagogisch medewerkers van 141 groepen (79.2%) en in 2022-2023 van 136 groepen (73.1%) de vragenlijst in. De vragenlijst voor leidinggevenden in 2021-2022 is uitgezet onder 125 locaties en in 2022-2023 onder 128 locaties. In 2021-2022 vulden leidinggevenden van 95 locaties (76%) en in 2022-2023 van 83 locaties (64.8%) de vragenlijst in.

Type kindcentrum

Het merendeel van de deelnemende locaties was onderdeel van een integraal kindcentrum (24%), een opvangvoorziening binnen een basisschool (22%) of een aparte voorschool/peuteropvang locatie (21%) (zie Figuur 3-1). Daarnaast werd op 24% van de kindcentra ook kinderdagopvang aangeboden, bij 15% was er ook een (kortdurende) reguliere peuteropvanggroep en op 27% van de locaties werd ook buitenschoolse opvang aangeboden.

Achtergrond pedagogisch medewerkers en leidinggevende

Tabel 3-1 geeft een overzicht van de achtergrond van de pedagogisch medewerkers en leidinggevenden die een EVENING-vragenlijst hebben ingevuld. De gemiddelde leeftijd van de pedagogisch medewerkers is bijna 45 jaar en de werktijd per week is gemiddeld 24 uur. De grootste groep pedagogisch medewerkers in de steekproef heeft een mbo-3 of -4 opleiding afgerond, maar ook een aanzienlijk deel (ruim 38%) heeft een hbo-opleiding of hoger afgerond. Dat aandeel is hoger dan in de reguliere kinderopvang (zie bijvoorbeeld het LKK project) en dat kan verklaard worden door het feit dat hbo-geschoolde medewerkers vaker werkzaam zijn op VE-groepen (en mogelijk ook vaker de vragenlijst invullen). De gemiddelde werkervaring van de pedagogisch medewerkers in de steekproef is 18 jaar.

Tabel 3-1 Achtergrond van pedagogisch medewerkers en leidinggevend

Achtergrond pedagogisch medewerker	N	%	Range	M	SD
Leeftijd	413		21-66	44.79	11.15
Werkdagen	409		1-5	3.71	0.93
Werktijd in uren	408		2-40	24.35	6.71
Opleidingsniveau (1-8)	284		5-8	5.80	1.00
mbo 3 of 4	171	60.2			
hbo of hoger	107	37.7			
Werkervaring in kinderopvang in jaren	273		0.5-43	18.09	9.50
Achtergrond leidinggevend	N	%	Range	M	SD
Leeftijd	176		26-65	48.15	9.96
Opleidingsniveau (1-8)	146		5-8	6.71	1.04
mbo 3 of 4	35	24.0			
hbo of hoger	111	76.0			
Werkervaring in kinderopvang in jaren	144		0-43	17.06	10.59

Noot. De tabel bevat gegevens over alle meetrondes gecombineerd. Zie Appendix C voor gegevens per meetronde.

De gemiddelde leeftijd van de leidinggevend is 48 jaar. Een meerderheid van de leidinggevend (76%) heeft een hbo-opleiding of hoger afgerond. De gemiddelde werkervaring in de kinderopvang is 17 jaar. Bijna 80% van de leidinggevend gaf aan dat zij locatiemanager waren van de desbetreffende locatie en 9% van de leidinggevend was een bestuurder vanuit de centrale organisatie of eigenaar. In sommige gevallen was de vragenlijst ingevuld door iemand in een andere leidinggevende/coördinerende functie, zoals de regiomanager of de VE-coördinator. Voor meer gedetailleerde informatie over de achtergrondgegevens van de deelnemende pedagogisch medewerkers en de leidinggevend per meetronde en de achtergrondkenmerken verwijzen we naar Appendix C.

3.2. Proceskwaliteit

Observatie-instrument

Om de proceskwaliteit binnen de voorschoolse opvanglocaties te meten, is gebruik gemaakt van het instrument Classroom Assessment Scoring System (CLASS) Toddler (La Paro et al., 2014). Dit is een veelvuldig gebruikt observatie-instrument dat ontwikkeld is om de interacties tussen pedagogisch medewerkers en kinderen binnen voorschoolse voorzieningen te beoordelen. In Nederland wordt de CLASS onder andere ingezet bij de Landelijke Kwaliteitsmonitor Kinderopvang (LKK). De CLASS werd ook gebruikt in het pre-COOL onderzoek. Analyses op basis van de pre-COOL data laten zien dat met name de educatieve proceskwaliteit zoals gemeten met de CLASS een positief effect heeft op de taalontwikkeling en op de ontwikkeling van de executieve functies (Leseman & van Huizen, 2022). Het instrument wordt regelmatig ingezet als basis voor verbeteringen in de interactievaardigheden van pedagogisch medewerkers en leerkrachten.

Het CLASS Toddler observatie-instrument meet de kwaliteit van de interacties tussen pedagogisch medewerkers en kinderen aan de hand van twee overkoepelende domeinen: 1) emotionele kwaliteit, bestaande uit de dimensies positieve sfeer, negatieve sfeer (omgekeerd gescoord), sensitiviteit, respect voor het kindperspectief, en gedragsregulering; 2) educatieve kwaliteit, met de dimensies faciliteren van ontwikkeling, kwaliteit van feedback en taalstimulering. Elke dimensie krijgt een eigen score toegekend op een schaal van 1 tot 7, waarbij 7 de hoogst mogelijke score aangeeft. De score op emotionele en educatieve kwaliteit wordt berekend door het gemiddelde te nemen van de bijbehorende dimensies.

Scores van emotionele en educatieve proceskwaliteit

Voor het EVENING-project hebben tussen 2020-2023 drie meetronden plaatsgevonden en zijn metingen van emotionele en educatieve proceskwaliteit uitgevoerd op 429 groepen in de deelnemende kindcentra.² Op elke groep zijn tijdens een dagdeel drie cycli van observaties uitgevoerd: tijdens een gerichte activiteit, een (vrij/begeleid) spelmoment en een eet- en drinkmoment.

In Tabel 3-2 zijn de gemiddelde scores, de standaarddeviaties en de scorering van de geobserveerde kwaliteit weergegeven op twee verschillende niveaus. In de linker kolommen wordt de gemiddelde kwaliteit van de verschillende VE-groepen weergegeven. In de rechter kolommen staan de gemiddelde kwaliteit van de observatiecycli binnen een groep; dit verwijst naar de drie verschillende typen activiteiten die binnen elke voorschool geobserveerd zijn. De gemiddelde scores zijn, zoals verwacht, vergelijkbaar omdat de scores per activiteit (rechts) op groepsniveau zijn geaggregeerd (links). Omdat er geaggregeerd (gemiddeld) wordt over drie activiteiten is ook de spreiding op groepsniveau lager dan op het activiteitenniveau.

De CLASS scores kunnen geclassificeerd worden als laag (onder de 3), midden-voldoende (tussen de 3 en 5) en hoog (boven de 5) (Slot et al., 2019). Uit Tabel 3-2 kan worden opgemaakt dat de emotionele kwaliteit zoals geobserveerd met de CLASS in de midden tot en met hoge range valt. Er is gemiddeld sprake van een hoge mate van positieve sfeer en een negatieve sfeer ontbreekt vrijwel in de meeste groepen. De pedagogische medewerkers laten een midden tot hoge mate van sensitiviteit zien en zijn in staat om het gedrag van de kinderen in de groep op een goede manier te begeleiden. De mate waarin de pedagogisch medewerkers uitgaan van de interesses, motivatie en perspectieven van kinderen in hun interacties, de dimensie respect voor het kindperspectief, scoort weliswaar iets lager, maar nog steeds in de middenrange. Op alle dimensies van de emotionele kwaliteit in groepsniveau zijn er hoge

² Aan het CLASS instrument is sinds de tweede meting een extra dimensie toegevoegd, waarop gescoord wordt in welke mate de pedagogisch medewerker samenwerking tussen kinderen bij gezamenlijk spel en groepsactiviteiten bevordert, de dimensie groepsocialisatie. De schaal is ontwikkeld in het kader van het LKK onderzoek en gebaseerd op eerder werk (Van Schaik et al., 2018). Deze gegevens worden niet gebruikt in de analyses in deze rapportage maar zullen in vervolganalyses naar de kwaliteit van voorschoolse educatie worden geanalyseerd.

beoordelingen gegeven, het grootste deel scoort boven de 5, en op geen enkele dimensie is de laagste score van 1 toegekend.

De educatieve kwaliteit scoort in de middenrange en deze score is over het algemeen lager dan de emotionele kwaliteit. Ook is de variatie in educatieve kwaliteit tussen de verschillende groepen groter. De score op de dimensie het stimuleren van taalontwikkeling tijdens interacties is vergelijkbaar met de score op de gemiddelde educatieve kwaliteit. De activiteiten die pedagogisch medewerkers aanbieden om het leren en de ontwikkeling te ondersteunen worden hoger gescoord. De kwaliteit van de feedback die zij kinderen geven scoort relatief laag vergeleken met die andere educatieve dimensies. Op de dimensies faciliteren van de begripsontwikkeling, stimuleren van taalontwikkeling en groepsocialisatie zijn ook hoge scores, boven de 5, toegekend aan groepen, terwijl op alle dimensies aan enkele groepen de laagste score van 1 zijn toegekend. Voor meer gedetailleerde informatie over de geobserveerde kwaliteit per meetronde verwijzen we naar Appendix C.

Tabel 3-2 Geobserveerde proceskwaliteit

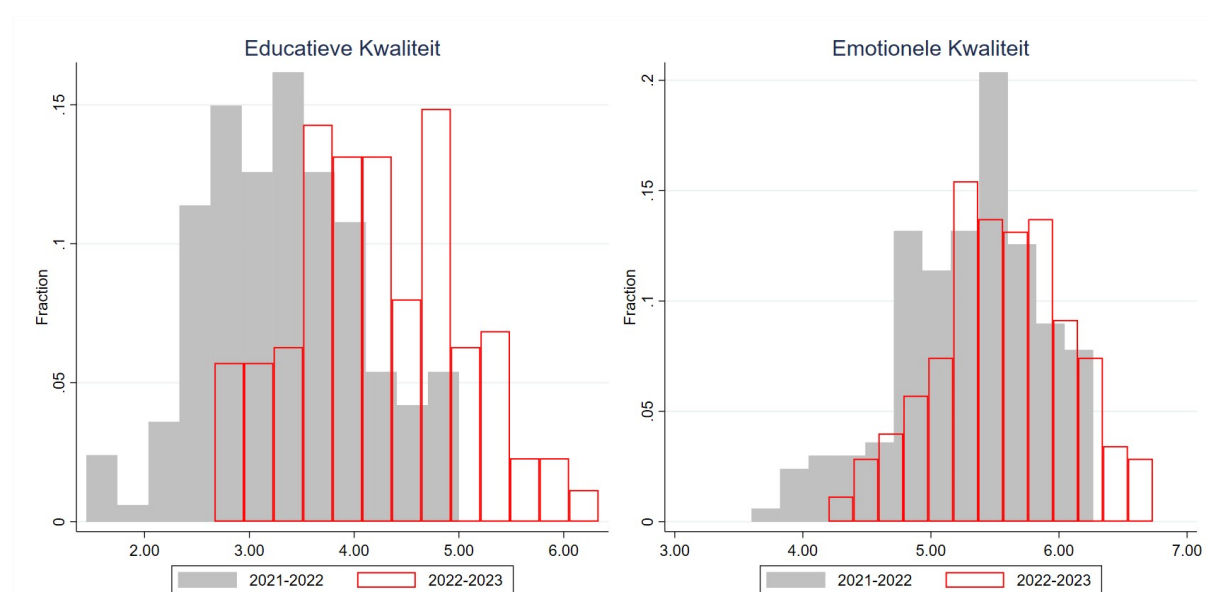
	Groepen (N=429)			Observatiecycli (N=1284)		
	Range	M	SD	Range	M	SD
Emotionele kwaliteit	3.60 – 6.73	5.41	0.56	2.8 – 7.0	5.41	0.67
Positieve sfeer	3.33 – 7.00	5.55	0.77	1.0 – 7.0	5.55	0.97
Negatieve sfeer (R)	5.33 – 7.00	6.84	0.30	3.0 – 7.0	6.84	0.41
Sensitiviteit	2.67 – 7.00	5.28	0.72	2.0 – 7.0	5.28	0.96
Kindperspectief	1.33 – 7.00	4.39	0.88	1.0 – 7.0	4.39	1.27
Gedragregulering	2.67 – 6.67	5.00	0.83	2.0 – 7.0	5.00	1.04
Educatieve kwaliteit	1.44 – 6.33	3.76	0.89	1.0 – 6.7	3.76	1.06
Faciliteren ontwik.	1.67 – 6.67	4.24	0.93	1.0 – 7.0	4.24	1.23
Kwaliteit feedback	1.00 – 7.00	3.33	1.16	1.0 – 7.0	3.32	1.33
Taalstimulering	1.00 – 6.00	3.72	0.93	1.0 – 7.0	3.72	1.13
Structurele kenmerken						
Aantal kinderen	3.7 – 16.0	9.58	2.68	1.0 – 16.0	9.58	3.19
Aantal volwassenen	1.0 – 4.0	1.97	0.48	1.0 – 4.0	1.97	0.61
Kind-staf ratio	2.0 – 12.0	5.05	1.56	1.0 – 15.0	5.12	1.88

Noot. De structurele kenmerken zijn gemeten tijdens de observaties en hebben betrekking op het aantal kinderen en aantal volwassenen dat deelneemt aan de activiteit waar de proceskwaliteit is gemeten. Zie Appendix C voor gegevens per meetronde.

De eerste meetronde vond plaats tijdens de COVID-pandemie; slechts een klein deel van de observaties kon plaatsvinden en deze ronde geeft waarschijnlijk geen representatief beeld van de kwaliteit van voorschoolse educatie. Wanneer we ons richten op de tweede en derde meetronde valt op dat de proceskwaliteit in de periode 2021-2023 is gestegen. Figuur 3-2 geeft de verdeling van de scores voor emotionele en educatieve kwaliteit tijdens de tweede (2021-2022) en derde (2022-2023) meetronde. Volgens de standaard CLASS classificering scoren de VE-groepen in 2021-2022 wat betreft emotionele kwaliteit in de midden-hoog range. Bijna alle scores van emotionele kwaliteit zijn boven de 4 en de meeste scores (88.0%) boven de 5, waarvan 10.3% een

score van 6 of hoger heeft. De scores voor educatieve kwaliteit zijn lager; in 34.1% van de groepen is er sprake van een lage educatieve kwaliteit en bij de meeste groepen (64.8%) ligt de score in de middenrange waarbij enkele groepen (1.1%) een score van 5 hebben. In 2022-2023 scoren de groepen wat betreft emotionele kwaliteit ook in de midden-hoog range, maar met een hoger percentage dat boven de 6 scoort. Alle scores van emotionele kwaliteit zijn boven de 4 en de meeste scores (85.2%) boven de 5, waarvan 23.2% een score van 6 of hoger heeft. De scores voor educatieve kwaliteit van de derde meetronde zijn significant hoger dan de scores van de tweede meetronde. In 5.4% van de groepen is er sprake van een lage educatieve kwaliteit, in de meeste groepen (75.3%) valt de score in de middenrange en 19.3% van de groepen heeft een score van 5 of hoger.

Figuur 3-2 Verdeling van de emotionele en educatieve kwaliteit in 2021-2022 en 2022-2023



Omdat in nagenoeg alle nationale en internationale studies de scores van educatieve kwaliteit lager liggen dan de scores voor emotionele kwaliteit, is het de vraag of voor beide domeinen dezelfde benchmark gehanteerd zou moeten worden (Leseman & Slot, 2023). Een vergelijking met andere studies gebaseerd op CLASS Toddler is daarom zinvol. In Tabel 3-3 zijn de resultaten van eerdere en recente grote landelijke onderzoeken opgenomen. Wanneer de CLASS scores van de EVENING groepen worden vergeleken met de CLASS scores uit de LKK metingen valt op dat de gemiddelde emotionele kwaliteit iets lager is in de steekproef uit 2021/2022 (en ook de scores uit 2020 liggen wat lager dan bij LKK). De gemiddelde educatieve kwaliteit is vergelijkbaar met de gemiddelde LKK scores, maar wat lager wanneer we het vergelijken met de LKK scores voor de peuteropvang. Ook bij de LKK metingen is er in de laatste meetronde een lichte stijging te zien in de emotionele kwaliteit en een sterkere stijging in de educatieve kwaliteit.

Wanneer we de scores voor educatieve kwaliteit vergelijken met internationale studies, is de score van de EVENING groepen relatief hoog. Vooral de scores van de meest recente meting voor educatieve kwaliteit liggen ruim boven die van andere metingen,

zowel in Nederland als daarbuiten. Hoewel CLASS-scores niet altijd direct vergelijkbaar zijn tussen contexten en studies, laat een vergelijking met andere stelsels zien dat de gemiddelde scores in EVENING relatief hoog zijn. Zo liggen de gemiddelde CLASS scores voor educatieve kwaliteit in een Vlaamse en Deense studie onder de 3 (Hulpia et al., 2016; Bleses et al., 2019). Relatief hoge scores voor educatieve kwaliteit worden gerapporteerd voor het Amerikaanse Early Head Start: 3.45 (Bandel et al., 2014) en 3.96 (Castle et al., 2016), en een Noorse studie met een score van 3.50 (Buøen et al., 2021); dit zijn de hoogste scores van alle internationale studies die zijn opgenomen in de LKK studie (Romijn et al., 2024). Tijdens de laatste EVENING meting scoren de VE-groepen echter nog hoger.

Tabel 3-3 Overzicht van CLASS Toddler scores in verschillende studies

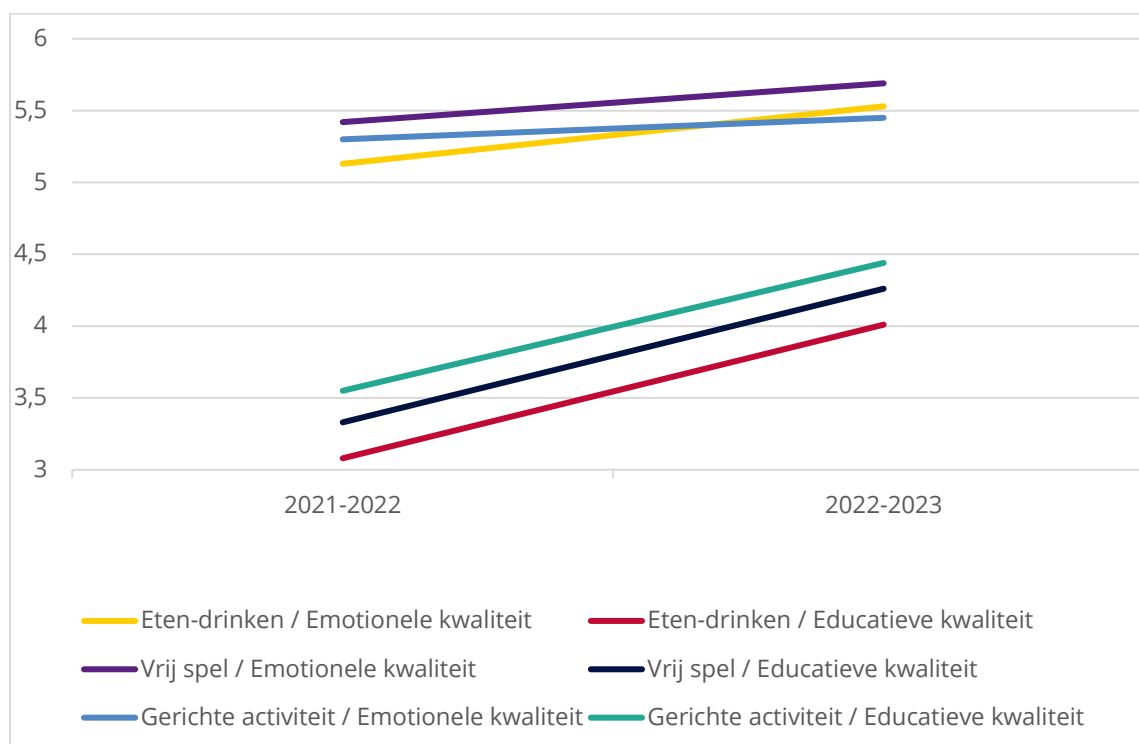
Studie	Voorziening	Jaar	N	Emotionele kwaliteit		Educatieve kwaliteit	
				M	SD	M	SD
EVENING	Voorschoolse educatie	2020	68	5.36	0.51	3.62	0.81
		2021-2022	177	5.28	0.54	3.32	0.76
		2022-2023	184	5.55	0.55	4.24	0.80
LKK (Romijn et al., 2023; Romijn et al., 2024)	Kinderdagopvang	2017-2019	93	5.71	0.52	3.40	0.82
	Peuteropvang	2017-2019	99	5.72	0.45	3.53	0.66
	Kinderdagopvang	2021-2022	40	5.82	0.46	3.78	0.79
	Peuteropvang	2021-2022	50	5.85	0.45	3.66	0.70
	Kinderdagopvang	2022-2023	47	6.02	0.50	4.15	0.96
	Peuteropvang	2022-2023	42	5.92	0.44	3.95	0.97
Pre-COOL (Leseman & Slot, 2013)	Kinderdagverblijven	2011	122	5.36	0.72	3.06	0.89
	Kinderdagverblijven	2012	97	4.97	0.95	2.63	1.01
	Peuterspeelzalen	2011	149	5.37	0.75	3.46	0.98
	Peuterspeelzalen	2012	121	5.08	0.86	3.09	0.98

Kwaliteit naar type activiteiten

Er zijn drie verschillende typen activiteiten geobserveerd met de CLASS: een eet- en drinkmoment, vrij/begeleid spel en een gerichte activiteit, in een ongeveer gelijke verhouding (zie Figuur 3-3). De gerichte activiteit is meestal een educatieve activiteit (bijvoorbeeld een activiteit uit een VVE-methode, een kringgesprek of voorlezen), maar in een aantal gevallen gaat het om een creatieve activiteit (bijvoorbeeld een knutselmoment). De figuur laat zien dat de emotionele kwaliteit het hoogst is tijdens spel, met name door een veel hogere score op de dimensie kindperspectief. In 2020-2021 en 2021-2022 wordt dit gevolgd door de gerichte activiteit; emotionele kwaliteit scoort het laagst tijdens eet- en drinkmomenten. In 2022-2023 scoort het eet- en

drinkmoment net iets hoger op de emotionele kwaliteit dan de gerichte activiteit. De educatieve kwaliteit is het hoogst bij de gerichte activiteit en het laagst bij het eet- en drinkmoment. Deze bevindingen zijn vergelijkbaar met die van de Landelijke Kwaliteitsmonitor Kinderopvang (Romijn et al., 2023; Romijn et al., 2024). Voor een beschrijving van de geobserveerde emotionele en educatieve kwaliteit per type activiteit per meetronde verwijzen we naar Appendix C.

Figuur 3-3 Geobserveerde kwaliteit naar type activiteit



Longitudinale analyses

Bij een aantal van de deelnemende EVENING-locaties ($n = 41$) is er een meting geweest op alle drie meetmomenten, namelijk in 2020-2021, in 2021-2022 en in 2022-2023. Gemiddeld genomen is de emotionele kwaliteit op meetmoment 2 ($M = 5.33$, $SD = 0.53$) ongeveer even hoog als de emotionele kwaliteit op meetmoment 1 ($M = 5.34$, $SD = 0.50$). De emotionele kwaliteit op meetmoment 3 is gemiddeld hoger dan op meetmomenten 1 en 2 ($M = 5.48$, $SD = 0.52$). Door middel van variantieanalyse voor herhaalde metingen is nagegaan of de gemiddelde emotionele kwaliteit statistisch significant verschilt tussen de jaren. De resultaten laten zien dat er geen significante toename is over de tijd ($F(2, 80) = .931$, $p = .398$).

Variantieanalyse laat zien dat er een statistisch significante toename is over de tijd van de educatieve kwaliteit ($F(2, 80) = 22.68$, $p < .001$, partial $\eta^2 = .54$). De post-hoc paarsgewijze vergelijking met Bonferroni-correctie laat zien dat de stijging van de educatieve proceskwaliteit tussen het 1e en 2e meetmoment enerzijds en het 3e meetmoment anderzijds statistisch significant is ($M_1 = 3.6$ en $M_3 = 4.3$, $p < .001$; $M_2 = 3.4$ en $M_3 = 4.3$, $p < .001$). De resultaten van de variantieanalyse wijzen dus op een significante stijging van de educatieve kwaliteit in 2022-2023 vergeleken met de

voorgaande meetmomenten (2020-2021 en 2021-2022). In hoofdstuk 7 analyseren we in hoeverre deze stijging toe te schrijven is aan de invoering van de urennorm voor de pedagogisch beleidsmedewerker.

3.3. Activiteitsaanbod

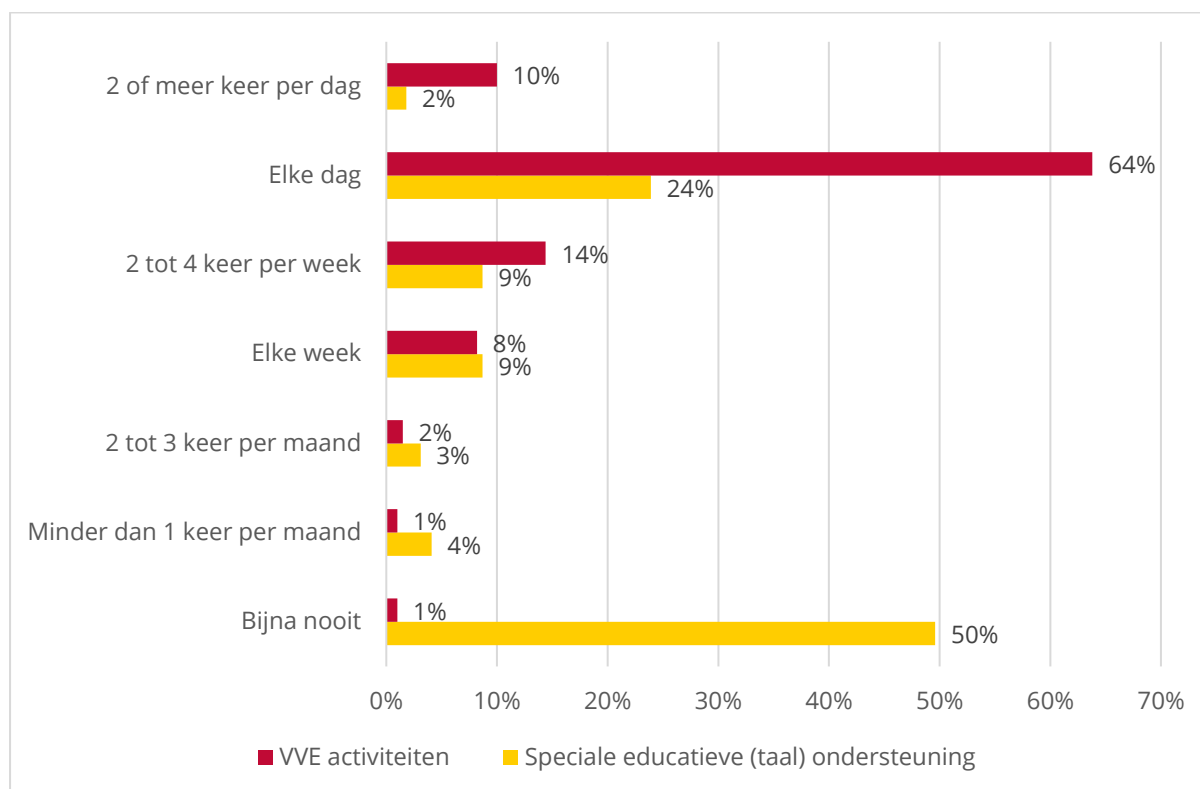
Met behulp van de vragenlijsten voor pedagogisch medewerkers is het activiteitsaanbod in kaart gebracht. Hierbij wordt onderscheid gemaakt tussen zes typen activiteiten, waarbij per type activiteit meerdere vragen over specifieke leeftijdsrelevante activiteiten worden uitgevraagd. Dit instrument is gebaseerd op het instrument dat binnen LKK wordt gebruikt, maar richt zich op activiteiten die relevant zijn voor de huidige leeftijdsgroep van 2 tot 4 jaar. In vergelijking met de LKK-instrumenten gaat EVENING ook uitgebreider in op specifieke VE-activiteiten.

De activiteiten zijn geclusterd in zes categorieën: 1) motorisch-ruimtelijk-fysiek exploratief (binnen) spel, sociaal spel en samenwerken; 2) creatief-beeldende vorming en muzikale vorming; 3) taal/geletterdheid en rekenen/wiskunde; 4) natuur, wetenschap en technologie; 5) buitenspelen en uitstapjes; 6) speciale educatieve activiteiten (VE-activiteiten). De categorieën bestaan uit verschillende items, waarbij gemeten wordt hoe vaak een bepaalde activiteit wordt aangeboden, variërend van “bijna nooit” tot “2 of meer keer per dag”. Voor de categorie “speciale educatieve activiteiten” wordt daarnaast ook uitgevraagd hoeveel tijd (in uren) per dag de VE-activiteit wordt uitgevoerd in de groep om zo een beter beeld te krijgen van de intensiteit van het VE-activiteitsaanbod. Tenslotte wordt gevraagd naar de manier waarop de VE-activiteit wordt aangeboden, bijvoorbeeld in een kleine groep of juist in de hele groep.

Allereerst is nagegaan hoe vaak kinderen een activiteit of les uit een educatieve (VVE) methode doen. Daarnaast is gevraagd hoe vaak kinderen speciale educatieve (taal) ondersteuning door een tutor of intern begeleider krijgen, bijvoorbeeld vanuit een VVE-programma. Deze gegevens zijn terug te vinden in Figuur 3-4. Van de deelnemende groepen doen ruim 74% minstens één keer per dag een activiteit of les uit een educatieve (VVE) methode. Ruim 88% van de groepen doen dit minstens 2 keer per week en 96% van de groepen doen dit minstens elke week.

Wanneer pedagogisch medewerkers rapporteerden minimaal één keer per week een VVE-activiteit of speciale taalondersteuning uitvoerden is ook gevraagd hoeveel tijd per dag aan deze activiteiten wordt besteed. We zien we dat van de 375 groepen (96% van alle groepen) die minimaal 1 keer per week VE activiteit aanbieden, 38% VVE-activiteiten gedurende minimaal 1 uur per dag aanbieden. In ongeveer 20% van de groepen is dat meer dan 2 uur per dag. Van de 168 groepen (43% van alle groepen) die minimaal 1 keer per week een speciale taalactiviteit aanbieden, biedt 33% een dergelijke activiteit gedurende meer dan 1 uur per dag aan en 16% gedurende meer dan 2 uur per dag.

Figuur 3-4 Frequentie van VE-activiteiten (alle meetronden, n=389)



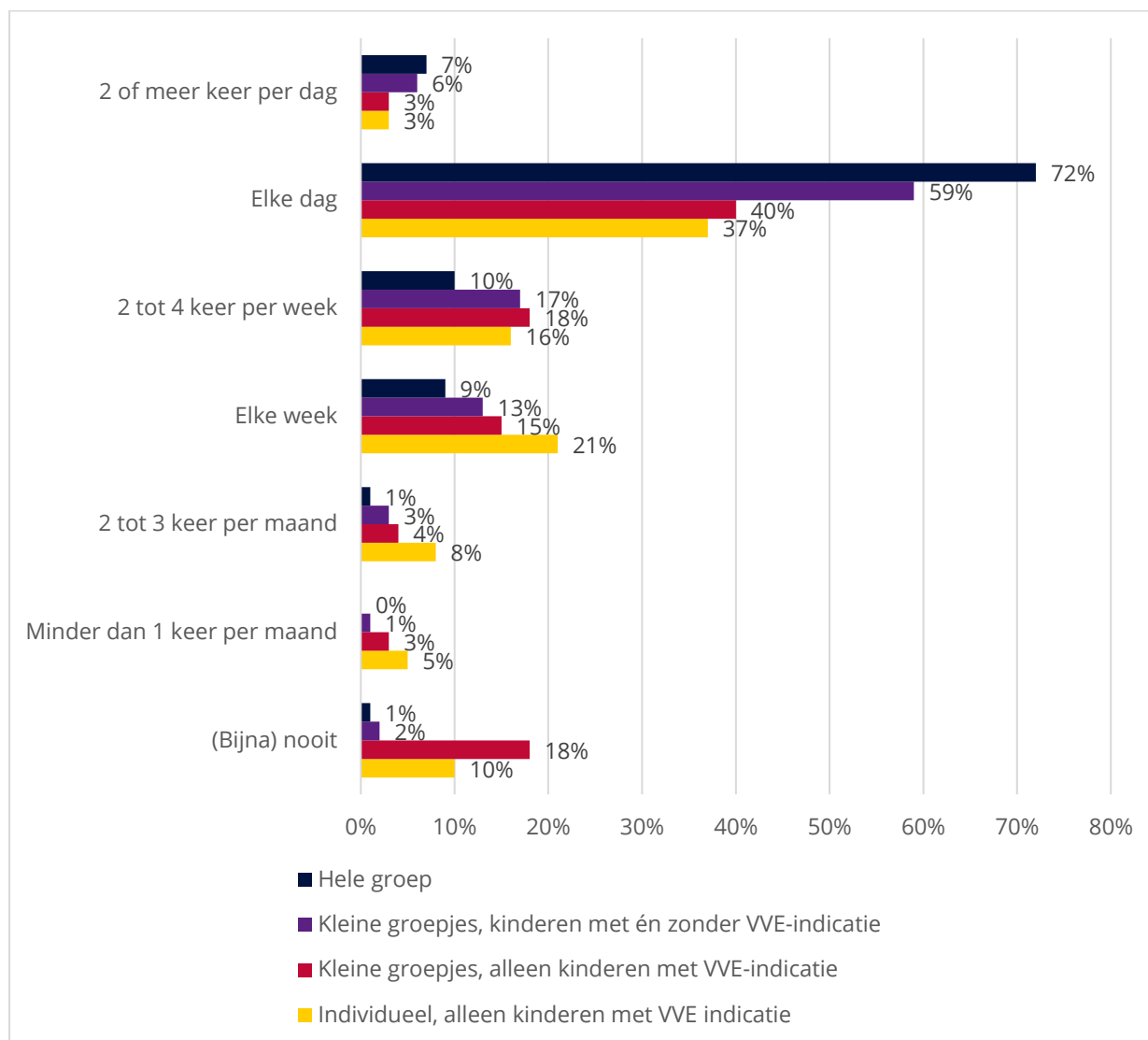
Speciale educatieve (taal) ondersteuning door een tutor of intern begeleider zijn gemiddeld genomen minder gebruikelijk. In ongeveer 44% van de groepen is dit aanbod redelijk frequent (minstens één keer per week) en in 24% van de groepen wordt dit minstens één keer per dag aangeboden. Echter, in 50% van de groepen krijgen kinderen bijna nooit speciale educatieve (taal) ondersteuning door een tutor of intern begeleider. Dit type activiteit wordt dus of relatief vaak aangeboden, of juist bijna niet. Dit is te verklaren door het feit dat de inzet van een tutor afhankelijk is van het gebruikte VVE-programma. Hier kan gemeentelijk beleid ook een rol spelen.

Als we kijken naar de VE activiteiten en de speciale educatieve (taal) ondersteuning tezamen, dan zien we dat in 88% van de groepen in ieder geval dagelijks één van deze speciale educatieve activiteiten wordt aangeboden.

VVE-activiteiten kunnen op verschillende manieren worden aangeboden. In het onderzoek is een onderscheid gemaakt tussen 1) het individueel, in kleine groepjes of in de hele groep aanbieden van de VVE-activiteiten; en 2) het aanbieden van de VVE-activiteiten aan alle kinderen of alleen aan kinderen met een VVE-indicatie. In Figuur 3-5 wordt een overzicht gegeven op welke manieren en in welke groepssamenstelling de VVE-activiteiten worden aangeboden, en wat de frequentie van deze VVE-activiteiten is. Hierbij valt op dat VVE-activiteiten niet alleen vaak in de hele groep worden aangeboden, maar ook regelmatig in kleinere groepjes of individueel worden aangeboden. VVE-activiteiten worden regelmatig op verschillende manieren aangeboden (bijvoorbeeld zowel in de hele groep als in kleinere groepen). Ook lijkt er

extra aandacht te zijn voor doelgroeppeuters, waarbij de activiteit gericht individueel of in kleine groepjes wordt aangeboden aan alleen kinderen met een VVE-indicatie.

Figuur 3-5 Groepssamenstelling en frequentie van VVE-activiteiten (alle meetronden, n=389)



Tenslotte is ook het reguliere activiteiten aanbod uitgevraagd (zie Appendix D). Hieruit blijkt dat op de voorscholen veel aandacht is voor spel en samenwerken, taal en beweging en buitenspelen, maar veel minder voor activiteiten die te maken hebben met gecijferdheid en vroege rekenvaardigheid. Ook activiteiten die te maken hebben met kennis van de wereld, natuur, wetenschap en technologie worden relatief weinig uitgevoerd. Hier lijken nog belangrijke kansen te liggen want deze typen activiteiten kunnen bijdragen aan de bredere ontwikkeling van peuters. Er is namelijk groeiend bewijs dat goed begeleide domein-specifieke activiteiten op het gebied van rekenen en geletterdheid de effectiviteit van voorschoolse programma's verhogen (Dickinson et al., 2019; Duncan et al., 2023; Bleses et al., 2021), terwijl met name activiteiten op het gebied van wetenschap en technologie vaak dialogen uitlokken met rijk academisch taalgebruik (Henrichs & Leseman, 2014).

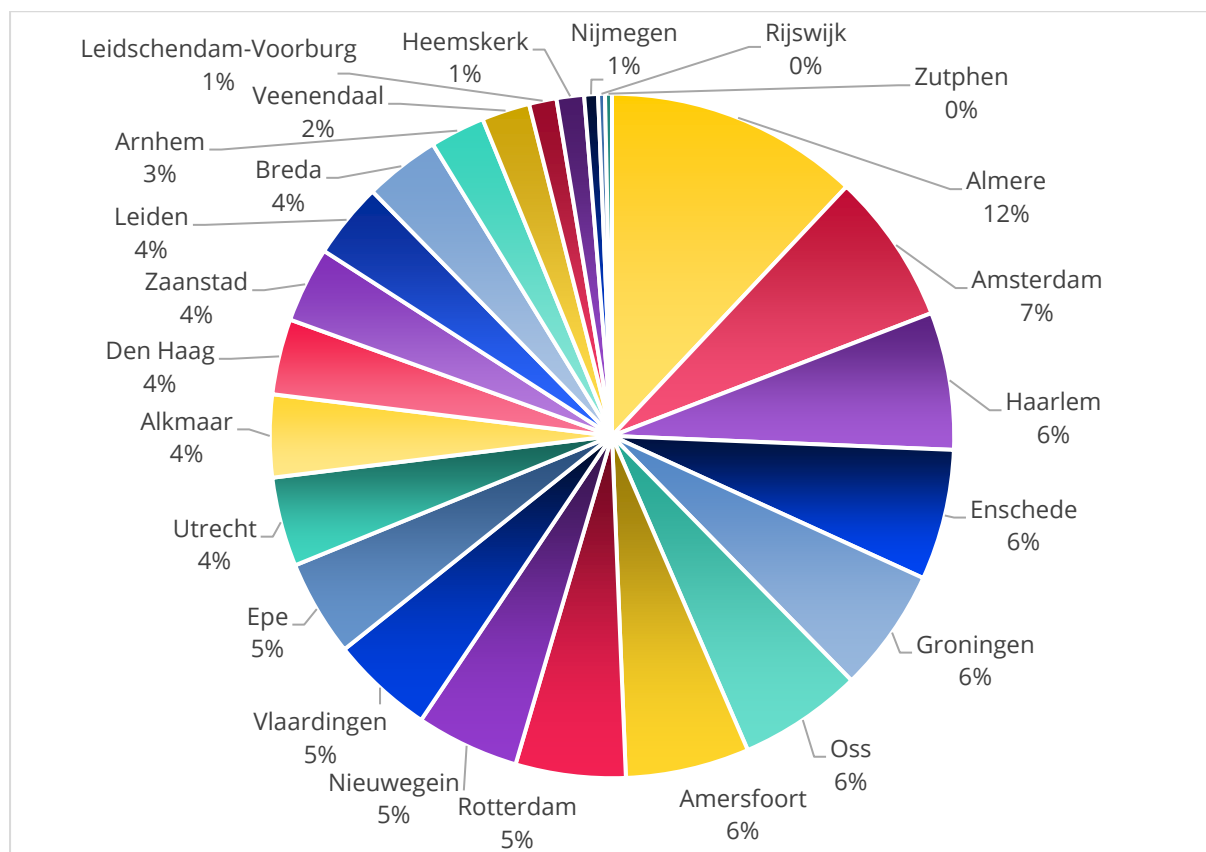
4. Gezinnen

In dit hoofdstuk worden de achtergrondkenmerken van de deelnemende gezinnen beschreven. We gaan in op de gemeenten waarin zij wonen, hun opleidingsniveau, herkomstland, thuistaal en de samenstelling van de gezinnen.

4.1. Steekproef

In totaal hebben er bij 314 gezinnen huisbezoeken plaatsgevonden in de periode december 2019 tot en met juni 2024. Van de 314 gezinnen hebben er bij 110 gezinnen twee bezoeken plaatsgevonden; één wanneer het kind ongeveer 2.5 jaar was (meestal wanneer het kind net was gestart in de voorschool), en de tweede afspraak als het kind bijna of net vier jaar oud was. Bij zes gezinnen is de oudervragenlijst bij het huisbezoek uiteindelijk niet afgenomen, waardoor in deze sectie een overzicht wordt gegeven van de achtergrond van 308 gezinnen.

Figuur 4-1 Woonplaats van de deelnemende gezinnen



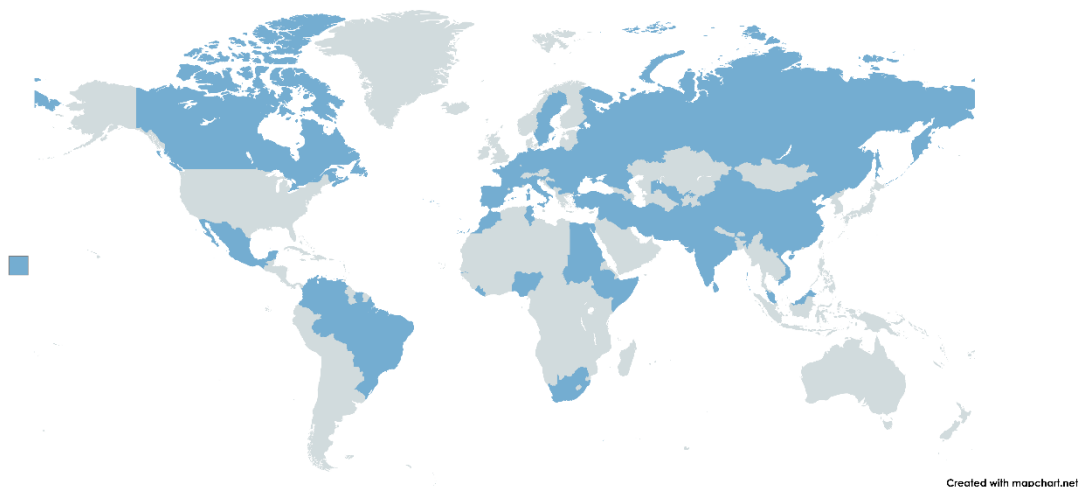
In totaal kwamen 62 gezinnen uit de G4 en 167 gezinnen uit de overige grotere gemeenten die vallen onder de G37.³ Verder waren er 79 gezinnen uit kleinere gemeenten. 171 gezinnen kwamen uit de Randstedelijke gemeenten en 137 gezinnen buiten de Randstad. Zie Figuur 4-1 voor de woonplaatsen van de deelnemende gezinnen.

4.2. Achtergrondkenmerken

In totaal hebben 253 vrouwen (82.1%) en 51 (16.6%) mannen de oudervragenlijst bij de eerste afspraak ingevuld. Bij de gezinnen waar een tweede afspraak plaatsvond, waren dit 87 vrouwen (79.1%) en 21 mannen (19.1%). De gemiddelde leeftijd van de ouder bij de eerste afspraak was 36.2 jaar, en bij de gezinnen waar de tweede afspraak plaatsvond was dit 36.6 jaar.

Van de 308 gezinnen gaven er 274 (89%) aan dat hun huishouden bestond uit twee verzorgers. In 27 gezinnen was er sprake van een eenoudergezin, waarbij meestal alleen de moeder aanwezig is in het huishouden. In 73% bestond het gezin uit meerdere kinderen.

Figuur 4-2 Land van herkomst van ouders



Van de 308 geïnterviewde ouders gaven er 163 (52.9%) aan in Nederland te zijn geboren. 140 ouders gaven aan in een ander land te zijn geboren, waarvan er 33 in een Westers land (grootste groep uit Polen, n = 6) en 107 ouders in een niet-Westers land

³ De G37 bestaan uit de G4 (Amsterdam, Rotterdam, Den Haag en Utrecht) en de 33 grotere vve-gemeenten (Eindhoven, Tilburg, Enschede, Dordrecht, Arnhem, Schiedam, Haarlem, 's-Hertogenbosch, Almere, Zaanstad, Breda, Helmond, Nijmegen, Amersfoort, Leiden, Venlo, Almelo, Ede, Apeldoorn, Deventer, Maastricht, Groningen, Lelystad, Heerlen, Delft, Emmen, Alkmaar, Zoetermeer, Sittard-Geleen, Zwolle, Leeuwarden, Hengelo en Haarlemmermeer).

(grootste groepen uit Syrië [n=18] en Turkije [n = 11]). Zie Figuur 4-2 voor de landen waar de geïnterviewde ouders geboren zijn, indien dit niet Nederland was. De ouders zijn tussen 1974 en 2021 naar Nederland verhuisd.

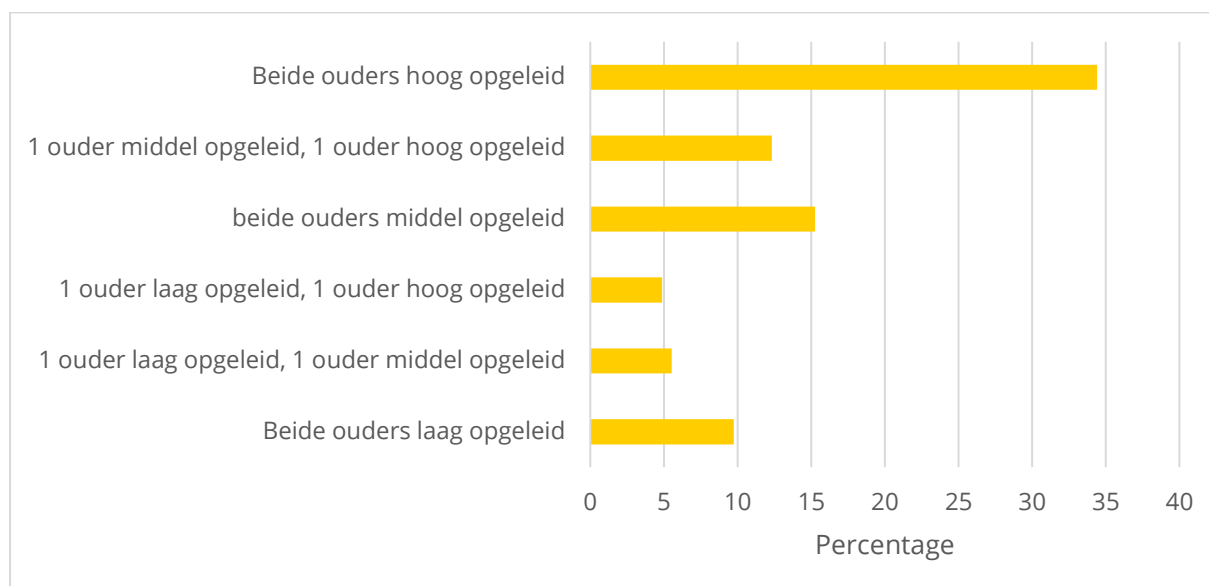
Ouders zijn gevraagd naar hun opleidingsniveau. Niet alle ouders wilden of konden deze vragen beantwoorden (ontbrekende gegevens voor ongeveer 10% van de steekproef). Hierdoor dienen deze resultaten met enige voorzichtigheid te worden geïnterpreteerd. In Figuur 4-3 zijn het opleidingsniveau van de geïnterviewde ouder en dat van de andere ouder weergegeven indien dit van toepassing was. In ongeveer 20% van de gezinnen is tenminste een van de ouders laag opgeleid. Omdat een groot deel van de ouders een migratieachtergrond heeft (zie hieronder) en een opleiding heeft gevolgd in het buitenland is het opleidingsniveau vastgesteld aan de hand van de internationale ISCED indeling. Hierdoor werd ook het diploma van de ouder in het land van herkomst meegenomen en gewogen.⁴ Het valt op dat het aantal gemiddeld en hoger opgeleide ouders in onze steekproef relatief groot is gegeven de onderzochte populatie. Dat kan door verschillende factoren verklaard worden. Ten eerste nemen ook niet-doelgroeppeuters deel aan het onderzoek. Daarnaast kan ook de gemeentelijke doelgroepdefinitie een rol spelen. Als bijvoorbeeld een andere thuistaal ongeacht het opleidingsniveau van de ouders tot een VVE-indicatie leidt, kunnen ook kinderen van hoogopgeleide ouders deelnemen aan VE. Ten slotte kan de vergelijkbaarheid van opleidingen in Nederland versus andere landen een rol spelen: het is mogelijk dat het niveau van (een deel van) de opleidingen gevolgd in het buitenland wordt overschat.

In veel onderzoek naar de ontwikkeling van kinderen wordt er specifiek gekeken naar het opleidingsniveau van de moeder. Als we specifiek naar moeders kijken die hun opleiding in Nederland hebben behaald (n = 182), zien we dat 22.5% van de moeders een laag opleidingsniveau heeft, 32.4% een gemiddeld opleidingsniveau, en 45.1% een hoog opleidingsniveau. In de hele groep, inclusief moeders die hun opleiding in het buitenland hebben behaald (n = 295) zien we 18.6% met een laag opleidingsniveau, 29.2% met een gemiddeld opleidingsniveau en 52.2% met een hoog opleidingsniveau. Zoals hierboven beschreven, is het mogelijk dat het niveau van opleiding in het buitenland wordt overschat. In lijn met de literatuur maken we bij de analyses verder in dit rapport ook gebruik van het opleidingsniveau van de moeder.

Ook hebben we ook gevraagd of de ouders een betaalde baan hadden; ruim 55.5% van de ouders gaven aan een betaalde baan te hebben. Indien ouders betaald werk hadden was dit gemiddeld 27 uur per week.

⁴ Ouders met een migratieachtergrond kregen de mogelijkheid om hun afgeronde opleiding (behaald in het land van herkomst) zelf te omschrijven.

Figuur 4-3 Opleidingsniveau van ouders



4.3. Taalomgeving

Bij 205 (68.6%) gezinnen wordt er door tenminste één ouder (onder andere) Nederlands met het kind gesproken, 94 (31.4%) ouders gaven aan dat beide ouders uitsluitend een andere taal praten met het kind. Er is vaak sprake is van een meertalige thuissituatie, zoals verwacht kan worden gegeven de doelgroepbevolking. Bij 164 gezinnen (53.2%) wordt er meer dan één taal door één of beide ouders thuis gesproken met het kind. Daarnaast hebben we ook gevraagd of het kind in de naaste omgeving ook nog een andere taal hoort (zoals door grootouders, op de voorschool, oppas etc.). Daaruit blijkt dat 69.5 % van de kinderen (214 kinderen) in onze steekproef opgroeit in een meertalige omgeving.

We hebben ook gevraagd naar de ervaren Nederlandse taalvaardigheid van de ouders, op drie taaldomeinen; de Nederlandse taal begrijpen, spreken of lezen. Ongeveer 9% van de ouders met een niet-Nederlandse achtergrond gaf aan vaak tot altijd moeite te hebben met al deze drie taaldomeinen, of gaf aan helemaal geen Nederlands te begrijpen, spreken of lezen. Ongeveer 17.5% van onze steekproef gaf aan moeite te hebben met minimaal één van de drie domeinen.

5. Kinderen

In dit hoofdstuk beschrijven we de kenmerken van de kinderen die aan het onderzoek hebben meegedaan (voor details over de gezinnen, zie hoofdstuk 4). Na het beschrijven van algemene kenmerken, beschrijven we de meetinstrumenten die we hebben gebruikt om de ontwikkeling van de kinderen in kaart te brengen. Vervolgens rapporteren we over de analyse van de betrouwbaarheid en validiteit van de metingen en hoe we tot een totaalscore komen. Dit doen we door het beschrijven van de vastgestelde meetinvariantie en schalen van scores van tests en vragenlijsten. Vervolgens beschrijven we de eigenschappen van de geschaalde score en beschrijven we het construeren van overkoepelende domeinscores en de eigenschappen van deze scores. Tenslotte laten we zien hoe deze scores samenhangen met een aantal achtergrondkenmerken.

5.1. Kenmerken van de kinderen

Er hebben in totaal 314 gezinnen deelgenomen aan het onderzoek. Van de kinderen die rond 2,5 jaar getest zijn hebben 110 kinderen voor een tweede keer deelgenomen rond de leeftijd van 3 jaar en 10 maanden. De overige kinderen (N=204) hebben één keer meegedaan; zie Tabel 5-1 voor de achtergrondgegevens van de deelnemende kinderen. De totale steekproef bevat daarmee gegevens over 135 kinderen die getest zijn rond de leeftijd van 2,5 jaar en 274 kinderen die zijn getest rond de leeftijd van 3 jaar en 10 maanden. Zie Figuur 5-1 voor een overzicht van de leeftijd van de deelnemende kinderen.

Tabel 5-1 Achtergrondkenmerken van de deelnemende kinderen

	Hele steekproef	Jongere groep	Oudere groep
Aantal kinderen	314	138	286
Leeftijd in maanden (SD)	-	3267 (2.57)	46.33 (1.60)
Meisje	51.4%	52.9%	50.7%
Meertalige kinderen	68.2%	69.6%	67.5%
VVE- indicatie	65.8%	64.0%	65.5%

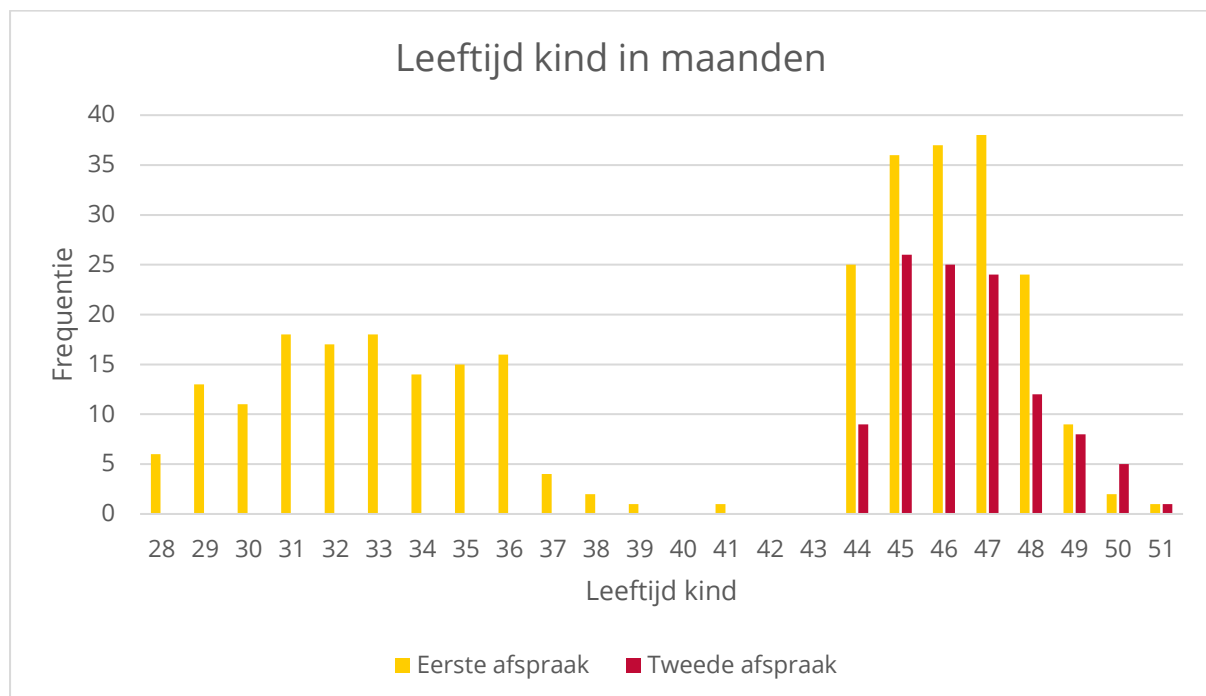
Noot. Onder hele steekproef wordt ieder kind maar één keer meegenomen. De steekproeven van de jongere groep en oudere groep overlappen voor een deels omdat een aantal kinderen twee keer hebben meegedaan. VE indicatie is bepaald op basis van rapportages van ouders en pedagogische medewerkers.

Voor het bepalen van de doelgroepstatus maken we gebruik van items over het kind die zijn beantwoord door pedagogisch medewerkers en ouders. Voor bijna de helft van de kinderen hebben pedagogisch medewerkers aangegeven of het kind een VVE-indicatie heeft: volgens die gegevens heeft 65 procent van de deelnemende kinderen een VVE-indicatie. Missende gegevens hebben we aangevuld door gebruik te maken van informatie uit de oudervragenlijst, waarbij we gebruik maken van informatie over de VVE-indicatie, het opleidingsniveau van de ouders, migratieachtergrond en de taalomgeving thuis. Op deze manier kan voor alle kinderen de doelgroepstatus bepaald worden. Deze indicator geeft aan dat bijna 66 procent van de kinderen een VVE-

indicatie heeft; dit percentage komt overeen met het percentage dat enkel gebaseerd is op de vragenlijst die pedagogisch medewerkers hebben ingevuld.

Ongeveer 89% van de ouders rapporteerden dat er geen bijzonderheden waren omtrent de ontwikkeling van hun kind. De overige 11% meldden vooral taalproblemen (4%), gehoorproblemen (3%), diverse medische problemen (3%) en motorische problemen (2%). De meeste kinderen zijn geboren na 37 weken zwangerschap; 6% is geboren vóór 37 weken en 2% is geboren vóór 34 weken zwangerschap. De leeftijd waarop de kinderen hebben leren lopen varieert van 10 tot 20 maanden. Dit is binnen de normale bandbreedte voor Nederlandse kinderen (van Baar et al., 2014).

Figuur 5-1 Overzicht van de leeftijd van de deelnemende kinderen (in maanden)



5.2. Metingen van de ontwikkeling

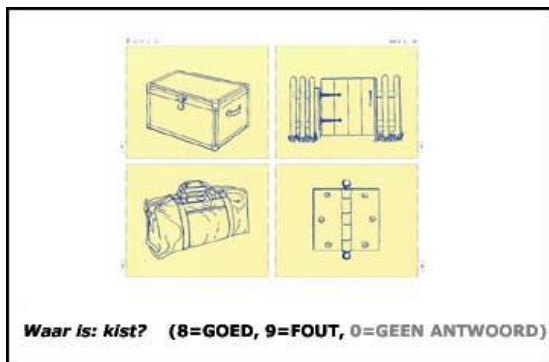
Om de ontwikkeling van kinderen te meten hebben we een set van tests en vragenlijsten gebruikt die grotendeels gelijk is aan de set die gebruikt is in het pre-COOL onderzoek. In Appendix A geven we een overzicht van alle onderdelen van deze set van speelse taakjes (of testen) en vragenlijsten. De taakjes en vragenlijsten zijn verdeeld naar de domeinen: taal; cognitieve of 'cool' executieve functies (EF); affectieve of 'hot' executieve functies; sociaal-emotioneel functioneren; en motorische ontwikkeling. Hieronder beschrijven we alle instrumenten per domein.

Taaltests

We hebben drie tests gebruikt voor verschillende aspecten van taal: een test voor receptieve woordenschat, een test voor productieve woordenschat (alleen afgenomen bij de kinderen die bijna 4 jaar zijn) en een test voor de herkenning van klanken.

Om begrip van taal (receptieve woordenschat) te meten gebruiken we een verkorte vorm van de Nederlandse versie van de Peabody Picture Vocabulary Test (PPVT-III-NL, Dunn et al., 2005). Kinderen zien in deze test vier afbeeldingen en horen een woord. De kinderen worden gevraagd naar het plaatje te wijzen dat bij het woord hoort. We gebruikten de verkorte versie van deze test die voor pre-COOL is ontwikkeld (Mulder et al., 2014; Veen et al., 2012). Deze versie omvat 24 items en 1 oefenitem. Zie Figuur 5-2A voor een voorbeeld van een item.

Figuur 5-2 Voorbeelden van items uit de taaltestjes



A. PPVT



B. CELF-Preschool



C. Foneemidentificatietaak

Om taalproductie te meten is gebruik gemaakt van de productieve woordenschat subtest uit de Nederlandse vertaling van de Clinical Evaluation of Language Preschool versie (CELF-Preschool 2 NL; Wiig et al., 2012; Nederlandse vertaling door De Jong en Pearson). In deze test zien kinderen een plaatje en krijgen de vraag 'wat is dat?' of 'wat doet het kindje?'. Kinderen krijgen 2 punten voor een correct antwoord en nul punten voor een incorrect antwoord of geen antwoord. Bij sommige items kunnen kinderen 1 punt scoren als ze een antwoord geven dat semantisch correct is, maar grammaticaal niet passend. Bijvoorbeeld, bij het item in Figuur 5-2B wordt er gevraagd 'wat doet dit meisje?'. Het antwoord 'fietsen' of 'zij fietst' krijgt 2 punten, maar het antwoord 'fietst' krijgt 1 punt. De test bevat 20 items en 3 oefenitems.

De foneemidentificatietaak is gebruikt om waarneming van spraakklanken te meten (Gerrits, 2003; Kuijpers, 1996). Deze taak is ontwikkeld voor het pre-COOL onderzoek (Mulder et al., 2014; Veen et al., 2012). Kinderen zagen twee plaatjes op een laptop van woorden die sterk op elkaar lijken, bijvoorbeeld een beer en een peer (zie Figuur 5-2C). Het kind hoort één van de twee woorden en moet het juiste plaatje aanwijzen. Kinderen kregen een punt voor ieder goed aangewezen plaatje. De taak omvat 12 items en 2 oefenitems.

Cool Executieve Functies

Drie taakjes werden gebruikt om de 'cool' executieve functies te meten. Dit zijn vaardigheden van hogere orde, waarbij er geen sterke affectieve component aanwezig is. Dat wil zeggen, de kinderen kunnen geen beloning of prijs verdienen die de emotionele motivatie van kinderen om de taak goed te doen kan verhogen.

Met een nonwoordrepetitietaak (NWR) werd de capaciteit van het korte-termijn verbaal geheugen van kinderen gemeten. Bij deze taak moeten kinderen niet-bestaande nonwoorden van één tot drie lettergrepen zo goed mogelijk nazeggen, wat een indicatie geeft van de geheugencapaciteit. Een soortgelijke taak is gebruikt in pre-COOL (Mulder et al., 2014; Veen et al., 2012). De pre-COOL versie bevatte nonwoorden die erg veel op het Nederlands lijken. Hierdoor was de taak moeilijker voor kinderen met een andere thuistaal dan het Nederlands. Daarom zijn een aantal woorden uit het pre-COOL taak vervangen door woorden afkomstig uit een quasi-universele nonwoordrepetitietaak (Chiat, 2015). De quasi-universele taak gebruikt alleen combinaties van fonemen die in veel verschillende talen voorkomen. Daarom zou de moedertaal van het kind geen invloed moeten hebben op de prestaties (Boerma et al., 2015). Kinderen kregen een punt voor ieder goed herhaalde woord. De taak bevat 12 items en 2 oefenitems.

Voor het meten van de selectieve aandacht, het vermogen om relevante informatie te selecteren en irrelevante informatie te negeren is een taak ontwikkeld die ook voor pre-COOL is gebruikt (Mulder et al., 2014; Veen et al., 2012). Dit is een visuele zoektaak waarbij kinderen zo snel mogelijk alle plaatjes van een olifantje aan moeten wijzen. Op het scherm zien ze meerdere plaatjes van dieren, die allemaal even groot zijn en in dezelfde kleur zijn afgebeeld. Kinderen krijgen 40 seconden per item om alle olifantjes aan te wijzen. Hiervoor moeten ze dus de relevante informatie (olifantjes) selecteren en de irrelevante informatie (andere dieren) negeren. De test bevat twee items met een doel/afleider ratio van 1:5, 1 item met een ratio van 1:8 en een item dat veel moeilijker was met ratio van 2:51, waarbij de plaatjes een stuk kleiner waren afgebeeld. Voordat de test items werden afgenomen, konden kinderen oefenen met het zoeken naar olifantjes en voor het laatste en moeilijkste item mochten ze nog een keer oefenen met kleinere plaatjes.

Om de capaciteit van het visueel-ruimtelijk werkgeheugen te meten, is de doosjes taak gebruikt. Deze taak is ontwikkeld voor pre-COOL (Mulder et al., 2014; Veen et al., 2012) en is gebaseerd op de 'six boxes task' van Diamond et al. (1997). Kinderen konden zien hoe tussen de 1 en 3 speelgoedjes onder 1 tot 3 omgekeerde en ondoorzichtige plastic bakjes van de in totaal 6 bakjes werden verstopt. Vervolgens werd er gedurende 3

seconden een scherm geplaatst tussen het kind en het plateau met de bakjes. Na 3 seconden mochten kinderen één voor één de bakjes optillen om de speelgoedjes te zoeken. Als er een stuk speelgoed was gevonden, werd het lege bakje weer omgekeerd teruggezet. Na opnieuw een afleiding van 3 seconden mochten de kinderen een ander stuk speelgoed zoeken. Kinderen moesten dus onthouden waar het speelgoed was verstopt, welke speelgoedjes zij al hadden gevonden, onder welk bakje, en zij moesten deze informatie gedurende de taak blijven bijhouden ('updating')n. De taak bevatte in totaal 5 testitems. Voor de eerste twee items werd telkens 1 stuk speelgoed verstopt onder een van de 6 bakjes. Voor het derde en vierde testitem werden, respectievelijk, 2 en 3 stukken speelgoed verstopt onder 2 respectievelijk 3 van de 6 tot 8 bakjes. De plaats van de bakjes op het plateau en de volgorde waarin en de bakjes waaronder de speelgoedjes werden verstopt, waren hetzelfde voor alle kinderen. Kinderen kregen een punt als ze alle stukken speelgoed terug hadden gevonden binnen een aantal zoekpogingen dat gelijk is aan het aantal verstopte stukken.

Hot Executieve Functies

In tegenstelling tot de cool executieve functies, is er bij 'hot' executieve functies een relatief sterke affectieve of emotionele component betrokken: een verwachte beloning die het verlangen en de motivatie van het kind beïnvloedt. In EVENING is de hot executieve functie, net als in pre-COOL, gemeten middels twee, zogenaamde 'uitstel van beloning' taken, kortweg wachttaken. Dit zijn taken die zelfcontrole, ook wel bewuste controle genoemd, meten. Zelfcontrole is het vermogen om een automatische of dominante respons te onderdrukken en een subdominante respons uit te voeren (Kochanska et al., 2000). Dit vermogen is van belang voor adaptief gedrag (de Ridder et al., 2006). Wij gebruikten de twee wachttaken die ook gebruikt zijn in het pre-COOL onderzoek (Mulder et al., 2014). Bij de eerste wachttak kregen kinderen een doosje rozijnen te zien. Vervolgens werd aan hen gevraagd om te wachten en te proberen om het doosje niet aan te raken terwijl de onderzoeker even iets anders ging doen. Bij de tweede taak werd hetzelfde gedaan, maar nu met een ingepakt cadeautje met een krullerig lint er om heen. Tijdens het wachten zat de onderzoeker schuin achter het kind om het gedrag te observeren, zonder oogcontact met het kind te maken. De onderzoeker scoorde of het kind het doosje rozijnen of cadeau aangeraakte. In EVENING is een kortere versie van de taak gebruikt, met een maximale wachttijd van 30 seconden, dan bij pre-COOL, waar de wachttijd 1 minuut was. Uit pre-COOL was gebleken dat als kinderen het voorwerp aanraakten, dat bijna altijd gebeurde tijdens de eerste 30 seconden.

Motorische ontwikkeling

Om de motorische ontwikkeling te meten zijn items uit de Ages and Stages Questionnaire (ASQ) bij de kinderen afgenomen (Squires et al., 1999; Nederlandse vertaling door van Baar et al., 2014). Kinderen werden gevraagd om 6 fijn-motorische activiteiten te doen (bijvoorbeeld een cirkel tekenen) en 4 grof-motorische. De items waren geselecteerd uit de deelverzameling van items die horen bij kinderen in de leeftijd van 2.5 tot 4 jaar. In de analyses zijn alleen de items voor fijne motoriek

opgenomen. De items voor grove motoriek lieten weinig variatie zien (deze vaardigheden werden door meer dan 80% van de kinderen beheerst).

Sociaal-emotioneel functioneren

Sociaal-emotioneel functioneren van de kinderen is gemeten middels vragenlijsten die ingevuld werden door zowel ouders als pedagogisch medewerkers. Deze vragenlijsten meten zelfregulatie (bestaande uit impulscontrole en volgehouden aandacht), externaliserend probleemgedrag en prosociaal gedrag.

Impulscontrole en volgehouden aandacht zijn gemeten middels vragen uit de temperamentvragenlijsten Early Childhood Behavior Questionnaire (ECBQ; Putnam et al., 2006) en Child Behavior Questionnaire (CBQ; Rothbart et al., 2001). Ouders en pedagogisch medewerkers kregen beschrijvingen van kindgedragingen voorgelegd en beoordeelden op een 7-punt schaal in hoeverre deze beschrijving van toepassing was op hun kind (1=helemaal niet van toepassing, 7=volledig van toepassing). In de vragenlijst voor ouders werden 6 items gebruikt, in de vragenlijst voor pedagogisch medewerkers 5 items om impulscontrole te meten. Het zesde item in de ouderlijst beschrijft een kind dat uit zichzelf, of als ouders er om vragen, makkelijk alles in gereedheid brengt om van huis te gaan. Dit item is minder van toepassing in kindcentra. Voor het meten van volgehouden aandacht zijn 6 items met gedragsbeschrijvingen voorgelegd aan zowel ouders als pedagogisch medewerkers.

Externaliserend probleemgedrag en prosociaal gedrag zijn gemeten middels de Strengths and Difficulties Questionnaire (SDQ; Goodman et al., 2010; Van Widenfelt et al., 2003). Ouders en pedagogisch medewerkers krijgen een aantal beschrijvingen van kinderen voorgelegd, zoals 'heeft tenminste één goede vriend of vriendin'. Per beschrijving geven ze aan in hoeverre dat past bij het kind. Dit doen ze op een drie-punt schaal variërend tussen 'niet waar', 'een beetje waar', 'zeker waar'. De schaal externaliserend probleemgedrag bevat 10 items die gaan over hyperactiviteit en gedragsproblemen als agressief gedrag. De schaal prosociaal gedrag bevat 5 items. De SDQ bevat ook een schaal over internaliserend probleemgedrag. Deze schaal hebben we niet gebruikt, omdat dit type problemen nog weinig voorkomt in de leeftijdsgroep van EVENING (Ezpeleta et al., 2013). De schaal externaliserend probleemgedrag is zo gecodeerd dat een hogere score correspondeert met minder probleemgedrag.

Procedure en betrokkenheid

De tests zijn afgenomen tijdens huisbezoeken aan de gezinnen volgens een vaste volgorde. De taaltoetsen, de nonwoordrepetitie- en de selectieve aandachtstaak zijn afgenomen met behulp van een laptop computer. De vragenlijsten voor de ouders zijn afgenomen in een gestructureerd interview, waarbij de onderzoeksassistenten konden helpen als het niveau van Nederlands bij de ouders niet toereikend was. Voor zover mogelijk hebben we onderzoeksassistenten ingezet die de thuistaal van de gezinnen ook spreken. De pedagogisch medewerkers hebben de vragenlijsten online ingevuld via een persoonlijke link.

Na afloop van de afname van iedere taak heeft de assistent de betrokkenheid van het kind beoordeeld op een 7-punt schaal, waarbij 1 staat voor 'het kind was niet betrokken bij de taak', 4 voor 'het kind was gematigd betrokken bij de taak of wisselend' en 7 voor 'het kind was enorm betrokken gedurende de hele taak'. De data laat zien dat de meerderheid van de kinderen betrokken waren bij de taken. Gemiddeld genomen kregen 70% van de kinderen de scores van 6 of 7 op de schaal en 80% van de kinderen scoorde minimaal een 4. De laagste betrokkenheid werd gerapporteerd voor de CELF (67% scoort 6 of 7 en 80% scoorde boven de 4) en de hoogste voor klankentaak (88% scoorde 6 of 7 en ca. 91% boven de 4).

5.3. Meetinvariantie en berekenen van geschaalde scores

In het EVENING onderzoek zijn dezelfde meetinstrumenten gebruikt voor kinderen van verschillende leeftijden, geslacht en meertaligheidstatus. Daarbij hebben sommige kinderen de toetsen één keer gemaakt en anderen twee keer. Om de scores te kunnen gebruiken in analyses is het belangrijk om eerst de meetinvariantie te toetsen. Dat wil zeggen: er moet vastgesteld worden dat de toetsen gelijk zijn voor kinderen van verschillende leeftijdsgroepen, geslachten, meertaligheidsgroepen en voor kinderen die één of twee keer hebben meegedaan (Agostinelli & Wiswall, 2025; Milsap, 2011; Leitgöb et al., 2024).

De meetinvariantie is getoetst middels een confirmatieve multigroep factoranalyse. In deze analyse wordt nagegaan of alle items uit de test of vragenlijst goed genoeg passen bij de totaalscore. Zo nodig kunnen er items eruit gehaald worden die minder goed hebben gewerkt. In Tabel 5-2 is een overzicht van het aantal items of vragen die in dit proces verwijderd zijn per toets en vragenlijst. De verwijderde items zijn bijvoorbeeld items die verwarrend bleken. Dit zijn bijvoorbeeld vragen die negatief geformuleerd zijn of testitems waarbij het plaatje onduidelijk bleek voor kinderen. Middels de analyse en na het verwijderen van sommige items en vragen is gebleken dat alle toetsen en vragenlijsten vergelijkbaar waren tussen leeftijdsgroepen, geslachtgroepen, meertaligheidsgroepen en voor kinderen die één of twee keer hebben meegedaan. De volledige details van de meetinvariantie analyse staan beschreven in Appendix E.

Tabel 5-2 Aantal verwijderde items per taak en vragenlijst

	Aantal items verwijderd	Percentage missende items /missende totaal scores
Klankentaak	6 van 12	9% / 4%%
PPVT receptieve woordenschat	2 van 24	9% / <1%
CELF expressieve woordenschat	6 van 20	59% / 1%
NWR nonwoord repetitie	1 van 12	26% / 6%
Selectieve aandacht	0 van 4	9% / 4%
Bakjestaak ruimtelijk werkgeheugen	2 van 5	7% / 5%
Wachttaak rozijn	0 van 1	<1% totaal
Wachttaak cadeau	0 van 1	<1% totaal

Motoriek	0 van 6	<1% / <1%
SDQ externalizerend gedrag - ouder	0 van 10	3% / 3%
SDQ prosociaal gedrag - ouder	1 van 5	3% / 3%
SDQ externalizerend gedrag - pm'er	3 van 10	36% / 36%
SDQ prosociaal gedrag - pm'er	0 van 5	36% / 36%
ECQB impuls controle - ouder	1 van 6	2% / 2%
ECBQ volgehouden aandacht - ouder	1 van 6	2% / 2%
ECQB impuls controle - pm'er	0 van 5	36% / 36%
ECBQ volgehouden aandacht - pm'er	0 van 6	36% / 36%

Noot. Als het percentage missende items en missende totaal scores gelijk is betekent dat bij de deelnemers waar de data ontbreken, de gehele test of vragenlijst ontbreken. Bij de wachttaken is er maar één item per taak; daarom wordt hier alleen het totale percentage missende waardes weergegeven.

Tabel 5-2 laat ook de percentages missende waarden zien per toets of vragenlijst. Over het algemeen zijn er weinig missende waarden. De meest voorkomende redenen voor het ontbreken van scores bij de testen waren taalproblemen (bij alle taken behalve de aandachttaak, en in minder mate de taak die motoriek meet), gedrag van het kind zoals huilen of weglopen (bij alle taken maar meer bij de motoriektaak en wachttaken). In enkele gevallen missen scores door storende factoren in de omgeving of door technische problemen.

Bij het berekenen van de totaalscores is rekening gehouden met deze missende waarden. Bij toetsen die taalvaardigheden meten (klankentaak, PPVT en CELF) komen de missende waarden vooral voor bij kinderen die een nog relatief lage taalvaardigheid in het Nederlands hebben (ca. 2-4% van de kinderen). Hierdoor kan het zijn dat ze niet altijd antwoord geven tijdens deze toetsen. Daarom zijn hier de ontbrekende waardes beschouwd als nul-scores bij het berekenen van een totaalscore. Voor deze taken ontbreekt een totaalscore alleen bij de kinderen waarbij de afname helemaal niet plaats heeft gevonden. Bij de CELF valt op dat er soms relatief veel missende waarden zijn voor afzonderlijke items maar niet voor andere items in de schaal, waarbij we de ontbrekende waarden konden imputeren en een totaalscore konden berekenen. In deze taak zien kinderen plaatjes en moeten het woord wat daarbij hoort hardop zeggen. Hierbij hebben veel kinderen niets gezegd, waarschijnlijk omdat ze het antwoord niet weten. Dit is dan een missende waarde (omdat kinderen niet zeggen 'weet ik niet'), maar kan wel beschouwd worden als een nul-score, omdat deze kinderen bij de meeste andere plaatjes wel antwoord hebben gegeven.

Bij toetsen die geen taalvaardigheden meten kunnen we niet goed inschatten of er ontbrekende waarden zijn omdat een kind het antwoord niet weet, of omdat ze de taak niet hebben begrepen. Dit komt omdat deze taken wel wat (Nederlandse) uitleg vragen, maar niet zijn bedoeld om taalvaardigheid te meten. Bij deze taken is ervoor gekozen om een totaalscore te berekenen door de som van de aanwezige items te delen door het aantal valide items, mits minstens de helft van de items een valide waarde had. Hier betekent het ontbreken van een totaalscore dat de taak helemaal niet is afgenomen of dat bij meer dan de helft van de items de scores ontbreken.

Op basis van deze gegevens kan er geconcludeerd worden dat de geschaalde scores een betrouwbare meting geven van de concepten en dat de meetkwaliteit niet verschilt op basis van achtergrondkenmerken. Om de validiteit van de testen wordt er in de volgende deel gekeken naar correlaties tussen testen en tussen testen en achtergrondkenmerken.

5.4. Kinduitkomsten

Scores

Tabel 5-3 presenteert beschrijvende statistieken van de scores op de verschillende taakjes en vragenlijsten. Voor de scores zijn gewogen gemiddelden berekend zoals hierboven beschreven. Dit houdt in dat de schaal van de scores aangepast is, waardoor de gemiddelde scores niet inhoudelijk geïnterpreteerd kunnen worden. De spreiding van alle scores lijkt in orde. Er is voldoende variantie om de uitkomsten te kunnen voorspellen en er zijn geen duidelijke bodem- of plafond-effecten zichtbaar.

Tabel 5-3 Totaalscores op taakjes en vragenlijsten

	Jonge groep (N=142)			Oude groep (N=286)		
	N	Gemiddelde (SD)	Range	N	Gemiddelde (SD)	Range
Klankentaak	128	0.30 (0.13)	0.00-0.48	278	0.35 (0.12)	0.00-0.48
PPVT	137	0.21 (0.13)	0.00-0.50	286	0.35 (0.13)	0.00-0.55
CELF	-	-	-	283	0.32 (0.22)	0.00-.076
NWR	97	0.33 (0.18)	0.00-0.66	258	0.45 (0.16)	0.00-0.68
Selectieve aandacht	126	2.57 (1.07)	0.13-5.02	280	3.78 (0.84)	0.33-6.11
Bakjestaak	128	0.32 (0.09)	0.09-0.52	247	0.38 (0.10)	0.13-0.58
Wachttaak (rozijn)	137	60.6% goed	-	284	78.9% goed	-
Wachttaak (cadeau)	136	56.6% goed	-	284	76.4% goed	-
Motoriektaak	142	0.35 (0.20)	0.00-0.66	275	0.56 (0.12)	0.00-0.66
SDQ externalizerend gedrag (ouder)	135	0.71 (0.15).	0.18-0.97	277	0.72 (0.17)	0.13-0.97
SDQ prosociaal gedrag (ouder)	135	1.15 (0.35)	0.00-1.47	277	1.22 (0.30)	0.00-1.47
SDQ externalizerend gedrag (pm'er)	109	1.01 (0.41)	0.10-1.54	164	1.19 (0.34)	0.20-1.54
SDQ prosociaal gedrag (pm'er)	109	0.76 (0.37)	0.00-1.43	164	1.07 (0.31)	0.11-1.43
ECQB impulscontrole (ouder)	135	2.47 (0.65)	0.86-3.89	281	2.51 (0.65)	0.55-3.83
ECBQ volgehouden aandacht (ouder)	135	2.34 (0.43)	1.11-3.20	281	2.34 (0.46)	0.84-3.26
ECQB impulscontrole (pm'er)	107	3.06 (0.99)	0.45-4.45	164	3.51 (0.67)	0.73-4.92
ECBQ volgehouden aandacht (pm'er)	107	2.58 (0.54)	1.22-3.74	164	2.82 (0.44)	1.69-4.03

Samenhang tussen uitkomsten

Tabel 5-4 presenteert de intercorrelaties tussen testcores en gedragsbeoordelingen. Over het algemeen zijn de correlaties in de verwachte richting en van de verwachte sterkte. Zo zijn de taaltaken (klanken, CELF en PPVT) sterker onderling gecorreleerd dan met andere taken. De scores op de selectieve aandacht taak zijn sterk gecorreleerd met scores uit verschillende andere testen; dit past bij het idee dat selectieve aandacht een basisvaardigheid is die noodzakelijk is om op andere domeinen goed te functioneren. De NWR laat hoge correlaties zien met zowel de EF indicatoren, zoals selectieve aandacht, als met de taalscores. NWR meet verbaal geheugen. Dit is niet direct een taalmeting, maar het gaat om een vaardigheid die zeer belangrijk is voor het leren van taal. Wat opvalt is dat de bakjestaak iets minder sterk samenhangt met de andere EF maten (selectieve aandacht en NWR) dan die twee EF maten onderling. Verder valt op dat de vragenlijsten die ingevuld zijn door ouders minder sterk samenhangen met de testjes dan de vragenlijsten die ingevuld zijn door de pedagogisch medewerkers. Dit kan verklaard worden door verschillen in gedrag van de kinderen in de thuisomgeving en op de kindcentra. Het kan dan bijvoorbeeld zijn dat het gedrag van kinderen op de voorschool meer voorspellend is voor de scores op de testjes. Ook is het mogelijk dat de vragenlijsten wat lastiger zijn voor ouders om in te vullen. De scores gebaseerd op de vragenlijsten van ouders hangen wel onderling samen. Ook de scores uit de vragenlijsten van de pedagogisch medewerker hangen onderling samen. De SDQ scores uit de oudervragenlijst hangen ook samen met die van de pedagogisch medewerkers (maar minder sterk dan de correlatie binnen ouders of pedagogisch medewerkers). De ECBQ scores gebaseerd op de vragenlijsten ingevuld door ouders hangen niet significant samen met de ECBQ scores uit de vragenlijsten voor pedagogisch medewerkers.

Tabel 5-4 Onderlinge correlaties tussen scores

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Klankentaak																
2. PPVT	.514***															
3. CELF	.542***	.766***														
4. NWR	.258***	.461***	.393***													
5. Selectieve aandacht	.277***	.515***	.279***	.300***												
6. Dozentaak	.128*	.254***	.139*	.265***	.293***											
7. Wachttak rozijn	.210***	.276***	.214***	.259***	.266***	.149**										
8. Wachttak cadeau	.208***	.334***	.295***	.241***	.230***	.108*	.582***									
9. Fijne motoriek	.197***	.490***	.251***	.274***	.538***	.246***	.247***	.231***								
10. SDQ exter. ouder	.110*	.132**	.162**	.118*	.057	.268***	.174***	.160**	.121*							
11. SDQ prosociaal ouder	.015	.141**	.095	.154**	.110*	.086	.076	.079	.212***	.338***						
12. SDQ exter. pmer	.200**	.345***	.266**	.195**	.291***	.241***	.281***	.296***	.286***	.322***	.247***					
13. SDQ prosociaal pmer	.236***	.371***	.157*	.242***	.379***	.209***	.288***	.284***	.443***	.231***	.364***	.619***				
14. ECBQ imp. contr. ouder	.053	.021	-.023	.125*	.055	.135**	.087	.098*	.077	.484***	.269***	.168**	.165**			
15. ECBQ aandacht ouder	.065	-.010	.019	-.073	.075	.119*	.044	.040	.070	.298***	.079	.049	-.031	.312***		
16. ECBQ imp contr pmer	.051	.216***	.159*	.101	.221***	.207***	.225***	.272***	.320***	.303***	.258***	.637***	.562***	.112	.048	
17. ECBQ aandacht pmer	.131*	.126*	.057	.125	.171**	.077	.102	.084	.235***	.136*	.101	.339***	.285***	.070	.030	.276***

Noot. * $p < .050$; ** $p < .010$; *** $p < .001$; exter.=externaliserend gedrag; imp. contr.= impuls controle; Correlaties met de wachttaken zijn Spearman's correlaties, de overige correlaties zijn Pearson correlaties.

Samenhang tussen testcores en achtergrondkenmerken

In Tabel 5-5 staan de correlaties van de testcores met achtergrondkenmerken (opleidingsniveau ouders, geslacht, meertaligheid status, migratieachtergrond, leeftijd) weergegeven. De tabel laat zien dat de meeste scores positief correleren met het opleidingsniveau van de moeder. Dat is te verwachten op basis van de literatuur (Hoff, 2013; Lawson et al., 2018). Verder zien we dat meisjes hoger scoren op een aantal aspecten: de expressieve taal, selectieve aandacht, wachttaken, motoriek, prosociaal gedrag zoals gerapporteerd door zowel ouders als pedagogisch medewerkers en ook op volgehouden aandacht zoals gerapporteerd door de pedagogisch medewerkers. Ook laten meisjes minder externaliserend probleemgedrag zien (gerapporteerd door de pedagogisch medewerkers). De verschillen zijn klein en zijn in lijn met de literatuur (Ezpeleta et al., 2013; Kokštejn et al., 2016; Lange et al., 2016; Veijalainen, 2021). Daarnaast zien we ook, zoals verwacht, dat eentalige (Nederlandse) kinderen hoger scoren dan meertalige kinderen op de talige taken. Echter, op de andere taken zijn er geen verschillen tussen een- en meertalige kinderen. Dit suggereert dat de uitvoering van taken die geen taal meten, niet sterk beïnvloed wordt door het taalniveau van de kinderen. Migratieachtergrond laat vergelijkbare correlaties zien met taken die taalniveau in het Nederlands meten, maar niet met de andere taken. Zoals verwacht zijn alle taken positief gecorreleerd met leeftijd. De correlatie met leeftijd is wat kleiner bij taken waarbij we een kleinere leeftijdsvariantie hebben (expressieve taal gemeten met de CELF – alleen afgenomen bij de oudere groep) en waarbij er minder variatie is in de scores bij de oudere kinderen (klankentaak, de twee wachttaken). De scores gebaseerd op vragenlijsten die ingevuld zijn door pedagogisch medewerkers correleren ook positief met leeftijd. Dit laat zien dat impulscontrole en volgehouden aandacht toenemen met leeftijd. Ook laten kinderen meer prosociaal gedrag zien en minder externaliserend probleemgedrag. De uitkomsten van de vragenlijsten die ingevuld zijn door de ouders hangen niet significant samen met leeftijd.

Op basis van alle informatie over de scores beschreven in dit hoofdstuk is te zien dat de tests en vragenlijsten een betrouwbare en valide meting geven van de vaardigheden die ze pogen te meten. Echter, de meting van visueel ruimtelijk geheugen middels de dozentaak laat lagere samenhang met andere taken zien. In deze taak hebben we ook slechts 3 items van de 5 overgehouden na de schaling. Ervaringen uit het veld lieten ook zien dat de taak moeilijk was om af te nemen en er veel variatie zat in de afname. Kinderen moesten worden afgeleid voor 3 seconden voordat ze mochten zoeken. Wellicht dat hier een te grote variatie in de werkelijke tijd zat waarop kinderen werden afgeleid. Gegeven deze overwegingen hebben we besloten om deze taak niet op te nemen in de analyses. De overige taken en vragenlijsten zijn meegenomen in de vervolganalyse waarin we vier domeinscores berekenen. Naast de vier domeinscores nemen we in de analyses ook fijne motoriek als een vaardigheid mee die niet onder één van de andere domeinen valt.

Tabel 5-5 Correlaties tussen taakjes en achtergrondkenmerken

	Opleiding moeder	Geslacht	Meertalig	Migratie	Leeftijd
Klankentaak	.298 ^{***}	.097	-.233 ^{***}	-.117 [*]	.201 ^{***}
PPVT	.246 ^{***}	.077	-.303 ^{***}	-.184 ^{***}	.462 ^{***}
CELF	.324 ^{***}	.135 [*]	-.358 ^{***}	-.260 ^{***}	.167 ^{**}
NWR	.190 ^{**}	.071	-.064	.054	.322 ^{***}
Selectieve aandacht	.124	.131 ^{**}	-.050	.089	.546 ^{***}
Doosjestaak	.084	.007	-.004	.005	.272 ^{***}
Wachttaak (rozijn)	.226 ^{***}	.139 ^{**}	-.026	.057	.170 ^{***}
Wachttaak (cadeau)	.176 ^{**}	.181 ^{***}	-.080	-.032	.144 ^{**}
Motoriektaak	.034	.259 ^{***}	.021	.100 [*]	.538 ^{***}
SDQ exter. (ouder)	.220 ^{***}	-.051	-.013	.081	.016
SDQ prosociaal gedrag (ouder)	-.011	.118 [*]	.025	.106 [*]	.101 [*]
SDQ exter. (pm'er)	.250 ^{**}	.209 ^{**}	-.118	.008	.235 ^{***}
SDQ prosociaal gedrag (pm'er)	-.089	.218 ^{**}	-.060	.023	.424 ^{***}
ECQB impulscontrole (ouder)	.038	.066	.090	.102 [*]	.021
ECBQ aandacht ouder	.069	-.047	.138 ^{**}	.067	.021
ECQB impulscontrole (pm'er)	.144	.073	-.057	.085	.248 ^{***}
ECBQ aandacht (pm'er)	.195 [*]	.130 [*]	-.022	.108	.226 ^{***}

Noot. * $p < .050$; ** $p < .010$; *** $p < .001$; exter.=externaliserend gedrag; Correlaties met leeftijd (met uitzondering van correlaties tussen leeftijd en de wachttaken) zijn Pearson correlaties, alle overige correlaties zijn Spearman correlaties. Geslacht: meisje=1, jongen=0; meertaligheid: meertalig=1, eentalig=0; migratie: ouders met migratieachtergrond=1; ouders zonder migratieachtergrond=0.

Domeinscores

De laatste stap is het berekenen van domeinscores. Dit zijn scores gebaseerd op een combinatie van tests en vragenlijsten die één ontwikkelingsdomein representeren. Het samenvoegen van scores tot één domeinscore verkleint de meetfout en zorgt voor efficiëntere modellen. Deze stap is gedaan middels een confirmatieve factoranalyse.

Op basis van pre-COOL (Mulder et al., 2014; Mulder et al., 2017; Veen et al., 2012) is er gekozen voor een 4-factor model. De 4 factoren zijn: taal (klankentaak, PPVT, CELF), cool executieve functies (NWR, selectieve aandacht), hot executieve functies (wachttaak rozijn, wachttaak cadeau), en sociaal-emotioneel functioneren (SDQ en ECBQ vragenlijsten ingevuld door ouders en pedagogisch medewerkers; 8 schalen in totaal). Bij pre-COOL was NWR soms ook ingedeeld onder het domein taal in plaats van onder executieve functies. Echter, bij EVENING is een aangepaste versie van de NWR taak gebruikt, waarbij de gebruikte non-woorden minder op het Nederlands lijken dan bij de pre-COOL NWR taak. De correlaties laten ook zien dat er geen significant verschil is

tussen eentalige en meertalige kinderen op deze taak. Daarom is ervoor gekozen om deze bij cool executieve functies in te delen. We hebben ook modellen getoetst met NWR ingedeeld bij taal of bij beide (taal en EF). De model fit was echter niet significant anders tussen deze modellen. Daarom is er, op theoretische gronden, gekozen voor het model waarbij NWR onder cool EF is ingedeeld.

Omdat de CELF alleen is afgenomen bij de kinderen die getest zijn rond de leeftijd van 3 jaar en 10 maanden, hebben we twee versies van het factormodel gemaakt. In de eerste versie (V1) zijn alleen de oudere kinderen meegenomen en daarin is de CELF in de taalfactor ondergebracht. In de tweede versie (V2) is data van alle kinderen meegenomen en daarom is in die versie de CELF niet opgenomen als een deel van de taalfactor. Verder bleek uit alle modellen dat de scores op volgehouden aandacht zoals gerapporteerd door ouders niet goed passen bij de modellen. Daarom is deze maat buiten beschouwing gelaten. De fit indices van de finale modellen geven aan dat beide modellen goed passen bij de data. Zie Appendix F voor de fit indices en factorladingen van de modellen.

Tabel 5-6 Beschrijvende statistiek domein scores

Domein scores	N	M(SD)	Range
Taal V1	286	-0.01(0.71)	-1.78 - 1.27
Taal V2	424	-0.02(0.72)	-2.01 – 1.16
Cool EF V1	281	-0.01(0.42)	-1.78 – 1.20
Cool EF V2	409	-0.03(0.59)	-2.25 – 1.53
Hot EF V1	284	1.58(0.33)	0.89 – 1.78
Hot EF V2	421	1.54(0.36)	0.90 – 1.80
Sociaal-emotioneel V1	286	-0.00(0.29)	-1.04 - 0.58
Sociaal-emotioneel V2	422	0.00(0.35)	-1.05 - 0.71

Noot. V1 =Versie 1 (met CELF, alleen oudere kinderen); V2 = versie 2 (zonder CELF, data van alle kinderen).

Tabel 5-6 laat de gemiddelden en standaarddeviaties van de factorscores zien; Tabel 5-7 en Tabel 5-8 laten de onderlinge correlaties tussen factoren en de correlaties met achtergrondkenmerken zien. Om totaalscores te berekenen van de verschillende factoren is er een gewogen gemiddelde van de test- en vragenlijstscores die horen bij de factor berekend, met de gestandaardiseerde factorladingen als gewichten. Als een score voor een toets of vragenlijst miste, is het gemiddelde berekend op basis van de resterende scores (zie Tabel 5-2 voor totaal aantal beschikbare scores per taak). In deze vorm van imputatie worden de scores van de taken die non-missing zijn binnen een domein gebruikt om de scores op de missende taak te schatten. Als de scores op alle taken binnen een domein missen dan is er geen score berekend voor dit domein. In Tabel 5-6 is te zien hoeveel scores uiteindelijk zijn berekend per domein.

Tabel 5-7 Onderlinge correlaties domeinscores

	1	2	3	4	5	6	7
1. Taal V1							
2. Cool EF V1	.469***						
3. Hot EF V1	.314***	.343***					
4. Sociaal-emotioneel V1	.161**	.214***	.280***				
5. Taal V2	.957***	.460***	.298***	.136*			
6. Cool EF V2	.434***	.950***	.335***	.294***	.543***		
7. Hot EF V2	.315***	.340***	1.00***	.279***	.322***	.364***	
8. Sociaal-emotioneel V2	.143*	.195***	.260***	.984***	.261***	.385***	.288***

Noot. * $p < .050$; ** $p < .010$; *** $p < .001$; V1 = Versie 1 (met CELF, alleen oudere kinderen); V2 = versie 2 (zonder CELF, data van alle kinderen).

Tabel 5-8 Correlaties tussen domeinscores en achtergrond kenmerken

	Opleiding moeder	Geslacht	Meertalig	Migratie	Leeftijd
Taal V1	.384***	.119*	-.384***	-.267***	.187**
Cool EF V1	.283***	.228***	-.079	.001	.125*
Hot EF V1	.248**	.218***	-.070	-.027	.026
Sociaal-emotioneel V1	.238**	.120*	-.032	.061	.030
Taal V2	.287***	.086	-.300***	-.190***	.422***
Cool EF V2	.214***	.155**	-.042	.105*	.555***
Hot EF V2	.236***	.182***	-.063	.005	.197***
Sociaal-emotioneel V2	.164**	.143*	-.027	.111*	.250***

Noot. * $p < .050$; ** $p < .010$; *** $p < .001$; Correlaties met leeftijd zijn Pearson correlaties, de overige correlaties zijn Spearman correlaties. Geslacht: meisje=1, jongen=0; meertaligheid: meertalig=1, eentalig=0; migratie; ouders met migratieachtergrond=1; ouders zonder migratieachtergrond=0; V1=Versie 1 (met CELF, alleen oudere kinderen); V2=versie 2 (zonder CELF, data van alle kinderen). Opleiding moeder is het opleidingsniveau van opleiding behaald in Nederland.

De gegevens in Tabel 5-6 laten geen indicaties van bodem- of plafond-effecten zien. Uit Tabel 5-7 blijkt dat, zoals verwacht, de scores binnen hetzelfde domein tussen de twee versies zeer sterk gecorreleerd zijn (alle correlaties zijn 0.95 of hoger). Ook zijn de scores voor de verschillende domeinen gecorreleerd zijn. Echter, de correlaties laten ook zien dat de domeinen andere vaardigheden meten. De scores hangen samen maar niet heel sterk (de correlaties tussen domeinen zijn bijna altijd onder de 0.5, vaak onder de 0.3). Het taaldomein en het cool EF domein laten hogere onderlinge correlaties zien dan de andere domeinen. Tabel 5-8 laat zien dat alle domeinscores positief correleren met het opleidingsniveau van de moeder. Dit is in lijn met de literatuur (Hoff; 2013; Lawson et al., 2018). Verder scoren meisjes hoger op de meeste domeinen. Dat is ook in lijn met de literatuur (Ezpeleta et al., 2013; Lange et al., 2016; Veijalainen, 2021). Meertaligheid correleert negatief met de taalscores maar niet met de andere domeinscores. Net als meertaligheid correleert migratieachtergrond negatief met taalscores, maar minder sterk. Bovendien zijn er ook indicaties dat migratieachtergrond positief correleert met andere domeinen. Ten slotte, de scores in de tweede versie, waar alle kinderen zijn meegenomen, laten positieve correlaties met leeftijd zien. Dat is

ook in lijn met de verwachting en laat zien dat oudere kinderen beter scoren dan jongere kinderen. De scores zoals berekend met de eerste modelversie laten kleinere of geen correlatie met leeftijd zien. Dit kan verwacht worden omdat in deze versie alleen oudere kinderen zijn meegenomen: de spreiding van leeftijd is daarom beperkt.

Samengenomen laten deze analyses zien dat de berekende domeinscores een valide en betrouwbare meting geven van de vaardigheden op deze domeinen. Daarom worden deze scores gebruikt in verdere analyses. Gegeven dat de totale steekproef relatief klein is kiezen we voor het tweede model (V2, zonder CELF), omdat we dan data over alle kinderen kunnen meenemen in de analyses. Bovendien zijn, zoals hierboven genoemd, de correlaties tussen de V1 en V2 scores zeer hoog, waardoor relatief weinig informatie verloren gaat wanneer we de CELF scores niet meenemen.

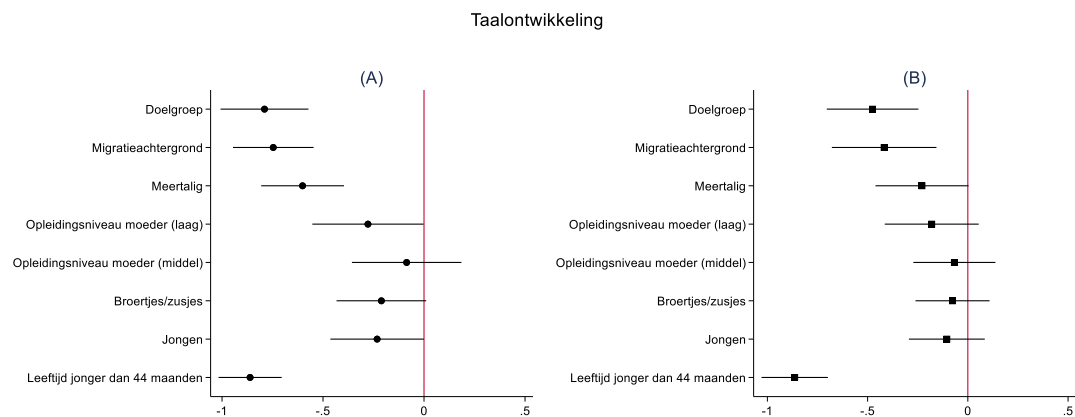
Domeinscores en doelgroepstatus

In Figuur 5-3 laten we zien hoe doelgroepstatus (wel of geen VVE-indicatie) en andere kenmerken gecorreleerd zijn met de verschillende domeinscores (de V2 scores zijn herschaald tot z-scores met een gemiddelde van 0 en een standaarddeviatie van 1). In de linker figuren (A) staan de coëfficiënten van bivariate regressies gepresenteerd, in de rechter figuren (B) zijn alle genoemde kenmerken opgenomen in de regressies. Zoals verwacht scoren doelgroepeuters lager dan peuters zonder VVE-indicatie op (Nederlandse) taal: dit verschil is komt overeen met ongeveer 12 maanden in leeftijd. Gegeven de jonge leeftijd van de kinderen is deze achterstand substantieel. Dit is in lijn met beleid: kinderen met een (hoog) risico op taal- of onderwijsachterstand krijgen een VVE-indicatie. Ook scoren doelgroepeuters significant lager op de twee domeinen van executieve functies. De verschillen in sociaal-emotionele ontwikkeling en fijne motoriek zijn minder duidelijk. Verder valt op dat achtergrondkenmerken als migratieachtergrond, meertaligheid en het opleidingsniveau van de moeder de achterstand van doelgroepeuters in taalontwikkeling en executieve functies maar deels kunnen verklaren. Dit wijst erop dat doelgroepstatus meer is dan enkel een aantal achtergrondkenmerken. Bij dezelfde achtergrondkenmerken krijgen sommige kinderen wel en andere kinderen niet een VVE-indicatie. Dit kan bijvoorbeeld zijn omdat ook andere kenmerken worden meegenomen bij de beslissing om een kind een VVE-indicatie te geven. Zo kan ook de thuisomgeving of de feitelijke ontwikkeling van het kind worden meegewogen bij het geven van de VVE-indicatie.

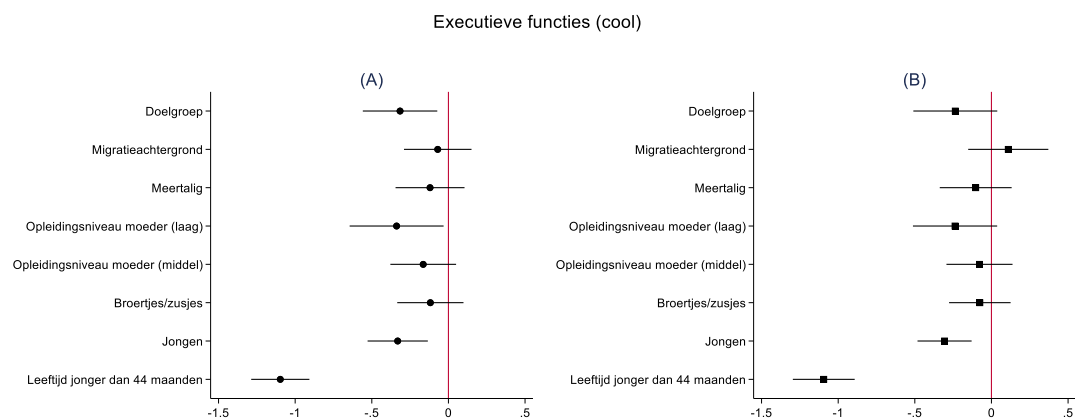
Wanneer we kijken naar de andere achtergrondkenmerken valt op dat naast de doelgroepstatus vooral leeftijd relatief sterk gecorreleerd is met alle scores (zoals hierboven al beschreven). Ook scores jongens significant lager op de verschillende domeinen.

Figuur 5-3 Domeinscores, doelgroep en achtergrondkenmerken

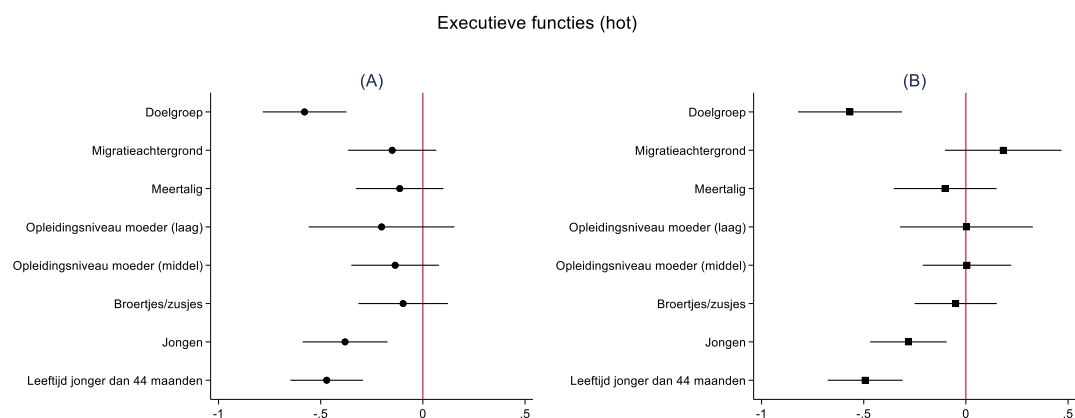
Figuur 5-3.1: Taalontwikkeling



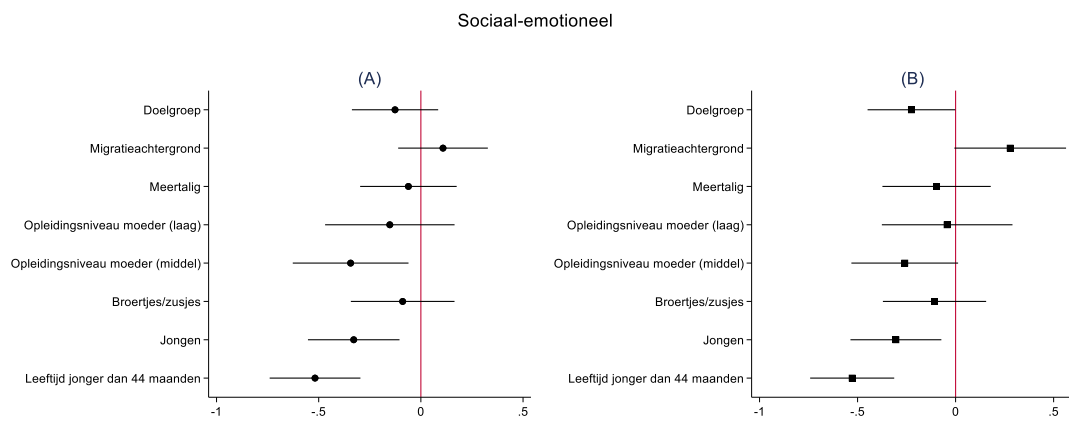
Figuur 5-3.2: Executieve functies (cool)



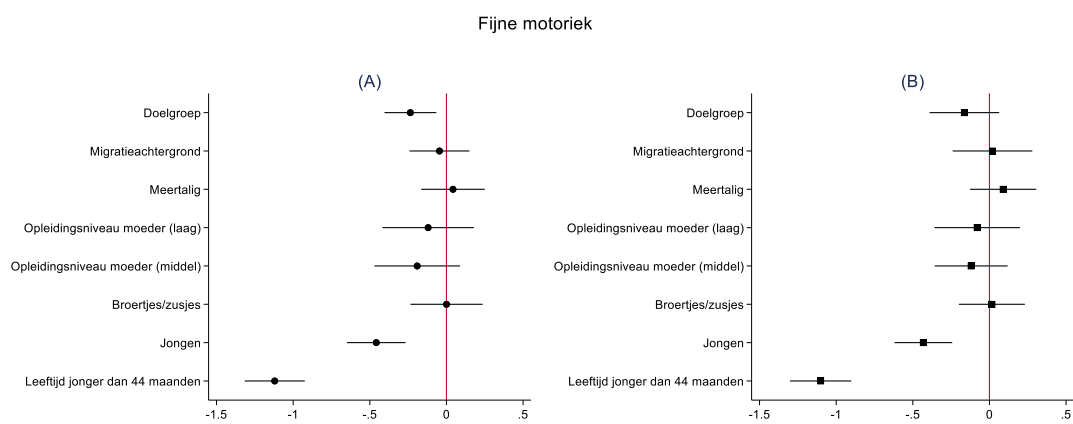
Figuur 5-3.3: Executieve functies (hot)



Figuur 5-3.4: Sociaal-emotioneel



Figuur 5-3.5: Fijne motoriek



6. Effecten 960 uur voorschoolse educatie

6.1. Inleiding

Sinds 1 augustus 2020 zijn gemeenten verplicht om kinderen met een VVE-indicatie 960 uur voorschoolse educatie te bieden tussen de leeftijd van 2,5 en 4 jaar. In de praktijk komt dit vaak neer op 16 uur per week tijdens de schoolweken. Dit is een toename van de wettelijke norm van 60 procent: voor augustus 2020 was een aanbod van 600 uur (10 uur per week) verplicht. In verschillende gemeenten is 960 uur al (ruim) voor augustus 2020 ingevoerd, bijvoorbeeld in januari 2020 of augustus 2019. Ook bestond in veel gemeenten al een aanbod van meer dan 10 uur per week, bijvoorbeeld 12 of 15 uur per week. In eerdere EVENING rapportages hebben we gedocumenteerd dat de beleidsmaatregel heeft geleid tot een toename in gerichte VVE-activiteiten, waarbij VE-activiteiten vaker worden aangeboden in kleine groepen met alleen doelgroepkinderen (van Huizen et al., 2023). De invoering van 960 uur betekende in de regel langere dagdelen (vaak van 2,5 naar 4 uur) en dit lijkt de pedagogisch praktijk op de groepen te hebben veranderd. Met 960 uur VE krijgen peuters dus niet alleen meer, maar deels ook andere ontwikkelingsmogelijkheden op de voorscholen.

In dit hoofdstuk beschrijven we de effecten van het verruimen van het urenaanbod voorschoolse educatie op het urengebruik en de ontwikkeling van peuters.⁵ We maken daarbij gebruik van een natuurlijke experiment: er is variatie in het (cumulatieve) urenaanbod van voorschoolse educatie door verschillen in de timing van implementatie van 960 uur en het urenaanbod voor de invoering van 960 uur. Deze variatie wordt gebruikt om de effecten op de ontwikkeling van doelgroeppeuters te schatten, waarbij effecten op (Nederlandse) taalvaardigheid, executieve functies, sociaal-emotionele ontwikkeling en fijne motoriek worden getoetst.

Er bestaat al veel internationaal onderzoek naar de effecten van deelname aan voorschoolse voorzieningen. De consensus is dat kinderen uit relatief kansarme gezinnen baat hebben bij deelname aan voorschoolse voorzieningen en dat de kwaliteit een belangrijke rol speelt (van Huizen & Plantenga, 2018; Leseman & van Huizen, 2022; Duncan et al., 2023). Er bestaat echter weinig onderzoek naar de effecten van een ruimer urenaanbod en de resultaten zijn niet eenduidig. In een studie naar een Duitse hervormingen laten Felfe & Lalive (2018) bijvoorbeeld zien dat een verschuiving van halve naar hele dag (universele) kleuterscholen een negatief effect had op de sociaal-emotionele ontwikkeling, vooral voor kinderen uit meer kansarme gezinnen. De studie laat tegelijkertijd positieve effecten zien op de schoolrijpheid van kinderen met een migratieachtergrond. Bijzonder aan de Nederlandse context is dat het gaat om de effecten van meer uren voorschoolse educatie binnen een specifieke range per week

⁵ In lijn met de hervorming richten we ons in dit hoofdstuk op de effecten van het urenaanbod tussen 2,5-4 jaar. Verschillende gemeenten hebben een aanbod van voorschoolse educatie voor de leeftijd van 2,5 jaar. De effecten van eerder starten met voorschoolse educatie wordt in een aparte rapportage behandeld.

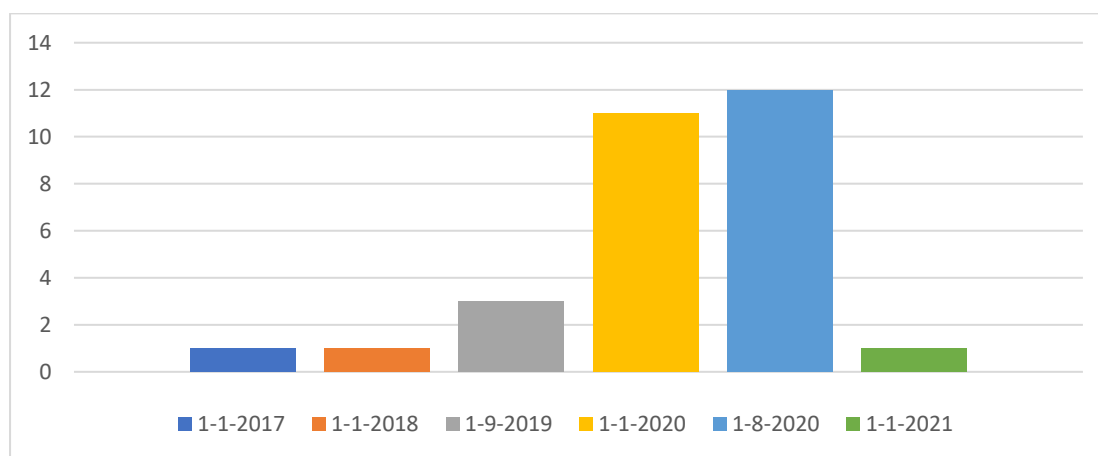
(10-16 uur per week) en per dag (vaak van 2.5 naar 4 uur per dag). Ook gaat het om kinderen uit relatieve kansarme gezinnen, vaak met een migratieachtergrond. Het is denkbaar dat juist in deze context extra uren voorschoolse educatie significant kunnen bijdragen aan de ontwikkeling van kinderen.

6.2. Implementatie van 960 uur voorschoolse educatie⁶

Gemeenten

Sinds augustus 2020 hebben Nederlandse gemeenten de taak om 960 uur (16 uur per week) voorschoolse educatie aan te bieden aan doelgroeppeuters tussen de leeftijd van 2.5 en 4 jaar. Binnen de deelnemende EVENING gemeenten zien we duidelijke verschillen in de invoeringsdatum. Met behulp van vragenlijsten hebben we gemeentelijke beleidsmedewerkers gevraagd sinds wanneer zij een aanbod van 960 uur (16 uur per week) hebben (Figuur 6-1). Van de 29 gemeenten die de bovengenoemde vragen hebben beantwoord, zijn sinds 1 august 2020 (de wettelijke invoering) 12 gemeenten gestart met 960 uur VE, 11 gemeenten sinds 1 januari 2020 en 3 gemeenten al sinds 1 september 2019. Ook zijn er twee gemeenten die al veel eerder (2017-2018) locaties hadden met een 16-uuraanbod. Er is ook een gemeente die de 16 uur later dan de wettelijke invoeringsdatum is gaan aanbieden (1 januari 2021).

Figuur 6-1 Datum invoering 960 uur VE



Naast variatie in de invoeringsdatum is er ook variatie in het urenaanbod vóór aanvang van de invoering van 960 uur VE. Het merendeel van de deelnemende gemeenten bood 10-12 VE uren per week aan. Er zijn echter enkele gemeenten die voorheen al een groter aanbod hadden voor sommige aanbieders, zoals 10.5 tot 14 uur per week, 10 tot 16 uur per week of 15 uur per week. Gemeentelijke beleidsmedewerkers is ook gevraagd naar de varianten binnen de gemeenten en de meeste voorkomende variant van 960 uur. Er zijn veel gemeenten die meerdere opties mogelijk maken en waar de aanbieders zelf kunnen beslissen hoe zij de uren spreiden. We zien echter wel dat

⁶ Deze sectie is grotendeels gebaseerd op eerdere (tussentijdse) rapporten van het EVENING project.

binnen onze deelnemende gemeenten de meest voorkomende variant '16 uur per week: 4 x 4uur per week gedurende 40 weken per jaar' is.

Verder is in de gemeentelijke vragenlijst ook gevraagd naar de aangeboden uren voor reguliere peuters. In het merendeel van onze deelnemende gemeenten kunnen reguliere peuters gebruik maken van de helft van de VE-uren die beschikbaar zijn voor doelgroeppeuters, dus maximaal 8 uur per week. Daarnaast zijn er een aantal gemeenten die dezelfde VE-uren aanbieden voor doelgroeppeuters en reguliere peuters (16 uur per week). Na de invoering van 960 uur zijn er in enkele gemeenten wijzigingen geweest in het urenaanbod voor reguliere peuters. Eén gemeente had een aanbod van 5 uur per week en heeft dit in 2023 verhoogd naar 8 uur per week; twee gemeenten die een aanbod hadden van 8 uur per week, verhoogden het aanbod naar 16 uur per week.

Aanbieders

Uit de vragenlijsten ingevuld door leidinggevendenden van 116 groepen komt naar voren dat vrijwel alle instellingen de wettelijke norm van 16 uur per week aanbieden aan doelgroepkinderen. Echter, er is wel enige variatie in hoe de uren verdeeld worden over de week en over het jaar. De meest voorkomende verdeling, te zien bij ruim de helft van de groepen, is het model van 4 uur per dag voor 4 dagdelen per week tijdens schoolweken (40 weken per jaar). Enkele groepen bieden 4x4 uur per week voor meer weken per jaar aan. De tweede meest voorkomende variant is het bieden van 5 dagdelen van 3 uur en 15 minuten, ook meestal alleen tijdens schoolweken. Bij een klein aantal centra zien we een aanbod van 3 dagdelen voor 5 uur en 20-30 minuten per dag gedurende 40 weken per jaar of 3 dagdelen van 5 uur voor meestal 43 of 52 weken per jaar. Bij een aantal groepen wordt een variërend aantal uren per dag aangeboden en bij de kinderopvanggroepen worden wel 16 uur VE aangeboden, maar het aantal uren per dag varieert en hangt soms af van het aantal dagdelen dat een kind naar de voorschool gaat.

De eerste online EVENING vragenlijsten zijn vanaf april 2020 uitgezet, vóór de landelijke invoering van 960 uur. Zo'n 64% van de groepen had op het moment van invullen al een aanbod van 960 uur; 36% van de groepen had nog geen aanbod van 960 uur. Aan de pedagogisch medewerkers die in groepen werkten waar op het moment van invullen al een VE-aanbod van 960 uur was, is de volgende stelling voorgelegd: Welk effect ziet u van de invoering van 960 uur voorschoolse educatie voor peuters met een VVE-indicatie? "De ontwikkeling van de peuters met een VVE-indicatie wordt versterkt". De meeste medewerkers (56.4%) gaven aan het eens te zijn met de stelling; een klein deel van de medewerkers (6.4%) was het oneens met de stelling. De overige medewerkers waren neutraal (33.3%) of hadden geen mening (3.8%).

Pedagogisch medewerkers die het eens waren met de stelling gaven aan dat er door de urenuitbreiding meer ruimte is om de ontwikkeling van de kinderen te stimuleren. Met name het extra taalaanbod in de Nederlandse taal en de extra tijd voor begeleiding en VVE-activiteiten worden vaak genoemd als positief effect van de urenuitbreiding. Ook werd regelmatig genoemd dat de langere dagdelen meer structuur en rust in de groep

brengt. Daartegenover gaven pedagogisch medewerkers die het oneens waren met de stelling vaak aan dat de dagdelen nu te lang zijn voor de kinderen. Aan het einde van de dagdelen (met name in de middagen) zijn de kinderen vaak erg vermoeid, waardoor die uren niet optimaal benut kunnen worden om de ontwikkeling van de kinderen te stimuleren.

Aan de pedagogisch medewerkers die in groepen werkten waar op het moment van invullen nog geen VE-aanbod van 960 uur was, is de volgende stelling voorgelegd: Welk effect verwacht u van de invoering van 960 uur voorschoolse educatie voor peuters met een VVE-indicatie: “De ontwikkeling van de peuters met een VVE-indicatie wordt versterkt”. De meeste medewerkers gaven aan het eens te zijn met de stelling (68.2%); slechts een klein deel van de medewerkers (6.8%) was het oneens met de stelling. De overige medewerkers waren neutraal (18.2%) of hadden geen mening (6.8%).

Tijdens alle meetronden hebben we aan de leidinggevenden van groepen waarbij de nieuwe urennorm werd ingevoerd, de volgende stelling voorgelegd: Welk effect ziet u van de invoering van 960 uur voorschoolse educatie voor peuters met een VVE-indicatie? Daarbij kregen ze verschillende mogelijkheden zoals ‘de kwaliteit van het voorschoolse aanbod voor peuters met een VVE-indicatie neemt toe’. We hebben in totaal gegevens over deze items van 166 verschillende locaties. De meest duidelijk positieve effecten die leidinggevenden benoemen zijn het versterken van de ontwikkeling van doelgroeppeuters en het verhogen van de kwaliteit van het aanbod voor doelgroeppeuters. Ook werd relatief vaak benoemd dat de deskundigheid van beroepskrachten voor voorschoolse educatie toeneemt. De meest genoemde negatieve effecten zijn dat het moeilijker is om pedagogisch medewerkers en beroepskrachten specifiek voor de voorschoolse educatie te werven en het stijgen van de kosten voor ouders van reguliere peuters en in iets mindere mate voor ouders van doelgroeppeuters.

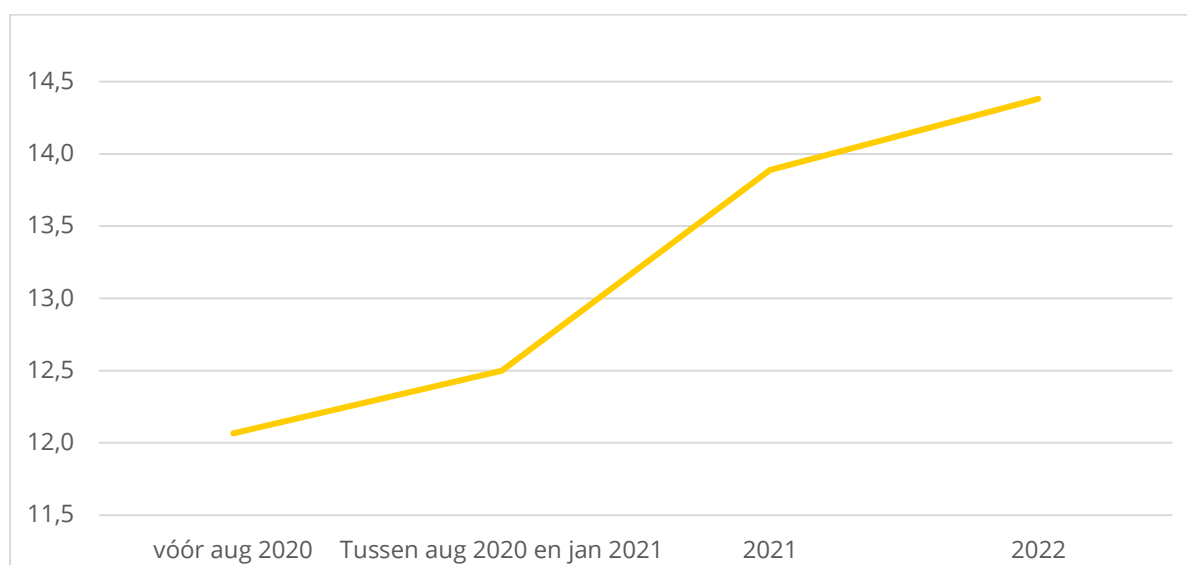
Aan de leidinggevenden werd gevraagd naar eventuele knelpunten bij het realiseren van 960 uur VE op hun locatie. De vier meest voorkomende knelpunten bij de locaties die het urenaanbod nog niet hadden verruimd ten tijde van de vragenlijst zijn: 1) Beperkte voorbereidingstijd in verband met de coronacrisis (79%); 2) Beperkte voorbereidingstijd algemeen (58%); 3) Toename van de kosten voor de ouders (47%); 4) Slaapbehoefte van peuters (waardoor peuters bijv. geen middagdagdeel kunnen deelnemen) (37%). Dit laat zien dat de coronacrisis bij veel aanbieders extra uitdagingen heeft gecreëerd bij het realiseren van 960 uur voorschoolse educatie.

Urengebruik

Later in dit hoofdstuk analyseren we de effecten van de uitbreiding van het aantal VE-uren. Hier bespreken we een aantal beschrijvende patronen: hoe heeft het VE-gebruik zich in de periode rond de invoering van 960 uur ontwikkeld? Figuur 6-2 laat zien dat het gemiddeld aantal uren per week dat kinderen naar de voorschool gaan is toegenomen in de afgelopen jaren. In deze grafiek zijn de gegevens te zien van de kinderen die een VVE-indicatie hebben en die alleen naar de voorschool gaan (en niet daarnaast nog naar een andere vorm van opvang gaan). De sub-sample die hier gebruikt is, bestaat uit 195

kinderen. De eerste kinderen zijn getest begin 2020. Zij gingen toen gemiddeld 12 uur per week naar de voorschool. Er was na augustus 2020, toen 960 uur landelijk is uitgebreid, al een kleine toename te zien, maar de stijging zette zich voort in de periode daarna. Gemiddeld ging deze groep kinderen in 2022 14.4 uur per week naar de voorschool. Als we dit uitsplitsen naar de leeftijd van de kinderen, zien we dat kinderen uit de jongere groep (wanneer het kind ongeveer 2.5 jaar was tijdens de testafname) in 2022 gemiddeld 13.2 uur naar de voorschool gingen, en de kinderen uit de oudere groep (als het kind bijna vier jaar oud was) gemiddeld 15.1 uur. Oudere kinderen gaan dus, zoals verwacht, meer uren naar de voorschool dan jongere kinderen.

Figuur 6-2 Gemiddeld aantal uren voorschool educatie per week over tijd



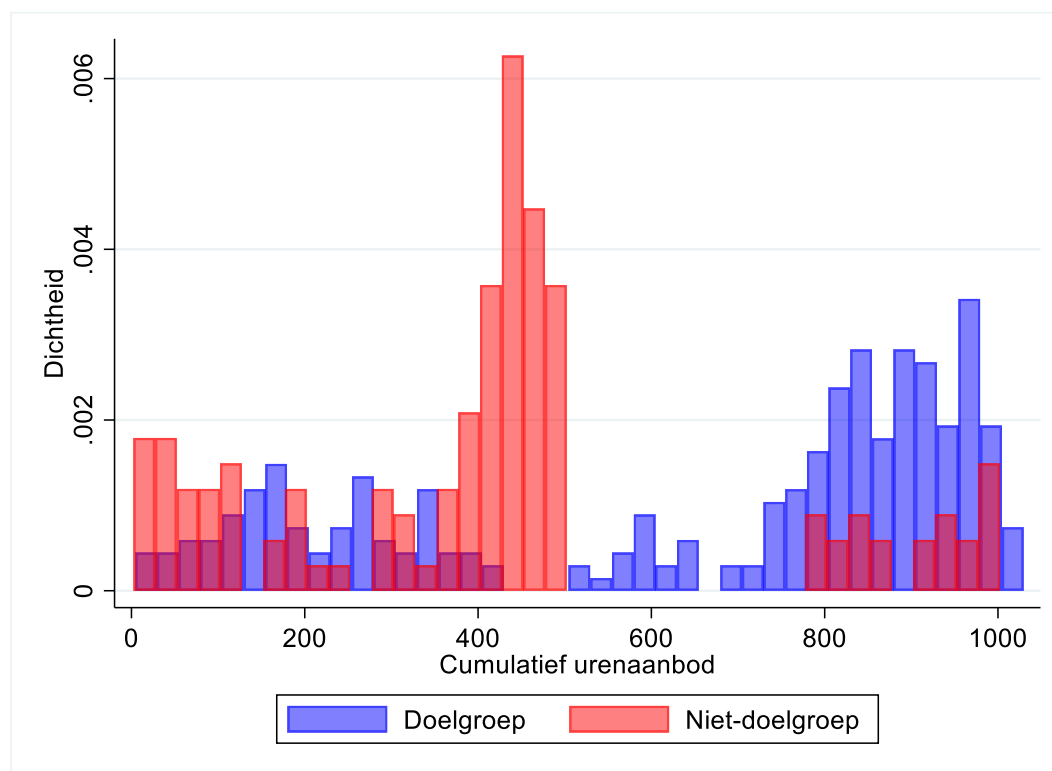
Sinds het urenaanbod is verruimd, gaan kinderen volgens hun ouders dus meer uren per week naar de voorschool. Het is een geleidelijke verandering, die mogelijk ook beïnvloed is door de coronasituatie. Deze beschrijvende resultaten suggereren ook dat niet alle ouders die hiervoor in aanmerking komen, gebruik van maken van het volledige aanbod van 16 uur. Opvallend is dat we geen stijging in het urengebruik zien bij de kinderen die geen VVE-indicatie hebben. Daarbij moet opgemerkt worden dat het gaat om een kleine sub-sample. In 2020, vóór de uitbreiding, gingen deze kinderen (N=27) 10.2 uur naar de voorschool. In 2022 gingen deze kinderen (N=38) 8.5 uur naar de voorschool. Tenslotte, ouders zijn in 2022 en 2023 meer tevreden over het aantal uren dat het kind naar de voorschool gaat dan in de jaren daarvoor. Ongeveer 59% van de ouders die in 2022 en 2023 zijn geïnterviewd geven aan tevreden te zijn over het aantal uren. In 2020 en 2021 waren deze percentages aanzienlijk lager (respectievelijk 35% en 40%).

6.3. Identificatiestrategie

Eén van de centrale vragen in EVENING is wat de effecten van een ruimer urenaanbod op de ontwikkeling van (doelgroep)peuters zijn. Het onderzoek maakt gebruik van een natuurlijk experiment om de effecten van een ruimer urenaanbod voorschoolse educatie te identificeren. Het feitelijk aantal gebruikte uren is een keuze van ouders.

Ouders maken niet altijd gebruik van het volledige voorschoolse aanbod (bijvoorbeeld 3 van de 4 dagen), kunnen ervoor kiezen later te starten met VE of een periode geen gebruik maken van voorschoolse educatie. Het urengebruik hangt daarom waarschijnlijk samen met (niet-geobserveerde) kenmerken van ouders en kinderen, die gecorreleerd kunnen zijn met de ontwikkeling van kinderen. Ook kan er sprake zijn van omgekeerde causaliteit. Zo is het denkbaar dat ouders meer de noodzaak zien van meer uren voorschoolse educatie wanneer er sprake is van een ontwikkelingsachterstand; de relatie tussen urengebruik en ontwikkeling is dan een onderschatting van het effect. In het onderzoek maken we daarom gebruik van het urenaanbod (het recht op een bepaald aantal uren) als “treatment” variabele om de effecten te schatten. Voor de analyses reconstrueren we voor elk kind het cumulatieve urenaanbod vanaf leeftijd 2.5 tot het testmoment. We baseren ons op gegevens over gemeentelijk beleid, waarbij we ons baseren op gemeentelijke beleidsdocumenten, vragenlijsten aan beleidsmedewerkers en persoonlijke interviews met beleidsmedewerkers. Ook is gebruik gemaakt van gegevens uit de GOAB monitor.

Figuur 6-3 Spreiding van het urenaanbod naar doelgroepstatus (cumulatief vanaf 2,5 jaar)

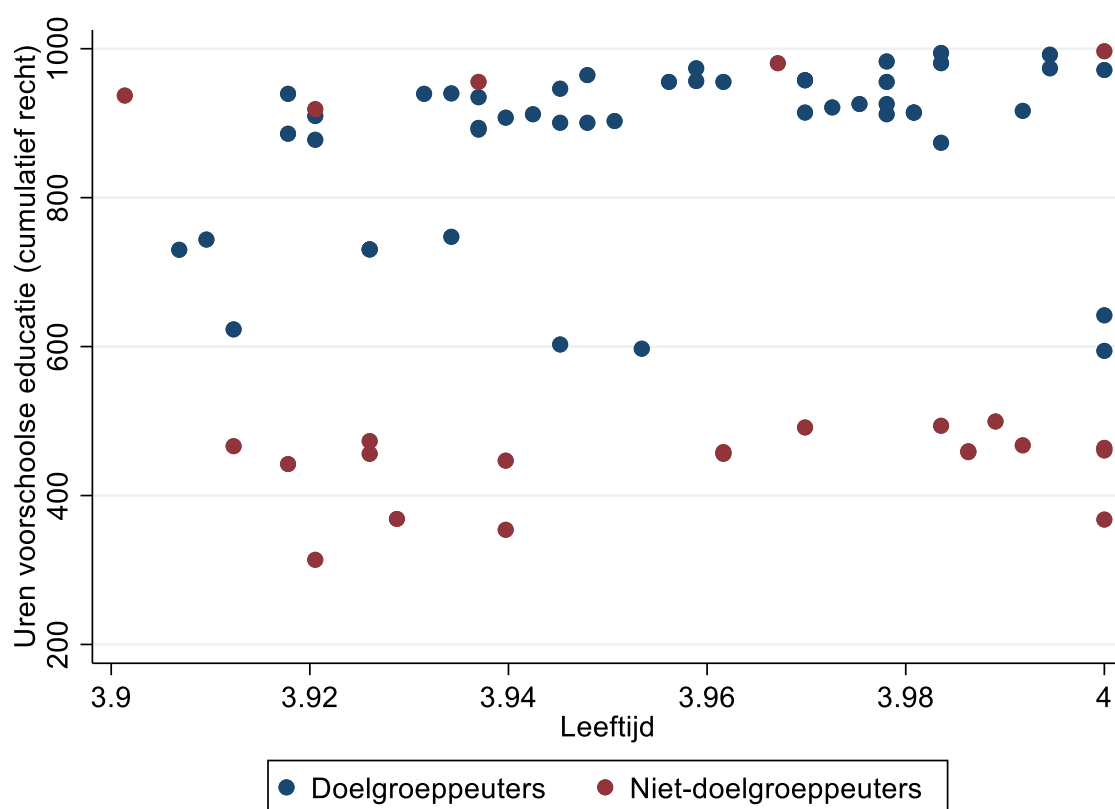


Het cumulatieve uren aanbod hangt af van de doelgroepstatus (wel of geen VVE-indicatie), gemeente, de geboortedatum van kind en de testdatum. In de meeste gemeenten zijn er grote verschillen tussen reguliere en doelgroepeuters in het uren aanbod (vaak hebben doelgroepeuters recht op het dubbele uren aanbod). Er zijn ook verschillende redenen waarom dit uren aanbod varieert tussen doelgroepeuters, ook als ze dezelfde leeftijd hebben. Ten eerste zijn er verschillen in de timing van de invoering van 960 uur voorschoolse educatie en in het uren aanbod voor de invoering 960 uur (16 uur per week) voorschoolse educatie. Zo heeft gemeente A 16 uur per week

ingevoerd in augustus 2020 en had daarvoor een aanbod van 10 uur per week; gemeente B 16 uur per week ingevoerd in januari 2020 en had daarvoor een aanbod van 10 uur per week; gemeente C 16 uur per week ingevoerd in augustus 2020 en had daarvoor een aanbod van 12 uur per week; gemeente D 16 uur per week ingevoerd in januari 2020 en had daarvoor een aanbod van 15 uur per week. Daarnaast zijn er verschillen in het cumulatieve urenaanbod vanwege COVID-sluitingen en door (regionale) verschillen in schoolvakanties. Gegeven een bepaalde geboortedatum en leeftijd kunnen doelgroeppeuters dus een ander cumulatief urenaanbod hebben, afhankelijk van de specifieke gemeente. En ook binnen een gemeente kunnen kinderen van dezelfde leeftijd een ander cumulatief urenaanbod hebben, wanneer zij op een andere datum geboren zijn.

Figuur 6-3 geeft de spreiding aan van het cumulatieve urenaanbod tussen 2.5 jaar en het testmoment in de EVENING steekproef. Er zijn grote verschillen in het urenaanbod, en dat is te verwachten gegeven de verschillen in leeftijd en het feit dat ongeveer een derde van de kinderen geen VVE-indicatie heeft. Zoals verwacht is het cumulatief recht van niet-doelgroeppeuters over het algemeen substantieel lager, maar zijn er ook niet-doelgroeppeuters met een aanbod rond de 900 uur in gemeenten waar hetzelfde aanbod geldt voor alle peuters. De figuur laat ook zien dat een groot deel van de doelgroeppeuters een aanbod tussen de 800-1000 heeft gehad.

Figuur 6-4 Cumulatief urenaanbod naar leeftijd 3,9-4 jaar

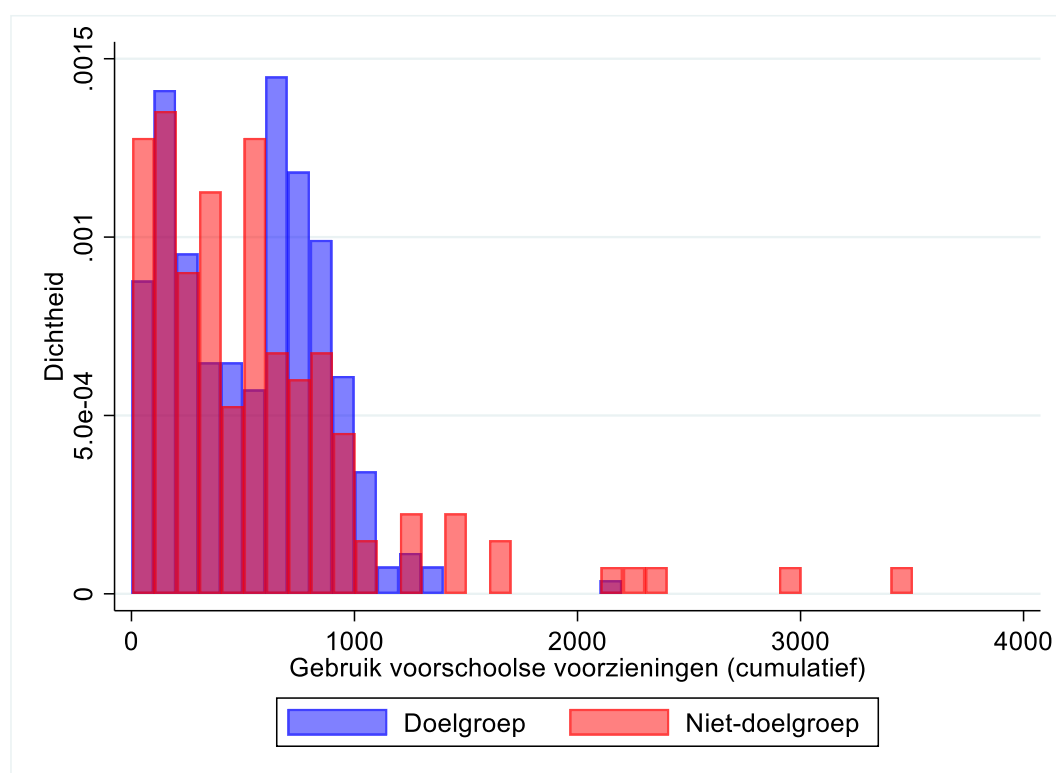


Om een beeld te geven van de variatie die we gebruiken in de analyses presenteren we in Figuur 6-4 het cumulatieve urenaanbod voor peuters van 3.9-4 jaar, uitgesplitst naar

doelgroepstatus. De figuur laat zien dat er ook bij dezelfde leeftijd variatie is in het urenaanbod. Zo hebben de drie doelgroepeuters die exact 4 jaar zijn op het testmoment recht gehad op 594, 652 respectievelijk 971 uur voorschoolse educatie.

Naast het cumulatieve urenaanbod reconstrueren we ook het cumulatieve *urengebruik* vanaf leeftijd 2.5 tot het testmoment. Hierbij maken we gebruik van een speciale, in EVENING ontwikkelde vragenlijst, die rekening houdt met verschillen in startleeftijd, urengebruik per week en veranderingen daarin en periodes van niet-gebruik. Het gaat hier om het totale gebruik van alle voorschoolse voorzieningen, inclusief kinderdagverblijf. Het urengebruik en urenaanbod komen dus via onafhankelijke bronnen tot stand: het urenaanbod wordt berekend op basis van gemeentelijk beleid, het urengebruik wordt berekend op basis van het ouderinterview. Figuur 6-5 laat de verdeling zien van het cumulatieve urengebruik. Hoewel voor veruit de meeste kinderen het urenaanbod tussen de 0 en 1000 ligt (in lijn met het urenaanbod), valt het op dat een klein deel van de peuters gebruik heeft gemaakt van (veel) meer dan 960 uur. Dit kan verklaard worden door het feit dat sommige peuters gebruik hebben gemaakt van kinderopvang, waardoor het urengebruik per week vaak ruim boven de 16 uur per week is. Ook hier wordt duidelijk dat er grote verschillen zijn in het urengebruik tussen peuters met en zonder VVE-indicatie. In veel gemeenten kunnen peuters zonder VVE-indicatie gebruik maken van veel minder uren (vaak de helft).

Figuur 6-5 Spreiding van het urengebruik (cumulatief vanaf 2,5 jaar)



We schatten de effecten van het cumulatieve urenaanbod op het urengebruik en de ontwikkeling van kinderen met het volgende model, waarbij we de modellen apart schatten voor doelgroepeuters en reguliere peuters:

$$y_{ict} = \alpha + \beta \cdot \text{CumulatiefAanbod}_{ict} + Z_{ict}'\pi + \mu_t + \theta_g + \varepsilon_{icgt}$$

waarin y_{ict} de uitkomst weergeeft voor kind i in voorschool c in periode t . De centrale parameter is β , het effect van extra uren voorschoolse educatie. In deze modellen gaat het om het effect van het urenaanbod en niet het effect van het urengebruik. Zoals gebruikelijk in veel beleidsanalyses, kunnen deze effecten daarom geïnterpreteerd worden als “intention-to-treat” effecten. Omdat extra aanbod niet altijd wordt gebruikt zijn deze effecten waarschijnlijk een onderschatting van de effecten van het urengebruik. In sectie 6.4 komen we hierop terug.

De modellen controleren voor Z_{ict} : kenmerken van het kind en gezin (leeftijd, geslacht, of het kind meertalig is, migratieachtergrond, opleidingsniveau van ouders, broertjes/zusjes) en kenmerken van het kindcentrum (omvang van het kindcentrum, marktaandeel van de aanbieder, aandeel centra van de aanbieder met een VE-aanbod). Daarnaast nemen we algemene tijdseffecten (μ_t) en (niet-geobserveerde) gemeentever verschillen (θ_g ; gemeente fixed effects) mee. In de analyses is het controleren voor leeftijd cruciaal, omdat dit sterk bepalend is voor het cumulatief gebruik en de ontwikkeling, en sterk samenhangt met het cumulatief urenaanbod. De standaardfouten worden geclusterd op centrumniveau.

We controleren voor leeftijdscategorieën, met voor elke leeftijdsmaand een aparte dummy variabele. Een dergelijke specificatie heeft de voorkeur boven een lineaire (of kwadratische) specificatie, omdat zo niet-lineaire effecten van leeftijd kunnen worden opgevangen. In de hoofdanalyses maken we gebruik van het leeftijdsgecorrigeerde (residualized) cumulatieve urenaanbod in plaats van het cumulatieve urenaanbod. Deze gecorrigeerde maat geeft aan in hoeverre het cumulatieve urenaanbod hoger of lager is dan verwacht gegeven een bepaalde leeftijd van het kind en gegeven de doelgroepstatus. We schatten de modellen afzonderlijk voor doelgroep- en niet-doelgroepeuters en passen deze leeftijdscorrectie daarom ook afzonderlijk toe voor de twee groepen. De belangrijkste reden om het leeftijdsgecorrigeerde cumulatieve urenaanbod te gebruiken is om multicollineariteit te voorkomen: de modellen corrigeren voor leeftijd en het cumulatief urenaanbod hangt sterk samen met leeftijd (we toetsen de modellen ook zonder deze correctie en deze laten vergelijkbare resultaten zien, zie sectie 6.4).

De centrale aanname in de identificatiestrategie is dat ouders geen invloed hebben op het cumulatieve urenaanbod (de variabele is exogeen), waardoor de resultaten kunnen worden geïnterpreteerd als causale effecten van een ruimer urenaanbod. Deze aanname is plausibel omdat het cumulatieve urenaanbod volgt uit gemeentelijk beleid, de geboortedatum en de testdatum. Hierbij is een belangrijke kanttekening dat de EVENING steekproef bestaat uit kinderen die gebruik (gaan) maken van voorschoolse educatie: de deelnemende gezinnen zijn geworven door kindcentra. Een mogelijk risico is dat systematisch andere typen ouders (en daarmee kinderen) kiezen voor deelname aan voorschoolse educatie wanneer er een ruimer urenaanbod is. Dit risico is waarschijnlijk beperkt omdat het in de regel gaat om het verlengen van een kort dagdeel, bijvoorbeeld van 2.5 naar 4 uur per dag. Dit betekent voor ouders meestal

geen fundamentele verandering. Om dit te toetsen voeren we balance tests uit. De schattingen laten zien dat het cumulatieve urenaanbod niet significant samenhangt met achtergrondkenmerken zoals het opleidingsniveau van ouders, arbeidsmarktstatus van ouders, migratieachtergrond, geslacht van het kind, en of het kind broers of zussen heeft.⁷ Deze resultaten wijzen erop dat de schattingen geïnterpreteerd kunnen worden causale effecten.

Tenslotte, in de oorspronkelijke opzet van het onderzoek zou het effect van extra urengebruik op de ontwikkeling van kinderen worden geschat met behulp van een Instrumentele Variabele (IV) methode. Deze methode is een veelgebruikte methode om causale effecten van beleid te schatten. Deze methode maakt gebruik van dezelfde variatie als hierboven beschreven, maar schat het effect van het aanbod via extra gebruik. In de eerste fase worden effecten van urenaanbod (het instrument) op het urengebruik geschat; in de tweede fase worden de effecten van het urengebruik op de ontwikkeling geschat. Deze benadering kon hier echter niet goed worden gevolgd vanwege de relatief kleine steekproef. In 6.4 bespreken we ook de resultaten op basis van de IV methode maar geven we aan dat deze resultaten terughoudend geïnterpreteerd moeten worden.

6.4. Resultaten

Effect op urengebruik

In hoeverre leidt een ruimer urenaanbod tot een hoger urengebruik? Wanneer het effect van het urenaanbod op urengebruik zwak is, is het waarschijnlijk dat er ook weinig effect is op de ontwikkeling van kinderen. Het effect van het urenaanbod op het urengebruik is daarom een cruciale schakel in de causale keten. Tabel 6-1 laat de resultaten zien voor vier verschillende samples: Panel A is voor doelgroeppeuters, Panel B voor peuters zonder VVE-indicatie; kolom (1) en (2) gaan over alle leeftijden, kolom (3) en (4) over kinderen aan het eind van de voorschoolse fase.

De resultaten voor doelgroeppeuters laten zien dat, zoals verwacht, het urenaanbod een sterk en significant effect heeft op het urengebruik. De schattingen wijzen op een toename van ruim 60 uur voor elke 100 uur extra aanbod. Het effect op het gebruik is groter wanneer we ons richten op het eind van de voorschoolse periode: ruim 90 uur voor elke 100 uur extra aanbod. Het verschil kan verklaard worden doordat de feitelijke startleeftijd een belangrijke rol speelt bij het bepalen van het cumulatieve urengebruik op jonge leeftijd. Wanneer kinderen ouder zijn heeft de startleeftijd een minder grote impact op het totaal cumulatief urengebruik, waardoor het urenaanbod een sterkere voorspeller wordt voor het urengebruik. Een klein aantal doelgroeppeuters (7) maakt gebruik van de dagopvang, en voor die groep zou het urenaanbod een beperkter effect kunnen hebben op het urengebruik. Het maakt weinig uit voor de schattingen wanneer

⁷ In lijn met de verwachting is slechts één van de twaalf coëfficiënten in Appendix Figuur G-1 is marginaal significant ($p < 0.10$).

we deze kinderen niet meenemen in de analyses. Voor reguliere peuters (Panel B) zijn de effecten een stuk kleiner. In geen van de specificaties zijn de effecten statistisch significant positief. Er wordt dus geen bewijs gevonden dat een ruimer urenaanbod voor reguliere peuters leidt tot meer urengebruik. Een verklaring hiervoor is dat relatief veel ouders ervoor kiezen minder dagdelen dan het recht af te nemen (bijvoorbeeld slechts 2 van de 4 dagdelen). Hierbij moet worden opgemerkt dat de sub-sample reguliere peuters een stuk kleiner is dan de sub-sample doelgroeppeuters waardoor de resultaten minder power hebben en minder betrouwbaar zijn.

Tabel 6-1 Effecten van cumulatief urenaanbod op urengebruik

	(1)	(2)	(3)	(4)
<i>Panel A: Doelgroep</i>				
Cumulatief urenaanbod	0.661*** (0.200)	0.612*** (0.182)	0.924** (0.365)	0.944*** (0.344)
N	262	255	181	174
R-squared	0.698	0.732	0.388	0.412
<i>Panel B: Niet-doelgroep</i>				
Cumulatief urenaanbod	0.441 (0.650)	0.159 (0.308)	-3.125* (1.718)	0.383 (1.122)
N	133	108	97	74
R-squared	0.581	0.765	0.532	0.773
Alleen peuteropvang		X		X
Eind voorschoolse periode			X	X

Noot: De analyses controleren voor kind-, gezins- en centrumkenmerken. Alle modellen nemen testjaar en gemeente fixed effects mee. Standaardfouten zijn geclusterd op centrumniveau.

***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.10$.

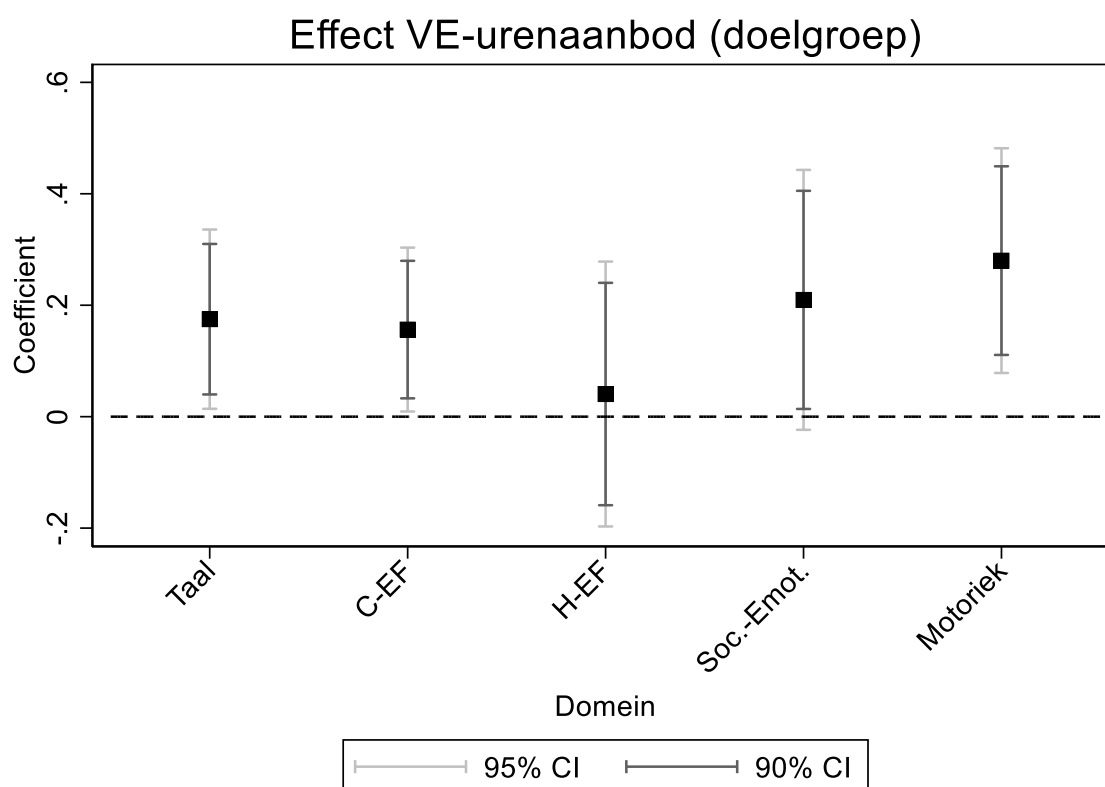
Effecten op ontwikkeling

In Figuur 6-6 (doelgroep) en Figuur 6-7 (niet-doelgroep) presenteren we de effecten van een ruimer urenaanbod op de vijf uitkomsten die centraal staan in dit onderzoek: taal, cool executieve functies, hot executieve functies, sociaal-emotionele ontwikkeling en fijne motoriek (zie Appendix G voor de bijbehorende tabellen). Voor de interpretatie van de resultaten zijn de scores getransformeerd naar z-scores, met een gemiddelde van 0 en een standaarddeviatie van 1. De coëfficiënten geven aan wat de effecten zijn van een toename van 100 uur cumulatief aanbod voorschoolse educatie.

De resultaten laten zien dat een ruimer urenaanbod een positief effect heeft op de brede ontwikkeling van doelgroeppeuters. Het effect voor taal is 0.175 SD per 100 uur extra voorschoolse educatie. Wanneer we dat vertalen naar 360 uur extra komt dit neer op 0.63 SD, wat kan worden gezien als een groot effect. Ter vergelijking: het verschil in taalscores tussen peuters met en zonder VVE-indicatie in de EVENING steekproef is 0.83. Dat zou betekenen dat met 360 uur extra voorschoolse educatie het verschil tussen doelgroep en niet-doelgroeppeuters voor een groot deel – zo'n 75 procent – wordt ingelopen. Daarbij moet worden opgemerkt dat het verschil in scores tussen peuters met en zonder VVE-indicatie in de EVENING steekproef wellicht kleiner is dan in

de gehele populatie. In de EVENING steekproef worden enkel peuters zonder VVE-indicatie meegenomen die deelnemen aan voorschools educatie; het gaat meestal om voorschoolse educatie in de peuteropvang en niet in de dagopvang. Het is denkbaar dat deze groep niet-doelgroeppeuters een achterstand heeft ten opzichte van de gemiddelde niet-doelgroeppeuter. Verder valt op dat de effecten voor meertalige doelgroeppeuters sterker zijn voor taalontwikkeling en fijne motoriek (zie Appendix G). De coëfficiënten voor cool executieve functies en sociaal-emotionele ontwikkeling zijn ook groter, maar minder precies geschat (kleinere steekproef) en statistisch niet significant.

Figuur 6-6 Effecten van VE-urenaanbod op de ontwikkeling van doelgroeppeuters



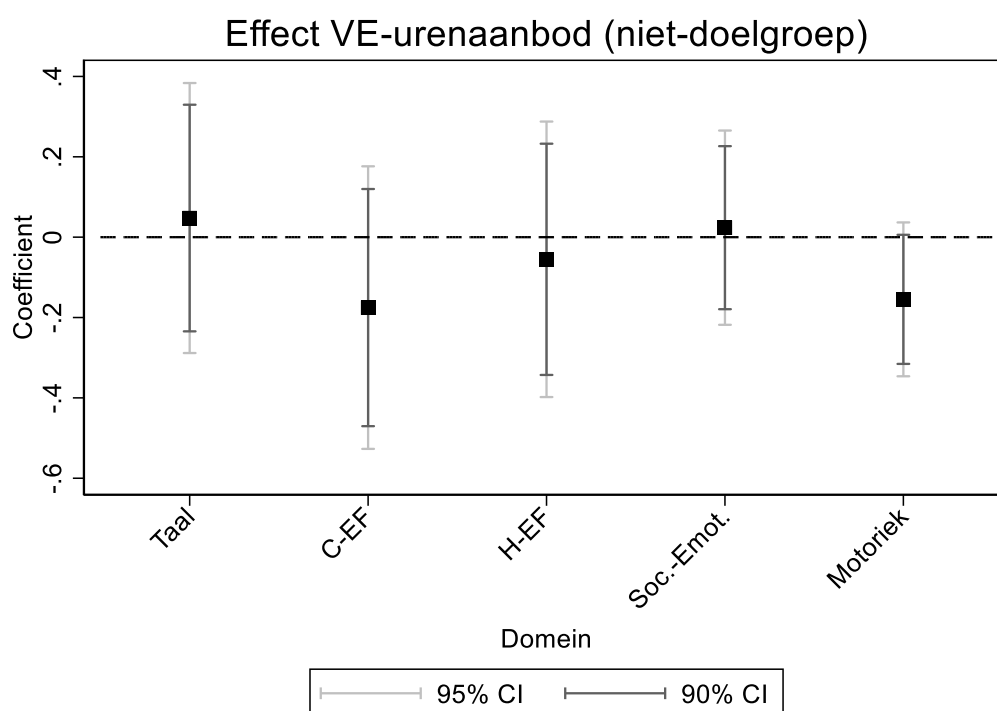
Noot: C-EF = cool (cognitieve) executieve functies; H-EF = hot (affectieve) executieve functies; Soc.-Emot. = sociaal-emotionele ontwikkeling. Uitkomsten zijn in z-scores (gemiddelde 0, SD 1). De coëfficiënten geven weer wat de effecten zijn van 100 uur extra voorschoolse educatie. De analyses controleren voor kind-, gezins- en centrumkenmerken. Alle modellen nemen testjaar en gemeente fixed effects mee. Standaardfouten zijn geclusterd op centrumniveau.

Naast effecten op taal worden ook significante effecten gevonden op cool executieve functies en sterke, significante effecten op de fijne motoriek. Ook zijn er indicaties van positieve effecten op sociaal-emotionele ontwikkeling, al zijn deze effecten minder nauwkeurig geschat ($p < 0.10$). Er worden geen effecten gevonden op de hot executieve functies. Er kunnen verschillende verklaringen zijn voor de verschillen in resultaten. In de eerste plaats is het aannemelijk dat pedagogisch medewerkers bepaalde domeinen meer stimuleren dan andere domeinen. Het stimuleren van taalontwikkeling is een centraal doel van voorschoolse educatie. Ook worden op voorschoolse educatie groepen regelmatig allerlei activiteiten uitgevoerd die de fijne motoriek stimuleren. Een

alternatieve verklaring is dat sommige domeinen accurater gemeten kunnen worden dan andere domeinen. Maar hoewel er geen significante effecten worden gevonden voor het executieve functies kan geconcludeerd worden dat extra uren voorschoolse educatie een significante bijdrage levert aan de brede ontwikkeling van doelgroeppeuters.

De schattingen voor reguliere peuters (panel B) laten geen significante effecten zien op de vijf domeinen: in geen van de specificaties is het effect statistisch significant en het teken is niet consistent in de vijf modellen. Waar positieve coëfficiënten worden gevonden (taal; sociaal-emotioneel) is de coëfficiënt veel kleiner dan in de modellen voor doelgroeppeuters. Gegeven de resultaten in Tabel 5.1 Panel B waren deze nul-effecten te verwachten: er is geen bewijs voor positieve effecten op het urengebruik van reguliere peuters, en dit effect is een centrale keten in het causale effect van een ruimer uren aanbod naar kindontwikkeling. De resultaten betekenen echter niet dat reguliere peuters geen baat hebben bij meer gebruik van voorschoolse educatie: het is alleen niet mogelijk om dit effect met de gevolgde identificatiestrategie te schatten.

Figuur 6-7 Effecten van VE-urenaanbod op de ontwikkeling van niet- doelgroeppeuters



Noot: Zie Figuur 6-7.

We hebben verschillende alternatieve modellen geschat om de robuustheid van de resultaten te toetsen (zie tabellen in Appendix G). Ten eerste, we hebben de modellen geschat met de cumulatieve aanbod variabele zonder leeftijdscorrectie. Ten tweede, zijn de modellen geschat met alleen kinderen in de peuteropvang (zonder dagopvang). Ten derde, we hebben de analyses uitgevoerd op basis van alleen de eerste

kindmetingen.⁸ Deze additionele analyses bevestigen het beeld uit de centrale resultaten: er worden positieve effecten gevonden op de ontwikkeling van doelgroeppeuters (behalve hot executieve functies), er worden sterkere effecten op taalontwikkeling van meertalige doelgroeppeuters en er worden geen effecten gevonden op de ontwikkeling van niet-doelgroeppeuters.

Wanneer een extra uur voorschools aanbod leidt tot minder dan een extra uur gebruik, is te verwachten dat de effecten van een extra uur aanbod op de ontwikkeling ook kleiner zijn dan de effecten van een extra uur gebruik. Een instrumentele variabele (IV) benadering herschaalt de effecten van het urenaanbod naar effecten van het urengebruik. Wanneer de variatie in het urenaanbod (conditioneel) exogeen is, zoals we in dit hoofdstuk beargumenteren, kan op deze manier het effect van het urengebruik worden geschat. Een belangrijke voorwaarde in de IV-methode is dat het instrument sterk genoeg is: het urenaanbod moet een sterke voorspeller zijn van het urengebruik. Hoewel het urenaanbod een significante voorspeller is van het urenaanbod is dat nog onvoldoende voor deze methode, waarschijnlijk door de relatief kleine steekproef.⁹ In Appendix G worden de effecten weergegeven van het urengebruik op de ontwikkeling van doelgroeppeuters. Gegeven de genoemde beperkingen zouden deze resultaten terughoudend geïnterpreteerd moeten worden. De resultaten komen overeen met de eerdere resultaten, maar zoals verwacht is het effect van extra gebruik groter dan het effect van extra aanbod; niet al het extra aanbod wordt immers opgenomen, zoals ook blijkt uit Tabel 6-1.

6.5. Conclusie en discussie

Het doel van de uitbreiding van het aantal uren voorschoolse educatie naar 960 uur was het vergroten van kansengelijkheid, door de (brede) ontwikkeling van doelgroeppeuters te stimuleren. Het EVENING onderzoek maakt gebruik van dit natuurlijke experiment om de effecten van een ruimer urenaanbod te toetsen op de ontwikkeling van kinderen. Omdat gemeenten verschillen in de timing van implementatie van 960 uur en het aantal voorschoolse uren voor de implementatie, ontstaat er variatie in het totale (cumulatieve) urenaanbod. De schattingen laten zien dat een ruimer urenaanbod een sterk effect heeft op het urengebruik van doelgroeppeuters. De centrale resultaten laten zien dat doelgroeppeuters baat hebben gehad van deze toename in het urengebruik. We vinden substantiële effecten op taal, (cognitieve) cool executieve functies en fijne motoriek. Ook worden er positieve effecten op de sociaal-emotionele ontwikkeling gevonden. Op (affectieve) hot executieve functies worden geen significante effecten gevonden. Er kan daarom geconcludeerd worden dat de extra investering in voorschoolse educatie een belangrijke bijdrage heeft geleverd aan het vergroten van kansengelijkheid. In tegenstelling tot de resultaten voor doelgroeppeuters vinden we

⁸ Een deel van de kinderen die rond leeftijd 2.5 jaar zijn gemeten opnieuw gemeten rond leeftijd 4 jaar. In deze alternatieve analyse is deze tweede meetronde niet meegenomen.

⁹ Een gebruikelijke indicator voor een relevant instrument is de F-test statistic van het instrument. Deze zou boven de 10 moeten zijn maar is rond de 7.

geen bewijs dat een ruimer urenaanbod voor reguliere peuters leidt tot meer urengebruik en (daarom) ook niet tot een verbetering in kinduitkomsten.

De positieve effecten op de ontwikkeling van doelgroeppeters kunnen verklaard door extra blootstelling aan voorschoolse educatie, een relatief hoge kwaliteit voorschoolse voorziening waarin doelgericht gewerkt wordt aan de ontwikkeling van peuters. Een ander mechanisme is dat het ruimere urenaanbod heeft geleid tot meer gerichte VE-activiteiten, die vaker in kleine groepjes met alleen doelgroeppeters worden aangeboden (zoals blijkt uit een eerdere EVENING studie). Met andere woorden: het aanbod is niet alleen ruimer, maar ook rijker geworden. Het is niet mogelijk met de bestaande gegevens om deze mechanismen te isoleren, maar het is plausibel dat beiden een rol hebben gespeeld. Ten slotte, de kwaliteit van voorschoolse educatie is substantieel toegenomen tussen 2021-2023. Dit kan voor een belangrijk deel worden toegeschreven aan de invoering van de norm voor de inzet van pedagogisch beleidsmedewerkers in 2022 (zie Hoofdstuk 6). Het is denkbaar dat de schattingen ook deels de effecten van hogere kwaliteit voorschoolse educatie opvangen. Of anders gezegd: de effecten waren wellicht kleiner wanneer de kwaliteit niet was toegenomen. Deze verklaring kan de resultaten echter niet volledig verklaren, omdat er enkel effecten van de pedagogisch beleidsmedewerker worden gevonden op taal.

7. Effecten van de pedagogisch beleidsmedewerker

7.1. Inleiding

Voorschoolse voorzieningen kunnen bijdragen aan de ontwikkeling van kinderen en de kansengelijkheid vergroten. Kinderen hebben echter niet altijd baat bij deelname aan voorschoolse voorzieningen. Er bestaat een brede consensus dat hoge kwaliteit een voorwaarde is voor positieve ontwikkelingseffecten van voorschoolse voorzieningen. Een belangrijke beleidsvraag is daarom hoe de kwaliteit effectief verhoogd kan worden. Een uitdaging is bovendien te zorgen voor gelijke toegang tot hoge kwaliteit, omdat – net als in bijvoorbeeld het primair onderwijs – zonder flankerend beleid het risico bestaat dat kinderen uit relatief kansarme gezinnen gebruik zullen maken van lagere kwaliteit voorzieningen. In het Nederlandse stelsel gelden additionele kwaliteitseisen (bovenop de eisen die gelden voor reguliere kinder- of peuteropvang) voor voorschoolse educatie. Kinderen uit relatief kansarme gezinnen kunnen daarom gebruik maken van relatief hoge kwaliteit voorzieningen.

In 2022 is een landelijke norm voor de inzet van de pedagogisch beleidsmedewerker ingevoerd. Het doel van deze norm is het verhogen van de kwaliteit van voorschoolse educatie. De pedagogisch beleidsmedewerker was geen nieuwe functie in de voorschoolse educatie. In veel gemeenten wordt al langer gewerkt met pedagogisch beleidsmedewerker of soortgelijke functies, en sinds 2019 (Wet IKK) bestaat er een landelijke norm voor de inzet van pedagogisch beleidsmedewerkers in de kinderopvang. De nieuwe norm stelt dat de minimum inzet van een pedagogisch beleidsmedewerker op een kindcentrum 10 uur per doelgroeppeter per jaar is. Voor centra met 16 doelgroeppeters per groep komt dat neer op 4 uur per week. In dat geval komt de inzet overeen met een kwart van de tijd die doelgroeppeters doorbrengen op de voorschool, uitgaande van een gebruik van 16 uur per week. Wettelijk heeft de pedagogisch beleidsmedewerker twee kerntaken: de 'totstandkoming en implementatie van beleidsvoornemens' en 'coaching van beroepskrachten' (zie Kader 7-1). De inzet wordt vastgelegd in het beleidsplan van het kindcentrum en toezichthouders van de GGD zien toe op het naleven van de eis.

De maatregel is in potentie kansrijk: de literatuur laat zien dat coaching en professionele ontwikkeling in (voorschoolse) educatie de kwaliteit kan verhogen en positief effecten kan hebben op kinduitkomsten (Kraft et al., 2018; Carneiro et al., 2022; Andrew et al., 2024). Echter, niet alle studies wijzen op significant positieve effecten en het is de vraag of opgeschaalde programma's vergelijkbare effecten hebben als relatief kleinschalige interventie-programma's. In Nederland bestaan er nog geen studies naar de effecten van de pedagogisch beleidsmedewerker of coaching in de voorschoolse educatie. De hervorming biedt daarom een unieke kans om de impact te toetsen in de Nederlandse context: draagt de inzet van de pedagogisch beleidsmedewerker bij aan de pedagogische kwaliteit en de ontwikkeling van kinderen?

In dit hoofdstuk presenteren we de effecten van de invoering van deze nieuwe norm op de kwaliteit, het activiteiten aanbod en de ontwikkeling van peuters. We maken daarbij gebruik van een verschillen-in-verschillen benadering, een veelgebruikte econometrische methode voor beleidsevaluaties. De educatieve proceskwaliteit is in 2021-2023 substantieel toegenomen, maar de vraag is of dat toe te schrijven is aan de pedagogisch beleidsmedewerker, andere maatregelen of een algemene trend weerspiegelt. Het doel van de verschillen-in-verschillen analyses is om het effect van de pedagogisch beleidsmedewerker te isoleren. Om een zo compleet mogelijk beeld te krijgen van de impact van de hervorming maken we gebruik van een combinatie van gegevens uit diverse bronnen: observaties op ve-groepen, vragenlijsten voor ouders, pedagogisch medewerkers en leidinggevendenden, testjes uitgevoerd tijdens huisbezoeken, CBS gegevens over buurten en LRK gegevens over kindcentra. De resultaten laten zien dat de nieuwe urennorm voor de pedagogisch beleidsmedewerker heeft bijgedragen aan het verhogen van de educatieve kwaliteit en de taalontwikkeling van doelgroeppeuters heeft versterkt.

Kader 7-1 De rol van de pedagogisch beleidsmedewerker volgens het Besluit basisvoorwaarden kwaliteit voorschoolse educatie

Artikel 2a. Inzet pedagogisch beleidsmedewerker

1. Onverminderd de bij of krachtens artikel 1.50 van de Wet kinderopvang voorgeschreven inzet van een pedagogisch beleidsmedewerker ten behoeve van de kinderopvang, zet de houder van een kindercentrum waar voorschoolse educatie wordt aangeboden een pedagogisch beleidsmedewerker in ten behoeve van de verhoging van de kwaliteit van de voorschoolse educatie.
2. De inzet van de pedagogisch beleidsmedewerker ten behoeve van de verhoging van de kwaliteit van voorschoolse educatie betreft de totstandkoming en implementatie van beleidsvoornemens met betrekking tot voorschoolse educatie of coaching van beroepskrachten voorschoolse educatie.
3. Deze inzet omvat per kindercentrum een minimaal aantal uren per jaar, dat jaarlijks wordt bepaald door het aantal kinderen waaraan in het kindercentrum op 1 januari van het betreffende jaar voorschoolse educatie wordt aangeboden te vermenigvuldigen met tien uur. Hierbij worden slechts kinderen meegeteld die tussen tweeëneenhalf en vier jaar oud zijn en behoren tot de op grond van artikel 167, eerste lid, onderdeel a, subonderdeel 2°, van de Wet op het primair onderwijs vastgestelde doelgroep.
4. De houder legt vast aan hoeveel van deze kinderen op 1 januari voorschoolse educatie wordt aangeboden.

7.2. Implementatie van de pedagogisch beleidsmedewerker

Gemeenten

We hebben met behulp van vragenlijsten voor gemeentelijke beleidsmedewerkers en beleidsdocumenten in kaart gebracht hoe het gemeentelijk beleid ten aanzien van de inzet van de pbm'er (of hbo'er meer algemeen) zich heeft ontwikkeld in de periode

2020-2023. Hier bespreken we de belangrijkste ontwikkelingen in de deelnemende gemeenten.

Een deel van de deelnemende gemeenten had voor de invoering van de nieuwe norm nog geen specifieke norm opgenomen voor de inzet van een pbm'er (of hbo'er) in de VE. Een aantal gemeenten had al voor 2022 een norm voor de inzet van de pbm'er. Dit gaat vooral om de G37, die sinds de bestuursafspraken een hbo'er in de VE hebben ingevoerd (van Teunenbroek & Vander Heyden, 2018). Er zijn echter grote verschillen in hoe die norm werd ingevuld. Zo gold in 2020 in de ene gemeente een norm van 100 uur per jaar, in andere gemeenten een norm van 4 uur per week of een norm van minimaal de helft van de dagdelen, of in weer een andere gemeente zelfs een norm van 16 uur per week. Opvallend is dat voor de invoering van de wettelijke norm er veel variatie was in gemeentelijk beleid, maar dat de meeste gemeenten in 2022 de wettelijke norm hebben overgenomen. Dat geldt ook voor veel gemeenten die al voldeden aan de wettelijke norm: wanneer de norm (minstens) 4 uur per week was, voldeden gemeenten al aan de norm. In die gevallen is de gemeentelijke norm dus minder strikt geworden. Met de invoering van de wettelijke norm lijkt er dus sprake van convergentie van gemeentelijk beleid.

We zien ook een verschuiving van een gemeentelijke norm die onafhankelijk is van het aantal doelgroeppetters (bijvoorbeeld: nog geen norm; 4 uur per week; 100 uur per jaar) naar een norm die sterk afhankelijk is van het aantal doelgroeppetters (de nieuwe wettelijke norm). In bepaalde centra kan er dus sprake zijn van een daling van de inzet van de pbm'er. Een opvallende casus is Rotterdam: dit is de enige deelnemende gemeente met stabiel ve-beleid ten aanzien van de inzet van de pbm'er/hbo'er (4 uur per week, onafhankelijk van het aantal doelgroeppetters). Verder lijkt er ook een ontwikkeling van een urennorm per week voor een hbo'er op de groep (intellig) naar een norm per jaar voor een boventallige coach. Deze overgang was meestal direct, met de invoering van de wettelijke norm. In Amsterdam bijvoorbeeld was deze verschuiving meer geleidelijk: daar gold in 2020 een norm van 16 uur per week; in 2021 een norm van 16 uur per week of 100 uren coaching per jaar; in 2022 een norm van 16 uur per week of minimaal 80 uren coaching per jaar, oplopend met 10 uur per doelgroepkind; en in 2023 minimaal 80 uren coaching per jaar, oplopend met 10 uur per doelgroepkind.

Aanbieders

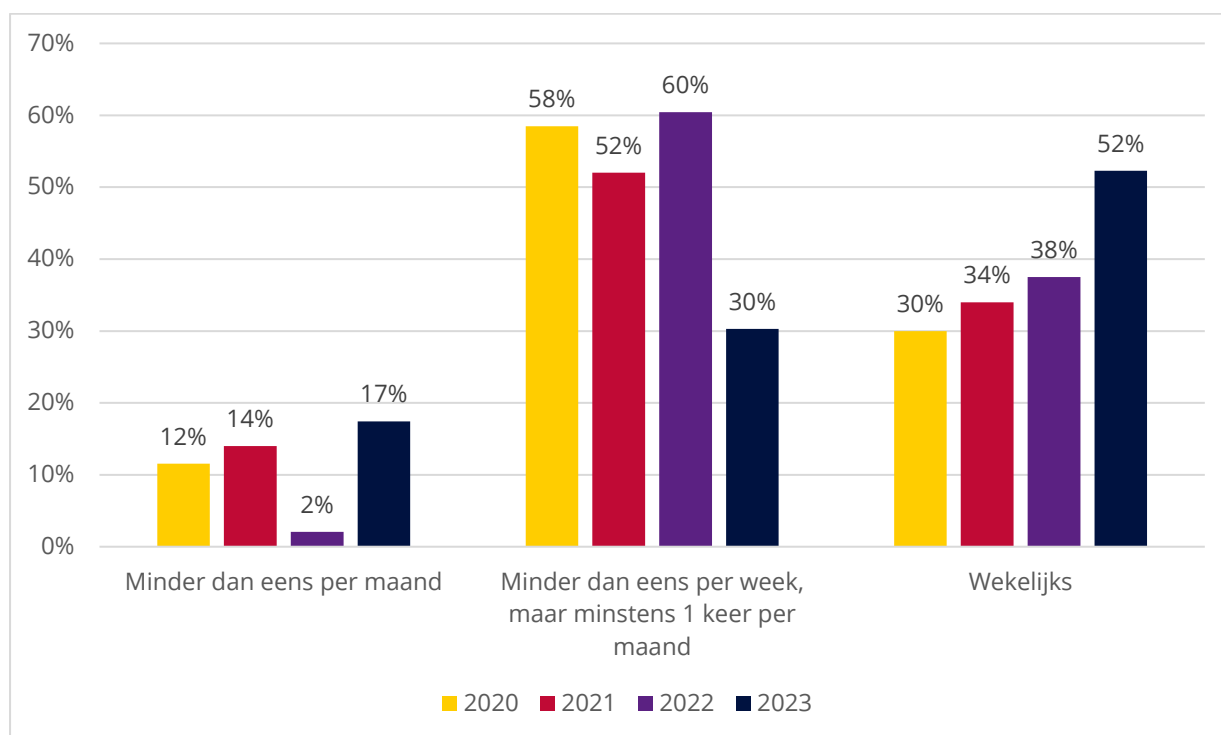
Inzet pedagogisch beleidsmedewerker

Om de inzet van pbm'ers in de voorschoolse educatie te meten, hebben we een aantal vragen gesteld aan de leidinggevendenden. Dit hebben we gedaan tijdens alle drie meetronden. In de periode van de tweede meetronde (tussen najaar 2021 en voorjaar 2022) werd de wettelijke urennorm voor de inzet van de pbm'er ingevoerd. Tijdens deze meetronde was het merendeel van de vragenlijsten (64%) ingevuld in de periode september tot en met december 2021, dus vóór de invoering van de wettelijke urennorm. De data uit de derde meetronde zijn verzameld na de invoering van de urennorm. Hier beschrijven we de data uit de derde meetronde en vergelijken we de

bevindingen met het beeld uit de eerdere meetronden. Bij het bespreken van de tweede meetronde, maken we onderscheid tussen data verkregen vóór en na de invoering van de urennorm.

In lijn met de wet IKK, in 2022-2023 rapporteerden alle 75 leidinggevendenden die de vragen over pbm'ers invulden, dat ze één (79%) of meerdere (21%) pbm'ers in dienst hadden. Deze verhoudingen komen overeen met de verhoudingen in onze vorige rapporten (in 2021-2022 had 75% één en 25% een team van pbm'ers en in 2020 was dat 70% en 28% en nog eens 2% die geen pbm'er in dienst hadden). In 2022-2023 was van bijna alle pbm'ers (78%) het opleidingsniveau hbo of wo. Bij een kleinere groep (17%) was dat mbo-4 in combinatie met aanvullende scholing of een associate degree. Deze verdeling is vergelijkbaar tussen locaties waar één pbm'er werkzaam is en locaties waar een team werkt. Dit is vergelijkbaar met de gegevens uit voorgaande metingen, behalve dat we in 2021-2022 zagen dat wanneer er één pbm'er werd ingezet op de locatie, het opleidingsniveau van de pbm'ers bijna altijd (94%) hbo of wo was. Wanneer er een team van pbm'ers werd ingezet op de locatie, was vaker het opleidingsniveau van tenminste één van de pmb'ers mbo-4 in combinatie met aanvullende scholing.

Figuur 7-1 Frequentie inzet van de pedagogisch beleidsmedewerker



Noot: De gegevens zijn gebaseerd op de volgende aantallen: 2020 (N=130); najaar 2021 (N=100); januari-juni 2022 (N=48); september 2022-mei 2023 (N=109).

Als we kijken naar de mate waarin de pbm'er wordt ingezet op de voorschoolse groepen vinden we grote verschillen. Figuur 7-1 laat de frequentie zien voor de drie meetronden. De data uit de tweede meetronde zijn onderverdeeld naar de locaties waarvoor de vragenlijst in 2021 (vóór de invoering van de urennorm) is ingevuld en locaties waarvoor

de vragenlijst in 2022 of 2023 (na de invoering van de urennorm) is ingevuld. De Figuur laat zien dat tijdens de eerste twee meetronden, de inzet in de meeste groepen minstens één keer per maand, maar minder dan één per week was. Tijdens de laatste meetronde is de inzet in de meeste groepen (52%) minstens één keer per week. De groepen waar de inzet lager was (minder dan een keer per maand) waren uitsluitend groepen waar zowel peuters met VE-indicatie als zonder VE-indicatie zitten (89%) of kinderopvanggroepen (11%). Het algehele beeld is dat de frequentie van de inzet is gestegen na de invoering van de nieuwe urennorm.

Wanneer de pbm'er wekelijks wordt ingezet op een groep, was dat gemiddeld gedurende 6.4 uur per week. Het aantal uren varieerde tussen 1 uur en 32 uur per week. Op 32% van de groepen was de inzet meer dan 4 uur inzet per week. Dit is vergelijkbaar met de gegevens uit 2022 (6.4 uur per week) en hoger dan in 2021 en 2020 (inzet van gemiddeld 4.6 uur per week). In de groepen waar de inzet van de pbm'er minimaal één keer per maand was, is het gemiddelde aantal uren per maand 4.9. Dit is gemiddeld lager dan in 2021 (5.4) en in 2022 (7.7) en 2020 (5.9). Bij 40% van de groepen is de inzet 4 uur of meer per maand, wat neerkomt op ongeveer één uur per week. De groepen met een lagere inzet zijn doorgaans gemengde groepen van peuters met en zonder VE-indicatie of groepen in de kinderdagopvang. Deze bevinding is vergelijkbaar met voorgaande jaren. In ruim 88% van de groepen (vergeleken met 80% in 21-22 en 90% in 2020) was de inzet van de pbm'er boventallig, dat wil zeggen, de pbm'er telt niet mee voor de beroepskracht-kind ratio.

Op de meeste deelnemende centra (63%) wordt de pbm'er op elke groep evenveel ingezet. Dit percentage was in de voorgaande meetronden hoger (ca. 80%). Bij de overige groepen werd de pbm'er vooral ingezet in de VE-groepen met een hoger aantal peuters met een VE-indicatie en in groepen waar een kwaliteitsverbetering nodig werd geacht. Dit sluit aan bij de verandering in gemeentelijk beleid, waarbij er een verschuiving heeft plaatsgevonden van onafhankelijke inzet naar inzet afhankelijk van het aantal doelgroeppetters op de locatie. Ook gaf een aantal leidinggevendenden aan dat de inzet van de pbm'er meer maatwerk was of dat de begeleiding van kinderen op de groep werd gedaan door VE-specialisten en dat de pedagogisch coach vooral werd ingezet voor coaching van pedagogisch medewerkers. Deze overige vormen van inzet zijn in lijn met het beeld uit eerdere meetronden. Dit laat zien dat relaties tussen aan de ene kant de inzet van pbm'ers (of hbo'ers in het algemeen) en aan de andere kant de geboden kwaliteit van de voorziening en mogelijk ook de ontwikkeling van kinderen complex zijn. Relaties tussen inzet van pbm'ers en relevante uitkomsten (kwaliteit, ontwikkeling van het kind) kunnen daarom niet worden geïnterpreteerd als causale effecten. Gegeven de selectieve inzet zou, zonder controle voor het selectie-effect, de relatie tussen de inzet van een pbm'er en de geboden kwaliteit zelfs negatief kunnen zijn, zelfs als het effect in werkelijkheid positief is. In de analyses (sectie 6.3) maken we daarom gebruik van het natuurlijk experiment dat met de invoering van de wettelijke norm is ontstaan.

Taken pedagogisch beleidsmedewerker

We hebben leidinggevenden ook gevraagd naar de taken van de pbm'er. De helft (53%) van de pbm'ers combineert coaching met beleidsontwikkeling en implementatie. Ongeveer 24% richt zich volledig op coaching en 20% op beleidsontwikkeling en implementatie. Bij ongeveer 14% van de locaties zijn er aparte functies voor deze twee aspecten. Tijdens de eerdere metingen was de verdeling van het takenpakket redelijk vergelijkbaar.

De meerderheid van de pbm'ers (54%) heeft geen andere overstijgende taken. Dit is vergelijkbaar met 2021-2022 (58%) en iets hoger dan in 2020 (44%). Van de overige 46% hebben de meesten taken als VE-coördinator of VE-specialist (16%). Een kleinere groep werkt ook als intern begeleider, is betrokken bij oudercontacten, doet administratief werk en planning, of heeft leidinggevende taken. Deze taakverdeling is redelijk vergelijkbaar met voorgaande meetronden. Ten slotte, op 44% van de locaties zijn de pbm'ers de enige medewerkers met een hbo-opleiding. Op de andere locaties werken ook andere medewerkers met een hbo-opleiding. Dit zijn meestal VE-coördinatoren of VE-specialisten (16%), pedagogische medewerkers met een hbo-diploma (14%), intern begeleiders of zorgcoördinatoren (14%), pedagogische coaches (9%) of leidinggevenden (2%). Dit beeld is redelijk vergelijkbaar met voorgaande metingen. Wel lijkt er een kleine verschuiving te zijn in het opleidingsniveau: er werken iets minder pedagogische medewerkers met hbo-opleiding op de groepen. Het is denkbaar dat deze pedagogisch medewerkers nu worden ingezet als pedagogisch beleidsmedewerker of coach.

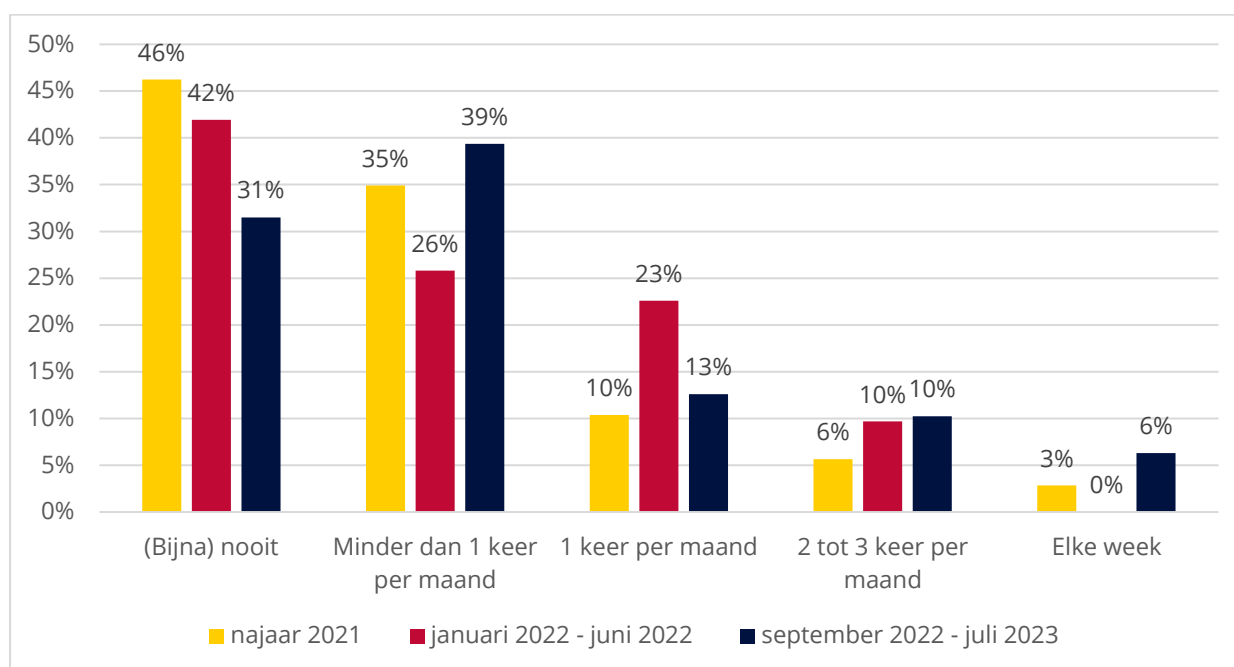
Een belangrijk doel van het beleid rondom de inzet van pbm'ers is de professionalisering van pedagogisch medewerkers door begeleiding op de werkvloer. Pedagogisch medewerkers zijn door middel van de online vragenlijst ook gevraagd naar hun ervaring met coaching. De resultaten laten zien dat in de periode januari 2022-juli 2023 ruim 70% van de pedagogische medewerkers regelmatig of vaak begeleiding hebben gekregen op de werkvloer van een pedagogische coach. In het najaar van 2021 was dat lager, ruim 60%. Hoewel coaching dus geen nieuw fenomeen is in de voorschoolse educatie is er wel een toename waarneembaar.

Hoewel begeleiding door een ervaren collega of leidinggevende een alternatief zou kunnen zijn voor de begeleiding door een coach, blijken de verschillende vormen van begeleiding positief gecorreleerd te zijn. Pedagogisch medewerkers die vaker door een coach worden begeleid, geven aan ook vaker door een ervaren collega of door de leidinggevende te worden begeleid. Een mogelijke verklaring hiervoor zou kunnen zijn dat de ervaren collega of leidinggevende dezelfde persoon is als de pedagogisch coach. Als we de drie typen begeleiding samen bekijken, valt op dat ongeveer 20% van de pedagogisch medewerkers (bijna) nooit of zelden begeleiding op de werkvloer krijgt. Dit percentage is vergelijkbaar met het percentage van de meetronde van 2021-2022, maar lager dan in 2020 (toen kreeg bijna 30% naar eigen zeggen geen begeleiding). Een mogelijke verklaring is dat de begeleidingstijd is toegenomen als gevolg van de verschillende maatregelen. Het lijkt er op dat een aanzienlijk deel van de pedagogisch medewerkers (bijna) nooit of zelden begeleiding krijgt, maar daarbij moet worden opgemerkt dat een aanzienlijk deel van de deelnemende pedagogisch medewerkers

een hbo-diploma heeft (bijna 38%, zie Tabel 3-1). Een deel van de deelnemende pedagogisch medewerkers is zelf de specialist of coach. De coaching kan dus meer gericht zijn op andere medewerkers.

Naast een item over begeleiding worden pedagogisch medewerkers ook meer expliciet gevraagd naar de frequentie van coaching als onderdeel van professionaliseringsactiviteiten binnen het team op het centrum: “Geobserveerd worden door een coach of leidinggevende en feedback op je werk krijgen”. Dit item is relevant omdat er ook gevraagd wordt naar feedback, een essentieel onderdeel van coaching. In Figuur 7-2 zijn geen grote verschuivingen te zien, maar het is wel opvallend dat het percentage dat (bijna) nooit coaching ervaart is gedaald van 46% in het najaar van 2021 naar 42% in het voorjaar van 2022 en later naar 31% vanaf najaar 2022. Ook deze cijfers wijzen dus op een toename van coaching.

Figuur 7-2 Frequentie geobserveerd worden door een coach/leidinggevende en feedback krijgen



Noot: De gegevens zijn gebaseerd op de volgende aantallen: najaar 2021 (N=106); januari-juni 2022 (N=31); september 2022-juli 2023 (N=127).

De beschrijvende resultaten laten zien dat er in alle drie meetronden veel variatie is in de inzet van de pbm'er. Dit kan verklaard worden door verschillen in gemeentelijk beleid, verschillen in aantal doelgroeppeuters en verschillen in leeftijden (de wettelijk norm geldt voor kinderen tussen de 2.5 en 4 jaar). Desalniettemin lijkt het erop dat sinds de invoering van de nieuwe norm in 2022 de inzet van de pbm'ers is toegenomen. De inzet lijkt in de laatste meetronde bovendien meer bepaald door het aantal doelgroepkinderen, waarbij de inzet hoger is op groepen met meer doelgroepkinderen. Dat sluit aan bij de wettelijke norm en de ontwikkelingen in gemeentelijk beleid, zoals hierboven beschreven. In de laatste meetronde lijkt ook de frequentie van begeleiding door een pedagogische coach te zijn toegenomen ten opzichte van de metingen vóór en

kort na de invoering van de nieuwe norm. Dit zou op een effect van de nieuwe norm kunnen duiden. Tenslotte is er een kleine verschuiving in hoe hbo'er worden ingezet. Er zijn minder hbo-opleide pedagogische medewerkers werkzaam naast de pbm'ers. Dit kan een resultaat zijn van de maatregel, waardoor meer hbo-opgeleide medewerkers zijn gaan werken als pbm'er of coach in plaats van als pedagogische medewerkers op de groep.

Verdiepend onderzoek

In het EVENING onderzoek is ook verdiepend (kwalitatief) onderzoek uitgevoerd onder pbm'ers na de invoering van de nieuwe norm (zie Wierdsma & Jepma, 2024). Het onderzoek laat zien dat de pbm'er op verschillende manieren wordt ingezet om de kwaliteit te verbeteren. We zien dat de pbm'ers vaak nog zoekende zijn naar hun rol, verantwoordelijkheden en taken. De functie is erg breed en de invulling divers. De pbm'ers voeren vaak een breed scala aan werkzaamheden uit op het gebied van beleid, pedagogiek, kwaliteit, coaching, wat kan leiden tot rol- en taakverwarring leidt. Verder valt op dat er grote verschillen zijn in de caseload in termen van centra, groepen en pedagogisch medewerkers. Hierdoor is er veel variatie in de tijd die beschikbaar is voor coaching. Uit focusgesprekken blijkt dat het cruciaal is dat er voldoende tijd en ruimte nodig is om een vertrouwensrelatie met de pedagogisch medewerkers op te bouwen. Niet altijd zijn die randvoorwaarden aanwezig. De pedagogisch beleidsmedewerkers werken aan het professionaliseren van pedagogisch medewerkers en het verbeteren van de kwaliteit. De indruk is dat dit vaak niet altijd heel systematisch gebeurt, bijvoorbeeld op basis van een individueel coachingsplan met duidelijke ontwikkeldoelen en feedback. Hier lijken nog kansen om de impact van de pedagogisch beleidsmedewerkers te vergroten, bijvoorbeeld via landelijke richtlijnen over evidence-informed werken.

7.3. Identificatiestrategie

Uitkomsten voor ve-groepen

Voor het schatten van de effecten van de pedagogisch beleidsmedewerker maken we gebruik van de landelijke hervorming, die de mogelijkheid van een natuurlijk experiment heeft gecreëerd. Voor de hervorming was er in veel gemeenten nog geen norm voor de inzet van de pedagogisch beleidsmedewerker. Sommige gemeenten voldeden al aan de nieuwe wettelijke norm. Echter, de inzet van de pedagogisch beleidsmedewerker was voor de hervorming niet afhankelijk van het aantal doelgroepkinderen op de locatie. Sinds 2022 hebben nagenoeg alle gemeenten de wettelijke norm overgenomen, of strikter beleid ingevoerd waarbij de inzet ook afhankelijk is van het aantal doelgroepkinderen.¹⁰

¹⁰ In één gemeente was tussen 2020-2023 de norm voor de inzet 4 uur per week, onafhankelijk van het aantal doelgroepkinderen. Deze gemeente voldeed dus al aan de wettelijke norm voor 2022 en de nieuwe wettelijke norm heeft geen invloed gehad op het beleid van deze gemeente. Omdat de hervorming geen

Om het effect van de norm voor pedagogisch beleidsmedewerkers te identificeren gebruiken we een verschillen-in-verschillen benadering, waarbij we gebruik maken van de variatie in het aandeel doelgroeppeuters tussen verschillende kindcentra. Omdat het aantal doelgroeppeuters bepalend is voor de minimale inzet van pedagogisch beleidsmedewerkers, verwachten we een sterker effect van de hervorming in centra met een hoger aandeel doelgroeppeuters ten opzichte van centra met een lager aandeel. Dit quasi-experimentele design stelt ons in staat om veranderingen in uitkomsten voor VE-groepen in centra met een hoog aandeel doelgroeppeuters vóór en na de beleidswijziging te vergelijken met veranderingen in uitkomsten voor VE-groepen met een laag aandeel doelgroeppeuters in dezelfde periode. Met andere woorden, we vergelijken de veranderingen in kwaliteit in groepen die meer werden beïnvloed door de hervorming met de veranderingen in kwaliteit in groepen die minder (of niet) werden beïnvloed door de invoering van de nieuwe landelijke norm. Formeel schatten we het volgende model:

$$q_{ict} = \alpha + \delta(AandeelDoelgroep_c * post_t) + \gamma(AandeelDoelgroep_c * impl_periode_t) + \beta AandeelDoelgroep_c + X_{ict}'\pi + \mu_t + \theta_g + \varepsilon_{ict}, \quad (7.1)$$

waarbij q_{ict} de uitkomst weergeeft van groep i in kindcentrum c in periode t . De focus van de analyses ligt op de educatieve en emotionele proceskwaliteit, maar we kijken ook naar effecten op het activiteitsaanbod en de wijze waarop VVE-activiteiten worden uitgevoerd. De parameter δ is de verschillen-in-verschillen parameter die het effect van de pedagogisch beleidsmedewerker weergeeft. $AandeelDoelgroep_c$ is een indicator voor het aandeel doelgroeppeuters in de wijk en is een maat voor "treatment intensity".¹¹ We maken gebruik van gegevens van het CBS, die informatie heeft over het percentage doelgroeppeuters op buurniveau voor elke voorschool. Op basis van de postcode van elke voorschool worden gegevens gematcht met CBS data. In de analyses gebruiken we naast deze continue treatment intensiteit maat als alternatief een dummy (0/1) die aangeeft of de voorschool in een buurt gevestigd is met een percentage doelgroeppeuters boven de mediaan (boven of onder 24%). Op deze manier evalueren we het effect door de ontwikkeling te vergelijken in voorscholen met een relatief hoog aandeel doelgroeppeuters versus een relatief laag aandeel.

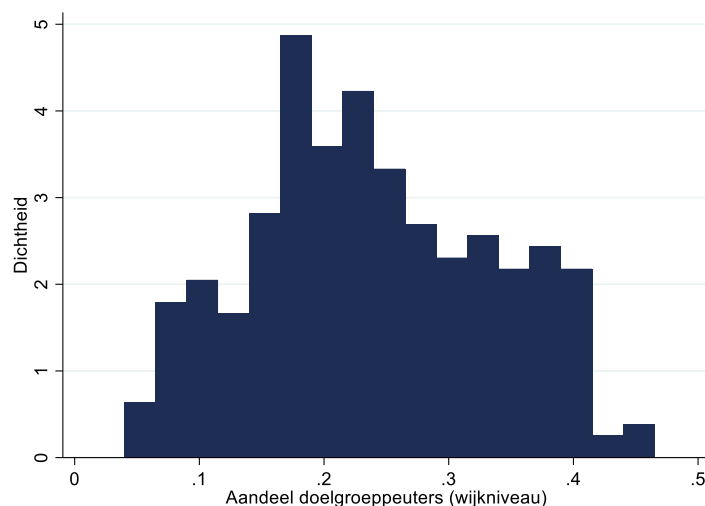
Figuur 7-3 geeft de verdeling aan van het aandeel doelgroeppeuters per wijk waarin EVENING centra gevestigd zijn. Volgens de CBS gegevens is in veel wijken waar EVENING centra zijn, het aandeel ruim boven de 20 procent, terwijl volgens de OCW norm 15 procent van de peuters tot de doelgroep behoort in heel Nederland. Omdat voorscholen vaker in wijken te vinden zijn met relatief veel doelgroeppeuters is het ook te verwachten dat de deelnemende voorscholen in wijken staan met een bovengemiddeld aandeel doelgroeppeuters.

effect heeft gehad op het gemeentelijk beleid wordt deze gemeente niet meegenomen in de hoofdanalyses.

¹¹ Hieronder bespreken we hoe het aandeel in de wijk samenhangt met het aantal doelgroeppeuters in de voorschool.

In de analyses gebruiken we naast deze continue treatment intensiteit maat als alternatief een dummy (0/1) die aangeeft of de voorschool in een buurt gevestigd is met een percentage doelgroeppeuters boven de mediaan (boven of onder 24%). Op deze manier evalueren we het effect door de ontwikkeling te vergelijken in voorscholen met een relatief hoog aandeel doelgroeppeuters versus een relatief laag aandeel.

Figuur 7-3 Aandeel doelgroeppeuters op wijkniveau (CBS)



Voor een groot deel van de EVENING centra hebben we ook gegevens over het daadwerkelijke aantal doelgroeppeuters in de voorschool. De gegevens zijn afkomstig uit EVENING vragenlijsten aan leidinggevendenden, gegevens uit GGD rapporten (het aantal doelgroeppeuters wordt in een deel van de rapporten vermeld) en een uitvraag bij aanbieders. Deze cijfers zijn echter niet compleet en de aantallen uit de verschillende bronnen komen niet altijd overeen met elkaar. Omdat we deze gegevens niet hebben voor alle centra en er onzekerheid is over de aantallen bij een deel van de centra gebruiken we deze data niet direct om de treatment intensity te meten. We gebruiken de gegevens wel om te verifiëren of het aandeel doelgroeppeuters per wijk een valide maat is van treatment intensiteit (zie Tabel 7-1). Zoals verwacht hangt het aantal doelgroeppeuters per centra significant samen met het aandeel doelgroeppeuters per wijk: voor elke 10 procentpunt toename in het percentage doelgroeppeuters in de wijk neemt het aantal doelgroeppeuters in het kindcentrum toe met ruim 4 ($p < .010$). Centra in wijken met veel doelgroeppeuters blijken ook vaker meerdere ve-groepen te hebben; dat kan verwacht worden omdat er in die wijken meer vraag is naar voorschoolse educatie. Deze resultaten bevestigen dat het aandeel doelgroeppeuters een relevante indicator is voor de treatment intensiteit: omdat het beleidseffect sterker is in centra met meer doelgroeppeuters, is te verwachten dat de effecten van de beleidsmaatregel gemiddeld groter zijn in centra die in wijken gevestigd zijn met relatief veel doelgroeppeuters.

Tabel 7-1 Relatie tussen aantal doelgroepeuters in centra en wijken

	Aantal doelgroepeuters (centrum)	Aantal ve-groepen (centrum)
Aandeel doelgroepeuters (wijkniveau)	42.856*** (9.993)	1.844*** (0.423)
N (centra)	114	145

Noot: Standaardfouten zijn geclusterd op kindcentrum-niveau.

*** significant op 1%; ** significant op 5%; * significant op 10%.

Hoewel de beleidsverandering per januari 2022 is ingegaan, zien we de periode van de derde meetronde (vanaf het najaar 2022) als de periode post-hervorming. De tweede meetronde heeft vooral plaatsgevonden in het najaar van 2021, maar deels ook het voorjaar van 2022 (op een enkele observatie in mei na in januari-april 2022).¹² De verwachting is dat in de eerste maanden van 2022 de extra inzet van de pedagogisch beleidsmedewerker geïmplementeerd wordt. Het is niet waarschijnlijk dat er al een duidelijk effect van coaching zichtbaar is op kwaliteit en de ontwikkeling van kinderen. Dat wil zeggen: hoewel de maatregel per 1 januari 2022 is ingegaan zullen pedagogisch medewerkers niet direct vanaf die datum op een fundamenteel andere manier zijn gaan werken. De verwachting is dat coaching enige tijd kost om effect te genereren. De eerste maanden van 2022 beschouwen we daarom als implementatie-periode. We nemen daarom ook de interactie ($AandeelDoelgroep_c * impl_periode_t$) mee, maar interpreteren deze coëfficiënt niet als beleidseffect.

Het model neemt tijd fixed effects (μ_t) en gemeente fixed effects (θ_g) mee, zodat gecontroleerd wordt voor niet-geobserveerde algemene tijdseffecten en niet-geobserveerde stabiele gemeentever verschillen. Omdat we een groot deel van de voorschoolse educatie groepen twee keer observeren in de periode 2021-2023, schatten we in een alternatieve specificatie het effect ook met groep fixed effects. Deze schattingen maken enkel gebruik van variatie binnen groepen over de tijd (op basis van herhaalde kwaliteitsmetingen van dezelfde groep). Tenslotte wordt gecontroleerd voor een aantal kenmerken van het centrum (X_{cg}): de omvang van het centrum (aantal kindplaatsen en aantal VE-groepen); het lokaal en nationaal marktaandeel kinderopvang van de aanbieder; het aandeel kindcentra van de aanbieder waar voorschoolse educatie wordt geboden.¹³ Standaardfouten worden geclusterd op centrumniveau.

Een belangrijk voordeel van de verschillen-in-verschillen methode is dat gecontroleerd wordt voor niet-geobserveerde gemeente- en tijdseffecten. Een eenvoudige analyse die enkel een voor- en nameting vergelijkt controleert niet voor de mogelijkheid dat kwaliteit (of een andere uitkomst) veranderd kan zijn om een andere reden. Voor

¹² Op een enkele observatie in mei 2022 na vonden alle observaties in het voorjaar 2022 plaats in januari-april 2022.

¹³ Deze gegevens zijn gehaald uit het Landelijk Register Kinderopvang (LRK), GGD inspectierapporten en vragenlijsten aan leidinggevenden.

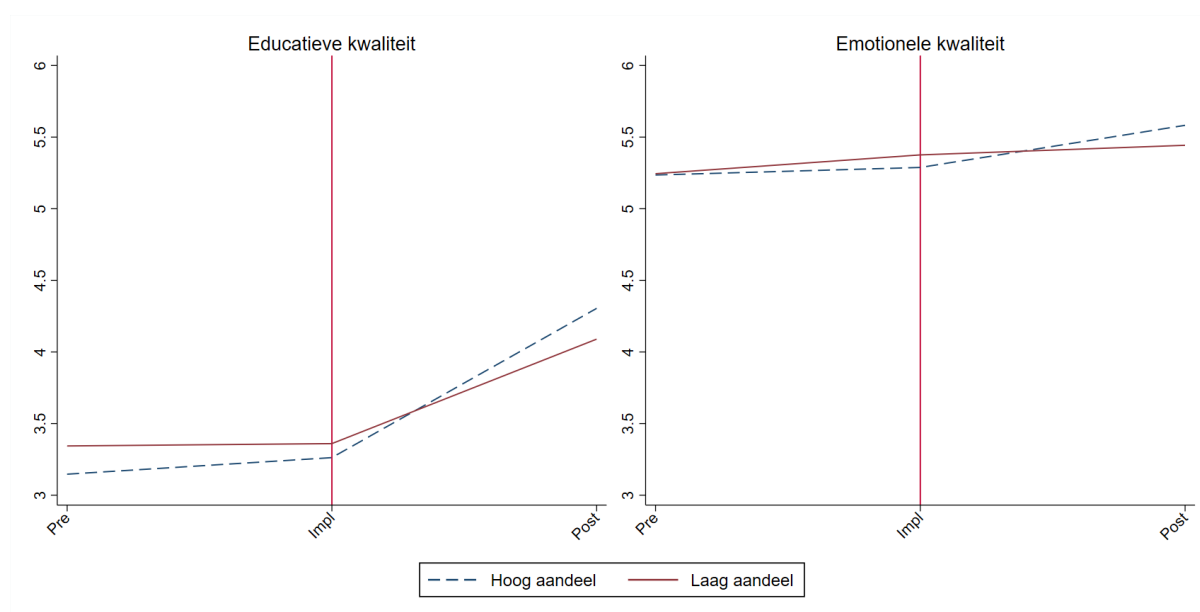
beleidsevaluaties is dit van belang omdat algemene niet-geobserveerde tijdseffecten uitgesloten moeten worden. In de context van dit onderzoek hebben deze tijdseffecten bijvoorbeeld te maken met personeelstekorten. Wanneer personeelstekorten bijvoorbeeld toenemen over de analyse-periode zou dat negatieve gevolgen kunnen hebben voor de kwaliteit, waardoor het effect van het beleid wordt onderschat. Ook is het denkbaar dat er sprake is van herstel na een dip in de kwaliteit tijdens de COVID-19 periode, waardoor het effect juist wordt overschat in een eenvoudige vergelijking van een voor- en nameting. Een ander belangrijk voordeel van de verschillen-in-verschillen methode is dat gecontroleerd wordt voor stabiele, niet-geobserveerde verschillen tussen gemeenten. Gemeenten met relatief veel centra met een hoog aandeel doelgroeppeuters kunnen ook niet-geobserveerde kenmerken hebben die samenhangen met kwaliteit. Zo is het denkbaar dat die gemeenten meer investeren in voorschoolse educatie.

Omdat de verschillen-in-verschillen methode controleert voor niet-geobserveerde tijdseffecten en niet-geobserveerde stabiele gemeentever verschillen kunnen de beleidseffecten van de pedagogisch beleidsmedewerker beter worden geïsoleerd. Het beleidseffect wordt geschat op basis van de interacties tussen het aandeel doelgroeppeuters (de treatment variabele) en een tijdseffect (periode post-hervorming). De centrale aanname in de verschillen-in-verschillen strategie is dat de controle- en treatment-centra zich parallel zouden hebben ontwikkeld wanneer de landelijke norm niet zou zijn ingevoerd (de parallele trend aanname). Verschillende toetsen laten zien dat deze aanname plausibel is. Zo zijn er voor de hervorming geen significante verschillen in educatieve en emotionele proceskwaliteit (noch in alle sub-dimensies van kwaliteit) tussen centra met een hoog versus laag aandeel doelgroeppeuters. Ook zijn er geen verschillen in de achtergrondkenmerken van pedagogisch medewerkers (leeftijd, ervaring, opleidingsniveau, migratieachtergrond, arbeidsuren) tussen centra met een hoog versus laag aandeel doelgroeppeuters voor de hervorming. Hieronder bespreken we meer in detail resultaten die suggereren dat de parallele trend aanname plausibel is.

Figuur 7-4 laat de ontwikkeling van educatieve en emotionele proceskwaliteit zien tussen 2021 en 2023 in centra in wijken met een laag versus hoog percentage doelgroeppeuters. Zoals beschreven in Hoofdstuk 2 is in de periode 2021-2023 de proceskwaliteit substantieel toegenomen, met name de educatieve proceskwaliteit. De toename is zichtbaar zowel in centra met een hoog als met een laag aandeel doelgroeppeuters. Het is mogelijk dat de invoering van de nieuwe norm voor de pedagogisch beleidsmedewerker een bijdrage heeft geleverd aan de ontwikkeling van de kwaliteit in beide typen centra. Echter, het kan verwacht worden dat de kwaliteit meer is toegenomen in centra met een hoog aandeel doelgroeppeuters. De figuur laat zien dat de educatieve kwaliteit in centra met een hoog aandeel doelgroeppeuters gemiddeld lager is in 2021, maar in het najaar 2022 is dat verschil ingelopen en scoren deze centra gemiddeld juist hoger. De ontwikkeling van emotionele kwaliteit is veel stabiel en laat weinig verschillen zien tussen de twee typen centra. Deze patronen duiden op een positief effect van de pedagogisch beleidsmedewerker op de educatieve

kwaliteit, maar geen effect op de emotionele kwaliteit. In dit hoofdstuk beschrijven we de resultaten van de verschillen-in-verschillen analyses.

Figuur 7-4 De ontwikkeling van educatieve proceskwaliteit (links) en emotionele proceskwaliteit (rechts)



Noot: de figuur geeft de ontwikkeling van educatieve en emotionele proceskwaliteit aan, tijdens de periode pre-hervorming (Pre; voor 2022), implementatie (Impl; voorjaar 2022) en de post-hervorming (vanaf najaar 2022).

De centrale aanname in verschillen-in-verschillen analyses is de parallelle trend aanname. In de context van dit onderzoek betekent deze aanname dat de ontwikkeling van uitkomsten zoals proceskwaliteit in centra met een hoog versus laag aandeel doelgroeppeuters vergelijkbaar zou zijn zonder de invoering van de norm voor de pedagogisch beleidsmedewerker in 2022. Met andere woorden: er hebben zich geen andere specifieke “schokken” voorgedaan die (meer) invloed hebben gehad op een centrum met een hoog (of juist laag) aandeel doelgroeppeuters. Een relatief sterke groei van kwaliteit in centra met een hoog aandeel doelgroeppeuters kan dan worden toegeschreven aan de hervorming. We hebben onvoldoende gegevens uit eerdere jaren om te toetsen of er structureel verschillen zijn in de ontwikkeling van proceskwaliteit tussen de verschillende typen centra. Echter, voor de hervorming is de kwaliteit redelijk vergelijkbaar tussen de twee typen centra. Verder zien we geen significante verschillen in centrum- en aanbieder-kenmerken tussen de twee typen centra (zoals aantal kindplaatsen in het centrum, het gemeentelijk en landelijk aandeel kindplaatsen). Ook zijn er geen verschillen in kenmerken tussen de beroepskrachten (bijvoorbeeld in termen van werkervaring, opleidingsniveau, migratieachtergrond, arbeidsuren). Tenslotte, zoals we hieronder beschrijven, blijken kenmerken van beroepskrachten en de samenstelling van de groepen niet op een verschillende manier te veranderen. Gecombineerd suggereren deze resultaten dat de parallelle trend aanname plausibel is.

Ontwikkeling van kinderen

Om de effecten van de hervorming op de ontwikkeling van kinderen te schatten, hanteren we een verschillen-in-verschillen benadering in lijn met de hierboven beschreven empirische aanpak. De verwachting is dat er sterkere positieve effecten worden gevonden voor kinderen in voorscholen met een hoger aandeel doelgroeppeuters. We vergelijken daarom veranderingen in kinduitkomsten voor kinderen in centra met een hoog aandeel met veranderingen voor kinderen in centra met een laag aandeel doelgroeppeuters.

Een belangrijk verschil met de methode die gebruikt wordt om effecten op kwaliteit te schatten is dat we voor analyses van kinduitkomsten ook gebruik maken van variatie in de *duur* dat kinderen worden blootgesteld aan hogere kwaliteit (dat wil zeggen, in centra met een hoger aandeel doelgroeppeuters). Om het beleidseffect te identificeren, gebruiken we gegevens uit de oudervragenlijst over de startdatum van kinderen in de voorschool. We gebruiken een maatstaf voor ‘maanden exposure’, die wordt gedefinieerd als de totale duur (in maanden) dat een kind na de hervorming in de voorschool heeft doorgebracht; er wordt de waarde nul wordt toegekend aan kinderen die vóór de hervorming zijn getest. We construeren een interactie-term van deze maatstaf met een binaire indicator voor de treatment intensity, die gebaseerd is op de variatie in het aandeel doelgroepkinderen tussen voorscholen, net zoals in de verschillen-in-verschillenmodellen voor kwaliteit hierboven beschreven. De intuïtie is dat het effect van hogere kwaliteit op kindontwikkeling toeneemt naarmate de peuter langer is blootgesteld aan hogere kwaliteit. Voor een peuter die ruim een jaar aan een hogere kwaliteit voorschool deelneemt is het effect groter dan voor een peuter die slechts een maand deelneemt aan een hogere kwaliteit voorziening. Formeel wordt het model als volgt gespecificeerd als:

$$y_{ict} = \alpha_1 + \alpha_2 \cdot \text{MaandenExposure}_{it} + \alpha_3 \cdot (\text{HoogAandeel}_c \cdot \text{MaandenExposure}_{it}) + \alpha_4 \cdot \text{HoogAandeel}_c + \alpha_5 \cdot (\text{HoogAandeel}_c \cdot \text{MaandenImplementatie}_{it}) + Z_{ict}'\pi + \mu_t + \theta_g + \varepsilon_{ict} \quad (7.2)$$

waarin y_{ict} de uitkomst weergeeft voor kind i in voorschool c tijdens periode t . De continue variabele $\text{MaandenExposure}_{it}$ geeft het totale aantal maanden weer dat het kind in de voorschool heeft doorgebracht in de periode na de beleidswijziging. In lijn met de analyses naar kwaliteit nemen we aan dat het tijd kost voordat de extra inzet van de pedagogisch beleidsmedewerker effect heeft gehad op de kwaliteit. In de hoofdanalyses tellen we $\text{MaandenExposure}_{it}$ daarom vanaf september 2022 (en januari 2022 tot september 2022 als implementatie-maanden, $\text{MaandenImplementatie}_{it}$, maar we laten zien hoe de resultaten veranderen wanneer we aannemen dat de maatregel eerder of juist later effect heeft gehad op de kwaliteit. HoogAandeel_c is een binaire treatment indicator die gelijk is aan 1 als de voorschool van kind een boven-mediaan aandeel doelgroeppeuters had, en anders 0. De centrale parameter is de coëfficiënt α_3 , die aangeeft of kinderen met een langere blootstelling in een voorschool met een hoger aandeel doelgroeppeuters (en dus meer inzet van de pedagogisch beleidsmedewerker) hogere ontwikkelingsscores laten zien.

De specificatie neemt een reeks controlevariabelen (Z_{ict}) mee met betrekking tot de achtergrondkenmerken van het kind en het gezin: het geslacht van het kind, de leeftijd op het moment van testen, doelgroepstatus, migratieachtergrond, het aantal broers en zussen, en het opleidingsniveau van de moeder. Daarnaast wordt gecontroleerd voor kenmerken van de voorschool: de grootte van de voorschool, het marktaandeel kinderopvang van de aanbieder op regionaal en nationaal niveau, en het aandeel kindcentra van de aanbieder waar voorschoolse educatie wordt geboden. We nemen daarnaast ook periode (μ_t) en gemeente (θ_g) fixed effects mee. De standaardfouten worden geclusterd op het niveau van de voorschool.

7.4. Resultaten

Effecten op proceskwaliteit

Tabel 7-2 presenteert de verschillen-in-verschillen schattingen van de impact van de norm voor pedagogisch beleidsmedewerkers op de proceskwaliteit. De eerste rij toont resultaten op basis van vergelijking (6.1), die het aandeel doelgroeppeuters gebruikt als de continue treatment maat. De tweede rij geeft schattingen weer gebaseerd op de binaire treatment maatstaf, die meet of voorscholen een aandeel doelgroeppeuters boven of onder de mediaan hebben.

Tabel 7-2 Effecten op educatieve proceskwaliteit

	(1)	(2)	(3)	(4)
	Educatieve kwaliteit	Faciliteren van ontwikkeling	Kwaliteit van feedback	Taalstimulering
A) Continue treatment				
AandeelDoelgroep × Post	2.145** (1.016)	0.924 (1.124)	1.809* (1.005)	3.002** (1.168)
B) Binaire treatment				
HoogAandeel × Post	0.443** (0.190)	0.332 (0.204)	0.293 (0.199)	0.579*** (0.215)
N	312	312	312	312
R-squared	0.313	0.218	0.361	0.210

Noot: de tabel geeft de coëfficiënten van de verschillen-in-verschillen analyses weer (standaard fouten geclusterd op centrum-niveau tussen haakjes). ***, significant op het 1% niveau; **, significant op het 5% niveau; *, significant op het 10% niveau.

De resultaten laten een significant positief effect op de educatieve proceskwaliteit zien. Voor elke 10 procent punt hoger aandeel doelgroeppeuters heeft de introductie van pedagogisch beleidsmedewerkers geleid tot een additionele toename van 0.2 standaarddeviatie in educatieve proceskwaliteit. Er zijn geen significante effecten op de educatieve proceskwaliteit tijdens de implementatieperiode (de eerste maanden van 2022). Schattingen met de binaire indicator komen overeen met deze bevindingen: de kwaliteit in groepen in centra met een hoog aandeel doelgroeppeuters is met ruim 0.4

standaarddeviatie verbeterd ten opzichte van groepen met een laag aandeel doelgroeppeuters. Kolommen 2–4 tonen de effecten op de drie sub-schalen van educatieve proceskwaliteit: facilitering van ontwikkeling, kwaliteit van feedback en taalmodellering. Deze resultaten laten positieve coëfficiënten zien voor alle sub-schalen, maar het effect op taalmodellering is het sterkst (en significant op het 1%-niveau).

De proceskwaliteit is gemeten tijdens drie typen momenten: eet- en drinkmomenten, (begeleid) spel en een gerichte educatieve activiteit. Wanneer we de effecten schatten per type moment zien we dat de educatieve proceskwaliteit vooral tijdens spelmomenten substantieel is toegenomen. Tijdens (begeleid) spel zijn verbeteringen zichtbaar in alle drie sub-schalen van educatieve kwaliteit. Er zijn ook indicaties van positieve effecten tijdens gerichte educatieve activiteiten, vooral in termen van taalmodellering. Tijdens eet- en drinkmomenten zijn er geen significante verbeteringen in de educatieve kwaliteit zichtbaar. Dat er vooral verbeteringen zijn in de interactiekwaliteit tijdens spel wijst erop dat de pedagogisch beleidsmedewerker niet enkel gericht is op het verbeteren van proceskwaliteit gedurende gerichte VVE-activiteiten, maar ook de kwaliteit van interacties tussen pedagogisch medewerkers en kinderen tijdens het spel versterkt.

Tabel 7-3 Effecten op educatieve kwaliteit

	(1) Emotionele kwaliteit	(2) Positieve sfeer	(3) Negatieve sfeer	(4) Sensitiviteit	(5) Kind- perspectief	(6) Gedrags- regulering
A) Continue treatment						
AandeelDoelgroep	-0.017	0.000	-0.243	0.972	-0.104	-0.712
× Post	(1.129)	(1.185)	(1.078)	(1.390)	(1.139)	(1.279)
B) Binaire treatment						
HoogAandeel	0.187	0.256	0.063	0.242	0.062	0.092
× Post	(0.261)	(0.259)	(0.254)	(0.284)	(0.267)	(0.248)
N	312	312	312	312	312	312
R-squared	0.138	0.111	0.044	0.106	0.173	0.070

Noot: de tabel geeft de coëfficiënten van de verschillen-in-verschillen analyses weer (standaard fouten geclusterd op centrum-niveau tussen haakjes). ***, significant op het 1% niveau; **, significant op het 5% niveau; *, significant op het 10% niveau.

De verschillen-in-verschillen resultaten voor emotionele proceskwaliteit worden gepresenteerd in Tabel 7-3. De resultaten tonen geen significante effecten op de emotionele kwaliteit. Ook wordt er voor geen van de vijf subschalen (positieve sfeer; negatieve sfeer; sensitiviteit; kindperspectief; gedragsregulering) een effect gevonden. Deze nulresultaten worden gevonden bij zowel de continue treatment indicator als de binaire treatment indicator. Deze bevindingen bieden geen sterk bewijs dat het beleid de emotionele kwaliteit in voorscholen heeft beïnvloed. Een mogelijke verklaring voor het algehele nul-effect is dat de emotionele proceskwaliteit al hoog was vóór de hervorming, waardoor er weinig ruimte was voor verdere verbetering. Ook is het

denkbaar dat de pedagogisch beleidsmedewerkers vooral inzetten op een verhoging van de educatieve kwaliteit (wellicht ook omdat daar de meeste ruimte voor verbetering is).

Om de robuustheid van de resultaten te toetsen schatten we de effecten ook met fixed effects voor VE-groepen. Deze schattingen maken enkel gebruik van ontwikkelingen over de tijd (binnen groepen). Deze analyses zijn mogelijk omdat we de meeste groepen twee keer observeren. Consistent met de hoofdanalyses zien we positieve en significante effecten van de hervorming op educatie proceskwaliteit. De schattingen voor zowel de continue als de binaire treatment indicator tonen significante effecten op de totale educatieve proceskwaliteit en taalmodellering. Bovendien zijn er indicaties van een positieve impact op de kwaliteit van feedback. Voor emotionele proceskwaliteit worden ook in modellen met fixed effects voor VE-groepen geen effecten gevonden: de geschatte coëfficiënten zijn klein en statistisch insignificant.

Al met al kan op basis van de resultaten geconcludeerd worden dat de hervorming heeft geleid tot verbeteringen in de educatieve proceskwaliteit. Er zijn geen aanwijzingen dat de emotionele proceskwaliteit significant beïnvloed is door de beleidsmaatregel.

Effecten op het activiteitenaanbod

Naast effecten op de kwaliteit van interacties is het ook denkbaar dat de pedagogisch beleidsmedewerker effect heeft gehad op het activiteitenaanbod (curriculumkwaliteit). We toetsen de effecten op het activiteitenaanbod met dezelfde modellen als hierboven in de analyses voor proceskwaliteit, maar nemen als afhankelijke variabele diverse indicatoren van het activiteitenaanbod.

Pedagogisch medewerkers zijn gevraagd naar de frequentie en de wijze van uitvoering van VVE-activiteiten, waarbij antwoordcategorieën varieerden van “(bijna) nooit” tot “2 of meer keer per dag”: (1) kinderen doen een activiteit of les uit een educatieve (VVE) methode; (2) kinderen krijgen speciale educatieve (taal)ondersteuning door een tutor of intern begeleider (bijv. vanuit het VVE-programma); (3) VVE activiteiten worden individueel aangeboden aan kinderen met een VVE-indicatie; (4) VVE activiteiten worden in kleine groepjes aangeboden aan alleen kinderen met een VVE-indicatie; (5) VVE activiteiten worden in kleine groepjes aangeboden aan kinderen met én zonder een VVE-indicatie; (6) VVE activiteiten worden in de hele groep aangeboden.

In Tabel 7-4 worden de resultaten gepresenteerd per item, waarbij de afhankelijke variabele aangeeft of de activiteit minstens één keer per dag plaatsvindt. In alternatieve specificaties analyseren we de effecten op de kans om twee tot drie keer per maand de activiteit uit te voeren. De resultaten laten geen effecten zien op de frequentie van activiteiten uit een VVE-methode of ondersteuning door een tutor of intern begeleider (kolom (1) en (2)). Daarbij moet opgemerkt worden dat bijna alle centra dagelijks activiteiten uit een VVE-methode doen, waardoor de variatie beperkt is. Wel blijkt er een significant effect te zijn op de wijze waarop VVE-activiteiten worden uitgevoerd: de activiteiten worden significant vaker uitgevoerd in kleine groepjes met alleen

doelgroepkinderen. Dit resultaat is robuust en wordt gevonden in verschillende specificaties. Het effect op de kans om dagelijks VVE-activiteiten in kleine groepjes met alleen doelgroepkinderen uit te voeren is aanzienlijk: de relatieve toename is bijna 30 procentpunt in centra met een hoog aandeel doelgroepkinderen. Alternatieve specificaties laten zien dat er ook indicaties zijn dat VVE-activiteiten vaker individueel worden aangeboden. Er worden geen effecten gevonden op de frequentie van VVE-activiteiten aangeboden in de hele groep of in kleine groepjes met peuters met en zonder een VVE-indicatie. Deze resultaten wijzen erop dat pedagogisch medewerkers VVE-activiteiten meer gericht aanbieden aan doelgroepkinderen, zowel in kleine groepjes als individueel. Het is denkbaar dat het werken in kleine groepjes (tutoring) kan bijdragen aan de impact van voorschoolse educatie.

Tabel 7-4 Effecten op VVE-activiteiten

	(1) VVE- activiteit	(2) Tutor/IB'er	(3) Individueel	(4) Kleine groepjes (doelgroep)	(5) Kleine groepjes (alle)	(6) Hele groep
A) Continue treatment						
AandeelDoelgroep	0.471	0.525	0.720	1.797**	0.621	0.084
× Post	(0.569)	(0.799)	(0.695)	(0.875)	(0.799)	(0.598)
B) Binaire treatment						
HoogAandeel	0.060	0.070	0.059	0.302**	0.023	0.065
× Post	(0.124)	(0.143)	(0.132)	(0.142)	(0.142)	(0.116)
N	232	232	232	232	232	232

Noot: de tabel geeft de coëfficiënten van de verschillen-in-verschillen analyses weer (standaard fouten geclusterd op centrum-niveau tussen haakjes). ***, significant op het 1% niveau; **, significant op het 5% niveau; *, significant op het 10% niveau.

Naast de effecten op VVE-activiteiten zijn pedagogisch medewerkers ook gevraagd naar het reguliere activiteitsaanbod. De items zijn afkomstig uit LKK vragenlijsten voor pedagogisch medewerkers in de kinderopvang en peuteropvang. Er wordt hierbij een onderscheid gemaakt tussen de volgende categorieën: 1) motorisch-ruimtelijk-fysiek (binnen); 2) spel en samenwerken; 3) creatief-beeldende vorming; 4) muzikale vorming; 5) taal en geletterdheid; 6) rekenen-wiskunde; 7) kennis van de wereld, natuur, wetenschap en technologie; 8) buitenspelen en uitstapjes (3.3 en Appendix D). Elke categorie bestaat uit meerdere items en er is voor elke categorie een index gemaakt met behulp van principale-componentenanalyses (met gemiddelde 0, standaard deviatie 1). De resultaten voor de effecten van de pedagogisch beleidsmedewerker op het activiteitsaanbod laten een significante toename zien in de frequentie van spel en samenwerken. Het gaat hier om vormen van begeleid spel, zoals het spelen van rollen, “doen alsof spel” en samen spelen van spelletjes met regels (zoals memory). Op andere domeinen wordt geen significant effect gevonden. De toename in frequentie van begeleid spel – en de hierboven beschreven toename in frequentie van VVE-activiteiten in kleine groepjes – lijken dus niet ten koste te gaan van andere activiteiten.

Ontwikkeling van kinderen

De resultaten laten zien dat de pedagogisch beleidsmedewerker de educatieve proceskwaliteit significant heeft verhoogd. Ook zijn er significante veranderingen in het activiteitenaanbod waarneembaar, waarbij vooral de toename van VVE-activiteiten in kleine groepjes met alleen doelgroepkinderen opvallend is. Vervolgens is de vraag in hoeverre de ontwikkeling van kinderen ook is beïnvloed door de beleidsmaatregel.

In Tabel 7-5 worden de effecten gerapporteerd op taal, executieve functies (hot en cool), sociaal-emotionele ontwikkeling en fijne motoriek. De tabel presenteert de resultaten voor de hele steekproef (panel A), voor alleen doelgroepeuters (panel B) en voor meertalige doelgroepeuters (panel C). De bevindingen laten zien dat er voor de hele steekproef geen significante effecten zijn op de verschillende ontwikkelingsuitkomsten; alleen het effect op taal is marginaal significant. De beleidsmaatregel heeft echter wel een positief effect op de taalontwikkeling van (meertalige) doelgroepeuters. De schattingen voor doelgroepeuters betekenen een effect van 0.3-0.4 standaard deviatie voor een half jaar in een hogere kwaliteit voorschool.

Tabel 7-5 Effecten op kinduitkomsten per maand exposure

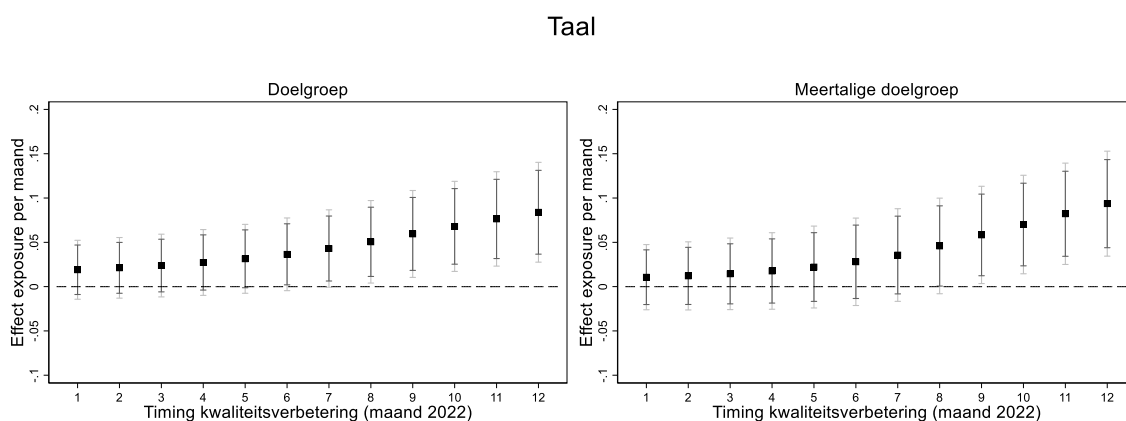
	(1)	(2)	(3)	(4)	(5)
	Taal	Hot EF	Cool EF	Sociaal- emotioneel	Fijne motoriek
Panel A: Alle kinderen					
HoogAandeel × MaandenExposure	0.032*	0.007	0.005	-0.001	-0.008
	(0.016)	(0.023)	(0.020)	(0.024)	(0.022)
N	388	386	374	388	387
Panel B: Doelgroepeuters					
HoogAandeel × MaandenExposure	0.059**	0.013	0.011	-0.033	0.024
	(0.025)	(0.031)	(0.035)	(0.033)	(0.032)
Number of observations	254	252	244	254	253
Panel C: Meertalige doelgroepeuters					
HoogAandeel × MaandenExposure	0.058**	0.029	0.006	-0.043	0.030
	(0.028)	(0.036)	(0.033)	(0.036)	(0.041)
N	198	197	190	198	198

Noot: de tabel geeft de coëfficiënten van de verschillen-in-verschillen analyses weer (standaard fouten geclusterd op centrum-niveau tussen haakjes). ***, significant op het 1% niveau; **, significant op het 5% niveau; *, significant op het 10% niveau.

De hoofdanalyses zoals hierboven beschreven gaan ervan uit dat de kwaliteit na de zomervakantie (vanaf september) 2022 meer is toegenomen in centra met een hoog aandeel doelgroepeuters. De eerste maanden van 2022 worden beschouwd als implementatieperiode, waarin er nog geen of beperkt effect op de kwaliteit is. Het is echter aannemelijk dat de kwaliteit geleidelijk is toegenomen sinds januari 2022; de

effecten op de ontwikkeling van kinderen zullen dan ook geleidelijk toenemen gedurende 2022. In Figuur 7-5 worden de beleidseffecten op taalontwikkeling getoond voor verschillende momenten van effectieve kwaliteitsverhogingen (zie Appendix H voor de resultaten op andere domeinen). Wanneer we aannemen dat het minder tijd kost voordat het interventie-effect zichtbaar is (de implementatie-periode is korter), zijn de effecten kleiner. Wanneer we aannemen dat het meer tijd kost voordat het interventie-effect zichtbaar is (de implementatie-periode is langer), zijn de effecten groter. Dit is in lijn met het idee dat investeringen in verbeteringen in kwaliteit tijd vergen om effect te hebben: eerst is tijd nodig om de kwaliteit significant te verbeteren, en vervolgens is er pas een effect op ontwikkeling te verwachten wanneer kinderen voldoende tijd zijn blootgesteld aan hogere kwaliteit. De maanden in de voorschoolse educatie die verder weg liggen van januari 2022 dragen daarom meer bij aan de ontwikkeling van kinderen. Voor de overige ontwikkelingsdomeinen maakt de timing niet uit: de effecten blijven klein en statistisch niet significant.

Figuur 7-5 Effecten op taalontwikkeling



Noot: De twee figuren geven het effect weer van de pedagogisch beleidsmedewerker op de taalontwikkeling van doelgroepeuters (links) en meertalige doelgroepeuters (rechts). Elk figuur geeft de resultaten weer van 12 regressies, waarbij de datum van implementatie varieert van januari 2022 tot december 2022.

Om te toetsen of de modellen geen verschillen in een structurele trend tussen de twee typen centra oppikken hebben we ook placebo-tests uitgevoerd. We nemen daarbij aan dat de norm voor 2022 is geïntroduceerd en schatten 12 regressies voor 12 (januari-december) hypothetische hervormingsmaanden in 2021. Deze laten allemaal kleine en statistisch insignificante effecten zien (zie Appendix Figuur H-6).

In augustus 2020 is het VE urenaanbod uitgebreid (hoofdstuk 6) en dat kan mogelijk gevolgen hebben voor de interpretatie van de resultaten. Hoewel het in theorie mogelijk is, is het niet waarschijnlijk dat de uitbreiding van voorschoolse uren de resultaten kan verklaren. Er is immers geen reden waarom kinderen binnen gemeenten in wijken met een hoger aandeel doelgroepeuters een hoger urenaanbod hebben gehad dan kinderen in wijken met een lager aandeel doelgroepeuters. Om dit uit te sluiten reproduceren we de analyses ook voor de groep kinderen die 2.5 jaar zijn geworden na 1 augustus 2020; deze groep heeft in alle gemeenten recht op minstens

960 uur tussen 2.5-4 jaar. De resultaten blijken opnieuw consistent met de hoofdanalyses.

In additionele analyses hebben we ook de effecten getoetst voor de subgroep van niet-doelgroepkinderen. Omdat het gaat om een kleine subgroep zijn deze resultaten minder betrouwbaar. De beleidseffecten zijn meestal statistisch insignificant en de schattingen laten geen consistent beeld zien. Wel zijn er indicaties van mogelijk negatieve effecten op de fijne motoriek (significant op 10% niveau). Een mogelijke verklaring hiervoor zou kunnen zijn dat er meer VVE-activiteiten in kleine groepjes met alleen doelgroepeuters worden gedaan, wat mogelijk ten koste gaat van motoriek-bevorderende activiteiten die worden gedaan met alle kinderen (inclusief niet-doelgroepkinderen). Met andere woorden, een intensievere inzet van de pedagogisch beleidsmedewerker zou kunnen leiden tot een verschuiving van de inzet van de pedagogisch medewerkers, waarbij relatief meer gerichte aandacht gaat naar doelgroepeuters. Echter, zoals hierboven beschreven, wordt deze verklaring niet ondersteund door de analyses naar de effecten op het activiteitsaanbod: VVE-activiteiten worden niet minder vaak met alle kinderen uitgevoerd, en activiteiten die bijvoorbeeld gericht zijn op creatief-beeldende vorming worden ook niet minder vaak uitgevoerd. Om hier meer betrouwbare uitspraken over te doen is aanvullend onderzoek nodig. In elk geval vinden we geen bewijs dat niet-doelgroepkinderen baat hebben gehad van de maatregel, ondanks dat de educatieve proceskwaliteit is gestegen.

Mogelijke mechanismen

We hebben mogelijke mechanismen getoetst die de gevonden effecten kunnen verklaren. Eerst analyseren we de effecten op (ervaren) professionele ontwikkeling en beleid op het gebied van kwaliteitszorg. We hebben het effect van de hervorming getoetst op ervaren coaching (gerapporteerd door de pedagogisch medewerker) en de inzet van de pedagogisch beleidsmedewerker (gerapporteerd door de leidinggevende). De resultaten laten, zoals verwacht, positieve effecten zien op deze uitkomsten. Het is ook denkbaar dat de pedagogisch beleidsmedewerker (team-)activiteiten faciliteert die te maken hebben met kwaliteitszorg. We gebruiken daarvoor de volgende items: "samen, als team, thema's en inhoudelijke (ontwikkelings)activiteiten kiezen en uitwerken"; "uitgevoerde thema's en activiteiten evalueren en bijstellen in teamverband"; "de ontwikkeling en het sociaal functioneren van de groepen met elkaar bespreken"; "samen nascholing volgen over de opvang en ontwikkelingsstimulering van jonge kinderen"; "studieboeken, vakbladen, blogs of internet sites met collega's bespreken"; "de algemene pedagogische doelen van het werk met elkaar bespreken". Op geen van deze dimensies kan een aantoonbaar effect worden gevonden.

Ten tweede is het denkbaar dat de beroepskracht-kindratio (BKR) gunstiger is geworden. Door de norm zijn er wellicht structureel een deel van de tijd meer beroepskrachten op de groep aanwezig (twee pedagogisch medewerkers plus een pedagogisch beleidsmedewerker). We maken gebruik van de gegevens verzameld tijdens de CLASS observaties om dit te toetsen; hierbij worden het aantal volwassenen en kinderen gemeten die deelnemen aan de activiteit. De resultaten laten geen of

negatieve effecten zien: tijdens spel en educatieve activiteiten is er geen effect op de BKR, maar tijdens eet- en drinkmomenten lijken er minder beroepskrachten aanwezig te zijn (significant op 10% niveau). Dit kan verklaren waarom er geen effecten worden gevonden op educatieve kwaliteit tijdens eet- en drinkmomenten, maar biedt geen verklaring voor de overige resultaten.

Tenslotte hebben we onderzocht of de hervorming heeft geleid tot veranderingen in de kenmerken van pedagogisch medewerkers en de samenstelling van de groepen. We hebben effecten op de volgende kenmerken van pedagogisch medewerkers getoetst: leeftijd, ervaring, opleidingsniveau, migratieachtergrond en gewerkte uren. De analyses laten geen significante veranderingen zien, wat suggereert dat de samenstelling van beroepskrachten stabiel is. De analyses laten ook geen effecten zien op de samenstelling van de groepen, bijvoorbeeld in termen van kinderen met een behoefte aan taalondersteuning of migratieachtergrond (zoals gerapporteerd door pedagogisch medewerkers en leidinggevenden).

7.5. Conclusie en discussie

Het doel van de invoering van de norm voor de inzet van de pedagogisch beleidsmedewerker was een verbetering van de kwaliteit de voorschoolse educatie, zodat de ontwikkeling van doelgroeppeuters (nog meer) kon worden ondersteund. Om de effecten van de pedagogisch beleidsmedewerker te toetsen hebben we gebruik gemaakt van een verschillen-in-verschillen methode. We maken daarbij gebruik van een centraal element van het beleid: de inzet van de pedagogisch beleidsmedewerker neemt toe met het aantal doelgroeppeuters in het centrum. Het is daarom te verwachten dat centra in wijken met relatief veel doelgroeppeuters meer zijn beïnvloed door de hervorming dan centra in wijken met relatief weinig doelgroeppeuters.

De resultaten laten een significant positief effect zien op de geobserveerde educatieve proceskwaliteit. De kwaliteitsverbeteringen worden gevonden tijdens (begeleid) spelmomenten en gestructureerde educatieve activiteiten, maar niet tijdens eet- en drinkmomenten. Er was geen verbetering zichtbaar in de emotionele proceskwaliteit, die al voor de invoering van de maatregel over de hele linie hoog was. Ook rapporteerden pedagogisch medewerkers vaker dat ze VVE-activiteiten in kleine groepjes met alleen doelgroepkinderen aanboden. In lijn met deze effecten worden ook positieve effecten op taalontwikkeling van doelgroeppeuters gevonden. Deze effecten zijn substantieel, maar dat is wellicht te verwachten gegeven het geheel aan empirische bevindingen: 1) de educatieve proceskwaliteit is sterk verbeterd, met vooral aanzienlijke effecten op de dimensie taalmodellering; 2) er wordt vaker in kleine groepjes met alleen doelgroeppeuters gewerkt, wat kan bijdragen aan de ontwikkeling van doelgroeppeuters; 3) effecten zijn klein en niet statistisch significant op andere ontwikkelingsdomeinen, de maatregel lijkt dus een vrij specifiek effect te hebben op taal en vooral op een specifieke groep kinderen. Omdat doelgroeppeuters baat hebben gehad bij de beleidsmaatregel wijzen de resultaten erop dat de kanselijkheid is vergroot.

8. Conclusies en aanbevelingen

Sinds 2020 zijn er twee landelijke maatregelen ingevoerd om de kwantiteit (aantal uren) en kwaliteit van de voorschoolse educatie te verhogen. Per 1 augustus 2020 zijn gemeenten verplicht doelgroeppeuters 960 uur voorschoolse educatie te bieden tussen de leeftijd van 2.5 en 4 jaar. Sinds januari 2022 geldt daarnaast een urennorm voor de inzet van pedagogisch beleidsmedewerkers in de voorschoolse educatie. Het doel van deze maatregelen is het vergroten van de ontwikkelingskansen van doelgroepkinderen. EVENING onderzocht de effectiviteit van beide maatregelen.

8.1. Effecten van meer uren voorschoolse educatie

De resultaten laten zien dat de extra uren voorschoolse educatie doelmatig worden ingezet voor educatieve activiteiten, meestal in een kleine tutoring groep. Analyses naar effecten op proceskwaliteit waren niet mogelijk omdat de meeste geplande voormetingen niet konden plaatsvinden in verband met corona. Meer uren voorschoolse educatie leidt tot hogere scores op de tests voor taal, executieve functies en fijne motoriek. Ook zijn er positieve effecten op sociaal-emotionele ontwikkeling. Gecombineerd geven de resultaten een beeld van het causale pad van de beleidsmaatregel naar kinduitkomsten: door het ruimere urenaanbod maken kinderen gebruik van meer uren voorschoolse educatie en doelgroepkinderen worden daardoor meer blootgesteld aan gestructureerde en intensief begeleide educatieve activiteiten.

8.2. Effecten van de pedagogisch beleidsmedewerker

Uit analyses blijkt een significant effect van de pedagogisch beleidsmedewerkers op de geobserveerde educatieve proceskwaliteit, dat wil zeggen op de wijze waarop pedagogisch medewerkers kinderen mogelijkheden bieden kennis te verwerven en hen uitdagen te redeneren, inhoudelijke feedback en waar nodig ondersteuning geven, en rijk en gevorderd taalgebruik voordoen en bij kinderen uitlokken. Ook rapporteerden pedagogisch medewerkers vaker in kleine groepen met doelgroeppeuters te werken. Er was geen significante verbetering van de emotionele proceskwaliteit; deze was over de hele linie al hoog in de voorschoolse educatie. Uit nadere analyses kan afgeleid worden dat de kwaliteitsverbetering waarschijnlijk door intensievere coaching tot stand is gekomen; er zijn geen aantoonbare veranderingen op andere dimensies van professionele ontwikkeling. Tenslotte wijzen de resultaten op positieve effecten op de taalvaardigheid van doelgroeppeuters, maar zijn er geen significante effecten op hun executieve functies, motorische vaardigheden of hun sociaal-emotioneel functioneren.

8.3. Aanbevelingen en vervolgonderzoek

De bevindingen laten zien dat doelgroeppeuters baat hebben bij een ruimer urenaanbod. Dit beleid dient daarom gecontinueerd te worden. Het verder uitbreiden van het urenaanbod, via verdere uitbreiding van het aantal uren per week dan wel via een start vóór de leeftijd van 2.5 jaar, lijkt een kansrijke strategie en vraagt nader onderzoek. In een vervolgrapportage zal EVENING aandacht besteden aan het starten

met voorschoolse educatie vanaf een leeftijd van 2 jaar. Ook zal in een vervolgrapportage worden getoetst of een ruimer uren aanbod effect heeft op de taalvaardigheden (begrijpend lezen en spelling) en rekenen in groep 3.

Het onderzoek toont ook aan dat de norm voor de pedagogisch beleidsmedewerker heeft bijgedragen aan hogere kwaliteit en de taalontwikkeling van doelgroepkinderen. Ook deze beleidsmaatregel verdient daarom continuering. De resultaten laten zien dat de maatregel sterkere effecten heeft gehad op centra met een groter aandeel doelgroepeuters. Dit kan wenselijk zijn omdat zo een groter aantal doelgroepeuters wordt bereikt (gegeven de beschikbare middelen). Maar deze uitkomst kan ook als onwenselijk worden beschouwd omdat centra met een lager aandeel doelgroepeuters achterblijven in termen van kwaliteit. Dat betekent dat kinderen in deze centra niet optimaal kunnen profiteren van voorschoolse educatie. Om dit effect te beperken is het aan te bevelen om naast een norm per doelgroepeuter een minimumnorm per centrum (of per voorschoolse educatie groep) te overwegen. Bij een minimumnorm van bijvoorbeeld 80 uur per VE-groep per jaar kan meer impact gemaakt worden met de pedagogisch beleidsmedewerker, ook wanneer er minder dan 8 kinderen met een VVE-indicatie deelnemen. In een aantal gemeenten wordt al gewerkt met een minimumnorm. Voor de wetswijziging bleek uit een quick scan onder de G37 dat een norm van 4 uur per week per groep als minimum werd beschouwd om een positief effect te genereren op de kwaliteit.¹⁴ In het conceptbesluit was daarom een norm opgenomen van 160 uur per jaar. Deze norm werd door een deel van de sector als te hoog ervaren en is daarom afgezwakt. Door een minimumnorm van bijvoorbeeld 80 uur per groep per jaar te introduceren zou in stappen toegewerkt kunnen worden naar de oorspronkelijk beoogde 160 uur per jaar (4 uur per week). Tenslotte is niet duidelijk hoe de pedagogisch beleidsmedewerker het meest effectief kan worden ingezet. Het is daarom aan te bevelen om via een gerandomiseerd experiment te onderzoeken hoe de pedagogisch beleidsmedewerker nog meer bij kan dragen aan de kwaliteit van voorschoolse educatie en de ontwikkeling van kinderen.

Met EVENING is er nieuw bewijs geleverd dat investeringen in voorschoolse educatie een belangrijke bijdrage kunnen leveren aan kansengelijkheid. Hoewel EVENING de effectiviteit van twee specifieke landelijke beleidsmaatregelen heeft geëvalueerd, kunnen de uitkomsten algemener gezien worden als bewijs dat voorschoolse educatie de ontwikkelingskansen van kinderen met een risico op taal- of onderwijsachterstand versterkt.

¹⁴ Staatsblad van het Koninkrijk der Nederlanden (2019), Besluit van 20 september 2019 tot wijziging van het Besluit basisvoorwaarden kwaliteit voorschoolse educatie in verband met de verhoging van het minimaal aantal uren aanbod voorschoolse educatie en de inzet van een pedagogisch beleidsmedewerker, jaargang 315, 1-24.

Literatuur

- Agostinelli, F., & Wiswall, M. (2025). Estimating the technology of children's skill formation. *Journal of Political Economy*, 133(3), 846-887.
- Akgündüz, Y. E., & Heijnen, S. (2018). Impact of funding targeted pre-school interventions on school readiness: Evidence from the Netherlands. *De Economist*, 166(2), 155-178.
- Andrew, A., Attanasio, O. P., Bernal, R., Sosa, L. C., Krutikova, S., & Rubio-Codina, M. (2024). Preschool quality and child development. *Journal of Political Economy*, 132(7), 2304-2345.
- Bailey, M. J., Sun, S., & Timpe, B. (2021). Prep school for poor kids: The long-run impacts of Head Start on human capital and economic self-sufficiency. *American Economic Review*, 111(12), 3963-4001.
- Bandel, E., Aikens, N., Vogel, C. A., Boller, K., & Murphy, L. (2014). Observed quality and psychometric properties of the CLASS-T in the Early Head Start Family and Child Experiences Survey (OPRE Technical Brief No. 2014-23). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.
- Bleses, D., Jensen, P., Højen, A., Slot, P., & Justice, L. (2021). Implementing toddler interventions at scale: The case of "We learn together". *Early Childhood Research Quarterly*, 57, 12-26.
- Bleses, D., Jensen, P., Slot, P., & Justice, L. (2020). Low-cost teacher-implemented intervention improves toddlers' language and math skills. *Early Childhood Research Quarterly*, 53, 64-76.
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech Language and Hearing Research*, 58, 1747-1760.
- Carneiro, P., Cruz-Aguayo, Y., Intriago, R., Ponce, J., Schady, N., & Schodt, S. (2022). When promising interventions fail: Personalized coaching for teachers in a middle-income country. *Journal of Public Economics Plus*, 3, 100012.
- Castle, S., Williamson, A. C., Young, E., Stubblefield, J., Laurin, D., & Pearce, N. (2016). Teacher-child interactions in early Head Start classrooms: Associations with teacher characteristics. *Early Education and Development*, 27, 259-274.
- Chiat, S. (2015). *Nonword repetition*. In: Armon-Lotem, S., Meir, N., & de Jong, J. (Eds.) *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment* (125-151). Bristol, VK: Multilingualism Matters.
- CPB (2020). Ongelijkheid van het jonge kind. Notitie Central Planbureau.
- De Ridder, J.G., Van der Stege, H.A., Van Agt, H.M.E., De Koning, H.J., & Verhoeven, L.T.W. (2006). *Handleiding VTO Taal 2-jarigen*. Amsterdam: Harcourt Assessment B.V.
- Diamond, A., Prevor, M. B., Callender, G., & Druin, D. P. (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. *Monographs of the Society for Research in Child Development*, 62(4), i.
- Dickinson, D.K., Collins, M.F., Nesbitt, K., Toub, T.S., Hassinger-Das, B., Hadley, E.B., Hirsh-Pasek, K., & Golinkoff, R.M. (2019). Effects of teacher-delivered book reading and play on vocabulary learning and self-regulation among low-income preschool children. *Journal of Cognition and Development*, 20(2), 136-164.
- Duncan, G., Kalil, A., Mogstad, M., & Rege, M. (2023). Investing in early childhood development in preschool and at home. *Handbook of the Economics of Education*, 6, 1-91.
- Dunn, M., Dunn, L.M. & Schlichting, L. (2005). Peabody Picture Vocabulary Test-III-NL. Amsterdam: Harcourt Test Publishers.
- Ezpeleta, L., Granero, R., de la Osa, N., Penelo, E., & Domènech, J. M. (2013). Psychometric properties of the strengths and Difficulties Questionnaire 3-4 in 3-year-old preschoolers. *Comprehensive Psychiatry*, 54(3), 282-291.
- Felfe, C., & Lalive, R. (2018). Does early child care affect children's development? *Journal of Public Economics*, 159, 33-53.
- Gerrits, E. (2003). Speech perception of young children at risk for dyslexia and children with specific language impairment. Proceedings of the 15th International Conference of Phonetic Sciences (ICPhS), Barcelona, Spain.
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38(8), 1179-1191.

- Henrichs, L.F., & Leseman, P.P.M. (2014). Early science instruction and academic language development can go hand in hand. *International Journal of Science Education*, 36, 17, 2978-2995.
- Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental psychology*, 49(1), 4.
- Hulpia, H., Vandenbroeck, M., Daems, M., Declercq, B., Janssen, J., Cleynenbreugel, C. van, & Laevers, F. (2016). MeMoQ Deelrapport 10: Emotionele en educatieve ondersteuning in de nulmeting. Kind en Gezin, Ugent, KU Leuven.
- Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, 36(2), 220-232. doi:10.1037//0012-1649.36.2.220
- Kokštejn, J., Musálek, M., & Tufano, J. J. (2017). Are sex differences in fundamental motor skills uniform throughout the entire preschool period?. *PloS one*, 12(4), e0176556.
- Kraft, A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547-588.
- Kuijpers, C. (1996). Perception of the voicing contrast by Dutch children and adults. *Journal of Phonetics*, 24, 367-382.
- La Paro, K. M., Williamson, A. C., & Hatfield, B. (2014). Assessing quality in toddler classrooms using the CLASS-Toddler and the ITERS-R. *Early Education and Development*, 25(6), 875-893.
- Lange, B. P., Euler, H. A., & Zaretsky, E. (2016). Sex differences in language competence of 3-to 6-year-old children. *Applied Psycholinguistics*, 37(6), 1417-1438.
- Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental science*, 21(2), e12529.
- Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., ... & van de Schoot, R. (2024). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social science research*, 110, 102805.
- Leseman, P., Mulder, H., Verhagen, J., Broekhuizen, M., van Schaik, S., & Slot, P. (2017). Effectiveness of Dutch targeted preschool education policy for disadvantaged children: Evidence from the pre-COOL study. In H. P. Blossfeld, N. Kulic, J. Skopek & M. Triventi (Eds.), *Childcare, early education, and social inequality – An international perspective* (pp. 173-193). Cheltenham, UK: Edward Elgar.
- Leseman, P., & Slot, P. (2023). Kwaliteit als sleutel tot maatschappelijk rendement in de kinderopvang. *Economisch Statistische Berichten*, 108(4827S), 54-59.
- Leseman, P. & Slot, P. (2013). Kwaliteit en curriculum van voorschoolse opvang en educatie in Nederland. Relaties met structurele kenmerken, organisatiekenmerken en gebruik van educatieve programma's. Utrecht: Afdeling Orthopedagogiek Cognitieve en Motorische Ontwikkelingsproblemen, Universiteit Utrecht.
- Leseman, P. & van Huizen, T. (2022). Effecten van kinderopvang op de cognitieve en sociaal-emotionele ontwikkeling van kinderen. Universiteit Utrecht.
- Leseman, P. & van der Veen, A. (2022). Het pre-COOL cohort tot en met groep 8. *Ontwikkeling van kinderen en relatie met kwaliteit in de voor-en vroegschoolse periode*. Amsterdam: Kohnstamm Instituut.
- Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. New York, NY: Routledge.
- Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P. P. (2014). Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds. *Frontiers in psychology*, 5, 733.
- Mulder, H., Verhagen, J., Van der Ven, S. H., Slot, P. L., & Leseman, P. P. (2017). Early executive function at age two predicts emergent mathematics and literacy at age five. *Frontiers in psychology*, 8, 1706.
- Putnam, S. P., Gartstein, M.A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development*, 29 (3), 386-401. (Nederlandse vertaling: Renee de Kruif, Tine Willekens, & Leen de Schuymer et al.).
- Romijn, B., Slot, P., Jepma, I., Muller, P., Bredeveld, M., & Leseman, P. (2023). *Kwaliteit van de Nederlandse kinderopvang. De kinderopvang, peuteropvang, buitenschoolse opvang en gastouderopvang in beeld – Gecombineerde metingen 2017-2022*. Utrecht: Universiteit Utrecht / Sardes.
- Romijn, B., Slot, P., Jepma, I., Muller, P., Bredeveld, M., & Leseman, P. (2024). *Kwaliteit van de Nederlandse kinderopvang. De kinderopvang, peuteropvang, buitenschoolse opvang en gastouderopvang in beeld – Gecombineerde metingen 2017-2023*. Utrecht: Universiteit Utrecht / Sardes.
- Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at 3-7 years: The Children's Behavior Questionnaire. *Child Development*, 72, 1394-1408. (Nederlandse vertaling: Mirjana Majdandzic).

- Slot, P., Jepma, I., Muller, P., Romijn, B., Bekkering, C., & Leseman, P. (2019). Ontwikkelingen in de kwaliteit van de Nederlandse kinderdagopvang, peuteropvang, buitenschoolse opvang en gastouderopvang. Gecombineerde metingen 2017-2019. Landelijke Kwaliteitsmonitor Kinderopvang. Utrecht: Universiteit Utrecht / Sardes.
- Squires, J., Potter, L., & Bricker, D. (1999). The ASQ user's guide for the Ages & Stages Questionnaires: A parent-completed, child-monitoring system (2nd ed.). Paul H. Brooks Publishing Co. (Nederlandse vertaling: Anneloes van Baar, Hedwig van Bakel, & Sabine Hunnius).
- Van Baar, A. L., Steenis, L. J. P., Verhoeven, M., & Hessen, D. J. (2014). Bayley-III-NL. Technische Handleiding. Amsterdam, NL: Pearson Assessment and Information.
- Vander Heyden, K. & van Teunenbroek, M. (2018). Inventarisatie hbo-geschoolde beroepskrachten in de we. Sardes.
- Van Huizen, T. (2018). The Evolution of Achievement Gaps from Early Childhood to Adolescence in the Netherlands. In: G. Passaretta and J. Skopek (eds.), *Roots and Development of Achievement Gaps. A Longitudinal Assessment in Selected European Countries*, pp. 50-87, ISOTIS Report (D 1.3).
- van Huizen, T., Leseman, P., Oudgenoeg - Paz, O., Francot, R., Appels, S., & Tiliopoulou, V. (2023). EVENING tussentijdsrapport, maart 2023.
- Van Huizen, T., & Plantenga, J. (2018). Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural experiments. *Economics of Education Review*, 66, 206-222.
- Van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry*, 12(6), 281-289.
- Veen, A., van der Veen, I., Heurter, A., Mulder, L. Paas, T., Leseman, P., Mulder, H., Verhagen, J., & Slot, P. (2012). Pre-COOL cohortonderzoek. Technisch rapport tweejarigencohort, tweede meting 2011-2012.
- Veijalainen, J., Reunamo, J., & Heikkilä, M. (2021). Early gender differences in emotional expressions and self-regulation in settings of early childhood education and care. *Early Child Development and Care*. 191(2), 173–186.
- Wierdsma, I. & Jepma, IJ. (2024). Coaching in de voorschoolse educatie. Verdiepend EVENING-onderzoek onder pedagogisch beleidsmedewerkers/coaches. Sardes.
- Wiig E.,H., Secord, W.A., Semel, E. (2012). CELF Preschool 2 Clinical Evaluation of Language Fundamentals Preschool – Nederlandse versie. Pearson. (Nederlandse vertaling: Jan de Jong, J. & Pearson Assessment and Information B.V.).

Appendix A. Kernconcepten en kerndomeinen

Tabel A-1 Kind kenmerken

Ontwikkelingsdomein	Construct	Instrumenten/bron
Nederlandse Taal	Foneemwaarneming en categorisatie Receptieve woordenschat Actieve woordenschat	Klankentaak (pre-COOL) pre-COOL (PPVT verkorte versie) actieve woordenschat schaal uit CELF Pre-School- afname alleen bij 4-jarigen
Cool Executieve Functies	Selectieve volgehouden aandacht Verbaal korte termijn geheugen Visueel-ruimtelijk geheugen	Olifantjestaak (pre-COOL) – 5 jarige versie Nonwoord repetitietask – aangepaste versie met items van pre-COOL en Crosslinguistic NWR Zes doosjestaak (pre-COOL) – aangepaste versie geschikt voor 2,5 en 4 jaar.
Hot Executieve Functies	Uitstel van beloning	Wachttaken rozijn en cadeau (pre-COOL)
Sociaal-emotioneel	Zelfregulatie: impuls controle en volgehouden aandacht thuis Zelfregulatie: impuls controle en volgehouden aandacht op de groep Prosociaal gedrag, externaliserend gedrag thuis Prosociaal gedrag, externaliserend gedrag op de groep	Items uit de (E)CBQ ingevuld door de ouders Items uit de (E)CBQ ingevuld door de pmer Strengths and Difficulties Questionnaire ingevuld door ouders Strengths and Difficulties Questionnaire ingevuld door pmer
Motorische ontwikkeling	Grove en fijne motoriek	Items uit Ages & Stages Questionnaire

Tabel A-2 Gezinskenmerken

Oudervragenlijst Domein	Construct	Instrumenten/bron
SES	Achtergrond gezin	pre-COOL/ISOTIS
	Opleidingsniveau	pre-COOL/ISOTIS
	Werksituatie en huishouden	pre-COOL/ISOTIS
Gezinssamenstelling	Aantal kinderen/gezinstype	pre-COOL
	Gebruik van voorschoolse voorzieningen	Ontwikkeld binnen EVENING
Taalachtergrond gezin	Taalgebruik thuis	VLOT/ DASH/ISOTIS
	Ervaren taalvaardigheid in het Nederlands	ISOTIS
	Ervaren taalvaardigheid in een andere taal	ISOTIS
Cultuurpedagogische gezinskenmerken	Taal- en rekenactiviteiten met kind, thuistaal, meertaligheid	pre-COOL/DASH/ ISOTIS/ VLOT
	Educatieve activiteiten buitenshuis	ISOTIS / DASH
	Betrokkenheid en relatie ouders bij peutergroep	ISOTIS

Tabel A-3 Kenmerken instellingen

Observaties	
<i>Groepsinteracties</i>	<i>Instrumenten/bron</i>
Emotionele ondersteuning	CLASS-Toddler
Educatieve ondersteuning	CLASS-Toddler
Vragenlijst leidinggevenden	
<i>Domein</i>	<i>Instrumenten/bron</i>
Werk en opleiding	pre-COOL en LKK
Kenmerken kinderopvangorganisatie	pre-COOL en LKK
Rol en bevoegdheid locatiemanager	pre-COOL en LKK
Achtergrond pedagogisch medewerker	pre-COOL en LKK
Samenstelling kindergroep(en)	pre-COOL en LKK
Beleid groepen	pre-COOL en LKK
Inzet van pedagogisch beleidsmedewerkers	Ontwikkeld binnen EVENING
Pedagogisch concept	pre-COOL en LKK
Diversiteitsbeleid	pre-COOL en LKK
Kwaliteitszorg en professionalisering	pre-COOL en LKK
Relatie met ouders	pre-COOL en LKK
Externe samenwerking	pre-COOL en LKK
Beleidsmaatregel: 16 uur voorschoolse educatie	Ontwikkeld binnen EVENING
Corona	Ontwikkeld binnen EVENING
Vragenlijst pedagogisch medewerkers	
<i>Domein</i>	<i>Instrumenten/bron</i>
Persoonsgegevens	pre-COOL en LKK
De groep	pre-COOL en LKK
Opleiding	pre-COOL en LKK
Geboorteland	pre-COOL en LKK
Samenstelling van de kindergroep	pre-COOL en LKK
Maatregel urenuitbreiding	Ontwikkeld binnen EVENING
Activiteitschema's	pre-COOL en LKK
Diversiteitspraktijken	pre-COOL en LKK
Gedrag en spel	pre-COOL en LKK
Werkbeleving	pre-COOL en LKK
Kwaliteitszorg en professionalisering	pre-COOL en LKK
Relatie met de ouders	pre-COOL en LKK

Appendix B. Non-respons

Tabel B-1 Redenen non-respons VE-groepen

Categorie	Voorbeelden redenen non-respons
Onderzoek	'geen interesse', 'ongunstige timing', 'te belastend', ' bezig met andere onderzoeken'
Opvanglocatie	'te hoge werkdruk', 'personeelstekort', 'personeelwisselingen', 'langdurige ziekte van leidinggevenden', 'onrustige situatie'
Beroepskrachten	'ziekte', 'wil niet bekeken worden'
Organisatie en beleid	'fusie van groepen', 'zitten in overname door een andere aanbieder', 'verhuizing', 'locatie is gesloten', 'stopt binnenkort met opvang'
Corona maatregelen	'lockdowns', 'te veel zieke kinderen/medewerkers', 'er zijn momenteel te veel besmettingen'
Ouders	'ouders willen dit niet in coronatijd', 'moeilijk om ouders te werven', 'ouders bezwaar tegen observatie'
Kinderen	'veel nieuwe kinderen in de groep', 'veel zorgkinderen', 'hectische groep'
Overig	'willen deelnemen na de zomervakantie', 'weigering zonder opgaaf van reden', 'niet ingepland krijgen'

Appendix C. Databeschrijving per meetronde

Tabel C-1 Geobserveerde kwaliteit 2020-2021. Eenheid: Groep (links) en observatiecyclus (rechts)

	Groepen (N=68)			Observatiecycli (N=204)		
Structurele kenmerken tijdens de observatie	Range	M	SD	Range	M	SD
Aantal kinderen	5.00 – 15.67	9.45	2.80	1.00 – 16.00	9.45	3.25
Aantal volwassenen	1.00 – 3.00	1.91	0.45	1.00 – 3.00	1.91	0.56
Kind-staf ratio	2.00 – 11.67	5.17	1.65	1.00 – 14.00	5.17	1.94
CLASS emotioneel	Range	M	SD	Range	M	SD
Positieve sfeer	4.00 – 7.00	5.57	0.69	3.00 – 7.00	5.57	0.90
Negatieve sfeer (R)	5.67 – 7.00	6.84	0.31	5.00 – 7.00	6.84	0.40
Sensitiviteit	3.33 – 6.33	5.17	0.67	3.00 – 7.00	5.17	0.91
Kindperspectief	3.00 – 6.00	4.37	0.76	1.00 – 7.00	4.37	1.24
Gedragregulering	3.33 – 6.00	4.87	0.72	2.00 – 7.00	4.87	0.96
Gemiddelde emotionele kwaliteit	4.13 – 6.13	5.36	0.51	3.20 – 6.80	5.36	0.63
CLASS educatief	Range	M	SD	Range	M	SD
Faciliteren ontwikkeling	2.33 – 6.00	4.11	0.81	2.00 – 7.00	4.11	1.08
Kwaliteit feedback	1.33 – 5.67	3.16	0.91	1.00 – 6.00	3.16	1.06
Taalstimulering	1.67 – 6.00	3.60	0.94	1.00 – 6.00	3.60	1.07
Gemiddelde educatieve kwaliteit	2.11 – 5.56	3.62	0.81	1.67 – 6.00	3.62	0.93

Tabel C-2 Geobserveerde kwaliteit 2021-2022. Eenheid: Groep (links) en observatiecyclus (rechts)

	Groepen (N=177)			Observatiecycli (N=530)		
Structurele kenmerken tijdens de observatie	Range	M	SD	Range	M	SD
Aantal kinderen	3.67 – 16.00	9.51	2.71	1.00 – 16.00	9.51	3.17
Aantal volwassenen	1.00 – 3.00	1.97	0.45	1.00 – 4.00	1.97	0.64
Kind-staf ratio	2.00 – 11.00	5.01	1.53	1.00 – 13.00	5.08	1.77
CLASS emotioneel	Range	M	SD	Range	M	SD
Positieve sfeer	3.33 – 7.00	5.48	0.80	1.00 – 7.00	5.47	1.04
Negatieve sfeer (R)	5.33 – 7.00	6.79	0.33	3.00 – 7.00	6.79	0.48
Sensitiviteit	2.67 – 6.67	5.19	0.68	2.00 – 7.00	5.19	0.97
Kindperspectief	1.33 – 6.33	4.05	0.80	1.00 – 7.00	4.05	1.26
Gedragregulering	2.67 – 6.67	4.92	0.82	2.00 – 7.00	4.92	1.07
Gemiddelde emotionele kwaliteit	3.60 – 6.27	5.28	0.54	2.80 – 6.60	5.28	0.68
CLASS educatief	Range	M	SD	Range	M	SD
Faciliteren ontwikkeling	1.67 – 6.33	3.89	0.91	1.00 – 7.00	3.89	1.25
Kwaliteit feedback	1.00 – 5.00	2.68	0.84	1.00 – 6.00	2.68	1.05
Taalstimulering	1.00 – 5.33	3.40	0.92	1.00 – 6.00	3.40	1.14
Groepssocialisatie	1.00 – 5.33	3.24	0.90	1.00 – 6.00	3.24	1.23
Gemiddelde educatieve kwaliteit	1.44 – 5.00	3.32	0.76	1.00 – 5.67	3.32	0.96

Tabel C-3 Geobserveerde kwaliteit 2022-2023. Eenheid: Groep (links) en observatiecyclus (rechts)

	Groepen (N=184)			Observatiecycli (N=550)		
Structurele kenmerken tijdens de observatie	Range	M	SD	Range	M	SD
Aantal kinderen	3.67 – 16.00	9.68	2.61	1.00 – 16.00	9.68	3.20
Aantal volwassenen	1.00 – 4.00	1.98	0.46	1.00 – 4.00	1.98	0.58
Kind-staf ratio	2.00 – 12.00	5.08	1.62	1.00 – 15.00	5.13	1.97
CLASS emotioneel	Range	M	SD	Range	M	SD
Positieve sfeer	3.67 – 7.00	5.61	0.75	3.00 – 7.00	5.62	0.92
Negatieve sfeer (R)	5.50 – 7.00	6.90	0.25	4.00 – 7.00	6.90	0.33
Sensitiviteit	3.00 – 7.00	5.40	0.77	3.00 – 7.00	5.41	0.95
Kindperspectief	2.67 – 7.00	4.73	0.86	2.00 – 7.00	4.73	1.20
Gedragregulering	2.67 – 6.67	5.12	0.85	2.00 – 7.00	5.13	1.03
Gemiddelde emotionele kwaliteit	4.20 – 6.73	5.55	0.56	4.00 – 7.00	5.56	0.64
CLASS educatief	Range	M	SD	Range	M	SD
Faciliteren ontwikkeling	2.67 – 6.67	4.63	0.84	2.00 – 7.00	4.63	1.13
Kwaliteit feedback	1.67 – 7.00	4.01	1.12	1.00 – 7.00	4.01	1.32
Taalstimulering	2.00 – 6.00	4.07	0.80	1.00 – 7.00	4.07	1.04
Groepssocialisatie	1.00 – 6.67	4.11	1.00	1.00 – 7.00	4.11	1.25
Gemiddelde educatieve kwaliteit	2.67 – 6.33	4.24	0.80	2.00 – 6.67	4.23	0.99

Tabel C-4 Geobserveerde emotionele en educatieve kwaliteit afhankelijk van type activiteit

2022-2023					
Activiteit	%	Emotionele kwaliteit		Educatieve kwaliteit	
		M	SD	M	SD
Eten/ drinken	33.82	5.53	0.62	4.01	0.94
Vrij spel	32.91	5.69	0.64	4.26	1.01
Gerichte activiteit	33.27	5.45	0.64	4.44	0.97
2021-2022					
Activiteit	%	Emotionele kwaliteit		Educatieve kwaliteit	
		M	SD	M	SD
Eten/ drinken	33.58%	5.13	0.65	3.08	0.91
Vrij spel	32.45%	5.43	0.70	3.34	0.95
Gerichte activiteit	33.96%	5.30	0.66	3.55	0.95
2020					
Activiteit	%	Emotionele kwaliteit		Educatieve kwaliteit	
		M	SD	M	SD
Eten/ drinken	34.48%	5.27	0.60	3.40	0.89
Vrij spel	31.53%	5.47	0.62	3.61	0.92
Gerichte activiteit	33.99%	5.36	0.67	3.86	0.95

Tabel C-5 Achtergrondkenmerken van pedagogisch medewerkers en leidinggevendenden

Variabelen	Schaal	Omschrijving
Leeftijd	Jaren	Leeftijd in jaren
Werkdagen	Dagen	Aantal dagen per week
Werktijd	Uren	Aantal uur per week dat de respondent werkt
Opleidingsniveau	1-8	Het maximum gehaald uit de vooropleiding en het vervolgopleidingsniveau
Vooropleiding	1-8	De vooropleiding (secundair onderwijs) van de medewerker naar onderwijsniveau: 1. Maximaal VMBO Kader/beroepsgericht 2. MBO 1 of 2 3. MAVO of VMBO Theoretisch 4. HAVO 5. MBO 3 of 4 6. VWO 7. HBO 8. Wetenschappelijk onderwijs
Niveau vervolgopleiding	1-8	De beroepsgerichte vervolgopleiding naar onderwijsniveau (zie schaal vooropleiding)
Werkervaring	Jaren	Aantal jaren werkervaring in de sector kinderopvang

Tabel C-6 Achtergrond van de pedagogisch medewerkers en leidinggevendenden (2020-2021)

Achtergrond pedagogisch medewerker	N	Range	M	SD
Leeftijd	134	21-65	44.50	11.11
Werkdagen	131	1-5	3.66	0.93
Werktijd in uren	131	8-40	23.58	6.98
Opleidingsniveau	127	5-8	5.74	0.96
Hoogste vooropleiding (1-8)	123	1-8	5.31	1.44
Niveau vervolgopleiding (1-8)	115	5-8	5.67	0.96
Werkervaring in kinderopvang in jaren	131	1-38	16.99	8.86
Achtergrond leidinggevendenden	N	Range	M	SD
Opleidingsniveau	84	5-8	6.88	0.97
Vooropleiding (1-8)	84	3-8	6.38	1.17
Niveau vervolgopleiding (1-8)	77	5-8	6.92	0.93
Werkervaring in kinderopvang in jaren	84	1-40	16.27	10.45

Tabel C-7 Achtergrond van de pedagogisch medewerkers en leidinggevenden (2021-2022)

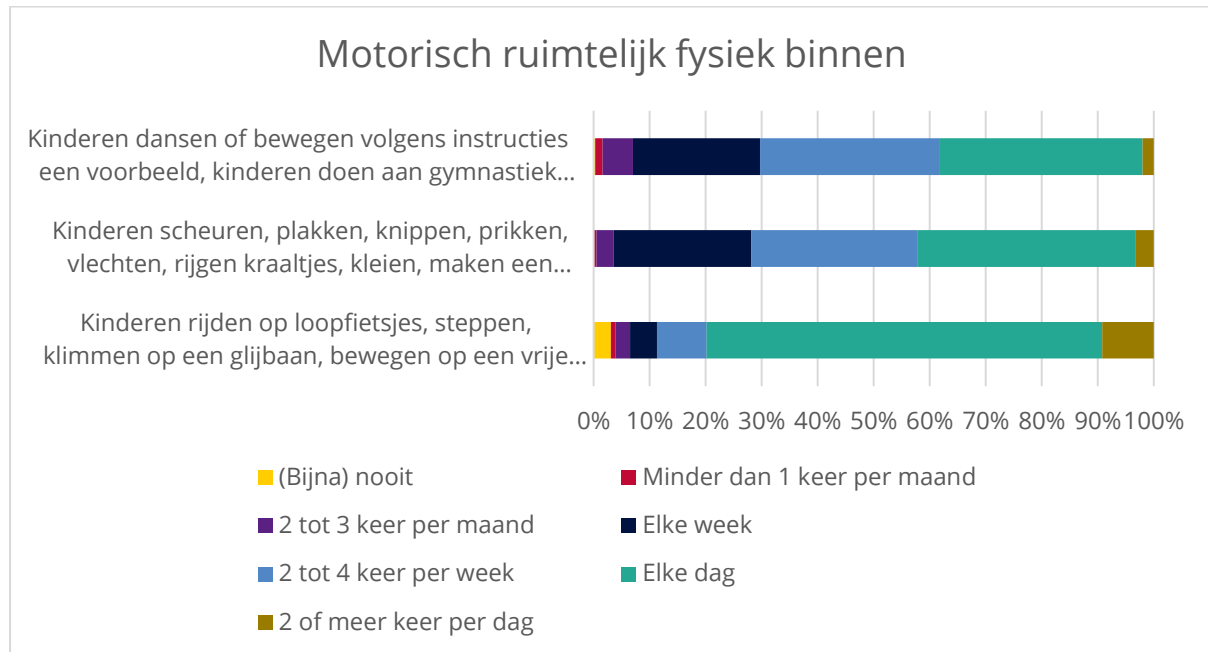
Achtergrond pedagogisch medewerker	N	Range	M	SD
Leeftijd	141	23-66	44.21	11.79
Werkdagen	141	1-5	3.68	0.98
Werktijd in uren	141	9-39	24.99	6.71
Opleidingsniveau	96	5-8	6.02	1.05
Hoogste vooropleiding (1-8)	92	1-8	5.77	1.43
Niveau vervolgopleiding (1-8)	89	5-8	5.90	1.03
Werkervaring in kinderopvang in jaren	92	0.5-43	18.79	10.53
Achtergrond leidinggevenden	N	Range	M	SD
Leeftijd	89	29-64	49.12	9.69
Opleidingsniveau	30	5-8	6.37	1.19
Vooropleiding (1-8)	30	3-8	6.17	1.18
Niveau vervolgopleiding (1-8)	28	5-8	6.21	1.20
Werkervaring in kinderopvang in jaren	29	0-43	20.41	11.00

Tabel C-8 Achtergrond van de pedagogisch medewerkers en leidinggevenden (2022-2023)

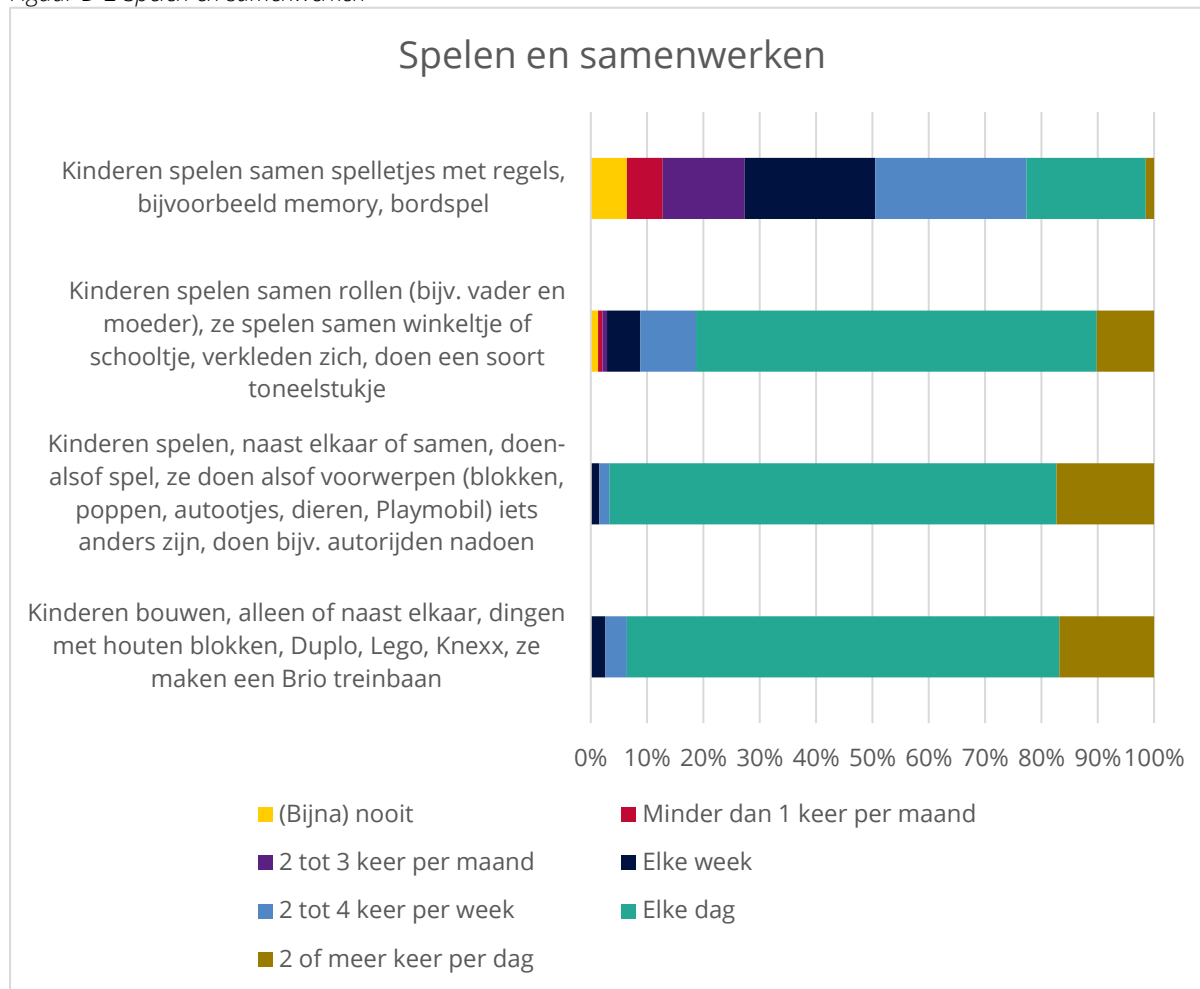
Achtergrond pedagogisch medewerker	N	Range	M	SD
Leeftijd	138	21-65	45.67	10.52
Werkdagen	137	1-5	3.79	0.90
Werktijd in uren	136	2-38	24.44	6.42
Opleidingsniveau	61	5-8	5.54	0.92
Hoogste vooropleiding (1-8)	56	1-8	5.07	1.48
Niveau vervolgopleiding (1-8)	57	2-7	5.37	0.94
Werkervaring in kinderopvang in jaren	50	2-40	19.69	8.95
Achtergrond leidinggevenden	N	Range	M	SD
Leeftijd	87	26-65	47.16	10.18
Opleidingsniveau	32	5-8	6.59	0.98
Vooropleiding (1-8)	32	3-8	6.25	1.39
Niveau vervolgopleiding (1-8)	22	5-8	6.41	1.14
Werkervaring in kinderopvang in jaren	31	0-35	16.06	10.31

Appendix D. Activiteitenaanbod

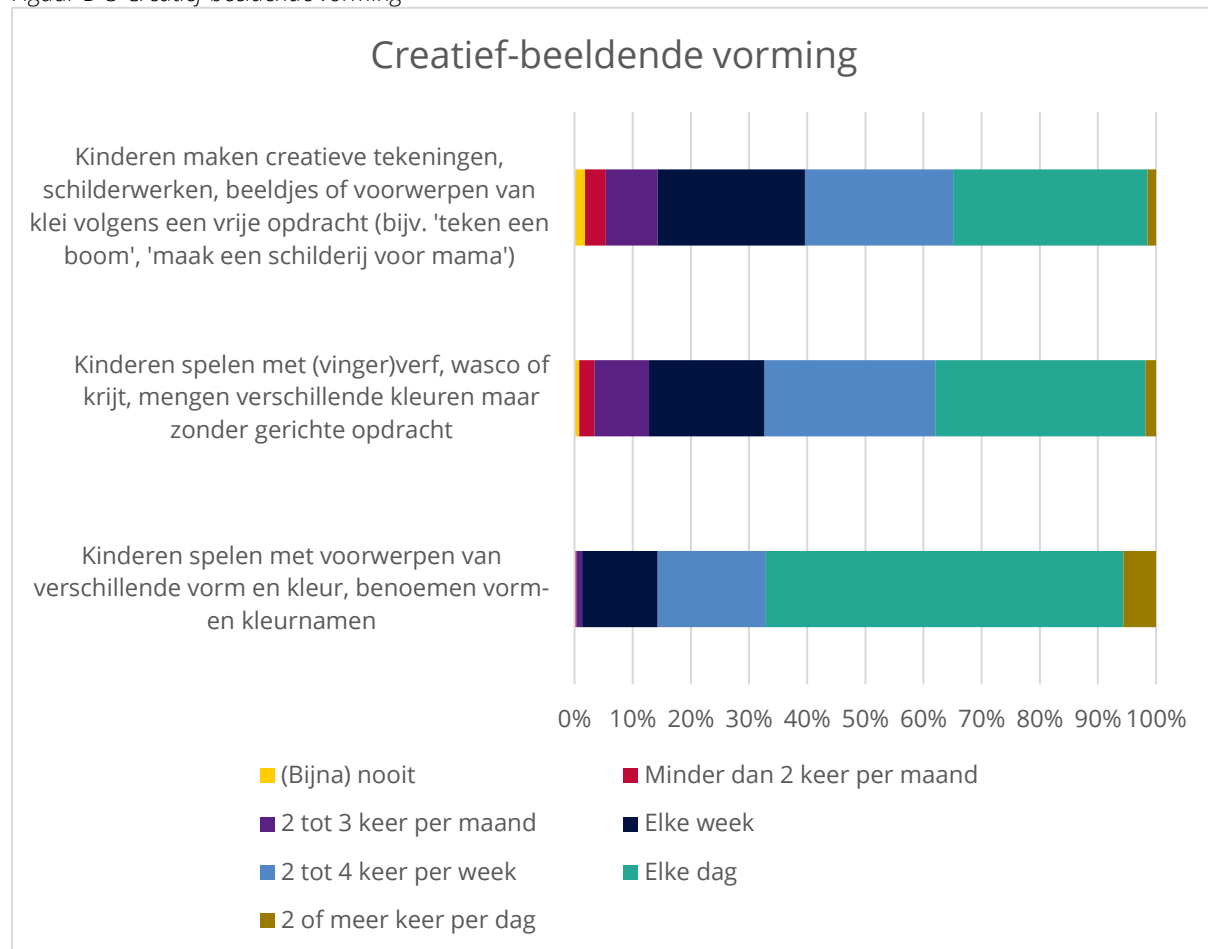
Figuur D-1 Motorisch ruimtelijk fysiek binnen



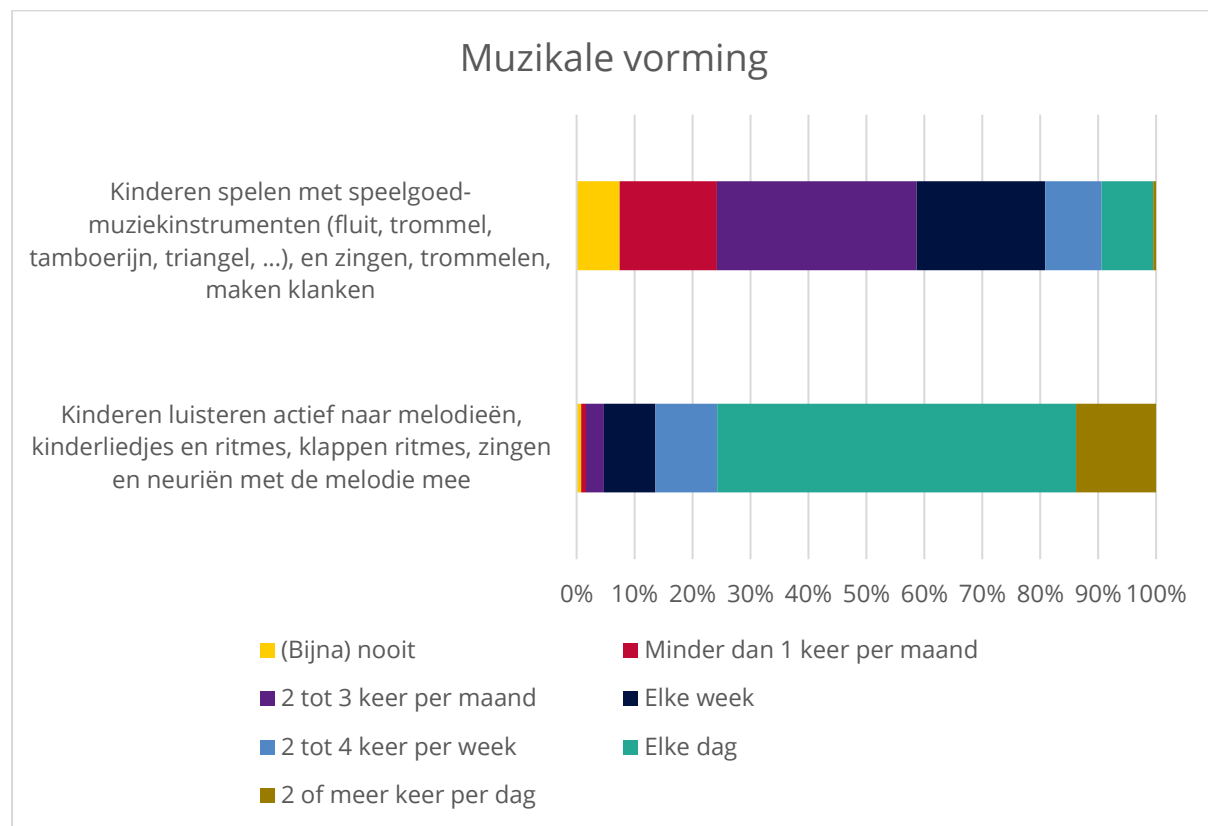
Figuur D-2 Spelen en samenwerken



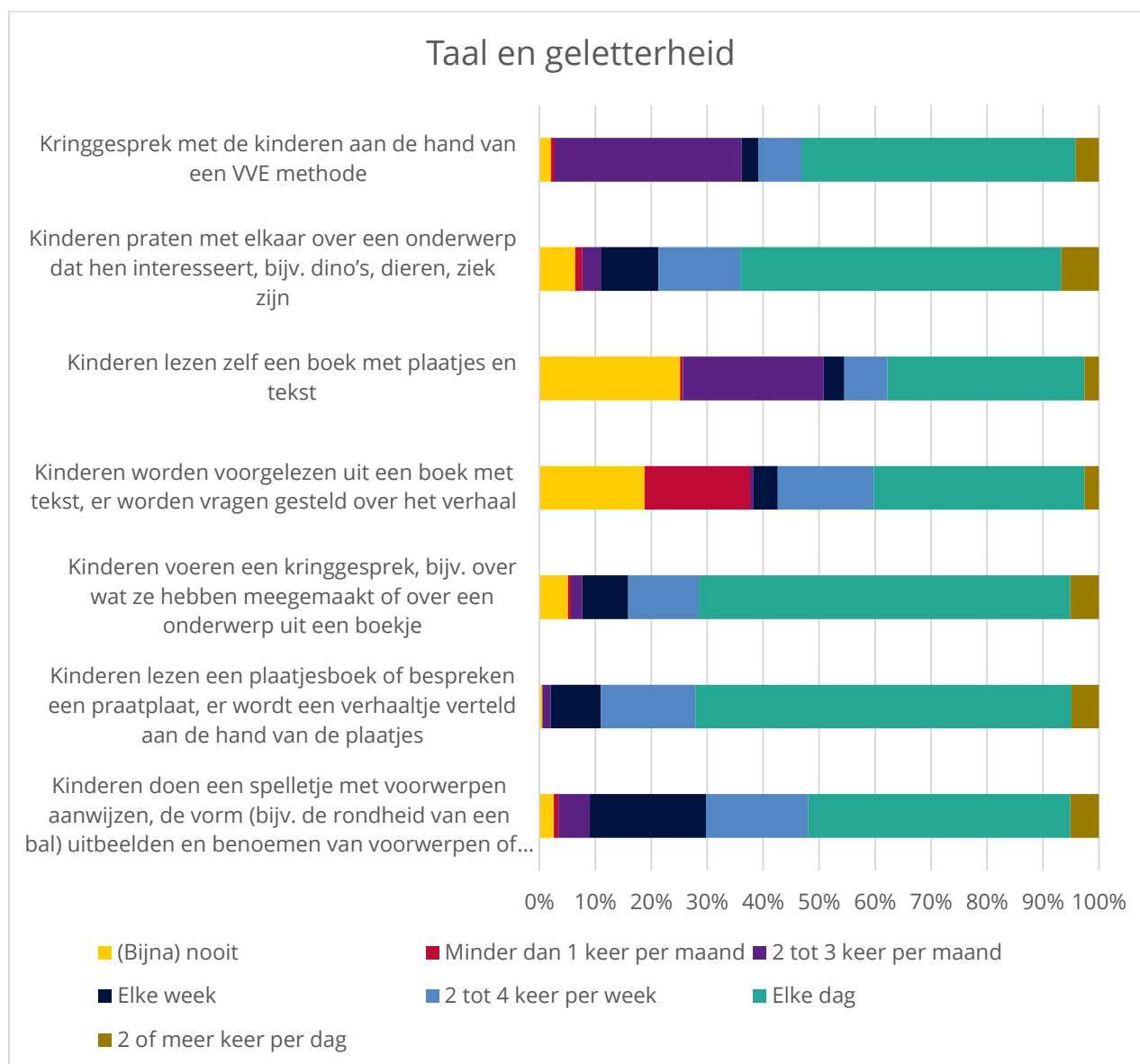
Figuur D-3 Creatief-beeldende vorming



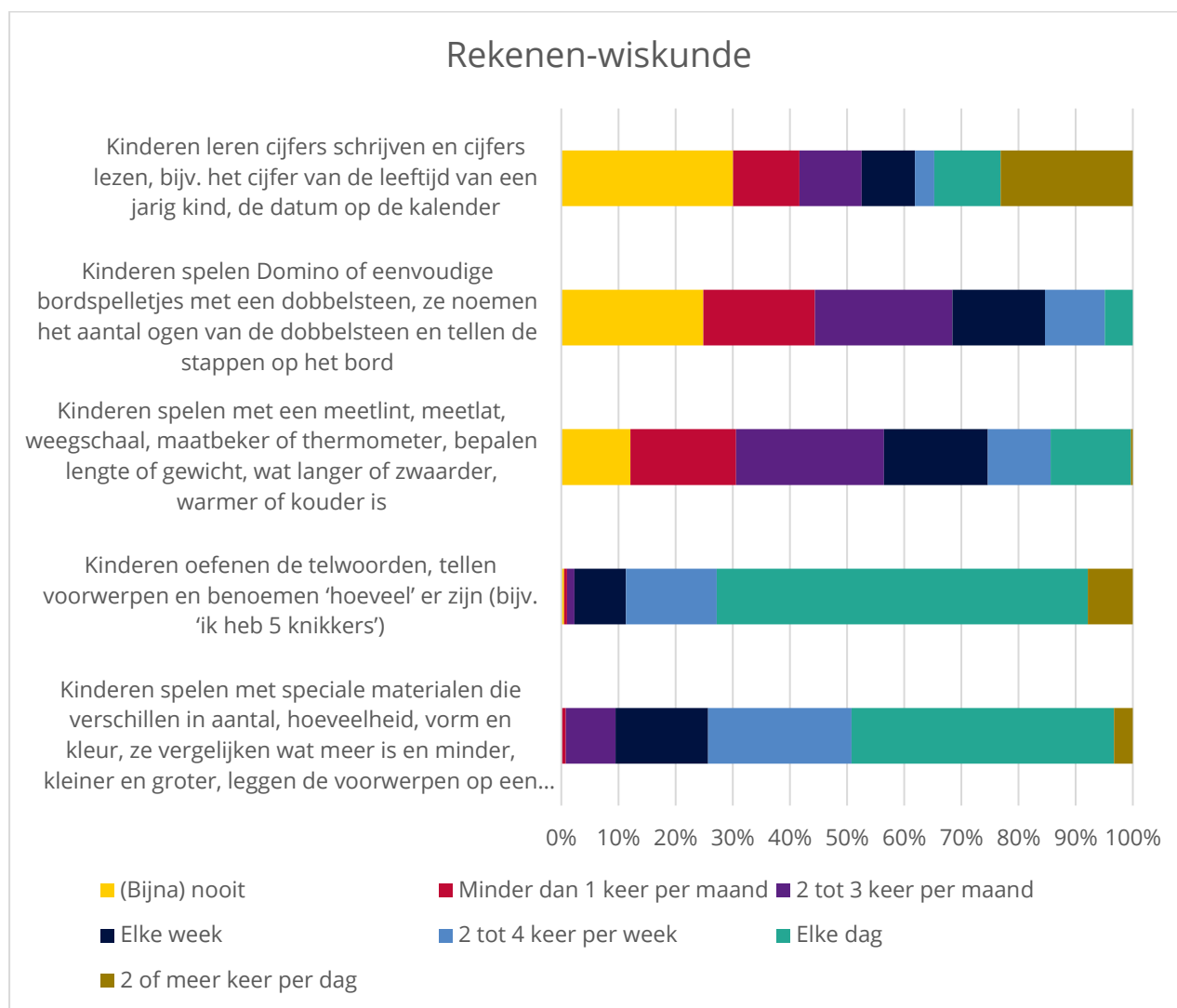
Figuur D-4 Muzikale vorming



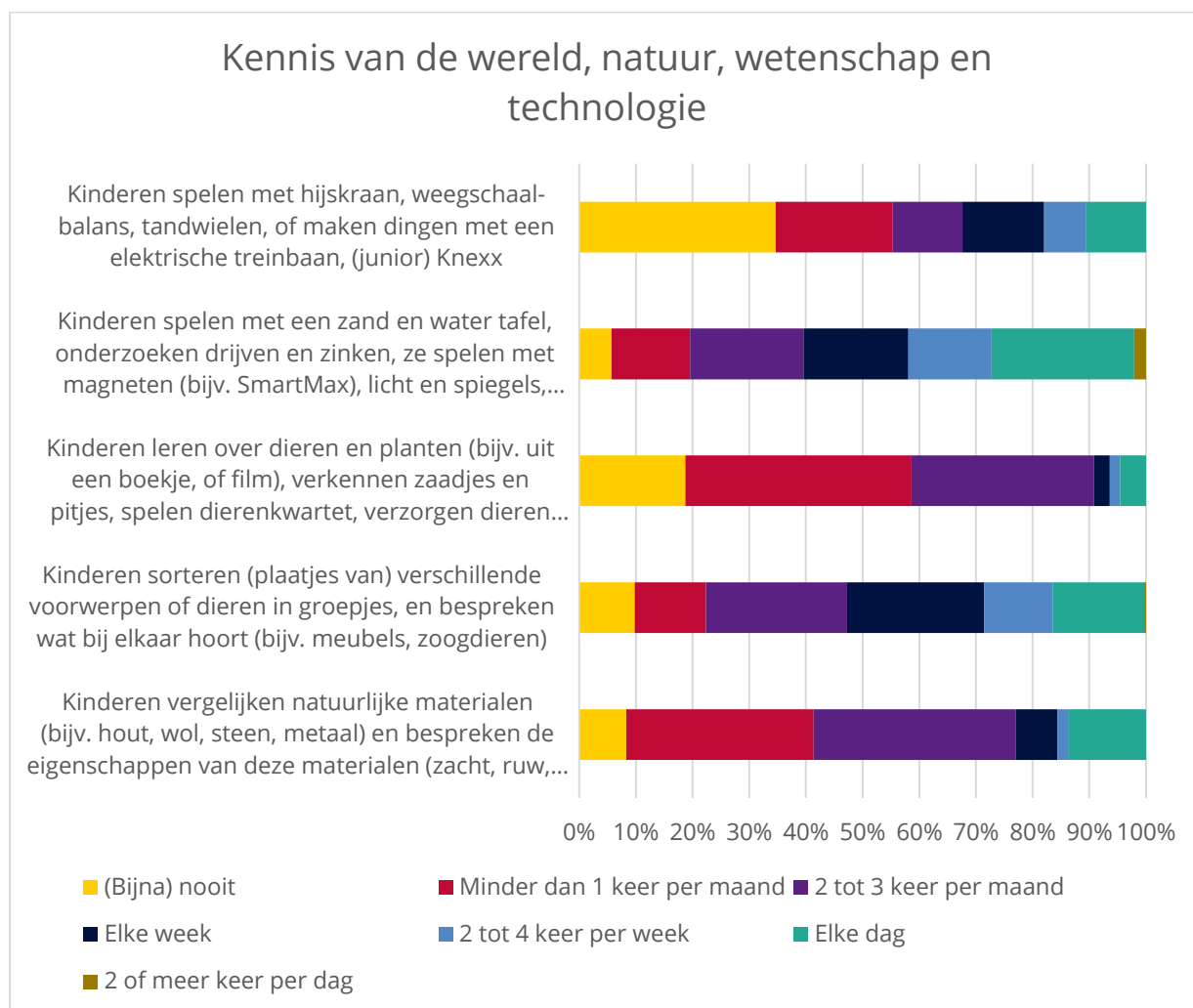
Figuur D-5 Taal en geletterdheid



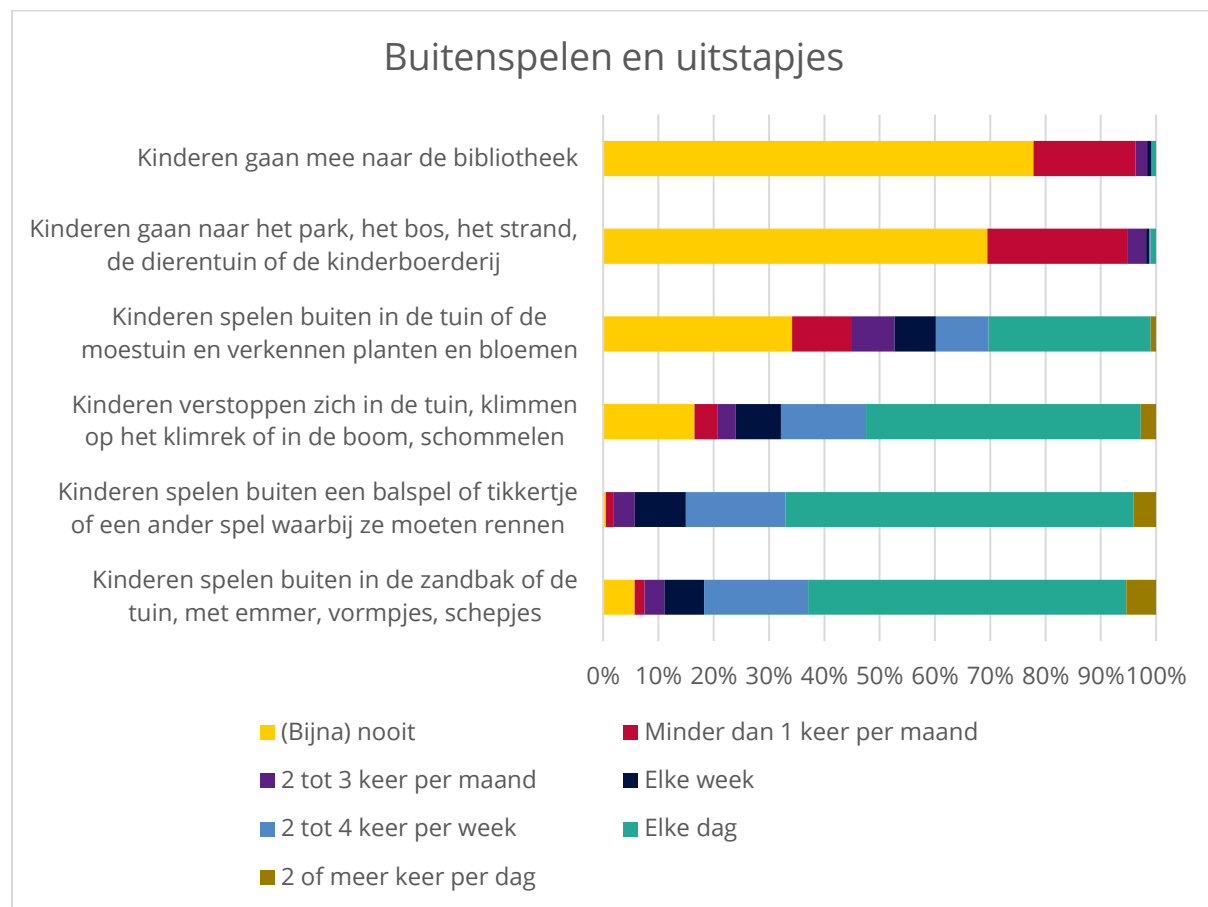
Figuur D-6 Rekenen-wiskunde



Figuur D-7 Kennis van de wereld, natuur, wetenschap en technologie



Figuur D-8 Buitenspelen en uitstapjes



Appendix E. Meetinvariantie en schalen van scores

In het EVENING onderzoek zijn dezelfde meetinstrumenten gebruikt voor kinderen van verschillende leeftijden, geslacht en meertaligheidstatus. Daarbij hebben sommige kinderen de toetsen één keer gemaakt en anderen twee keer. Voordat de scores gebruikt kunnen worden in de analyses moeten eerst de meetinvariantie worden vastgesteld. Meetinvariantie betekent dat een meetinstrument (toets of vragenlijst) dezelfde eigenschap op dezelfde manier meet in verschillende groepen. Dit is belangrijk om ervoor te zorgen dat de resultaten van een onderzoek vergelijkbaar zijn tussen verschillende groepen (Agostinelli & Wiswall, 20125; Milsap, 2011). In het onderzoek hebben we eerst vastgesteld dat de toetsen hetzelfde meten bij kinderen van verschillende leeftijdsgroepen, geslachten, meertaligheidsgroepen en voor kinderen die één of twee keer hebben meegedaan.

Om meetinvariantie vast te stellen moet, per meetinstrument (toets of vragenlijst) een factormodel getoetst worden. In de analyse wordt gekeken of hetzelfde factormodel past bij verschillende groepen (bijvoorbeeld twee leeftijdsgroepen, jongens en meisjes etc.). Wanneer nodig kunnen bepaalde toets-items of vragen uit de vragenlijst worden verwijderd om zo een vergelijkbaar meetinstrument te krijgen (van der Schoot et al., 2012; Leitgöb et al., 2024).

De analyses zijn uitgevoerd in het programma Mplus (Muthén & Muthén, 2017). In de eerste stap wordt per meetinstrument gekeken of een factormodel waarbij alle items (toets-items of vragen uit de vragenlijst) laden op één factor. Wanneer nodig kunnen al in deze fase bepaalde items verwijderd worden. Vervolgens wordt de multigroep-optie gebruikt om het factormodel te vergelijken tussen twee groepen (bijvoorbeeld twee leeftijdsgroepen). Bij deze vergelijking kan er gesproken worden van verschillende niveaus van meetinvariantie. Ten eerste: *configural invariantie*. Dit is het basisniveau waarbij wordt getoetst of dezelfde items op de factor laden in beide groepen. Dit geeft aan dat de structuur van de factor hetzelfde is. In andere woorden, dezelfde items meten dezelfde onderliggende constructen in beide groepen. Ten tweede: *metrisch invariantie*, waarbij wordt getoetst of de factorladingen van de items gelijk zijn tussen de groepen. Dit houdt in dat de relatie tussen de items en de onderliggende constructen die het instrument beoogt te meten hetzelfde zijn in elke groep. Als laatste: *scalar invariantie*. Naast de factorladingen moeten ook de intercepten van de items gelijk zijn tussen de groepen. De intercepten zijn de verwachte gemiddelde scores op een item wanneer de onderliggende factor een score nul heeft. Scalar invariance is belangrijk omdat het ons in staat stelt om de gemiddelde scores op de onderliggende factoren eerlijk te vergelijken tussen verschillende groepen. Zonder scalar invariance zouden verschillen in gemiddelde scores kunnen worden toegeschreven aan verschillen in de meeteigenschappen van de items, in plaats van aan werkelijke verschillen tussen de groepen. Alleen als de scalar invariance kan worden vastgesteld kunnen er gemiddelden worden vergeleken tussen de groepen (van der Schoot et al., 2012; Leitgöb et al., 2024).

Sommige toetsen bevatten categorische items. Dat wil zeggen, items die maar 2 of 3 niveaus hebben (bijvoorbeeld “goed” of “fout”; of “niet waar”, “een beetje waar”, “helemaal waar”). Bij dit soort type items is het vaststellen van metrisch invariantie zonder scalar invariantie niet mogelijk. Bij toetsen en vragenlijst met dit type items hebben we de configural en scalar invariantie getoetst. Aangezien de scalar invariantie metrisch invariantie vereist geeft dit een goed beeld van invariantie (Muthén & Muthén, 2017).

Per gebruikt instrument is eerst de meetinvariantie getoetst in deze volgorde: leeftijd, geslacht, meertaligheid, heeft het kind één of twee keer meegedaan. Alleen voor de CELF (maat voor productieve woordenschat) is geen meetinvariantie voor leeftijd en herhaling van de toets getoetst, omdat de CELF alleen tijdens de meting rond de leeftijd van 4 jaar is afgenomen. Voor de vragenlijsten ECBQ en SDQ zijn twee factoren gemaakt, omdat de vragenlijsten twee constructen beogen te meten (zie beschrijving hieronder). Zoals boven genoemd, is het mogelijk dat tijdens het fitten van de verschillende modellen blijkt dat sommige items verwijderd moeten worden om invariantie te bewerkstelligen. In het geval dat items zijn weggehaald is de analyse voor de voorgaande stappen herhaald zonder de weggehaalde items. Op die manier ontstaat het uiteindelijke model dat invariant is op alle dimensies. Bij alle analyses is er rekening gehouden met de nesting van de data in kinderen, omdat sommige kinderen twee keer hebben meegedaan. Dit is gedaan door model type “complex” te gebruiken in Mplus waarbij de data wordt geclusterd op kindniveau (Muthén & Muthén, 2017).

In onze data is scalar invariantie vastgesteld voor alle meetinstrumenten en alle groepsverschillen. Alleen voor de schaal grove motoriek kon dit niet worden vastgesteld. De vier items die behoren tot de schaal grove motoriek waren redelijk makkelijk voor de meeste kinderen (behaald door over 80% van de kinderen). Dat resulteert in een niet goed fittend model. Daarom is grove motoriek niet verder meegenomen in het schalen van de scores en de analyses. Tabellen E-1 t/m E-4 geven de fit indices weer van het uiteindelijke model voor iedere test en vragenlijst voor de vier groepsvergelijkingen. Ook wordt vermeld hoeveel items verwijderd zijn. We rapporteren een aantal fit indices. De Chi kwadraat is niet een goede maat om te gebruiken bij steekproeven van deze omvang (Zheng, & Bentler, 2024). Echter, als de omvang van de Chi kwadraat minder dan 2 keer de vrijheidsgraden is wordt dit gezien als goede fit (Byrne, 2010). Verder wordt goede fit weergegeven door CFI en TLI waarden van .90 of hoger, RMSEA waarden kleiner dan .08 en SRMR waarden van .09 of kleiner. De laatste kolom (“Aantal items verwijderd”) is alleen opgenomen in Tabel E-1 omdat dit aantal gelijk is gehouden voor alle vergelijkingen in het eindmodel (zoals hierboven uitgelegd).

De wachttaken bevatten maar één item per taak. Daarom kan er geen factormodel gemaakt worden voor deze taken. Wel is onderzocht of een factor waarbij beide wachttaken op één factor laden invariant is. Dit model kan niet getoetst worden als dusdanig, omdat de factor maar twee items heeft die erop laden. Dit betekent dat het model niet is geïdentificeerd. Om het model te identificeren zijn de variabelen geslacht en meertaligheid als voorspellers meegenomen. Dit model is getoetst op invariantie

voor leeftijd en voor kinderen die aan één of twee meetrondes hebben meegedaan. De invariantie is vervolgens vastgesteld.

De tabellen laten over het algemeen zien dat model fit van de eindmodellen in orde is. In enkele gevallen is de waarde van SRMR net te hoog. Dit duidt op wat hogere correlaties tussen residuen bij de items. Dit kan bijvoorbeeld komen door gecorreleerde meetfouten. Echter, aangezien de waarden net boven het grenswaardes zijn en de andere fit indices wel allemaal goed zijn wordt de fit van deze modellen ook als acceptabel beschouwd.

Tabel E-1 Fit indices eindmodel (scalar invariantie) bij vergelijking leeftijdsgroepen

	Chi kwadraat (df)	CFI	TLI	RMSEA	SRMR	Aantal items verwijderd
Klankentaak	17.09 (22)	1.00	1.00	.00	.06	6 van 12
PPVT	514.54 (438) **	0.97	0.97	.03	.09	2 van 24
CELF	NVT	NVT	NVT	NVT	NVT	6 van 20
NWR	127.36 (97) *	0.96	0.95	.04	.11	1 van 12
Selectieve aandacht	11.02 (10)	1.00	1.00	.02	.04	0 van 4
Doosjestaak	7.10 (4)	.96	.94	.06	.03	2 van 5
Wachtaken	4.65 (3)	.99	.98	.05	.05	0 van 2
Fijne motoriek	32.66 (18) *	.96	.94	.06	.09	0 van 6
SDQ ouder	225.64 (166) **	.95	.95	.04	.08	Externaliserend gedrag 0 van 10; prosociaal gedrag 1 van 5
SDQ pm'er	187.78 (120) ***	.98	.97	.07	.07	Externaliserend gedrag 3 van 10; prosociaal gedrag 0 van 5
ECQB ouder	107.15 (82) *	.95	.95	.04	.07	Impulscontrole 1 van 6; volgehouden aandacht 1 van 6
ECQB pm'er	164.68 (98) ***	.94	.93	.07	.09	Impulscontrole 0 van 5; volgehouden aandacht 0 van 6

Noot. * $p < .050$; ** $p < .010$; *** $p < .001$

Tabel E-2 Fit indices laatste model bij vergelijking tussen jongens en meisjes

	Chi kwadraat (df)	CFI	TLI	RMSEA	RSMR
Klankentaak	15.73(22)	1.00	1.00	0.00	0.06
PPVT	504.83(438) *	0.98	0.98	0.03	0.08
CELF	192.75(180)	10.00	10.00	0.02	0.07
NWR	104.64(97)	0.99	0.99	0.02	0.09
Selectieve aandacht	14.78(10)	0.99	0.99	0.05	0.03
Doosjestaak	4.04(4)	10.00	10.00	0.01	0.03
Fijne motoriek	21.38(16)	0.99	0.99	0.04	0.05
SDQ ouder	230.78(166) ***	0.94	0.94	0.04	0.08
SDQ pm'er	188.36(120) ***	0.98	0.98	0.07	0.08
ECQB ouder	97.78(81)	0.98	0.97	0.03	0.06
ECQB pm'er	174.15(98) ***	0.94	0.93	0.08	0.09

Noot. *p<.05 **p<.01 ***p<.001

Tabel E-3 Fit indices laatste model bij vergelijking tussen eentalige en meertalige kinderen

	Chi kwadraat (df)	CFI	TLI	RMSEA	RSMR
Klankentaak	15.73(22)	1.00	1.00	.00	.06
PPVT	515.06(438) **	.96	.96	.03	.10
CELF	172.84(166)	1.00	1.00	.02	.10
NWR	105.54(97)	.99	.99	.02	.09
Selectieve aandacht	16.01(10)	.99	.99	.06	.04
Doosjestaak	21.79(16)	.99	.99	.04	.05
Fijne motoriek	3.85(4)	1.00	1.00	.00	.04
SDQ ouder	235.53(166) ***	.94	.93	.05	.08
SDQ pm'er	157.13(120) *	.99	.99	.05	.07
ECQB ouder	124.89(88) **	.92	.91	.05	.07
ECQB pm'er	178.72(98) ***	.93	.93	.08	.10

Noot. *p<.05 **p<.01 ***p<.001

Tabel E-4 Fit indices bij vergelijking tussen kinderen die één of twee keer hebben meegedaan

	Chi kwadraat (df)	CFI	TLI	RMSEA	RSMR
Klankentaak	23.93(22)	.99	.99	.02	.07
PPVT	524.91(438)**	.97	.97	.03	.08
CELF	102.57(97)	.99	.99	.02	.09
NWR	13.72(10)	.99	.99	.04	.03
Selectieve aandacht	2.07(4)	1.00	1.00	.00	.02
Doosjestaak	1.08(3)	1.00	1.00	.00	.02
Fijne motoriek	22.82(14)	.99	.98	.06	.04
SDQ ouder	209.11(166)*	.96	.96	.04	.07
SDQ pm'er	184.33(120)***	.98	.98	.06	.07
ECQB ouder	105.98(82)*	.96	.85	.04	.07
ECQB pm'er	191.73(98)***	.93	.92	.08	.10

Noot. *p<.05 **p<.01 ***p<.001

Schalen van scores

Na het vaststellen van meetinvariantie hebben we de scores op de toetsen en vragenlijsten berekend op basis van de factormodellen. In vervolganalyses konden we niet werken met latente variabelen (zoals de factors). Daarom hebben we de factorladingen gebruikt om een gewogen gemiddelde te berekenen per meetinstrument van alle items die bij het uiteindelijke model zijn meegenomen. Op deze manier behouden onze scores de meetstructuur van de factors, waarvoor we meetinvariantie hebben vastgesteld. Om de methode consistent te houden zijn standaard factorladingen voor de jongere groep gebruikt vanuit het eindmodel waar meetinvariantie voor leeftijd is vastgesteld. Voor de CELF (waar geen leeftijd vergelijking is gemaakt) zijn de factorladingen gebruikt vanuit het basismodel, zonder groepsvergelijking. Voor de wachttaken is geen geschaalde score berekend. Hier is per taak de dichotome score van behaald/niet behaald aangehouden. Omdat de taken maar één item bevatten, is schalen niet mogelijk. Echter, zoals hierboven beschreven zijn er sterke indicaties voor voldoende meetinvariantie.

Tabellen E-5 t/m E-9 laten de factorladingen per meetinstrument zien voor de gekozen items. Alle ladingen zijn significant en meestal boven de .30.

Tabel E-5 Factorladingen Taaltoetsen

PPVT items	PPVT factor ladingen	Klankentaak Items	Klankentaak factorladingen
BLIJ	0.573 ^{***}	BLAD	0.425 ^{**}
BOREN	0.531 ^{***}	DAK	0.712 ^{***}
CACTUS	0.746 ^{***}	KAARS	0.500 ^{***}
GROEPJE	0.414 ^{***}	POT	0.414 ^{**}
HAAI	0.538 ^{***}	TONG	0.403 ^{**}
INSCHENKEN	0.669 ^{***}	VEER	0.444 ^{***}
KANGOEROE	0.557 ^{***}		
KIST	0.552 ^{***}		
		CELF items	CELF factorladingen
KNAGEN	0.519 ^{***}	AFWASSEN	0.533 ^{***}
KRUK	0.692 ^{***}	BRANDWEER	0.822 ^{***}
ONDER_DE_TAFEL	0.469 ^{***}	FIETSEN	0.775 ^{***}
POST	0.633 ^{***}	GITAAR	0.769 ^{***}
STOPCONTACT	0.468 ^{***}	HORLOGE	0.817 ^{***}
TOT_ZIENS	0.506 ^{***}	INKTVIS	0.631 ^{***}
TROMMEL	0.722 ^{***}	INSCHENKEN	0.786 ^{***}
TROMPET	0.667 ^{***}	KALF	0.594 ^{***}
VAREN	0.757 ^{***}	KASTEEL	0.870 ^{***}
VERREKIJKER	0.580 ^{***}	PIANO	0.702 ^{***}
VIERKANT	0.557 ^{***}	ROLSTOEL	0.768 ^{***}
VLOT	0.423 ^{***}	VERREKIJKER	0.868 ^{***}
VUILNIS	0.458 ^{***}	VLAG	0.911 ^{***}
ZIEK	0.556 ^{***}	VOETSTAP	0.777 ^{***}

Noot. *p<.05 **p<.01 ***p<.001

Tabel E-6 Factorladingen Executieve functies taken

Selectieve aandacht items	Selectieve aandacht factor ladingen	NWR Items	NWR factorladingen	Dozentaak items	Dozentaak factorladingen
Item 1	0.791 ^{***}	BAAMOEDIE	0.756 ^{***}	Item 3	0.604 ^{***}
Item 2	0.912 ^{***}	HOLIN	0.736 ^{***}	Item 4	0.554 ^{***}
Item 3	0.782 ^{***}	JAAT	0.625 ^{***}	Item 5	0.586 ^{***}
Item 4	0.513 ^{***}	KEEPON	0.592 ^{***}		
		LIPOETAAN	0.729 ^{***}		
		MIEKAA	0.558 ^{***}		
		NAALIDOE	0.725 ^{***}		
		NATEP	0.700 ^{***}		
		NOELIE	0.497 ^{***}		
		PEELANOT	0.709 ^{***}		
		SIEBOE	0.673 ^{***}		

Noot. *p<.05 **p<.01 ***p<.001

Tabel E-7 Factorloadingen (E)CBQ

	Factor loadingen	
	Ouders	Pm'er
Impulsecontrole items		
Zoeken naar wat het nodig heeft voor een uitje	0.281 ^{***}	NVT
Stoppen als nee wordt gezegd	0.711 ^{***}	0.879 ^{***}
Geduldig wachten op iets wat het graag wil	0.437 ^{***}	0.864 ^{***}
Moeite met stil zitten en mond dicht houden	-	0.583 ^{***}
Stoppen als er stop gezegd wordt	0.848 ^{***}	0.906 ^{***}
Voorzichtig doen als opvoeder 'voorzichtig' zegt	0.457 ^{***}	0.869 ^{***}
Volgehouden aandacht items		
Spelen langer dan 10 minuten met favoriete speelgoed	0.806 ^{***}	0.801 ^{***}
Snel wisselen van speelgoed	0.309 ^{***}	0.736 ^{***}
Met hetzelfde spelen langer dan 5 minuten	0.659 ^{***}	0.790 ^{***}
Snel afgeleid bij luisteren naar verhaal	-	0.620 ^{***}
Makkelijk afgeleid bij kijken naar boeken met plaatjes	0.216 ^{***}	0.706 ^{***}
Erg geconcentreerd tijdens tekenen of kleuren	0.341 ^{***}	0.728 ^{***}

Noot. *p<.05 **p<.01 ***p<.001

Tabel E-8 Factorloadingen SDQ

	Factor loadingen	
	Ouders	Pm'er
Externaliserend probleemgedrag items		
Vaak driftbuien of woede-uitbarstingen	0.442 ^{***}	
Doorgaans gehoorzaam	0.555 ^{***}	0.861 ^{***}
Vecht of pest vaak	0.636 ^{***}	
Maakt vaak ruzie met volwassenen	0.708 ^{***}	
Kan gemeen doen tegen anderen	0.371 ^{***}	0.684 ^{***}
Rusteloos, overactief, kan niet lang stilzitten	0.431 ^{***}	0.864 ^{***}
Constant aan het wiebelen of friemelen	0.491 ^{***}	0.682 ^{***}
Gemakkelijk afgeleid, heeft moeite om zich te concentreren	0.363 ^{***}	0.764 ^{***}
Kan stoppen en nadenken voor iets te doen	0.394 ^{***}	0.791 ^{***}
Maakt opdrachten af, kan de aandacht goed vasthouden	0.437 ^{***}	0.747 ^{***}
Prosociaal gedrag items		
Houdt rekening met gevoelens van anderen	0.783 ^{***}	0.767 ^{***}
Deelt makkelijk	0.549 ^{***}	0.801 ^{***}
Behulpzaam	0.901 ^{***}	0.535 ^{***}
Aardig tegen jongere kinderen		0.874 ^{***}
Biedt vaak vrijwillig hulp aan anderen	0.711 ^{***}	0.602 ^{***}

Noot. *p<.05 **p<.01 ***p<.001

Tabel E-9 factorladingen Motorische ontwikkeling

Items Fijn motoriek	Factor ladingen Fijne motoriek
Rechte lijn tekenen	.61 ^{***}
Rondje tekenen	.80 ^{***}
Knippen	.70 ^{***}
Potloodgrip	.38 ^{***}
Brug bouwen met blokken	.63 ^{***}
Kruis tekenen	.82 ^{***}

Noot. *p<.05 **p<.01 ***p<.001

Testen van structurele invariantie

In de laatste stap voordat algemene domeinscores worden berekend hebben we onderzocht of de correlaties tussen de geschaalde scores onderling overeen komen met de scores tussen de niet-geschaalde somscores. Als deze niet significant van elkaar verschillen hebben we een zogenaamd structurele variantie waarbij de variantie en samenhang structuur in de data is behouden, ook na het schalen van de scores. Omdat we hier correlaties vergelijken die weliswaar uit dezelfde steekproef komen, maar geen overlap hebben in variabelen, doen we dit middels de "Fisher's r to Z" toets (Hoerger, 2013). In Tabel E-10 staan de correlaties tussen de geschaalde scores en tussen de somscores. In Tabel E-11 staan de waardes van Fisher's r to z toets die deze correlaties vergelijken. Tabel E-11 laat zien dat de meeste correlaties niet significant van elkaar verschillen. Wel zijn er een klein aantal correlaties die wel significant van elkaar verschillen. Dit zijn vooral correlaties met de vragenlijsten (ECBQ en SDQ). Echter, het verschil in correlaties is meestal (zeer) klein (Tabachnik & Fidell, 2013). In sommige gevallen is het verschil groter, namelijk de correlatie tussen de ECBQ impulscontrole schaal / SDQ externaliserend gedrag schaal en de ECBQ volgehouden aandacht schaal (allemaal ingevuld zijn door de pedagogisch medewerker). Dit kan komen door problemen met methode-gebonden correlaties, correlaties die ontstaan omdat sommige mensen de vragenlijsten op dezelfde manier invullen (bijvoorbeeld altijd aan de randen van een schaal of midden op een schaal). Deze problemen worden deels opgelost door het schalen van de scores. Op basis van deze analyse kunnen we ook concluderen dat er sprake is van acceptabele structurele invariantie. Daarom wordt er in de volgende fase gewerkt met deze scores om de domeinscores te berekenen.

Tabel E-10 Correlaties tussen somscores en geschaalde scores

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Klankentaak		.518***	.536***	.277***	.278***	.132**	.215***	.214***	.175***	.117*	.022	.216***	.246***	.045	.068	.127*	.193**
2. PPVT	.514***		.760***	.442***	.499***	.251***	.280***	.233***	.467***	.133*	.142**	.342***	.373***	.014	.118*	.247***	.333***
3. CELF	.542***	.766***		.365***	.273***	.140*	.212***	.294***	.210***	.163**	.089	.264***	.156*	.043	.121*	.254**	.261***
4. NWR	.258***	.461***	.393***		.313***	.240***	.246***	.251***	.294***	.124*	.147**	.207**	.278***	.128*	.020	.232***	.254***
5. Selectieve aandacht	.277***	.515***	.279***	.300***		.286***	.242***	.213***	.499***	.049	.082	.266***	.381***	.042	.134**	.233***	.292***
6. Dozentaak	.128*	.254***	.139*	.265***	.293***		.141**	.125*	.233***	.094	.079	.120	.263***	.083	.121*	.111	.173**
7. Wachtttaak rozijn	.210***	.276***	.214***	.259***	.266***	.149**		.582***	.266***	.247***	.078	.234***	.199**	.121*	.130**	.278***	.263***
8. Wachtttaak cadeau	.208***	.334***	.295***	.241***	.230***	.108*	.582***		.250***	.174***	.073	.277***	.283***	.068	.037	.254***	.289***
9. Fijne motoriek	.197***	.490***	.251***	.274***	.538***	.246***	.247***	.231***		.124*	.212***	.269***	.432***	.075	.164***	.277***	.284***
10. SDQ exter. ouder	.110*	.132**	.162**	.118*	.057	.268***	.174***	.160**	.121*		.322***	.327***	.212***	.477***	.396***	.369***	.248***
11. SDQ prosociaal ouder	.015	.141**	.095	.154**	.110*	.086	.076	.079	.212*	.338***		.241***	.366***	.287***	.164**	.231***	.200**
12. SDQ exter. pmer	.200**	.345***	.266**	.195**	.291***	.241***	.281***	.296***	.286***	.322***	.247***		.600***	.171**	.206***	.764***	.775***
13. SDQ prosociaal gedrag pmer	.236***	.371***	.157*	.242***	.379***	.209***	.288***	.284***	.443***	.231***	.364***	.619**		.158*	.070	.535***	.458***
14. ECBQ impuls controle ouder	.053	.021	-.023	.125*	.055	.135**	.087	.098*	.077	.484***	.269***	.168**	.165**		.296***	.156*	.074
15. ECBQ aandacht ouder	.065	-.010	.019	-.073	.075	.119*	.044	.040	.070	.298***	.079	.049	-.031	.312***		.144*	.195**
16. ECBQ impuls controle pmer	.051	.216***	.159*	.101	.221***	.207***	.225***	.272***	.320***	.303***	.258***	.637***	.562***	.112	.048		.639***
17. ECBQ aandacht pmer	.131*	.126*	.057	.125	.171**	.077	.102	.084	.235***	.136*	.101	.339***	.285***	.070	.030	.276***	

Noot. Correlaties aan de bovenkant van de diagonaal zijn correlaties tussen somscores. Correlaties aan de onderkant zijn correlaties tussen geschaalde scores. ext. = externaliserend;

pros=prosociaal; imp=impuls controle; v.a.=volgehouden aandacht.

*p<.05. **p<.01. ***p<.001

Tabel E-11 Z waardes van Fisher's r to z toets

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Klankentaak																
2. PPVT	0.077															
3. CELF	0.099	0.170														
4. NWR	0.271	0.319	0.370													
5. Selectieve aandacht	0.015	0.305	0.076	0.191												
6. Dozentaak	0.056	0.045	0.012	0.352	0.106											
7. Wachtttaak rozijn	0.074	0.063	0.025	0.185	0.363	0.115										
8. Wachtttaak cadeau	0.089	1.586	0.013	0.142	0.253	0.242	NVT									
9. Fijne motoriek	0.323	0.432	0.512	0.291	0.756	0.123	0.294	0.291								
10. SDQ exter. ouder	0.099	0.015	0.012	0.080	0.112	2.507	1.090	0.205	0.044							
11. SDQ prosociaal ouder	0.098	0.015	0.070	0.095	0.395	0.098	0.029	0.086	0.000	0.257						
12. SDQ exter. pmer	0.190	0.039	0.019	0.135	0.308	1.414	0.583	0.239	0.214	0.064	0.073					
13. SDQ prosociaal gedrag pmer	0.120	0.027	0.009	0.416	0.027	0.646	1.096	0.013	0.158	0.229	0.026	0.351				
14. ECBQ impuls controle ouder	0.113	0.101	0.774	0.041	0.183	0.735	0.493	0.433	0.029	0.130	0.279	0.035	0.083			
15. ECBQ aandacht ouder	0.042	1.843*	1.201	1.236	0.838	0.028	1.242	0.043	1.368	1.594	1.233	1.834*	1.160	0.253		
16. ECBQ impuls controle pmer	0.862	0.376	0.884	1.442	0.142	1.101	0.648	0.221	0.541	0.842	0.325	2.902**	0.444	0.509	1.099	
17. ECBQ aandacht pmer	0.720	2.536**	1.879*	1.442	1.450	1.098	1.926*	2.455**	0.607	1.328	1.156	7.869***	2.334*	0.046	1.914*	5.435***

Noot. * $p < .05$. ** $p < .01$. *** $p < .001$. ext. = externaliserend; pros=prosociaal; imp=impuls controle; v.a.=volgehouden aandacht. De scores van de twee wachttaken zijn niet geschaald.

Appendix F. Domeinscores

In deze appendix staan de fit indices en factorladingen van de vier modellen die gebruikt zijn om de domeinscores op de vier domeinen (Taal, Cool Executive Functies, Hot Executieve Functies, Sociaal-emotioneel) te berekenen. Zoals in hoofdstuk 4 vermeld staat is de dozentaak niet meegenomen in de berekening van de domeinscores. Omdat de CELF alleen is afgenomen bij de oudere groep (rond leeftijd 3 jaar en 10 maanden) zijn er twee modellen gefit. In één model zijn alleen de oudere kinderen meegenomen en is de CELF in het domein taal opgenomen en in het andere model zijn alle kinderen (dus ook de jongere kinderen) meegenomen en is CELF niet opgenomen. Bij analyses die alleen over de oudere groep gaan wordt gebruik gemaakt van het eerste model en bij analyses die over het hele steekproef gaan, wordt er gebruik gemaakt van het tweede model. Verder bleek dat de modellen niet goed pasten wanneer de antwoorden van ouders op de volgehouden aandacht schaal van de ECBQ werden meegenomen. Daarom is deze schaal niet meegenomen. Uit ervaringen van test-assistenten bleek ook dat ouders de vragen bij deze schaal vaak lastig vonden. Dit kwam vooral omdat vragen soms ook negatief worden verwoord.

In Tabel F-1 staan de fit indices van de twee versies van het eindmodel van de vier domeinen. We rapporteren een aantal fit indices. De Chi kwadraat is niet een goede maat om te gebruiken bij steekproeven van deze omvang (Zheng, & Bentler, 2024). Echter, wanneer de Chi kwadraat statistiek minder dan 2 keer de vrijheidsgraden is wordt dit gezien als goede fit (Byrne, 2010). Verder wordt goede fit weergegeven door CFI En TLI waardes van .90 of hoger, RMSEA waarden kleiner dan .08 en SRMR waarden van .90 of kleiner. Tabel F-1 laat zien dat beide modellen voldoen aan deze criteria; er is sprake van een goede model fit. In Tabel F-2 staan de factorladingen van de verschillende toetsen en vragenlijsten per domein. Deze ladingen zijn gebruikt om een totaalscore te berekenen, zoals in hoofdstuk 4 wordt uitgelegd. Ladingen van .30 of hoger worden gezien als voldoende. Alle ladingen voldoen hieraan behalve de ladingen van de ECBQ impulsief gedrag schaal gerapporteerd door ouders. Echter, omdat de lading wel significant is, de model fit goed is en de indicator theoretisch past bij het domein is er gekozen om deze te behouden. Door de lage lading zal het niet veel bijdragen aan het totaalscore op het domein.

Tabel F-1 Fit indices van eindmodellen voor de vier domeinen

	Model met CELF (oudere kinderen)	Model zonder CELF (hele steekproef)
X ² (df)	77.72(70)	100.41(58) ^{***}
CFI	0.99	0.95
TLI	0.98	0.94
RMSEA	0.02	0.04
SRMR	0.05	0.05

Noot. *p<.05. **p<.01. ***p<.001

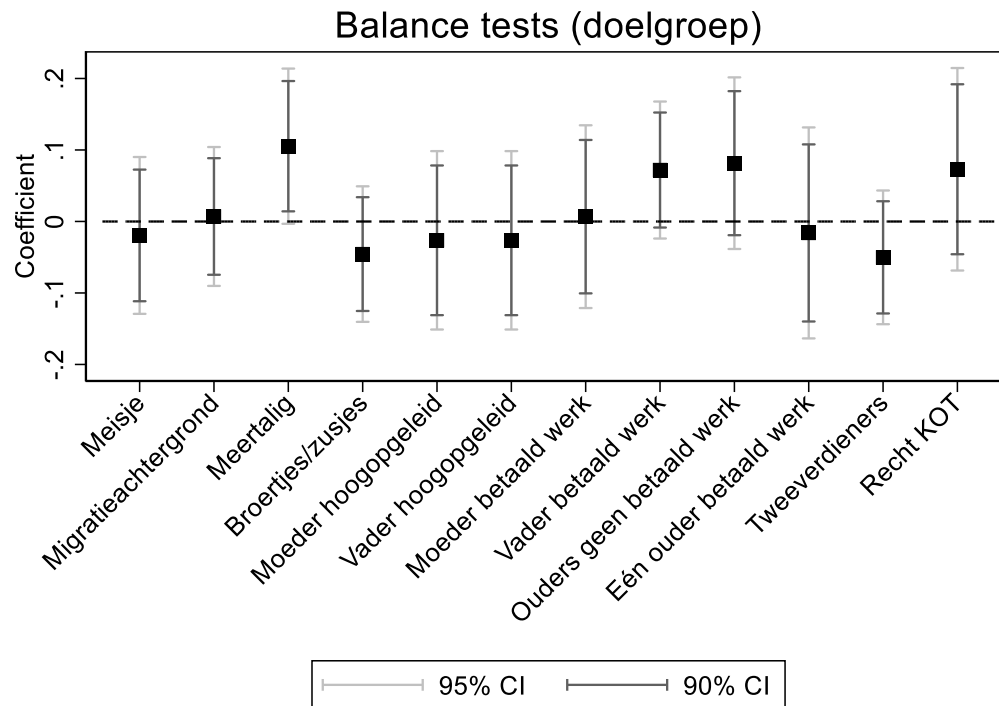
Tabel F-2 Standard factorloadingen van de indicatoren voor de vier domeinen oudere kinderen

	Model met CELF (oudere kinderen)	Model zonder CELF (hele steekproef)
Taal		
Klankentaak	.65***	.51***
PPVT	.88***	.99***
CELF	.86***	NVT
Cool EF		
NWR	.56***	.55***
Selectieve aandacht	.43***	.61***
Hot EF		
Wachttaak rozijn	.93***	.89***
Wachttaak cadeau	.85***	.91***
Sociaal-emotioneel		
SDQ ext. ouder	.49***	.44***
SDQ pros. ouder	.34***	.41***
SDQ ext. pm'er	.81***	.80***
SDQ pros. pm'er	.60***	.80***
ECQB imp. ouder	.16*	.25***
ECBQ imp. pm'er	.60***	.68***
ECBQ v.a. pm'er	.34***	.37***

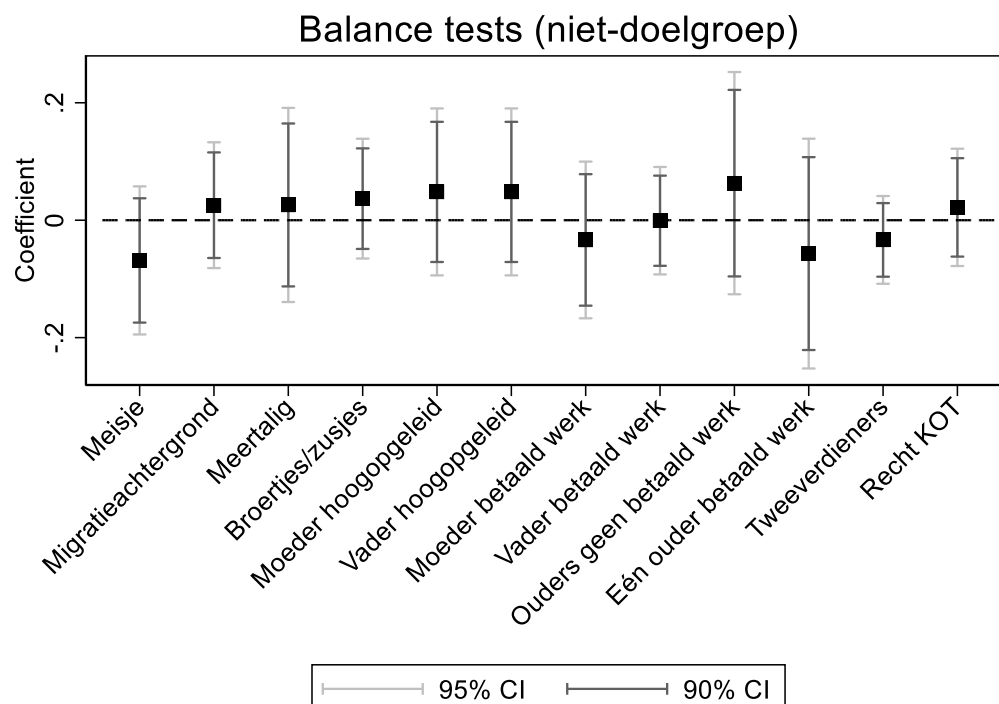
Noot. ext. = externaliserend; pros=prosociaal; imp=impulscontrole;
v.a.=volgehouden aandacht. *p<.05, **p<.01, ***p<.001

Appendix G. Additionele resultaten hoofdstuk 6

Figuur G-1 Balance tests (doelgroep)



Figuur G-2 Balance tests (niet-doelgroep)



Tabel G-1 Effecten van cumulatief urenaanbod op ontwikkeling

	(1) Taal	(2) Cool-EF	(3) Hot-EF	(4) Soc.-emot.	(5) Motoriek
<i>Panel A: Doelgroep</i>					
Cumulatief urenaanbod	0.175** (0.0821)	0.156** (0.0750)	0.0406 (0.121)	0.210* (0.119)	0.280*** (0.103)
N	269	259	267	267	268
R-squared	0.422	0.507	0.292	0.375	0.549
<i>Panel B: Doelgroep meertalig</i>					
Cumulatief urenaanbod	0.281** (0.136)	0.183 (0.111)	-0.0124 (0.150)	0.226 (0.155)	0.364*** (0.130)
N	208	199	207	207	208
R-squared	0.434	0.541	0.293	0.377	0.623
<i>Panel C: Niet doelgroep</i>					
Cumulatief urenaanbod	0.0476 (0.171)	-0.175 (0.179)	-0.0552 (0.175)	0.0236 (0.123)	-0.155 (0.0977)
N	134	130	133	134	133
R-squared	0.595	0.755	0.398	0.520	0.583

Noot: Uitkomsten zijn in z-scores (gemiddelde 0, SD 1). De coëfficiënten geven weer wat de effecten zijn van 100 uur extra voorschoolse educatie. De analyses controleren voor kind-, gezins- en centrumkenmerken. Alle modellen nemen testjaar en gemeente fixed effects mee. Standaardfouten zijn geclusterd op centrumniveau.

***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.10$.

Tabel G-2 Effecten van cumulatief urenaanbod op ontwikkeling: alternatieve specificatie met niet-leeftijdgecorrigeerde cumulatief aanbod

	(1) Taal	(2) Cool-EF	(3) Hot-EF	(4) Soc.-emot.	(5) Motoriek
<i>Panel A: Doelgroep</i>					
Cumulatief urenaanbod	0.192** (0.0768)	0.191** (0.0744)	0.0658 (0.121)	0.226* (0.118)	0.281*** (0.0984)
N	269	259	267	267	268
R-squared	0.424	0.510	0.292	0.377	0.549
<i>Panel B: Doelgroep meertalig</i>					
Cumulatief urenaanbod	0.286** (0.128)	0.208** (0.103)	0.0153 (0.143)	0.208 (0.156)	0.324*** (0.122)
N	208	199	207	207	208
R-squared	0.435	0.544	0.293	0.376	0.619
<i>Panel C: Niet doelgroep</i>					
Cumulatief urenaanbod	0.0584 (0.189)	-0.175 (0.195)	-0.0234 (0.185)	0.0587 (0.122)	-0.121 (0.0963)
N	134	130	133	134	133
R-squared	0.596	0.755	0.396	0.521	0.579

Noot: Uitkomsten zijn in z-scores (gemiddelde 0, SD 1). De coëfficiënten geven weer wat de effecten zijn van 100 uur extra voorschoolse educatie. De analyses controleren voor kind-, gezins- en centrumkenmerken. Alle modellen nemen testjaar en gemeente fixed effects mee. Standaardfouten zijn geclusterd op centrumniveau.

***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.10$.

Tabel G-3 Effecten van cumulatief urenaanbod op ontwikkeling: subsample zonder kinderdagopvang

	(1) Taal	(2) Cool-EF	(3) Hot-EF	(4) Soc.-emot.	(5) Motoriek
<i>Panel A: Doelgroep</i>					
Cumulatief urenaanbod	0.189** (0.0789)	0.158** (0.0751)	0.0493 (0.124)	0.215* (0.121)	0.289*** (0.101)
N	262	253	261	260	261
R-squared	0.439	0.505	0.296	0.371	0.569
<i>Panel B: Doelgroep meertalig</i>					
Cumulatief urenaanbod	0.306** (0.122)	0.175 (0.112)	-0.0260 (0.149)	0.250 (0.161)	0.396*** (0.109)
N	203	195	203	202	203
R-squared	0.459	0.541	0.299	0.379	0.642
<i>Panel C: Niet doelgroep</i>					
Cumulatief urenaanbod	0.0519 (0.166)	-0.225 (0.206)	0.0244 (0.233)	0.150 (0.107)	-0.158 (0.123)
N	109	106	108	109	108
R-squared	0.772	0.781	0.434	0.595	0.690

Noot: Uitkomsten zijn in z-scores (gemiddelde 0, SD 1). De coëfficiënten geven weer wat de effecten zijn van 100 uur extra voorschoolse educatie. De analyses controleren voor kind-, gezins- en centrumkenmerken. Alle modellen nemen testjaar en gemeente fixed effects mee. Standaardfouten zijn geclusterd op centrumniveau.

***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.10$.

Tabel G-4 Effecten van cumulatief urenaanbod op ontwikkeling: subsample eerste meting

	(1) Taal	(2) Cool-EF	(3) Hot-EF	(4) Soc.-emot.	(5) Motoriek
<i>Panel A: Doelgroep</i>					
Cumulatief urenaanbod	0.221** (0.0857)	0.191** (0.0860)	0.0159 (0.140)	0.255* (0.136)	0.338*** (0.115)
N	200	193	199	198	199
R-squared	0.440	0.584	0.331	0.438	0.587
<i>Panel B: Doelgroep meertalig</i>					
Cumulatief urenaanbod	0.359*** (0.120)	0.170 (0.135)	-0.0371 (0.157)	0.301* (0.176)	0.410*** (0.125)
N	154	148	154	153	154
R-squared	0.518	0.617	0.347	0.442	0.679
<i>Panel C: Niet doelgroep</i>					
Cumulatief urenaanbod	0.00539 (0.213)	-0.193 (0.218)	-0.0968 (0.236)	-0.0729 (0.177)	-0.121 (0.126)
N	93	89	92	93	92
R-squared	0.640	0.839	0.501	0.549	0.659

Noot: Uitkomsten zijn in z-scores (gemiddelde 0, SD 1). De coëfficiënten geven weer wat de effecten zijn van 100 uur extra voorschoolse educatie. De analyses controleren voor kind-, gezins- en centrumkenmerken. Alle modellen nemen testjaar en gemeente fixed effects mee. Standaardfouten zijn geclusterd op centrumniveau.

***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.10$.

Tabel G-5 Effecten van cumulatief urengebruik op ontwikkeling van doelgroepeuters

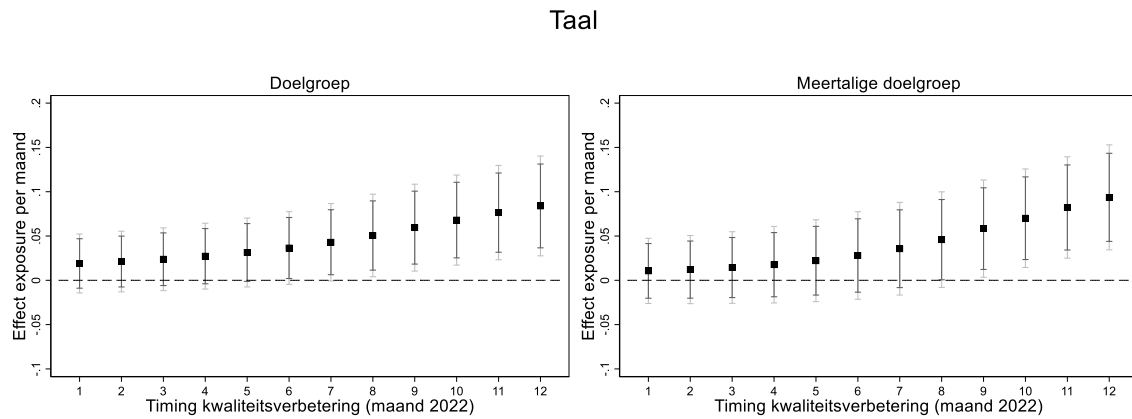
	(1) Taal	(2) Cool-EF	(3) Hot-EF	(4) Soc.-emot.	(5) Motoriek
Cumulatief urengebruik	0.254* (0.132)	0.228* (0.123)	0.0588 (0.160)	0.318* (0.164)	0.399** (0.182)
N	262	253	260	262	261
R-squared	0.203	0.405	0.299	0.066	0.051
F-test instrument	7	7	7	7	7

Noot: De IV-schattingen moeten terughoudend geïnterpreteerd worden omdat het instrument niet sterk genoeg is (lage waarde F-test). Uitkomsten zijn in z-scores (gemiddelde 0, SD 1). De coëfficiënten geven weer wat de effecten zijn van 100 uur extra voorschoolse educatie. De analyses controleren voor kind-, gezins- en centrumkenmerken. Alle modellen nemen testjaar en gemeente fixed effects mee. Standaardfouten zijn geclusterd op centrumniveau.

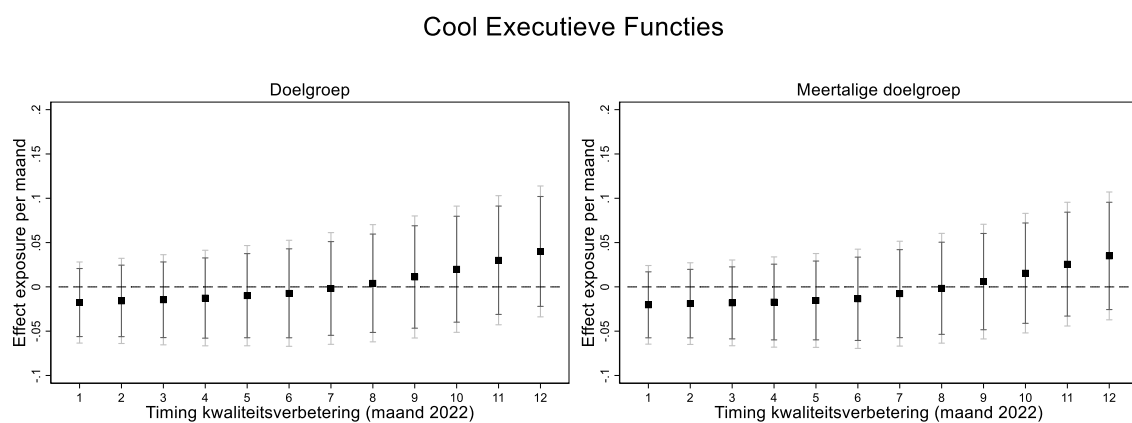
***, $p < 0.01$; **, $p < 0.05$; *, $p < 0.10$.

Appendix H. Additionele resultaten hoofdstuk 7

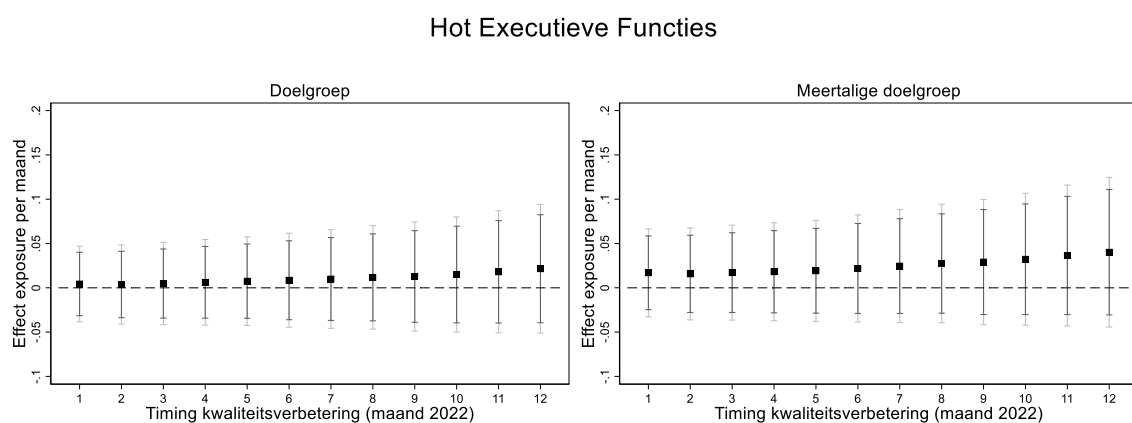
Figuur H-1 Effecten van de pedagogisch beleidsmedewerker op taal



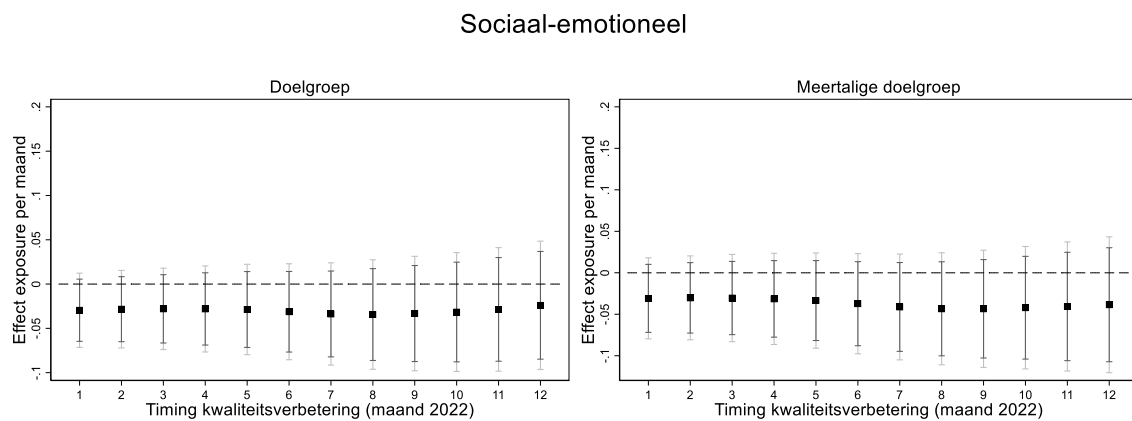
Figuur H-2 Effecten van de pedagogisch beleidsmedewerker op cool executieve functies



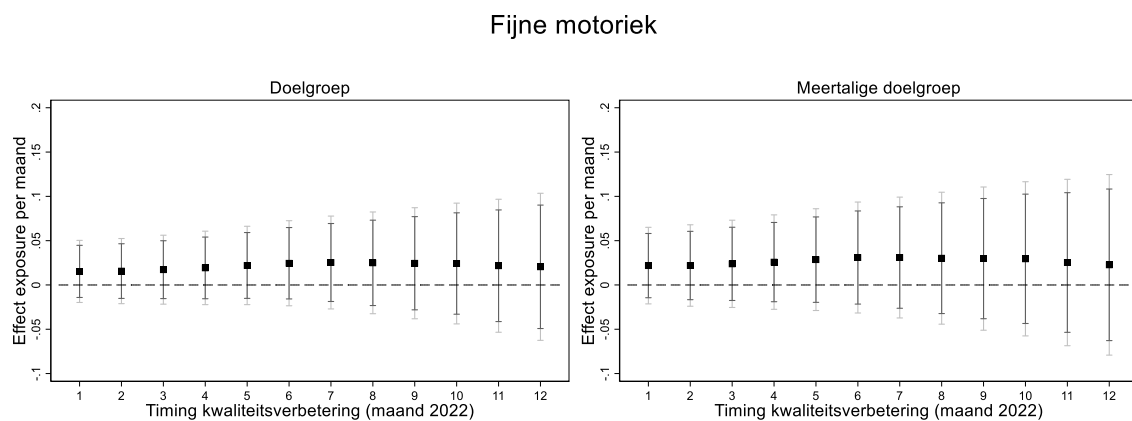
Figuur H-3 Effecten van de pedagogisch beleidsmedewerker op hot executieve functies



Figuur H-4 Effecten van de pedagogisch beleidsmedewerker op sociaal-emotioneel functioneren



Figuur H-5 Effecten van de pedagogisch beleidsmedewerker op fijne motoriek



Figuur H-6 Placebo test: Effecten van de pedagogisch beleidsmedewerker op taal

