

# **Doubleren in groep 3 t/m groep 5 van het basisonderwijs**

Shelley de Wolf

S3491390

Begeleider: dr. N. Frans

2e beoordelaar: prof. L. Baams

Master Orthopedagogiek

Faculteit der Gedrags- en Maatschappijwetenschappen

Rijksuniversiteit Groningen

26 mei 2024

7998 woorden

### Abstract

This research provides insights into which factors best predict the chance of grade retention between first grade and third grade of primary education in the Netherlands. There is a bigger chance of grade retention in the Netherlands than in many other countries. International policy is often implemented nationally and a grade retention only follows if certain conditions are met. In the Netherlands, schools can make their own policy regarding grade retention and the decision seems to be more dependent on the opinion and vision of a teacher. Grade retention appears to be ineffective and research has shown negative long-term effects on school progress and psychosocial aspects for the student. The sample concerns 1402 students from 59 different schools in the Netherlands. The analyses are performed in R, version 4.2.0.. Based on bivariate relationships, students from low-educated parents, with low scores for mathematics or reading comprehension, or with little didactic growth in mathematics midyear are more likely to be retained. When correcting for differences in skill level in a random intercept model, only the mathematics score emerges as a significant predictor of grade retention. Correcting for didactic growth does not improve the model. The results of the random intercept models show little to no variation between the chance of grade retention in different schools. The results may be slightly biased, because of the relatively low number of students with a grade retention in the sample. It is therefore important for follow-up research to ensure a representative sample.

Dit onderzoek biedt inzicht in welke factoren bijdragen aan het voorspellen van de kans op doublure tussen groep 3 en groep 5 van het basisonderwijs in Nederland. De kans om te doubleren is in Nederland groter dan in veel andere landen. Internationaal beleid rondom doublure is vaak landelijk vastgelegd, maar in Nederland kunnen scholen zelf het beleid rondom doublure maken en daardoor lijkt de beslissing afhankelijker te zijn van de mening en visie van een leerkracht. Doublure als interventie blijkt niet effectief en in onderzoek wordt meermaals gerapporteerd over nadelige effecten op lange termijn voor het schoolverloop en psychosociale aspecten voor de leerling. De steekproef betreft 1402 leerlingen van 59 verschillende scholen in Nederland. De analyses zijn uitgevoerd in R, versie 4.2.0.. Op basis van de bivariate relaties blijken leerlingen van laagopgeleide ouders, met een lage reken- en taalvaardigheid of met weinig didactische groei op rekenen-wiskunde halverwege het jaar meer kans te hebben op doublure. Als er wordt gecorrigeerd voor verschillen in het vaardigheidsniveau in het random intercept model, blijkt enkel de rekenvaardigheid een significante voorspeller voor de kans op doublure. Corrigeren voor didactische groei leidt niet

tot een beter model. Bij het random intercept model is er geen variatie tussen scholen zichtbaar. De resultaten bevatten mogelijk een kleine bias, doordat er relatief weinig leerlingen in de steekproef doubleren. Voor eventueel vervolgonderzoek wordt aanbevolen om een representatieve steekproef te gebruiken.

## **Inleiding**

Het percentage leerlingen in Nederland dat doubleert is hoog vergeleken met andere landen: in het basisonderwijs is 22,4% van de leerlingen gedoubleerd, tegenover 7,3% van de leerlingen in andere OECD-landen (OECD, 2010). Hierin is echter ook kleuterverlenging opgenomen. Volgens recentere cijfers is 10% van de leerlingen tussen groep 3 en groep 8 minimaal één keer gedoubleerd (OCO, 2023). Er zijn echter twijfels over de effectiviteit van het doubleren en over de objectiviteit van degene die de beslissing moeten maken. Internationaal is hiervoor veelal landelijk beleid. In Nederland bepalen scholen zelf het beleid. Dit onderzoek richt zich op welke factoren de kans op doublure tussen groep 3 en groep 5 van het basisonderwijs in Nederland voorspellen.

### **Internationaal beleid rondom doublure**

Uit een onderzoek van Eurydice Network (2011) blijkt dat er internationaal wordt aangenomen dat leerlingen een extra kans krijgen om hun leerproces en vaardigheden te verbeteren door een schooljaar over te doen. In bijna alle Europese landen is centraal vastgelegd welke criteria worden gehanteerd bij het nemen van de beslissing of een leerling moet doubleren of niet. In enkele landen is doubleren niet mogelijk. In de structuur van het onderwijs is dan opgenomen dat het aanbod passend moet zijn voor de leeftijd en het vermogen van een kind. Er wordt meer gedifferentieerd, waardoor doubleren niet nodig is. Daarnaast zijn er een aantal landen die leerlingen in vaste klassen laten overgaan, maar aan het einde van de basisschooltijd een aantal leerlingen een jaar extra laten doen als hun niveau nog niet passend is voor een vervolgopleiding.

Criteria die wel kunnen leiden tot doublure zijn onder andere absentie, familieomstandigheden, het gedrag, de didactische gegevens en eerdere interventies (Eurydice Network, 2011). In veel landen blijft een leerling zitten bij veel gemiste onderwijstijd, variërend tussen 25% en 50%. In sommige landen krijgen leerlingen ook cijfers voor gedrag. Als deze cijfers opeenvolgend ondergemiddeld zijn, resulteert dit in doublure. De beoordeling van didactische scores varieert: soms moeten leerlingen op alle leergebieden voldoende scores, terwijl in andere landen één leergebied mag achterblijven. Het moment van beoordeling varieert van halverwege het jaar tot aan de start van de vakantie, en de beoordelingsperiode kan een half schooljaar of alle voorgaande jaren omvatten.

In veel landen verloopt de procedure volgens landelijke richtlijnen. In sommige gevallen wordt dit binnen de school afgehandeld, maar moet er een unanieme beoordeling zijn van een 'expertgroep'. In andere landen beoordeelt een onafhankelijke organisatie een

eventuele doublure. In veel van deze procedures is een vergelijking van de didactische scores met het intelligentieprofiel opgenomen. Indien de scores passen bij de totale intelligentie, is een doublure uitgesloten.

### **Doubleren in Nederland**

In Nederland zijn de regels rondom doubleren niet centraal vastgelegd, waardoor elk schoolbestuur eigen regels kan opstellen (OCO, 2022). Hierdoor varieert het beleid per school. Uit een analyse van beleidsplannen, uitgevoerd door OCO (2023), blijkt dat bij de meeste scholen de didactische gegevens en werkhouding leidend zijn bij een keuze tot doubleren. Vaak is het de leerkracht die, in overleg met een intern begeleider of orthopedagoog, beslist of een leerling een schooljaar over moet doen. Volgens Eurydice Network (2011) zijn er in landen waar een positief idee heerst over doubleren, zoals in Nederland, significant meer doublures dan in landen waar de nadelen en negatieve effecten de overhand hebben in het algemene discours. Nederlandse leerkrachten geven aan dat zij een leerling laten doubleren bij lage leerprestaties, psychosociale problemen, lage intelligentie of als leerlingen ‘nog niet klaar zijn’ voor de volgende stap (Bonvin, 2003).

In Nederland komt zittenblijven dus relatief vaak voor. Bonvin (2003) stelt dat in landen waar de beslissing veelal is gebaseerd op subjectieve leerkrachtbeoordelingen, zoals in Nederland, de percentages hoger liggen dan in landen waar de beslissing wordt genomen op basis van een landelijk protocol. Bij subjectieve leerkrachtbeoordelingen leidt de persoonlijke overtuiging en visie tot verschillende uitkomsten voor een leerling, zelfs nadat wordt gecorrigeerd voor intelligentie. Om leerlingen gelijke kansen te geven, is het daarom belangrijk dat er gebruik wordt gemaakt van protocollen waardoor de persoonlijke visie van de leerkracht minder bepalend is voor het schoolverloop van een leerling.

### **Effecten van doubleren**

Als er wordt gekeken naar het effect van doubleren zijn er grote twijfels over het nut van deze interventie. Vooral op langere termijn heeft doubleren een negatief effect op de schoolprestaties, het welbevinden en de schoolloopbaan van deze leerlingen (Pameijer & Van Minderhout, 2017). In de systematische literatuurstudie van Goos et al. (2013) zijn de verschillen onderzocht tussen leerlingen met grotendeels dezelfde achtergronden, kenmerken en didactische scores, waarvan de ene groep is blijven zitten en de andere groep niet. Op korte termijn heeft doubleren een positief effect op de schoolprestaties, maar deze voorsprong neemt geleidelijk af tot de gedoubleerde leerlingen op langere termijn opnieuw achterlopen ten opzichte van hun klasgenoten. Er worden negatieve effecten gerapporteerd binnen het

psychosociaal functioneren: gedoubleerde leerlingen hebben op lange termijn minder zelfvertrouwen, kwalitatief minder goede relaties met leeftijdsgenoten en minder zin in school (Goos et al., 2013). Daarnaast laten zij meer hyperactief en asociaal gedrag zien en kunnen zij minder goed zelfstandig werken. Enkele longitudinale onderzoeken die zijn opgenomen in de literatuurstudie rapporteren als negatief effect ook dat gedoubleerde leerlingen minder remediëring of ondersteuning hebben ontvangen dan vergelijkbare leerlingen die wel zijn doorgestroomd. Betreft de uitstroom naar het voortgezet onderwijs, blijken de gedoubleerde leerlingen niet meer kans te hebben om naar een hoger niveau voortgezet onderwijs te gaan dan de vergelijkbare leerlingen die niet zijn gedoubleerd.

Er is volgens Hattie (2013) nauwelijks een andere interventie die zo vaak toegepast wordt en tegelijkertijd zo ineffectief is als doubleren. Dit baseert hij op een meta-analyse van 52.600 effectstudies. In een onderzoek van Fraine et al. (2012) wordt doubleren vergeleken met het herhalen van een behandeling die eerder al niet bleek aan te slaan. Hierin wordt benadrukt hoe belangrijk het is om de effectiviteit van doubleren beter onder de aandacht te brengen bij leerkrachten en schoolbesturen.

### **Voorspellende factoren**

Er is al veel onderzoek gedaan naar welke kenmerken van Nederlandse leerlingen samenhangen met doubleren. Een laag beheersingsniveau op het gebied van rekenen of taal is het vaakst de reden dat een leerling doubleert, stelt Fossen (2005) in haar promotieonderzoek naar 604 gedoubleerde leerlingen. In dit onderzoek bleek ook het intelligentiequotiënt van een leerling een significante voorspeller voor de kans op doublure. Een lager IQ betekent een grotere kans op doublure. Driessen en Langen (2010) concluderen in hun onderzoek naar de onderwijspositie van jongens dat er op basis van de leeftijdsverschillen in groep 8 geen sprake is van sekseverschillen als gevolg van doubleren. Op basis van de gebruikte dataset met 13888 leerlingen lijkt geslacht daarom geen voorspeller voor doublure te zijn. In onderzoek van Jeronimus et al. (2015) ( $n=2935$ ) naar de leeftijdsverschillen binnen een klas blijkt dat de relatief jonge leerlingen een grotere kans hebben op doublure. Voor elke extra maand was de relatieve leeftijd geassocieerd met een 17% lagere kans op doublure. In onderzoek van Kloosterman en De Graaf (2010) ( $n=59156$ ) blijkt dat leerlingen uit een gezin met een lage sociaal-economische status (SES) meer kans hebben op doublure. In het tweede, kwalitatieve deel van dit onderzoek lijkt dit een gevolg van de visie van ouders. Gezinnen met een niet-Nederlandse culturele achtergrond maken een groot deel uit van de gezinnen met een lage SES. In deze gezinnen met een andere culturele achtergrond wordt doubleren als minder erg ervaren

dan mogelijk uitstromen naar een lager niveau vervolgonderwijs, dus stemmen ouders eerder in met de optie om hun kind te laten doubleren. De invloed van ouders komt in meer onderzoeken naar voor: Cabus en Ariës (2016) concluderen in hun onderzoek ( $n=9126$ ) dat leerlingen met minder betrokken ouders een grotere kans hebben op doublure. De betrokkenheid bestaat uit contact tussen ouder en leerkracht en de mate van begeleiding bij het inoefenen in de thuissituatie.

Ook kenmerken van de school kunnen invloed hebben op de kans om te doubleren. Doubleren gebeurt niet op alle scholen even vaak; de verschillen tussen scholen zijn erg groot (Centraal Planbureau, 2015). Op sommige scholen komt doubleren helemaal niet voor, terwijl op andere scholen meer dan de helft van de leerlingen minimaal één keer in de schoolloopbaan doubleert. Leerlingen op grotere basisscholen en in goed presterende klassen, hebben een grotere kans op doubleren of een verwijzing naar een ander type onderwijs (Jepma & Meijnen, 2001). De analyses werden gebaseerd op de gegevens van 11101 leerlingen in groep 4 uit het meest recente PRIMA-cohortonderzoek. De kans op doublure is volgens dit onderzoek ook groter op scholen met minder ervaren teams.

Uit internationale onderzoeken blijken meer schoolkenmerken een voorspellende waarde te hebben voor de kans op doublure. In een meta-analyse van Jimerson (2001) blijkt dat een grotere klas ook de kans op doublure vergroot. Een grootschalig onderzoek van Hong & Raudenbush (2005) naar de verschillen tussen doublures op ruim 1200 basisscholen laat zien dat de kans op doublure toeneemt op scholen waar weinig voorzieningen zijn voor zorgleerlingen, waar weinig contact is met ouders en waar een veilig leerklimaat niet altijd kan worden gewaarborgd. In een analyse van de PISA gegevens concluderen Dupriez et al. (2008) dat de kans op doublure groter is als een school weinig aandacht besteedt aan opbrengstgericht werken en er weinig mogelijkheden zijn tot adaptief onderwijs of differentiëren in de klas.

### ***Didactische groei***

Het gebruik van didactische gegevens om te beslissen over het laten doubleren van leerlingen wordt zowel in de beleidsplannen van scholen als in de literatuur benadrukt. Om de resultaten en voortgang van leerlingen bij te houden, zijn Nederlandse basisscholen verplicht om met een leerlingvolgsysteem te werken (Ministerie van Algemene Zaken, 2023a). Deze systemen bestaan uit jaarlijkse toetsen en een eindtoets die per 2023-2024 als de ‘doorstroomtoets’ bekend staat (Ministerie van Algemene Zaken, 2023b). Met deze nieuwe naam wordt beter duidelijk dat de leerling zich blijft ontwikkelen. Deze keuze past bij de verandering in focus van de aanbieders van de leerlingvolgsystemen. Deze is in de laatste jaren

namelijk steeds meer gericht op de doorgaande groei die een leerling laat zien en minder op het vaardigheidsniveau op één specifiek meetmoment (Cito, z.d.-a). Welke invloed deze groei heeft op een keuze om wel of niet te doubleren, wordt nog weinig over gepubliceerd. Het is echter waardevol om verder te onderzoeken in hoeverre het ‘terecht’ is dat lagere scores of weinig groei vaker resulteren in doublure.

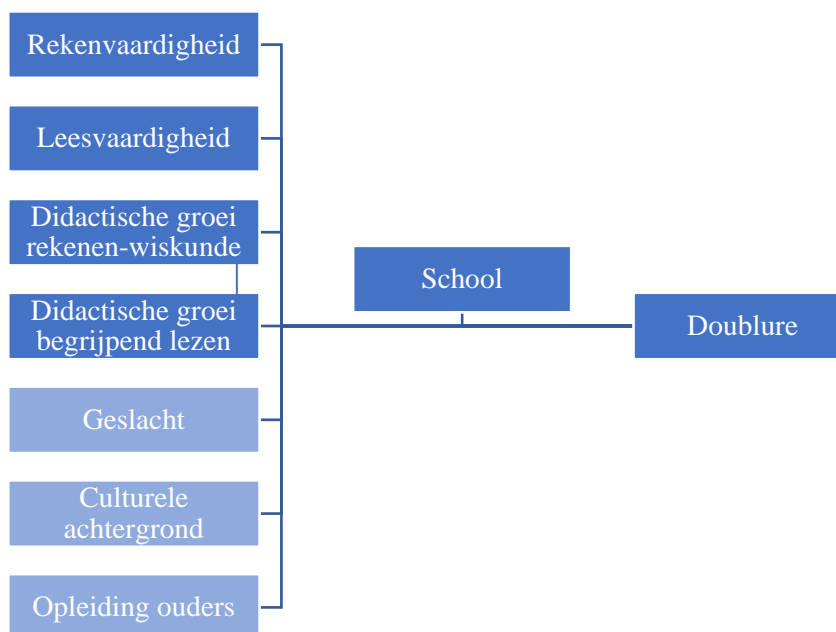
### Onderzoeksvraag

Doubleren komt in Nederland dus relatief veel voor, maar tegelijkertijd wordt de effectiviteit van doubleren in twijfel getrokken. Voor onder andere schoolbesturen en de onderwijsinspectie is het daarom waardevol om meer inzicht te krijgen in de keuzes rondom het laten doubleren van leerlingen en hoe de verschuiving van de focus naar didactisch groei daar invloed op zou kunnen hebben. Dit onderzoek kijkt naar de samenhang tussen verschillende factoren en doublure en er wordt onderzocht of groei ook meespeelt in de keuze. Dit wordt geanalyseerd op basis van de gegevens van groep 3 tot en met 5 in het basisonderwijs. Dit zijn, afgezien van de kleuterverlenging, de groepen waarin doubleren het meest voorkomt (Onderwijsinspectie, 2015). Ook is dit de fase waarin zich specifieke leerproblemen voordoen die vragen om snelle signalering en passende begeleiding (Kievit et al., 2008). De hoofdvraag van dit onderzoek luidt: *Welke factoren wegen mee in de beslissing om een leerling te laten doubleren tussen groep 3 en groep 5?* De subvraag die daarbij gesteld wordt is: *In hoeverre draagt didactische groei bij aan het voorspellen van een doublure tussen groep 3 en groep 5?*

Figuur 1 geeft inzicht in het conceptuele model.

### Figuur 1

*Conceptueel model*





## **Methode**

### **Onderzoeksdesign**

Om de onderzoeksvraag te beantwoorden, wordt in dit kwantitatieve onderzoek bestaande data geanalyseerd uit een eerder onderzoek naar stagnaties in rekenontwikkeling (Frans et al., 2020). Deze data betreffen demografische informatie en toetsgegevens van 2011 tot en met 2014 en bestaan uit meerdere metingen bij dezelfde leerlingen. Dit maakt het huidige onderzoek een retrospectief, longitudinaal cohortonderzoek. Er wordt gebruikgemaakt van een correlatieel onderzoeksdesign, waarbij gekeken wordt naar verbanden tussen leerlingkenmerken, vaardigheidsscores en doublure zonder dat deze variabelen worden gemanipuleerd.

### **Populatie en steekproef**

De doelpopulatie van dit onderzoek zijn onderbouwleerlingen die worden getoetst met de LOVS-toetsen van Cito voor rekenen-wiskunde en begrijpend lezen. Voor de oorspronkelijke steekproef zijn door middel van een random steekproef scholen benaderd. Het gaat hierbij om een geclusterde steekproef, waarbij scholen werden benaderd en leerlingen binnen een school de steekproef vormen. Omdat scholen deelname konden weigeren, is er ook sprake van een gelegenheidssteekproef. Uiteindelijk bestond de steekproef van het oorspronkelijke onderzoek uit 1402 leerlingen van 59 verschillende scholen. In september 2014 zijn deze leerlingen allemaal gestart in groep 6. De verzamelde data betreft LOVS-gegevens uit groep 3, 4 en 5 van de betreffende leerlingen. Bij toetsafnames waar minder dan twee maanden tussen zat, is enkel gekeken naar de eerste toetsafname, om de periodes van groei zo vergelijkbaar mogelijk te houden. De leerlingen waarvan maar één toetsafname beschikbaar was en geen groei vastgesteld kon worden, zijn ook niet opgenomen in de steekproef.

### **Variabelen en instrumenten**

Binnen dit onderzoek wordt de afhankelijke variabele doublure voorspeld aan de hand van ‘geslacht’, ‘culturele achtergrond’, ‘opleiding ouders’ en de vaardigheidsscore en didactische groei van rekenen-wiskunde en begrijpend lezen.

### ***Doublure***

Bij het aanleveren van de data, heeft de school aangegeven of en wanneer een leerling tussen groep 3 en groep 5 is gedoubleerd. De analyses gaan over toetsen die leerlingen maken voorafgaand aan een doublure. De gegevens van leerlingen die niet doubleren worden vergeleken met gegevens van leerlingen die aan het einde van het jaar gaan doubleren. Indien

een leerling is gedoubleerd, worden de toetsgegevens van na de doublure niet meer meegenomen in de analyses.

### ***Rekenen-Wiskunde***

Voor het analyseren van de resultaten en de didactische groei voor rekenen, wordt gebruikgemaakt van de tweede generatie Cito LOVS-toetsen Rekenen-Wiskunde (Janssen et al., 2010). Deze toetsen richten zich op het meten van fundamentele rekenvaardigheden. Voor de groepen 3 tot en met 5 zijn er twee afnamemomenten: halverwege het schooljaar (de M-toetsen) en aan het einde van het schooljaar (de E-toetsen). De toetsen bestaan voornamelijk uit open opgaven waarbij van de leerling een kort antwoord in de vorm van een getal verwacht wordt. De ruwe toetsscores kunnen door gebruik van een IRT-model worden omgezet in vaardigheidsscores, hierna ook de rekenvaardigheid genoemd. Hierdoor is het mogelijk verschillende leerlingen en toetsen met elkaar te vergelijken. De rekenvaardigheid is een afhankelijke variabele op interval niveau. De schaal waarop deze scores worden gemeten loopt van 0 bij M3 tot 101 bij E5 (N. Frans, persoonlijke communicatie, 16 mei 2024). Een hogere rekenvaardigheid betekent een betere beheersing van het rekenaanbod. De opbouw van deze toetsen is te vinden in de wetenschappelijke verantwoording (Janssen et al., 2010).

De COTAN (2015a) heeft de betrouwbaarheid en validiteit van de toetsen beoordeeld. De geschatte betrouwbaarheidscoëfficiënt van elke toets was minstens .91, waardoor deze als ‘goed’ beoordeeld is. Ook de validiteit is met een ‘goed’ beoordeeld. De begripsvaliditeit wordt gewaarborgd door de hoge correlaties tussen de vaardigheden van leerlingen en de verschillende toetsen. De inhoudsvaliditeit van de toetsen wordt gegarandeerd door de wijze waarop de opgaven ontwikkeld zijn. Hierbij is oog geweest voor landelijke richtlijnen van de SLO en verschillende gangbare rekenmethodieken.

### ***Begrijpend lezen***

Voor het analyseren van de vaardigheidsscores en de didactische groei voor begrijpend lezen wordt gebruikgemaakt van de Cito LOVS-toetsen Begrijpend Lezen (Feenstra et al., 2010). Deze toetsen of leerlingen geschreven teksten begrijpen, hierna ook de leesvaardigheid genoemd. Ook de leesvaardigheid is een afhankelijke variabele op interval niveau. De schaal loopt van -87 bij M3 tot 101 bij M5 (N. Frans, persoonlijke communicatie, 16 mei 2024). Een hogere leesvaardigheid betekent een betere beheersing van het leesbegrip. De opbouw van deze toetsen is te vinden in de wetenschappelijke verantwoording (Feenstra et al., 2010). In veel aspecten lijkt dit instrument op het hierboven beschreven instrument voor rekenen-wiskunde,

omdat de meetwijze en de wijze van scoring en verwerking in grote mate overeenkomen. De verschillen worden hieronder besproken.

In tegenstelling tot de rekentoetsen heeft dit toetspakket geen M3-toets en geen E5-toets. Halverwege groep 3 is de technische leesvaardigheid nog niet voldoende om na te gaan of leerlingen ook met begrip kunnen lezen. Vanaf groep 5 wordt verondersteld dat één toetsmoment per jaar genoeg informatie geeft over de ontwikkeling. De toetsmodules bevatten elk 25 meerkeuze opgaven. Alle leerlingen beginnen met dezelfde startmodule. Hierna gaan de zwakke lezers door met module 1 en de sterke lezers met module 2. De werking van het meetmodel is hetzelfde als bij de Cito LOVS Rekenen-Wiskunde. Daarom kan er voor begrijpend lezen wederom worden gekeken naar de didactische groei als afhankelijke variabele en kunnen toetsscores met elkaar worden vergeleken.

De COTAN (2015b) heeft ook de betrouwbaarheid en begripsvaliditeit van deze toetsen beoordeeld. De geschatte betrouwbaarheidscoëfficiënt van elke toets lag tussen de .84 en .93, waardoor deze als hoog tot zeer hoog beoordeeld zijn. De toetsen LOVS Begrijpend Lezen sluiten nauw aan bij het doel en de inhoud van het leesonderwijs op de basisschool, waardoor deze ook als valide worden beschouwd.

### ***Categorische variabelen***

De variabelen geslacht, culturele achtergrond en opleiding ouders zijn ook aangeleverd door de scholen. Deze gegevens hadden zij al in bezit. De aangeleverde opleiding van ouders bestond uit de categorieën ‘minstens vmbo-tl/gl’, ‘laag’ en ‘zeer laag’, maar de laatstgenoemde zijn samengevoegd tot ‘maximaal vmbo-kb’, omdat de originele groepen te klein waren voor de analyses.

### **Procedure**

De geselecteerde scholen zijn in 2014 via de mail en/of telefonisch benaderd om mee te doen aan een promotietraject naar de stabiliteit en voorspellende waarde van de Cito LOVS Kleutertoetsen. Van de 1116 benaderde scholen reageerden er 84 positief. Meerdere scholen die deelname aan het onderzoek afwezen, gaven aan niet genoeg tijd te hebben. Dezelfde reden gaven de 25 scholen die de opgevraagde data niet tijdig konden leveren. De 59 scholen in de uiteindelijke steekproef hebben bij het participeren toestemming gegeven voor het promotieonderzoek en verwante onderzoeken. Na de selectie hebben zij, in samenwerking met een PhD-student, de gewenste data uit het gebruikte leerlingvolgsysteem verzameld. Dit betrof toetsgegevens van 2011 tot en met 2014 van leerlingen die op het moment van verzamelen in groep 6 waren gestart. Ook is van deze leerlingen aanvullende informatie verkregen, om te

controleren of deze steekproef representatief is voor de doelpopulatie. Deze data zijn geanonimiseerd door namen van leerlingen en scholen te verwijderen en geboortedata af te ronden op maanden. Omdat de data uit een reeds bestaande database komen, heeft het onderzoek geen invloed gehad op het onderwijs van leerlingen. De ethische commissie PedOn van de Rijksuniversiteit Groningen heeft toestemming gegeven voor het promotieonderzoek en verwante onderzoeken.

### **Analyseplan**

De data is geanalyseerd in de analysesoftware R, versie 4.2.0 (R Core Team, 2022). Er bleken enkele databewerkingen noodzakelijk. Zo is de didactische groei tussen twee toetsmomenten afgeleid door het verschil te berekenen tussen twee achtereenvolgende vaardigheidsscores, de categorieën van de variabele ‘opleiding ouders’ zijn samengevoegd en de doubluregegevens zijn gecontroleerd aan de hand van de toetsgegevens. Tot slot is een afgeleide van de variabele doublure toegevoegd aan de dataset.

Om de verdeling van belangrijke kenmerken van de steekproef te omschrijven, is gebruikgemaakt van frequentietabellen voor de categorische variabelen ‘geslacht’, ‘opleiding ouders’ en ‘culturele achtergrond’. Bij de analyse hiervan is gebruikgemaakt van een kruistabel en de chikwadraattoets. De continue variabelen zijn in beeld gebracht met boxplots per toetsmoment. Vervolgens zijn de bivariate relaties voor doublure en didactische gegevens berekend met de onafhankelijke  $t$ -toets. Hiervoor is als vuistregel voor de significantie een alfa van .05 gehanteerd.

Het verband tussen de kans (logodds) op doublure en de onafhankelijke variabelen is geanalyseerd aan de hand van een random intercept model (Snijder & Bosker, 1999). Dit model is gebruikt om bij de geclusterde data onderscheid te maken in variantie op schoolniveau en leerlingniveau. Er wordt gestart met een leeg model, zonder verklarende variabelen om te kijken naar de splitsing van variantie op leerling- en schoolniveau. Aan dit model worden de vaardigheidsscores als voorspellers toegevoegd. Vervolgens wordt ook de didactische groei toegevoegd en tot slot worden de leerlingkenmerken ‘geslacht’, ‘opleiding ouders’ en ‘culturele achtergrond’ toegevoegd. Met de deviance test wordt berekend of het model door het toevoegen van meer variabelen ook daadwerkelijk beter voorspelt. Het model wordt gefit op de dataset bestaande uit leerlingen die geen missende gegevens hebben op een van de gebruikte variabelen. Voordat deze leerlingen eruit worden gehaald, worden de modellen ook gerund om te controleren voor systematische uitval. Omdat de standaardfouten hierbij

maximaal .002 veranderen, lijkt hiervan geen sprake te zijn. De regressiecoëfficiënten worden tot slot getransformeerd naar odds, omdat deze beter te interpreteren zijn.

## **Resultaten**

### **Problemen in de dataset**

Tijdens het in kaart brengen van de steekproef is kritisch gekeken naar de gegevens in de dataset. Zo bleek dat het gegeven of leerlingen wel of niet waren gedoubleerd in elf gevallen niet overeenkwam met de toetsgegevens. Zes van deze leerlingen hadden van elk toetsmoment slechts één resultaat waarbij werd getoetst op reguliere intervallen van een half jaar, dus van deze leerlingen is de variabele ‘doublure’ aangepast van ‘wel gedoubleerd’ naar ‘niet gedoubleerd’. Vijf andere leerlingen hadden op een bepaald toetsmoment wel meerdere resultaten (bijvoorbeeld twee achtereenvolgende jaren M4 en E4) waarna weer getoetst werd op reguliere intervallen van een half jaar. Van deze leerlingen is de variabele ‘doublure’ aangepast van ‘niet gedoubleerd’ naar ‘wel gedoubleerd’. Dit betreft 0.8% van het uiteindelijke aantal leerlingen in de dataset. Het matchen van toetsgegevens van begrijpend lezen en rekenen-wiskunde op dezelfde periode bleek lastig voor de gegevens van één school en vijf losse leerlingen. Deze leerlingen ( $n=37$ ) zijn omwille van tijd uit de dataset gehaald.

### **Verdeling van de steekproef en bivariate relaties**

De steekproef bestaat na de aanpassingen uit 1355 leerlingen. Hiervan zijn 101 leerlingen (7.5%) wel gedoubleerd en 1254 leerlingen (92.5%) niet. De verdeling van de categorische kenmerken van deze steekproef is weergegeven in Tabel 1. Hieruit blijkt dat er ongeveer evenveel jongens als meisjes zijn opgenomen in de steekproef. Daarnaast blijkt dat het merendeel van de leerlingen een Nederlandse culturele achtergrond heeft (90.6%) en de meeste leerlingen ouders hebben die minimaal een vmbo-tl/gl opleiding hebben afgerond (87.5%).

Voor elke categorische variabele is gekeken naar de relatie met een eventuele doublure. In Tabel 1 zijn ook de resultaten hiervan samengevat. Er is enkel een significant verband tussen doubleren en de opleiding van de ouders ( $\chi^2(1) = 6.342, p=.012$ ). Hieruit blijkt dat leerlingen van laag opgeleide ouders significant vaker doubleren (16.8%) dan andere leerlingen (8.7%). Bij de variabelen ‘geslacht’ en ‘culturele achtergrond’ is er geen significant verschil voor het wel of niet doubleren.

**Tabel 1***Verdeling categorische variabelen en bijbehorende bivariate relaties*

Variabele		Aantal (kolompercentage)			Chi-kwadraat	
		Geen doublure	Wel doublure	Totaal	$\chi^2$	$p$
Geslacht	Jongen	621 (49.5)	49 (48.5)	670 (49.4)	0.008	.927
	Meisje	633 (50.5)	52 (51.5)	685 (50.6)		
	Missing			0		
Culturele achtergrond	Nederlands	1133 (90.6)	94 (93.1)	1227 (90.6)	0.402	.526
	Niet-Nederlands	117 (9.4)	7 (6.9)	124 (9.2)		
	Missing			4 (0.3)		
Opleiding ouders	Min. vmbo-tl/gl	1101 (91.3)	84 (83.2)	1185 (87.5)	6.342	.012
	Max. vmbo-kb	105 (8.7)	17 (16.8)	122 (9.0)		
	Missing			43 (3.2)		

Hoe de vaardigheidsscores in de steekproef zich verhouden ten opzichte van de normscores van Cito is weergegeven in Tabel 2. Hierin is zichtbaar dat de gemiddelde rekenvaardigheid in de steekproef op alle toetsmomenten hoger ligt dan in de populatie (Cito, z.d.-b). Dit verschilt 3 punten bij M3 en neemt af tot 1 punt bij E5. Bij de leesvaardigheid liggen de gemiddelden bij elk toetsmoment ook 1 tot 3 punten hoger. De didactische groei ligt met uitzondering van E4 bij rekenen-wiskunde en M5 bij begrijpend lezen op elk toetsmoment onder het landelijk gemiddelde.

**Tabel 2***Gemiddelde reken- en leesvaardigheid*

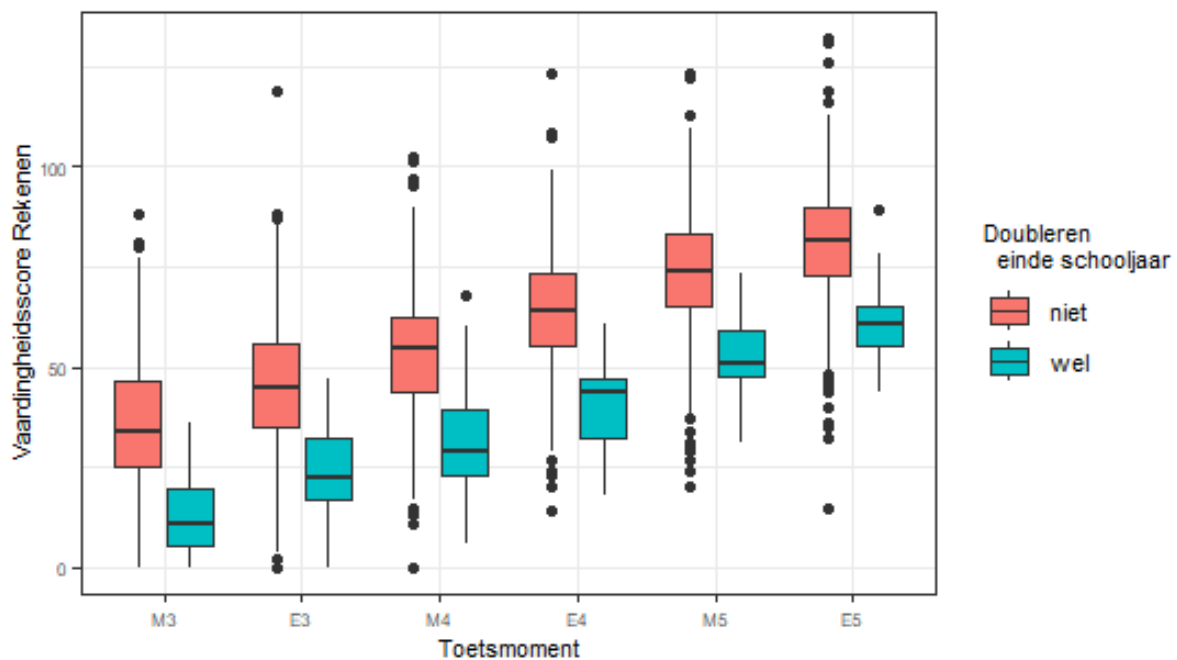
	Rekenen-wiskunde				Begrijpend lezen			
	Vaardigheidsscore		Didactische groei		Vaardigheidsscore		Didactische groei	
	$\bar{x}$	$\mu$	$\bar{x}$	$\mu$	$\bar{x}$	$\mu$	$\bar{x}$	$\mu$
M3	35	32						
E3	45	43	10	11	1	-2		
M4	54	52	8	9	14	13	13	15
E4	64	62	11	10	20	19	5	6
M5	73	72	9	10	27	25	7	6
E5	81	80	7	8				

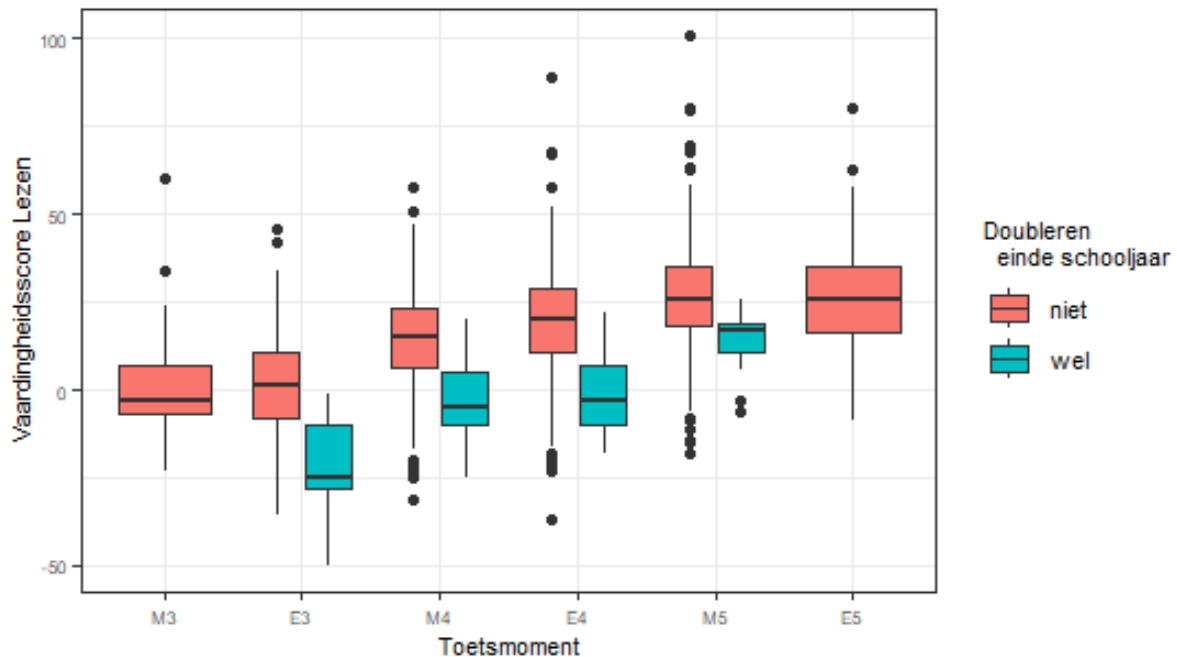
De verdeling van de vaardigheidsscores is weergegeven in boxplots per toetsmoment in Figuur 2 en 3. Voor begrijpend lezen zijn er geen toetsen voor M3 en M5, enkel een kleine groep leerlingen heeft een andere toets gemaakt op deze momenten. Voor elk opeenvolgende moment wordt de vaardigheidsscore hoger, ongeacht wel of geen doublure. De leerlingen die niet gaan doubleren behalen op elk toetsmoment gemiddeld een hogere vaardigheidsscore dan leerlingen die wel gaan doubleren. Het gaat hierbij bij elk toetsmoment om een significant verschil ( $p < .001$ ). Dat betekent dat leerlingen met een lagere vaardigheidsscore een grotere kans hebben om te doubleren.

Voor elk toetsmoment zijn er meerdere uitbijters. Deze extreme waarde komen zowel voor bovengemiddeld als ondergemiddeld ( $>1,5 \cdot \text{IQR}$ ). De extreme waarden ondergemiddeld komen vaker voor, maar de extreme waarden bovengemiddeld zijn extremer. De leerlingen die deze extreme waarden laten zien, doen dat in de meeste gevallen op meerdere toetsmomenten. Ook is er overlap in welke leerlingen dit zijn bij de verschillende vaardigheidsgebieden.

## Figuur 2

### Rekenvaardigheid

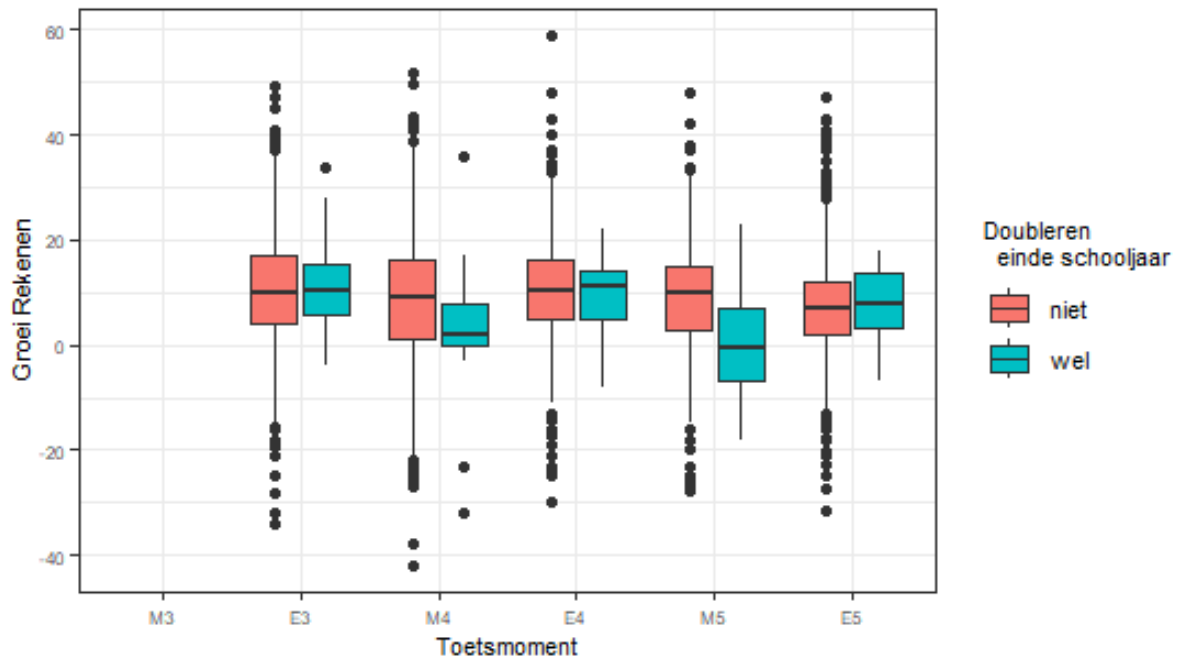
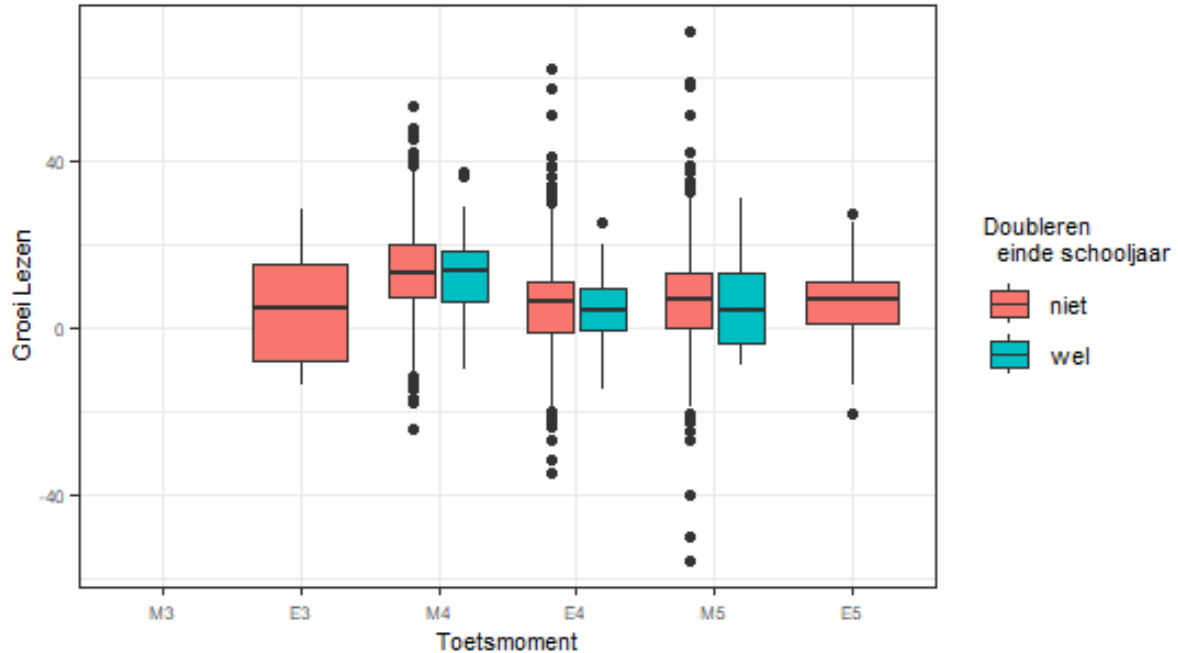


**Figuur 3***Leesvaardigheid*

De didactische groei is weergegeven in de boxplots in Figuur 4 en 5. Omdat er voor de M3 toetsen nog geen groei berekend kon worden ten opzichte van een eerdere toets, zijn er voor dat moment geen gegevens over didactische groei. Voor de meeste toetsmomenten van rekenen-wiskunde liggen de gemiddelde waarden voor groei van leerlingen met en zonder doublure binnen 1.5 punt van elkaar. Twee momenten zijn hierop een uitzondering. Bij rekenen-wiskunde lijkt de groei van leerlingen die gaan doubleren het eerste deel van het jaar (M4 en M5) lager dan de groei die deze leerlingen de tweede helft van het jaar laten zien (E3, E4 en E5). Enkel op de toetsmomenten M4 en M5 is het verschil in groei ook significant lager ten opzichte van de leerlingen die niet gaan doubleren (M4:  $p=.047$ ; M5:  $p=.003$ ). Bij M4 groeiden leerlingen met doublure gemiddeld 5.6 punten lager en bij M5 was dit 9 punten lager. Dit betekent dat leerlingen die een lagere groei laten zien op de M4 of M5 toetsen van rekenen-wiskunde een grotere kans hebben om te doubleren. Bij begrijpend lezen liggen de gemiddelde waarden voor groei voor leerlingen met en zonder doublure binnen 0.2 punten van elkaar. Het gaat bij geen enkel toetsmoment om een significant verschil.

De extreme waarden komen bij de didactische groei voornamelijk voor in de groep leerlingen die niet doubleren. Het gaat hier wederom om extreme waarden bovengemiddeld en ondergemiddeld ( $>1,5 \cdot \text{IQR}$ ). De grootte van de extreme waarden lijkt beide kanten op ongeveer gelijk te zijn. De leerlingen die extreem veel groeien ( $>3 \cdot \text{IQR}$ ) komen slechts van twee verschillende scholen.



**Figuur 4***Didactische groei rekenen-wiskunde***Figuur 5***Didactische groei begrijpend lezen***Random intercept model**

In Tabel 3 worden de uitkomsten van verschillende random intercept modellen weergegeven. In model 0 is enkel gekeken naar de afhankelijke variabele 'leerlingen geclusterd binnen scholen' en 'leerlingen' zonder predictoren. Er zijn hierbij geen systematische

verschillen in de kans op doublure op verschillende scholen, de variatie wordt geschat op 0.00. De deviance test geeft weer dat een variant van model 0, waarbij het random effect voor scholen wordt weggelaten, de kans op doublure niet beter kan voorspellen ( $p=.919$ ). De school waar een leerling op zit heeft dus weinig tot geen invloed op de kans om te doubleren. Daarom worden de schoolkenmerken buiten beschouwing gelaten en wordt een model gebruikt waarbij variantie alleen uitgesplitst wordt naar leerlingen.

In model 1 zijn de vaardigheidsscores toegevoegd. Hier is te zien dat hogere vaardigheidsscores leiden tot een afname van de kans op doublure. Wanneer beide vaardigheidsscores worden opgenomen in het model, is het effect voor rekenvaardigheid sterker en significant ( $p<.001$ ) en voor de leesvaardigheid is er geen significant effect ( $p=.328$ ). Met de odds ratio is berekend wat deze effecten inhoudelijk zeggen voor een leerling over de kans om te doubleren in dit specifieke model. Voor een leerling met een lage rekenvaardigheid ( $QI=43$ ) is de kans om te doubleren 4.8 keer groter dan voor iemand met een gemiddelde vaardigheidsscore ( $M=58.7$ ). Voor een leerling met een lage leesvaardigheid ( $QI=5$ ) is de kans om te doubleren 1.4 keer groter dan voor iemand met een gemiddelde vaardigheidsscore ( $M=16.2$ ). De deviance test toont dat in vergelijking met Model 0, Model 1 significant beter de kans op doublure voorspelt ( $\chi^2(1) = 33.92, p<.001$ ).

In model 2 is de didactische groei toegevoegd. Hierdoor neemt het effect van de rekenvaardigheid iets af, maar het effect van de leesvaardigheid neemt juist iets toe. Voor de didactische groei is het effect voor beide schoolvakken ongeveer even groot, maar tegenovergesteld. Meer didactische groei bij rekenen-wiskunde leidt volgens het model tot een afname van de kans op doublure, terwijl meer didactische groei bij begrijpend lezen leidt tot een toename van de kans op doublure. Beide hebben geen significant effect. In dit model is de kans om te doubleren voor een leerling met weinig groei voor rekenen-wiskunde ( $QI=3$ ) 1.1 keer groter dan voor een leerling met een gemiddelde didactische groei ( $M=9.1$ ). Voor een leerling met een gemiddelde groei voor begrijpend lezen ( $M=8.1$ ) is de kans om te doubleren juist 1.1 keer groter dan voor iemand met weinig groei ( $QI=3$ ). De deviance test geeft geen significant verschil tussen Model 1 en Model 2 ( $\chi^2(2) = 0.38, p<.827$ ). Dit betekent dat het toevoegen van de didactische groei de voorspelling van een mogelijke doublure niet significant verbetert.

In model 3 zijn tot slot enkele leerlingkenmerken toegevoegd. In eerste instantie betrof dit de variabelen ‘geslacht’, ‘culturele achtergrond’ en ‘opleiding ouders’. Dit leidde echter tot problemen met de modelfitting vanwege een klein aantal leerlingen met een niet-Nederlandse culturele achtergrond en doublure ( $n=7$ ), dus is ‘culturele achtergrond’ niet toegevoegd aan het

model. Er is veel samenhang tussen model 2 en 3 en het gaat enkel om kleine toenames en afnames van de effectsterkte. De richting blijft overall gelijk. In model 3 hebben vrijwel alle variabelen een negatief effect op de kans op doublure, behalve de didactische groei voor begrijpend lezen en het zijn van een meisje. Dat betekent dat bij een hogere reken- en leesvaardigheid, een hogere groei op rekenen-wiskunde of bij ouders met een opleidingsniveau van minimaal vmbo-tl/gl de odds voor doublure lager worden. Enkel de rekenvaardigheid is in dit model significant ( $p=.012$ ). De deviance test toont ook geen significante verbetering tussen Model 1 en Model 3 ( $\chi^2(4) = 0.56, p<.968$ ) of Model 2 en Model 3 ( $\chi^2(2) = 0.17, p<.916$ ). Dat betekent dat Model 1, met daarin enkel de vaardigheidsscores als voorspellers, het beste gebruikt kan worden als model om doublure te voorspellen.

Omdat het effect van de didactische groei op begrijpend lezen opvallend is en niet passend bij de verwachting, is deze variabele nogmaals geanalyseerd in een model waarbij geen andere variabelen werden meegenomen. Hier komt een coëfficiënt uit van 0.015 ( $SE = 0.022$ ). De richting blijft dan hetzelfde en de sterkte komt vrijwel overeen met Model 1. De andere variabelen hebben hier dus geen invloed op.

### Tabel 3

#### Modelbouw

	Model 0		Model 1		Model 2		Model 3	
Fixed	Coëfficiënt (SE)		Coëfficiënt (SE)		Coëfficiënt (SE)		Coëfficiënt (SE)	
Intercept	-10.777	(0.791)	-4.992	(1.220)	-5.363	(1.522)	-5.468	(1.791)
VS RW			-0.100	(0.028)*	-0.090	(0.035)*	-0.088	(0.035)*
VS BL			-0.031	(0.032)	-0.043	(0.040)	-0.046	(0.041)
Groei RW					-0.018	(0.033)	-0.018	(0.033)
Groei BL					0.012	(0.031)	0.013	(0.032)
Geslacht (V)							0.234	(1.112)
Opleiding (laag)							-0.528	(1.613)
Random	sd		sd		sd		sd	
School	0.00							
Leerling	10.37		9.578		9.651		9.511	

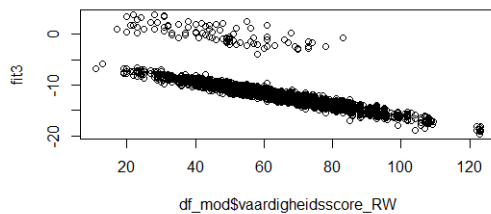
Note: \*Coëfficiënt is groter dan twee keer de standaardafwijking, wijst op een significant effect

## Assumpties

In Figuur 6, 7, 8 en 9 is te zien dat er geen duidelijke curvelinaire relaties zijn tussen de voorspelde waarden (y-as) en de voorspellende variabelen (x-as) die zijn gebruikt in Model 3. Daarmee wordt aan de assumptie van lineariteit voldaan.

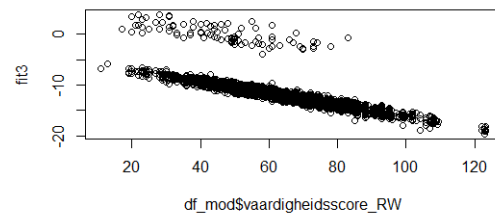
**Figuur 6**

*Vaardigheidsscore RW*



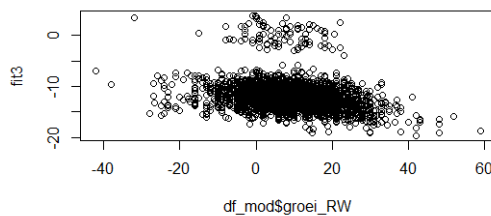
**Figuur 7**

*Vaardigheidsscore BL*



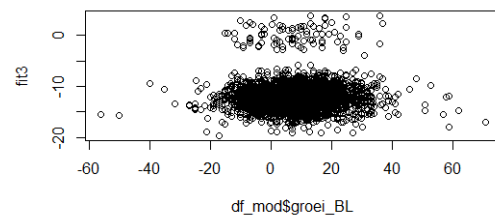
**Figuur 8**

*Groei RW*



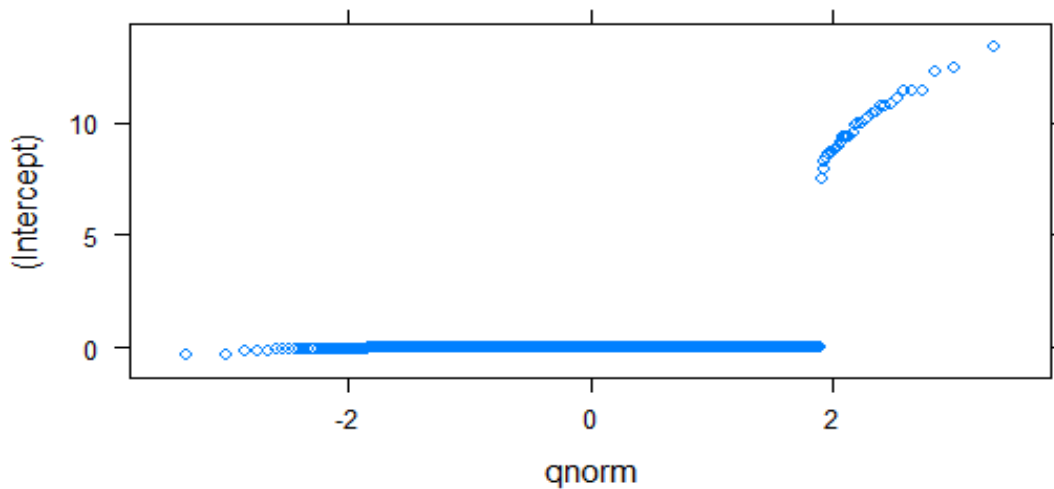
**Figuur 9**

*Groei BL*



In Figuur 10 is er een afwijkende groep te zien in de normaliteit van de random effecten. Deze groep laat extreem afwijkende effecten zien en bestaat uit 33 leerlingen, van 21 verschillende scholen, waarbij er maximaal 5 leerlingen van dezelfde school komen. Al deze leerlingen zijn gedoubleerd. De groep bestaat uit 12 jongens en 21 meisjes. De meeste leerlingen hebben ouders die minimaal vmbo-tl/gl hebben behaald (n=29) en alle leerlingen hebben een Nederlands culturele achtergrond (n=33).

Om te controleren welke effect deze afwijkende groep heeft op de modellen in Tabel 3, werd Model 3b gerund. De data bestond wederom uit alle leerlingen zonder missende gegevens en de extreme groep uit Figuur 10 werd ook buiten beschouwing gelaten. Er bleken daarna echter geen leerlingen met doublure over te blijven in de dataset, waardoor het runnen van het model niet mogelijk was.

**Figuur 10***Normaliteit random effecten*

### Conclusie

Het doel van dit onderzoek was het in kaart brengen welke factoren er mogelijk bijdragen aan het voorspellen van de kans op doublure tussen groep 3 en groep 5. Uit de resultaten blijkt dat 7.5% van de leerlingen tussen groep 3 en groep 5 zijn gedoubleerd. Leerlingen van ouders met een opleidingsniveau van maximaal vmbo-kb, leerlingen met lagere vaardigheidsscores of leerlingen met weinig didactische groei voor rekenen-wiskunde op toetsmoment M4 of M5 hebben een significant grotere kans om te doubleren. Als er wordt gecorrigeerd voor vaardigheidsscores in het random intercept model, lijkt geen van de andere variabele nog een significante voorspeller voor de kans op doublure. In het model is er geen variatie tussen scholen zichtbaar, dus lijkt de school waar de leerling heen gaat geen systematisch verschil te maken voor de kans op doublure.

De didactische groei lijkt vergelijkbaar tussen de leerlingen die wel en niet gaan doubleren. Mogelijk is het startniveau van een leerling bepalend en kan worden gesteld dat leerlingen die doubleren vanaf het eerste meetmoment een lager niveau hebben, maar vervolgens wel een vergelijkbare groei doormaken. In de steekproef werd er bij rekenen-wiskunde een verschil zichtbaar tussen de groei die leerlingen die wel gaan doubleren maken in de eerste helft van het jaar (de M-toetsen) en de groei die zij het tweede helft van het jaar maken (de E-toetsen). Dit kan mogelijk worden verklaard doordat leerlingen die minder groei laten zien bij de M-toets extra ondersteuning krijgen, waardoor de groei bij de E-toets weer hoger is. In het model zijn de variabelen als geheel meegenomen en werd er, vanwege tijdsrestricties, niet gekeken naar interactie-effecten met specifieke perioden. Dit zou nog invloed kunnen hebben op de resultaten en de uitkomst hiervan verder kunnen specificeren.

De resultaten komen gedeeltelijk overeen met de verwachtingen vanuit de literatuur. Volgens Fossen (2015) draagt een zwak prestatieniveau van Nederlandse leerlingen bij aan een grotere kans op doublure. Dit komt overeen met de *t*-toetsen voor zowel rekenen-wiskunde als begrijpend lezen en met de modellen voor alleen rekenen-wiskunde. In onderzoek van Driesen en Langen (2010) leek geslacht bij Nederlandse leerlingen geen voorspellende functie te hebben voor doublure. Dit komt overeen met de resultaten uit de steekproef. In onderzoek van Kloosterman en De Graaf (2010) bleek een lage SES een significante voorspeller voor de kans op doublure. Zowel de culturele achtergrond als de opleiding van ouders zijn bepalend voor de SES van een gezin. In deze steekproef wijst de bivariate analyse van de Nederlandse culturele achtergrond op een kleinere kans om te doubleren, maar dit is geen significante voorspeller. Het opleidingsniveau van ouders lijkt significant, maar zodra er wordt gecorrigeerd voor de vaardigheidsscore van de leerling, maakt een lage opleiding van ouders de kans op doublure iets groter, maar is dit niet significant. In het onderzoek van Kloosterman en De Graaf (2010) wordt niet gecorrigeerd voor didactisch niveau en er wordt voor de SES naar meer variabelen gekeken dan alleen de opleiding van ouders en de culturele achtergrond. Dit verklaart mogelijk een deel van de verschillende resultaten.

In de inleiding kwam naar voren dat, in tegenstelling tot het internationale beleid, de regelgeving rondom doubleren in Nederland niet centraal is vastgelegd. Scholen zijn vrij om een eigen beleid op te stellen (OCO, 2022). Bonvin (2003) stelt dat bij subjectieve leerkrachtbeoordelingen de persoonlijke overtuiging en visie op doubleren tot verschillende uitkomsten kan leiden. Het Centraal Planbureau (2015) baseert op basis van gegevens van de onderwijsinspectie dat het aantal doublures tussen scholen erg verschilt. Toch bleek uit de steekproef dat er weinig variatie is tussen de kans op doubleren op verschillende scholen en lijken er ondanks het zelf opstellen van het beleid geen systematische verschillen voor deze kans te zijn.

Goos et al. (2013) stellen in eerder onderzoek dat doubleren op lange termijn weinig effectief is. Daarom is het belangrijk de leerlingen die een grotere kans hebben op doublure goed in beeld te hebben, zodat er vroegtijdig en adequaat kan worden ingezet op extra ondersteuning. Daarnaast is het belangrijk om te bedenken wat het doel is van doubleren als interventie en wie daar de meeste voordelen van ervaart (Hattie, 2013). Heeft de leerling er op lange termijn ook daadwerkelijk baat bij of wordt doublure als interventie ingezet omdat leerkrachten keuzes maken op basis van verkeerde informatie? Wanneer er desondanks wordt overwogen een leerling te laten doubleren, zijn er meer variabelen van belang dan in dit

onderzoek aan bod zijn gekomen, denk onder andere aan het intelligentiequotiënt, de sociaal-emotionele ontwikkeling of andere interventiemogelijkheden.

### **Discussie**

De steekproef van dit onderzoek komt gedeeltelijk overeen met de populatie. De verdeling van geslacht is ongeveer gelijk. In de steekproef is 49.4% een jongen en in de populatie is dit 50.5% van alle basisschoolleerlingen in 2014 (CBS Statline, 2022). Het percentage leerlingen met een niet-Nederlandse culturele achtergrond ligt in de steekproef lager dan in de populatie: in de steekproef heeft 9.2% van de leerlingen een niet-Nederlandse culturele achtergrond, terwijl dit in de populatie ongeveer 24% van alle basisschoolleerlingen in 2014 betreft (CBS Statline, 2022). Een mogelijke verklaring hiervoor is dat scholen uit het noorden van Nederland oververtegenwoordigd zijn in de steekproef (Frans, 2019). Omdat de steekproef en de populatie hierbij niet volledig overeenkomen, heeft deze variabele in werkelijkheid mogelijk een andere invloed op de kans op doublure dan in deze steekproef duidelijk werd. De opleiding van ouders is gebaseerd op de ‘gewichtenregeling’, omdat deze gegevens al bekend waren bij de scholen. In de literatuur wordt echter veelal gerapporteerd over de proporties binnen scholen en die zijn niet opgenomen in de dataset, waardoor een minder goede vergelijking te maken is (Frans, 2019; ESIS, z.d.).

Met betrekking tot doublure lijkt de steekproef niet geheel representatief voor de populatie. Van alle leerlingen in de steekproef heeft 7.5% een keer gedoubleerd. Volgens DUO (2022) is in de periode 2015-2021 ongeveer 11% van de leerlingen tijdens de basisschooltijd tussen groep 3 en groep 8 minimaal één keer gedoubleerd. Een verklaring hiervoor is dat de cijfers van DUO uitgaan van doublures tijdens de basisschoolperiode van groep 3 tot en met groep 8. In de steekproef is echter alleen gekeken naar doublures van groep 3 tot en met groep 5. Een deel van het verschil zou kunnen worden verklaard doordat de scholen in het noorden oververtegenwoordigd zijn en er weinig leerlingen met een niet-Nederlandse culturele achtergrond in de steekproef zijn opgenomen. Van deze leerlingen is bekend dat zij relatief vaak doubleren.

De gemiddelde rekenvaardigheid in de steekproef is op alle toetsmomenten hoger dan de normscores (Cito, z.d.-b). In onderzoek van Keuning et al. (2014) waarin leesprestaties in de populatie worden onderzocht op basis van Cito dataretour blijkt dat deze hogere vaardigheidsscores bij begrijpend lezen in de populatie vaker voorkomen. Een andere verklaring is dat de hogere gemiddelde vaardigheidsscores verband houden met het relatief lage aantal doublures in de steekproef. Als er meer leerlingen met doublure waren meegenomen

en ervan uitgegaan wordt dat zij over het algemeen een lagere vaardigheidsscore hebben, zou dat het gemiddelde wat doen dalen. De verschillen in de gemiddelde scores zijn slechts minimaal, dus naar verwachting heeft dit geen invloed op de uitkomst van het onderzoek. De didactische groei ligt, met uitzondering van E4 bij rekenen-wiskunde en M5 bij begrijpend lezen, op elk toetsmoment onder het landelijk gemiddelde. Dit laatste is te verklaren doordat leerlingen met een hogere vaardigheidsscore gemiddeld minder hard groeien (Verhaeghe & Van Damme, 2004).

De missing data is onderzocht voor de variabelen afhankelijk van doublure. De missende gegevens waren daarbij gelijk verdeeld over de subgroepen, dus op basis van de gemeten variabelen is er geen systematische uitval en worden de resultaten nauwelijks beïnvloed. De power van het onderzoek is wellicht wel iets afgenomen door het weglaten van leerlingen met missende gegevens. Daarentegen maakt het onderzoeksdesign mogelijk dat de grootte waaruit de steekproef bestaat zorgt voor genoeg power. Door de samenwerking met scholen kon er in relatief korte tijd veel data worden verzameld van veel leerlingen over een langere periode. Een nadeel van het onderzoeksdesign is echter dat alleen gegevens van leerlingen zijn opgenomen in de dataset die op het moment van verzamelen nog op school zaten. Leerlingen die eerder zijn uitgestroomd naar een andere vorm van onderwijs of een verhuizing hebben meegemaakt, zijn daarom niet opgenomen in de dataset. Voor de verhuizende leerlingen is geen reden om aan te nemen dat zij de resultaten sterk zouden beïnvloeden, want er zijn geen aanwijzingen dat al deze leerlingen systematisch dezelfde gegevens hebben. Het weglaten van deze leerlingen zal naar verwachting daarom geen specifieke invloed hebben op de verdeling van de steekproef. De leerlingen die uitstromen naar een andere vorm van onderwijs hebben naar verwachting wel invloed. Er is een aannemelijke kans dat het hierbij gaat om leerlingen die wel heel hoog of laag scoren, want daarom is een andere onderwijssetting gewenst. Door de gegevens van deze leerlingen niet op te nemen in de steekproef, lijkt de spreiding van toetsgegevens in de steekproef kleiner dan deze in werkelijkheid mogelijk is. Als deze leerlingen wel waren opgenomen in de steekproef en daadwerkelijk veel hogere of lagere vaardigheidsscores zouden hebben, komen er meer uitbijters bij.

Om de constructvaliditeit te waarborgen is er in dit onderzoek gebruikgemaakt van gevalideerde instrumenten. De Cito LOVS-toetsen hebben een hoge betrouwbaarheid en validiteit volgens de beoordeling van de COTAN (Egberink, 2015a; Egberink, 2015b). Door het bijbehorende IRT-model was het ook mogelijk de didactische groei als continue variabele te beoordelen. Dit verhoogt de constructvaliditeit van het onderzoek, met name voor scholen



die ook met Cito werken. Dit betreft ongeveer 81% van de scholen (Geurtsen, 2023). Bij andere veelgebruikte methodes zijn de toetsscores niet te vergelijken met de vaardigheidsscores van Cito (Huijsmans, 2024). Ook internationaal zullen de resultaten van dit onderzoek met voorzichtigheid moeten worden geïnterpreteerd. Door het variërende internationale beleid rondom doublure, zullen de resultaten uit deze steekproef, gericht op het Nederlandse stelsel, minder waardevol zijn.

Een minder positief aspect van de constructvaliditeit zijn de demografische gegevens die zijn aangeleverd door de school. De correctheid hiervan kon bij veel van de variabelen niet worden gecontroleerd. Het is aannemelijk dat de variabele ‘geslacht’ van leerlingen bij de scholen goed in beeld is. Kenmerken die verder van de leerling af staan en meer met de gezinssituatie te maken hebben, zijn mogelijk minder goed in beeld. Mogelijk is er sprake van verschillende interpretaties tussen scholen. De gegevens over de variabele ‘doublure’ zijn op dezelfde manier aangeleverd, maar met behulp van de toetsgegevens en leeftijd ten tijde van de toets, kon deze variabele wel worden gecontroleerd. Daarbij werd zichtbaar dat er toch enkele fouten in de gegevens zaten. Daarom bestaat de kans dat er bij de andere variabelen ook enkele foutieve antwoorden in de dataset zijn opgenomen. Bij doublure ging dit slechts om 0.8% van het uiteindelijke aantal leerlingen in de dataset. Indien er bij andere variabelen ook sprake is van verkeerde gegevens, zal dit ook gaan om een klein aantal gegevens. Daarom is de verwachting dat dit geen grote invloed heeft op de resultaten.

De externe validiteit van dit onderzoek ligt lager, doordat de steekproef en populatie niet volledig hetzelfde waren. In eventueel vervolgonderzoek is het daarom belangrijk dat er wordt gelet op een voldoende grote steekproef uit de populatie leerlingen die doubleren. Een meer aselechte steekproef zou hierbij helpend kunnen zijn. Ook andere schoolse vaardigheden kunnen in een eventueel vervolgonderzoek worden meegenomen, zoals sociaal-emotionele ontwikkeling of executieve functies. Een vervolgonderzoek naar keuzes rondom doubleren van leerkrachten kan meer inzicht geven over de invloed en het gevolg van leerkrachtvisies. Voor beleidsmakers is het eventueel waardevol om meer te verdiepen in het internationale beleid van landen waar doubleren niet nodig is vanwege de goede mate van differentiatie binnen het onderwijs om zo in Nederland ook te komen tot minder doublures.

Dit onderzoek biedt inzichten in de factoren die mogelijk wel of juist niet bijdragen aan de kans op doublure tussen groep 3 en 5 van het basisonderwijs. Ondanks dat het beleid rondom doubleren in Nederland niet landelijk is vastgelegd, lijken er in dit onderzoek geen systematische verschillen te zijn tussen de kans op doubleren op scholen. Met name de rekenvaardigheid lijkt een belangrijke voorspeller voor de kans op doublure. Daarom is het

belangrijk leerlingen met een laag vaardigheidsniveau adequaat te ondersteunen met als doel een regulier schoolverloop. Doublure als interventie blijkt namelijk niet effectief en in onderzoeken wordt meermaals gerapporteerd over nadelige effecten op lange termijn. Het is positief dat leerkrachten niet bevooroordeeld zijn bij beslissingen rondom doublure en deze baseren op schoolse vaardigheden in plaats van op kindkenmerken zoals geslacht of culturele achtergrond. Toch blijft het de vraag of een lage vaardigheid altijd moet leiden tot doublure of dat er betere alternatieve benaderingen zijn om het onderwijs aan te passen aan de behoeften van het kind. Ongeacht welke factoren wel of niet een rol spelen: het is essentieel dat bij elke beslissing het welzijn centraal staat en dat leerkrachten handelen in het belang van het kind.

### Literatuurlijst

- Bonvin, P. (2003). The role of Teacher Attitudes and Judgement in Decision-Making: The case of grade retention. *European Educational Research Journal*, 2(2), 277–294. <https://doi.org/10.2304/eerj.2003.2.2.6>
- Cabus, S., & Ariës, R. (2016). What do parents teach their children? – The effects of parental involvement on student performance in Dutch compulsory education. *Educational Review*, 69(3), 285–302. <https://doi.org/10.1080/00131911.2016.1208148>
- CBS Statline. (2022). *Leerlingen (speciaal) basisonderwijs*. <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83295NED/table?dl=8C08>
- Centraal Planbureau. (2015). *CPB Policy Brief: Zittenblijven in het Nederlandse primair en voortgezet onderwijs*. <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-policy-brief-2015-01-zittenblijven-het-primair-en-voortgezet-onderwijs.pdf>
- Cito. (z.d.-a). *Leerling in beeld - LVS*. <https://cito.nl/onderwijs/primair-onderwijs/leerling-in-beeld-lvs/>
- Cito. (z.d.-b). *Tabellen tussenopbrengsten 2.0 CITO LOVS*. CITO <https://www.bazaltgroep.nl/wp-content/uploads/sites/2/2021/02/2.0-Tabellen-tussenopbrengsten-Cito-LOVS-leerlingniveau.pdf>
- Driessen, G., & Langen, A. van (2010). *De onderwijsachterstand van jongens. Omvang, oorzaken en interventies*. ITS. [https://www.researchgate.net/publication/235963374\\_De\\_onderwijsachterstand\\_van\\_jongens\\_Omvang\\_oorzaken\\_en\\_interventies](https://www.researchgate.net/publication/235963374_De_onderwijsachterstand_van_jongens_Omvang_oorzaken_en_interventies)
- DUO. (2022). *Leerlingen basisonderwijs - zittenblijven 2016-2022* [Dataset]. [https://duo.nl/open\\_onderwijsdata/primair-onderwijs/aantal-leerlingen/zittenblijvers.jsp](https://duo.nl/open_onderwijsdata/primair-onderwijs/aantal-leerlingen/zittenblijvers.jsp)
- Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils' heterogeneity? *Comparative Education Review*, 52, 245-273.
- Egberink, I.J.L. & Leng, W.E. de. (2015a). COTAN beoordeling 2015, Rekenen-wiskunde groep 3 t/m 8 (papieren versie). [www.cotandocumentatie.nl](http://www.cotandocumentatie.nl)
- Egberink, I.J.L. & Leng, W.E. de. (2015b). COTAN beoordeling 2015, Begrijpend lezen groep 3 t/m 6 (papieren versie). [www.cotandocumentatie.nl](http://www.cotandocumentatie.nl)
- ESIS. (z.d.). Gewichtenregeling. <https://support.rovict.nl/hc/nl/articles/360002241958-Gewichtenregeling>
- Eurydice (2011). *Grade retention during compulsory education in Europe: Regulations and statistics*. European Commission.

<https://eric.ed.gov/?redir=http%3a%2f%2fdx.publications.europa.eu%2f10.2797%2f50570>

- Frans, N. (2019). *A captivating snapshot of standardized testing in early childhood: on the stability and utility of the Cito preschool/kindergarten tests*. [Proefschrift]. Rijksuniversiteit Groningen. <https://doi.org/10.33612/diss.95431744>
- Frans, N., Post, W. S., Oenema-Mostert, C., & Minnaert, A. (2020). Signalering met de Cito kleutertoetsen: Ondergemiddeld is niet gelijk aan problematisch. *Tijdschrift voor Orthopedagogiek*, 59(2), 20–27. <https://doi.org/10.1080/1743727X.2020.1736024>
- Fossen, M. W. E. B. (2005). *Invloeden op de onderwijspositie van leerlingen bij de overgang van het basisonderwijs naar het voortgezet onderwijs*. [Proefschrift]. SCO Kohnstamm Instituut. <https://hdl.handle.net/11245/1.269880>
- Fraine, B. De, Goos, M., Van Landeghem, G., Juchtmans, G., & Vandenbroucke, A. (2012). Zittenblijven: Praktijken en overtuigingen in het Vlaams onderwijs. *Caleidoscoop*, 24(5), 25-31. <https://api.semanticscholar.org/CorpusID:157154054>
- Feenstra, H., Kamphuis, F., Kleintjes, F., & Krom, R. (2008). *Wetenschappelijke verantwoording. Begrijpend lezen voor groep 3 tot en met 6*. Cito. <https://docplayer.nl/370648-Wetenschappelijke-verantwoording-begrijpend-lezen-voor-groep-3-tot-en-met-6.html>
- Geurtsen, K. (2023). *Rapportage Onderzoek LVS en eindtoets*. DUO Onderwijsonderzoek & Advies. <https://www.duo-onderwijsonderzoek.nl/wp-content/uploads/2023/03/Rapportage-Pointer-onderzoek-toetsen-maart-2023.pdf>
- Goos, M., Belfi, B., de Fraine, B., van Damme, J., Onghena, P., & Petry, K. (2013). Effecten van zittenblijven in het basis- en secundair onderwijs in kaart gebracht: Een systematische literatuurstudie. *Pedagogische Studiën*, 90(5), 17–30. [https://www.researchgate.net/publication/241040469\\_Effecten\\_van\\_zittenblijven\\_in\\_het\\_basis-en\\_secundair\\_onderwijs\\_in\\_kaart\\_gebracht\\_Een\\_systematische\\_literatuurstudie](https://www.researchgate.net/publication/241040469_Effecten_van_zittenblijven_in_het_basis-en_secundair_onderwijs_in_kaart_gebracht_Een_systematische_literatuurstudie)
- Hattie, J. (2013). *Leren zichtbaar maken. Nederlandse vertaling van Visible Learning for Teachers (2012)*. Bazalt Educatieve Uitgeverijen.
- Hong, G., & Raudenbush, S. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, 27(3), 205-224. <https://www.jstor.org/stable/3699569>

- Huijsmans, M. (2024). *Een nieuw Leerlingvolgsysteem: de huidige generatie LVS naast elkaar*. Edux. [https://edux.nl/\\_asset/\\_public/Uit-onze-pen/Overzicht-leerlingvolgsystemen.pdf](https://edux.nl/_asset/_public/Uit-onze-pen/Overzicht-leerlingvolgsystemen.pdf)
- Janssen, J., Verhelst, N., Engelen, R. & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8*. Cito.
- Jepma, J., & Meijnen, G.W. (2001). Risicoleerlingen in het regulier basisonderwijs: handhaving of verwijzing? *Pedagogische Studiën*, 78(5), 313-329.  
<https://pedagogischestudien.nl/article/view/14794>
- Jeronimus, B. F., Stavrakakis, N., Veenstra, R., & Oldehinkel, A. J. (2015). Relative Age Effects in Dutch Adolescents: Concurrent and Prospective Analyses. *PLoS ONE*, 10(6). <https://doi.org/10.1371/journal.pone.0128856>
- Jimerson, S. (2001). Meta-analysis of grade retention research: Implications for practice in The 21st century. *School Psychology Review*, 30, 420–437.  
<https://doi.org/10.1080/02796015.2001.12086124>
- Keuning, J., Hilde, M., & Weekers, A. (2014). Begrijpend leesprestaties onderzocht – een analyse op basis van Cito Dataretou. *Tijdschrift Voor Orthopedagogiek*, 53, 2–13.  
[https://mail.wij-leren.nl/userfiles/files/cito/cito\\_begrijpend\\_lezen\\_onderzocht-analyse-dataretour.pdf](https://mail.wij-leren.nl/userfiles/files/cito/cito_begrijpend_lezen_onderzocht-analyse-dataretour.pdf)
- Kievit, T., Tak, J., & Bosch, J. D. (2008). *Handboek psychodiagnostiek voor de hulpverlening aan kinderen* (7e, herz. druk). De Tijdstroom.
- Kloosterman, R., & de Graaf, P. M. (2010). Non-promotion or enrolment in a lower track? The influence of social background on choices in secondary education for three cohorts of Dutch pupils. *Oxford Review of Education*, 36(3), 363–384.  
<http://www.jstor.org/stable/25699590>
- Ministerie van Algemene Zaken. (2023a). *Hoe legt de basisschool de ontwikkeling van mijn kind vast?* Rijksoverheid.  
<https://www.rijksoverheid.nl/onderwerpen/basisonderwijs/vraag-en-antwoord/hoe-legt-de-basisschool-de-prestaties-van-mijn-kind-vast#:~:text=asisscholen%20werken%20verplicht%20met%20een,kind%20en%20van%20de%20school>
- Ministerie van Algemene Zaken. (2023b). *Eindtoets basisschool wordt doorstroomtoets*. *Schooladvies en doorstroomtoets basisschool*. Rijksoverheid.

<https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-doorstroomtoets-basisschool/wetsvoorstel-doorstroomtoetsen-po>

OCO (2022). *De regels voor zittenblijven op de basisschool*. OCO.

<https://www.onderwijsconsument.nl/zittenblijven-op-de-basisschool/>

OCO (2023). *Wat betekent zittenblijven of doubleren voor uw kind?* OCO. <https://www.onderwijsconsument.nl/wat-betekent-doubleren-zittenblijven/#:~:text=In%202019%20bleef%20tien%20procent,schoolgids%20en%20verschillen%20per%20school>

OECD (2010). *Results: What makes a school successful? Resources, policies and practices*. OECD.

Onderwijsinspectie (2015). *De staat van het onderwijs*. [www.onderwijsinspectie.nl](http://www.onderwijsinspectie.nl)

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>

Snijders, T.A.B. & Bosker R.J. (1999) *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Sage Publications.

Verhaeghe, J., & Van Damme, J. (2004). *Longitudinaal onderzoek in het basisonderwijs: Methodescholen in Vlaanderen*. KU Leuven.  
<https://lirias.kuleuven.be/retrieve/346895>