

Stagnatie in de woordenschatontwikkeling

Ontwikkelingsprobleem en/of meetfout?

Student: M.H.F. Faasen (s3673731)

Begeleider: dr. N. Frans

2e beoordelaar: dr. J. Brouwer

Rijksuniversiteit Groningen

Faculteit der Gedrags- en Maatschappijwetenschappen

Bachelorwerkstuk Pedagogische Wetenschappen

Juni 2023

Abstract

The new pupil monitoring system and corresponding tests recently introduced by Cito, focus on growth of primary education students, including vocabulary growth. Lack of growth, indicates a potential development problem. It has not previously been investigated what characteristics of the learner are exactly related to growth as determined by Cito tests. The purpose of this study is to determine the extent to which stagnation in Cito vocabulary growth of students can be predicted in students in grades 1-3, based on characteristics of the learner and context. An existing dataset has been used to investigate the relationship between stagnation and other student characteristics through χ^2 -tests and *t*-tests. With a logistic regression analysis, a model has been drawn up that best predicts stagnation. The results show that stagnation occurs in 84,4% of the sample, with higher starting levels and more tests taken as significant predictors. In 91,2% of cases, this model correctly predicts if a student in the sample stagnates or not. The relationship between starting level and number of tests taken on the one hand, and stagnation on the other hand, can be explained statistically. In combination with the extraordinary high percentage of stagnating students, it is very likely that stagnation is not only observed in students with a developmental problem, but that measurement errors also play a considerable role. This study shows that Cito vocabulary proficiency scores cannot be blindly relied upon, when identifying students with developmental problems.

Inleiding

Het Centraal Instituut voor Toetsontwikkeling (Cito) heeft een leerlingvolgsysteem (LVS) ontwikkeld voor het systematisch en longitudinaal monitoren van de leerprogressie van leerlingen in het basisonderwijs. Dit maakt aanpassing van onderwijs aan de behoefte van leerlingen mogelijk, om zo bij te dragen aan hun optimale ontwikkeling (Vlug, 1997). Met het LVS en bijbehorende toetsen kan het niveau en de groei van leerlingen worden gemeten en worden afgezet tegen de gemiddelde leerling. Groei is meer geschikt als maat om ontwikkeling te voorspellen dan niveau, omdat bij stagnatie in groei het risico bestaat dat achterstand toeneemt (Frans et al., 2020). Daarom staat groei meer centraal in de nieuwe generatie toetsen (Cito, z.d.). Stagnatie in groei wordt gedefinieerd als eenzelfde of lagere vaardigheidsscore dan de eerdere toetsafname (Frans, 2019).

Woordenschat, één van de domeinen binnen het onderwijs en LVS, is van belang bij het opdoen van kennis en vaardigheden (Schrooten & Vermeer, 1994) en hangt samen met de ontwikkeling van andere schoolse vaardigheden (Duff et al., 2015; Sénéchal et al., 2006). Een gebrek aan groei in de woordenschatontwikkeling wijst op een mogelijk ontwikkelingsprobleem. Vroegtijdig ingrijpen is van belang om toename van problemen te voorkomen (Schrooten & Vermeer, 1994). Dit geldt niet alleen op het gebied van woordenschat. In 2019 is daarom onderzoek uitgevoerd naar de voorspellende waarde van Cito LVS Kleutertoetsen voor de latere reken- en taalontwikkeling. Meetfouten in de toetsresultaten lijken snel tot onterechte conclusies te leiden als risicoleerlingen worden geïdentificeerd op basis van stagnaties in groei. In groep 2-5 vertoonde meer dan 60% van de steekproef stagnatie, terwijl deze leerlingen vaak geen structureel benedengemiddelde groei vertoonden (Frans, 2019). Een structurele samenhang tussen kenmerken van leerlingen en stagnatie in vaardigheidsgroei, kan duiden op systematische ontwikkelingsproblemen, terwijl een statistisch verklaarbare samenhang eerder duidt op willekeurige meetfouten. Het is relevant om te onderzoeken in hoeverre stagnatie in vaardigheidsgroei willekeurig is, of samenhangt met specifieke leerlingkenmerken, aangezien deze informatie onderwijsprofessionals kan helpen bij het zoeken naar verklaringen voor een waargenomen stagnatie op basis van toetsscores en het voorkomen van onnodig remediëren als gevolg van meetfouten. In dit onderzoek zal specifiek op woordenschatvaardigheid worden gefocust.

Eerder onderzoek wijst op factoren die mogelijk samenhangen met groei en stagnatie in woordenschat. Een eerste factor is leeftijd. Met de leeftijd vlakt de aanvankelijk snel groeiende woordenschat af, aangezien de kans kleiner wordt om woorden tegen te komen die nog niet tot de woordenschat van de persoon behoren (Keuleers, 2015). In het onderzoek van

Verhoeven en Van Leeuwe (2008) bij basisschoolkinderen, neemt vooral tussen groep 4 en 5 de woordenschatvaardigheid sterk toe, waarna de groei afneemt. Huttenlocher et al. (1991) wijzen op de rol van niveauverschillen. Leerlingen met een grotere woordenschat, lokken mogelijk meer taalinput uit, wat hun woordenschatontwikkeling bevordert. Zij vonden tot twee jaar ook een verschil tussen jongens en meisjes in woordenschatgroei. Uit onderzoek van Bauer et al. (2002) volgt dat meisjes woordenschat sneller verwerven dan jongens, wat volgens Waber (1976) grotendeels verklaard kan worden door een verschil in neurologische ontwikkeling.

Uit onderzoek blijkt dat meer en meer diverse taalinput vanuit ouders samenhangt met een snellere groei in woordenschat van het kind (Huttenlocher et al., 1991; Jones & Rowland, 2017). Lager opgeleide moeders, met een lagere SES, zouden een kleinere woordenschat en minder gevarieerde en complexe zinsstructuren gebruiken. Basisschoolleerlingen met een migratieachtergrond hebben een aanzienlijke taalachterstand ten opzichte van autochtone leerlingen. Deze achterstand is afhankelijk van de specifieke etnische achtergrond en het opleidingsniveau van de ouders, en wordt met de leeftijd kleiner (Dagevos et al., 2003). Keuleers (2015) laat zien dat bij kinderen die meertalig opgroeien, de indirecte woordenschatverwerving via de tweede taal opweegt tegen het relatief kleine verlies in woordenschatgroei in de eerste taal als gevolg van minder blootstelling daaraan. Echter groeit niet ieder kind met een migratieachtergrond tweetalig op.

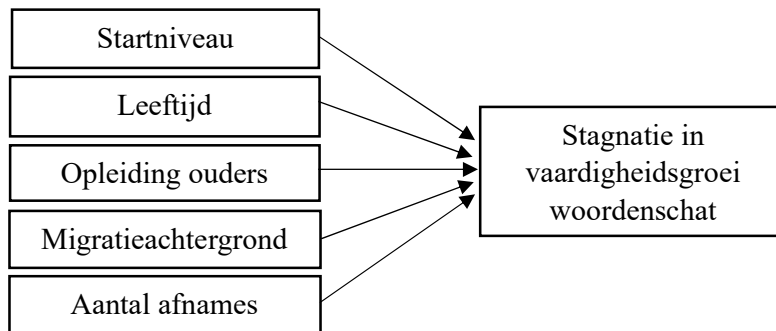
Het Cito toetsysteem wordt door meer dan 80% van de Nederlandse scholen gebruikt (Gelderblom et al., 2016) en Cito heeft recent nieuwe toetsen op de markt gebracht met de focus op groei (Cito, z.d.). Echter is niet onderzocht waar de groei in Cito vaardigheidsscores precies mee samenhangt en in welke mate. In dit onderzoek staat dan ook de vraag centraal: In hoeverre is, op basis van kenmerken van de leerling en zijn/haar context, stagnatie in Cito vaardigheidsgroei op gebied van woordenschat bij leerlingen in groep 3-5 te voorspellen?

Naar verwachting komen stagnaties systematisch vaker voor bij leerlingen met specifieke kenmerken (Figuur 1). De zojuist beschreven studies bieden aanwijzingen dat een lager startniveau, lager opgeleide ouders en een migratieachtergrond samenhangen met stagnaties in vaardigheidsgroei op gebied van woordenschat. Deze kenmerken zouden echter ook samen kunnen hangen met een aanvankelijk achterblijvende groei, die bijtrekt in de klas. Uit onderzoek van Mayo & Leseman (2008) bij kinderen van 3;10 en 4;2 jaar volgt dat klassenervaring van invloed is op de groei van woordenschat. Mogelijk kan er sprake zijn van een snellere groei binnen de leeftijdscategorie van de doelgroep van dit onderzoek, die met de leeftijd afneemt. Naar verwachting treedt stagnatie vaker op wanneer leerlingen vaker getoetst

worden, aangezien de kans op stagnatie groter is bij meer en dichter op elkaar gelegen meetmomenten. Het aantal afnames is daarom meegenomen om hiervoor te kunnen corrigeren.

Figuur 1

Conceptueel model



Methode

Onderzoeksdesign

Een retrospectief correlatieel onderzoek is uitgevoerd, waarbij gebruik is gemaakt van data verzameld voor een onderzoek naar de stabiliteit en voorspellende waarde van Cito kleutertoetsen (Frans, 2019). Data met betrekking tot de afhankelijke variabele – stagnatie van leerlingen in woordenschatontwikkeling – is in dit onderzoek in verband gebracht met de variabelen startniveau, leeftijd, opleiding ouders, migratieachtergrond en aantal afnames.

Populatie en steekproef

De doelpopulatie van dit onderzoek bestond uit kinderen in groep 3-5 van het basisonderwijs in Nederland. Reguliere basisscholen die het Cito LVS gebruiken, specifiek ook de kleutertoetsen, zijn willekeurig benaderd voor het onderzoek van Frans (2019). Van 59 deelnemende scholen zijn 1402 leerlingen die in 2014 in groep 6 zijn gestart, opgenomen in de steekproef. Voor het huidige onderzoek bestond de steekproef enkel uit leerlingen bij wie de Cito Woordenschattoets is afgenomen. Leerlingen met leerlinggebonden financiering ($n=17$) en met één toetsafname ($n=34$) zijn geëxcludeerd, respectievelijk vanwege de kleine omvang van de groep en omdat met één afname geen groei vastgesteld kon worden. Om leereffecten bij herhaaldelijk toetsen zoveel mogelijk te voorkomen, is bij leerlingen met twee toetsafnames binnen twee maanden, alleen de eerste afname behouden. De definitieve steekproef bestond uit 1063 leerlingen.

Onderzoeksvariabelen en instrumenten

Woordenschat is het geheel van labels waarover men beschikt voor taalgebruik (productief) en taalbegrip (receptief). Midden en eind van het leerjaar kunnen Cito LVS Woordenschattoetsen worden afgenomen, om het niveau en de ontwikkeling van woordenschat vast te leggen. Cito toetst enkel de receptieve woordenschat – het identificeren van betekenissen en betekenisrelaties – met als voordeel dat toetsscores eenvoudiger en objectiever bepaald kunnen worden (Berkel et al., 2010). De woordenschattoetsen van midden groep 3 (M3) tot en met eind groep 5 (E5) worden per leerling in de groep afgenomen door de leerkracht en bestaan uit 50-70 meerkeuzeopgaven. Aan de ontwikkeling van de opgaven ligt een woordfrequentielijst ten grondslag en de toetsen bevatten een voor de Nederlandse taal representatieve verdeling van woordsoorten, wat de inhoudsvaliditeit garandeert. Cito beschikt over met Item Respons Theorie (IRT)-modellen gekalibreerde opgavenbanken, op basis waarvan toetsen zijn samengesteld (Berkel et al., 2010). IRT neemt aan dat één onderliggende vaardigheid bepalend is voor de kans om een opgave goed te maken. Hoe meer vaardigheid een leerling bezit, hoe groter de kans dat de leerling de opgave goed maakt. Door opgaven te combineren in een toets, wordt de vaardigheidsschatting nauwkeuriger. De toetsen zijn op eenzelfde vaardigheidsschaal geijkt, waardoor vaardigheidsscores tussen toetsen kunnen worden vergeleken en stagnatie of groei kan worden vastgesteld. Op basis van resultaten van de woordenschattoetsen M3-E5 is per leerling bepaald hoeveel stagnaties (eenzelfde of lagere vaardigheidsscore dan het vorige meetmoment) plaatsgevonden hebben.

De globale meetnauwkeurigheid van de toetsen is geschat op een MAcc-waarde hoger dan ,80 (Berkel et al., 2010). Dit duidt op een goede betrouwbaarheid, volgens de criteria van COTAN (Evers et al., 2010, p. 33). Uit onderzoek blijken vier aanwijzingen voor begripsvaliditeit: het is aannemelijk dat de opgavenbank die ten grondslag ligt aan de toetsen de woordenschat meet, de opgaven Woordenschat correleren hoger met vaardigheden waar correlatie mee te verwachten is dan met onderdelen waarin semantische vaardigheden minder domineren, er is sprake van groei in vaardigheidsniveau zoals verwacht mag worden en de toetsen zijn voldoende responsief voor het meten van verandering. Op basis van de criteria van COTAN wordt de begripsvaliditeit als voldoende aangemerkt (Berkel et al., 2010).

In dit onderzoek stonden vijf onafhankelijke variabelen centraal die mogelijk samenhangen met woordenschatgroei: aantal toetsafnames, startniveau (vaardigheidsscore M3), leeftijd in jaren bij aanvang groep 6, migratieachtergrond (Nederlandse ouders [0] of minstens één niet-Nederlandse ouder [1]) en opleidingsniveau ouders. Leerlingen hebben gewichten toegekend gekregen op basis van de opleiding van de ouders: 0 bij minimaal één

ouder (of verzorgende) met niveau vmbo-g/tl of hoger, 0,3 als beide ouders laag opgeleid zijn (maximaal lbo/vbo, praktijkonderwijs of vmbo-b/k) en 1,2 bij één zeer laag opgeleide (maximaal basisonderwijs of (V)SO-ZMLK) en één zeer laag of laag opgeleide ouder (Werkgroep IBO, 2017). Overige variabelen die ter exploratie en beschrijving van de steekproef zijn meegenomen, zijn geboortemaand (1-12), afkomstig uit een eenoudergezin (1) of niet (0), geslacht (jongen [0] of meisje [1]), en of de leerling gedoubleerd heeft voor start groep 6 (1) of niet (0).

Procedure

Via e-mail en telefoon zijn scholen benaderd en geïnformeerd over het onderzoek van Frans (2019). Deelnemende scholen hebben vrijwillig toestemming gegeven voor het gebruik van de data, ook voor gerelateerd vervolgonderzoek. Dataverzameling vond plaats van september 2014 tot november 2015. Data is anoniem door scholen, eventueel in samenwerking met de onderzoekers, vanuit het leerlingadministratiesysteem aangeleverd. De anonimiteit van leerlingen is gewaarborgd door namen van leerlingen en scholen niet te rapporteren en leeftijd op maanden af te ronden. De ethische commissie PedOn heeft toestemming gegeven voor het onderzoek (Frans, 2019).

Analyseplan

De gegevens zijn met behulp van SPSS versie 28.0.1.0 bewerkt en geanalyseerd. Kenmerken van de steekproef en missing data zijn weergegeven in frequentietabellen en een bar plot. Patronen in missing data zijn nader geanalyseerd. De afhankelijke variabele is gedichotomiseerd in wel (1) of geen (0) stagnatie. De variabele geboortemaand is gedichotomiseerd in tot (0) en vanaf (1) oktober, aangezien leerlingen geboren in oktober, november of december zogenaamde ‘herfstkinderen’ zijn. Scholen gaan hier verschillend mee om: ofwel kleuterbouwverlenging met als gevolg dat een herfstkind ouder is dan veel klasgenoten, ofwel geen kleuterbouwverlenging met als gevolg dat een herfstkind jonger is dan veel klasgenoten (Driessen et al., 2014). Bivariate relaties tussen de afhankelijke variabele en de variabelen migratieachtergrond, opleiding ouders, eenoudergezin, geboortemaand, doubleren en geslacht zijn in kruistabellen weergegeven. Met chi-kwadraattoetsen is de samenhang tussen categorische variabelen en stagnatie onderzocht ($\alpha=,05$), waarbij de assumptie werd gecontroleerd dat iedere verwachte frequentie minimaal 5 bedraagt. Bij schending wordt uitgeweken naar Fisher’s exact test. Independent samples *t*-toetsen zijn uitgevoerd om na te gaan of wel en niet stagnerende leerlingen significant verschillen in leeftijd, startniveau en aantal afnames ($\alpha=,05$). De assumptie van gelijke variantie (homoscedasticiteit) wordt niet geschonden wanneer de standaarddeviaties van de

groep die wel en niet stagneert minder dan een factor twee verschillen. Bij schending zijn de ‘equal variances not assumed’ resultaten gerapporteerd. Normaliteit werd gecheckt met boxplots.

Een logistische regressie is uitgevoerd om na te gaan of stagnatie in scores op Cito Woordenschattoetsen voorspeld kon worden door de overige gemeten variabelen. Hierbij werd nagegaan of de voorspellers significant samenhangen met stagnatie ($\alpha=,05$) en zijn modeluitkomsten geïnterpreteerd. Bij $p <,050$ werden variabelen opgenomen in het model, dat via forward selection werd opgebouwd. Middels plots waarin de logit van stagnatie werd afgezet tegen de continue variabelen in het definitieve model, is de assumptie van lineariteit gecheckt. De VIF-statistic is berekend, waarbij $VIF \geq 5-10$ duidt op multicollineariteit (Shrestha, 2020). Ook is er gecontroleerd op invloedrijke waarden (Cook’s distance >1).

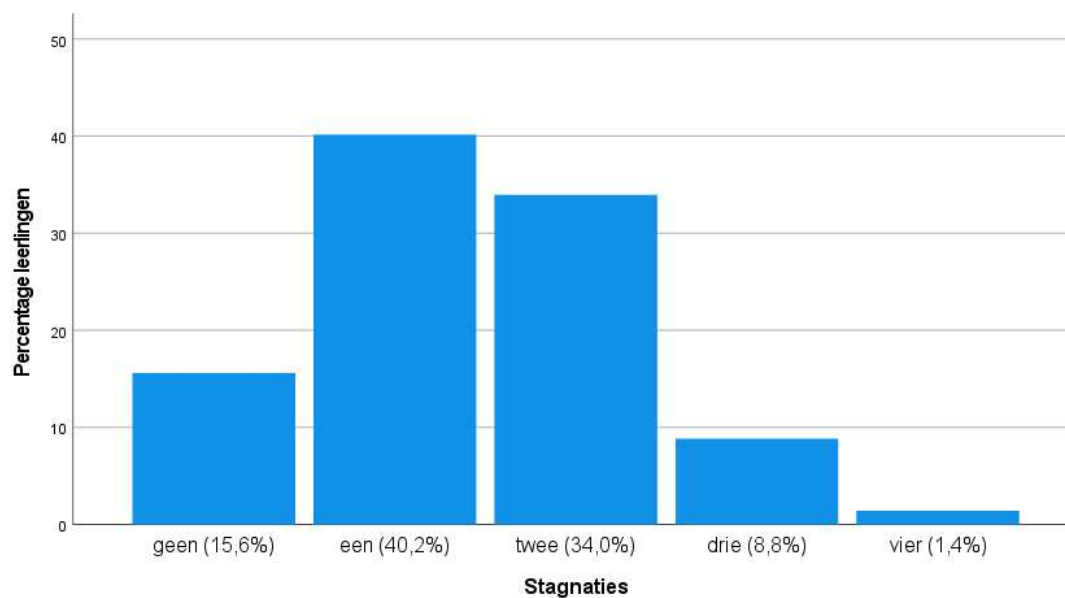
Resultaten

In de analyses zijn van 51 scholen 1063 leerlingen meegenomen, die aan de inclusiecriteria van dit onderzoek voldeden. Tabel 1 geeft de verdeling van de steekproef op categorische kenmerken weer, inclusief missing data. Opvallend is dat een klein percentage leerlingen zeer laag en/of laag opgeleide ouders heeft (10,9%). De ruime meerderheid heeft minstens één Nederlandse ouder (89,5%), groeit niet op in een eenoudergezin (87,6%) en/of heeft niet gedoubleerd (91,3%). Het percentage herfstkinderen in de steekproef is 24,6%. Bij 84,4% komt stagnatie in vaardigheidsgroei van Woordenschat in groep 3-5 voor. Driekwart is één of twee keer gestagneerd (Figuur 2). De verhouding leerlingen die wel en niet stagneert, verschilt significant tussen scholen, aangezien uit Fisher’s exact test een $p <,001$ ¹ volgt. Op sommige scholen stagneert geen enkele leerling in de steekproef, op andere scholen (bijna) alle leerlingen.

¹ Door technische restricties is de p -waarde berekend op basis van Monte-Carlo simulatie

Tabel 1*Beschrijving steekproef categorische variabelen (N=1063)*

Variabele	Categorieën	<i>n</i>	%
Geslacht	Jongen	536	50,4%
	Meisje	527	49,6%
Opleidingsniveau ouders	Zeer laag	45	4,2%
	Laag	71	6,7%
	Gemiddeld	903	84,9%
	Missing	44	4,1%
Migratie	Wel	112	10,5%
	Niet	951	89,5%
Eenouder	Wel	130	12,2%
	Niet	931	87,6%
	Missing	2	0,2%
Doublure	Wel	91	8,6%
	Niet	970	91,3%
	Missing	2	0,2%
Geboortemaand	Jan-sept	801	75,4%
	Okt-dec	261	24,6%
	Missing	1	0,1%

Figuur 2*Steekproefverdeling aantal stagnaties (N=1063)*

Tabel 2*Beschrijving steekproef continue variabelen (N=1063)*

Variabele	<i>n</i>	Gem.	<i>SD</i>	Mediaan	Minimum	Maximum
Leeftijd	1062	9,83	0,67	9,67	8,08	11,92
Aantal afnames	1063	5,36	1,20	6,00	2,00	10,00
Startniveau	731	40,46	18,26	40,05	-16,00	94,65

Tabel 2 toont de beschrijvende gegevens van de continue variabelen binnen de steekproef van dit onderzoek. De leerlingen zijn tussen de 8 en 12 jaar oud. De leeftijd is bimodaal verdeeld, met een piek tussen de 9 en 10 jaar en rond de 11 jaar. Dit kan verklaard worden doordat leerlingen die wel gedoubleerd hebben in de steekproef significant ouder zijn dan leerlingen die niet gedoubleerd hebben $t(1058) = -11,479; p <,001$. De mediane leeftijd van de niet gedoubleerde leerlingen, ligt tussen de 9 en 10 jaar en kan de eerste piek verklaren. De mediane leeftijd van de gedoubleerden, ligt tussen de 10 en 11 jaar, en kan in combinatie met de outliers binnen de groep die niet gedoubleerd heeft, de tweede piek verklaren.

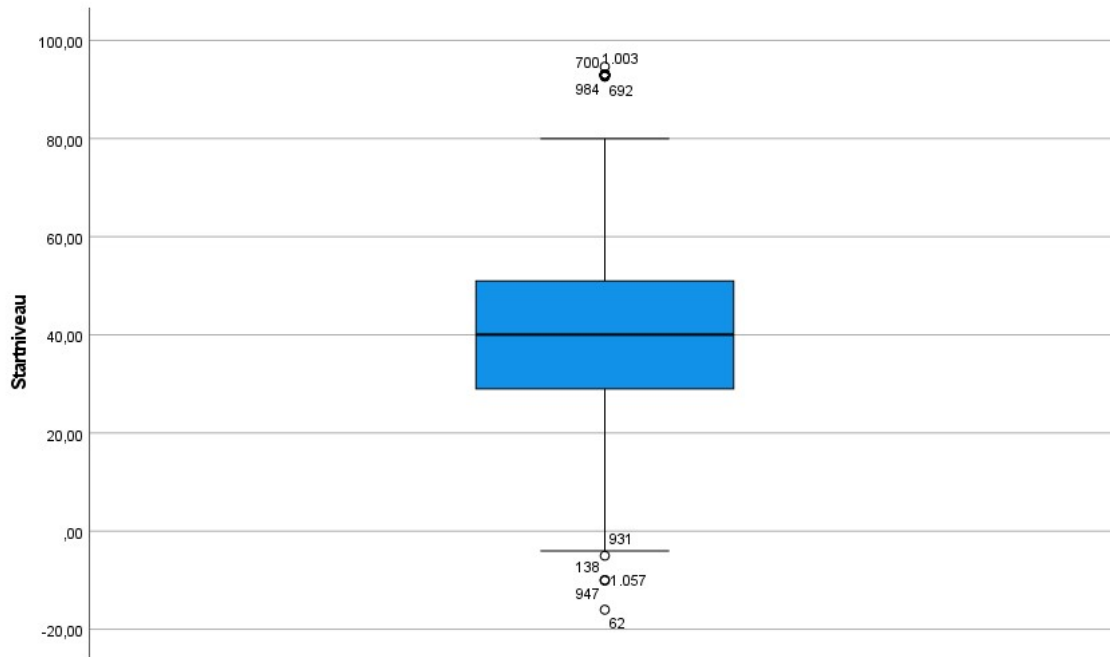
Conform verwachting hebben de meeste leerlingen 6 afnames gehad, aangezien de resultaten van groep 3-5 zijn geanalyseerd en er gewoonlijk 2 afnames per jaar plaatsvinden. Leerlingen bij wie meer dan 6 afnames plaatsgevonden hebben, zijn gemiddeld significant ouder zijn dan leerlingen met minder toetsafnames ($r =,099$). Bij gedoubleerde leerlingen vinden significant meer afnames plaats $t(1059) = -6,617; p <,001$.

Het startniveau beschrijft een vrij normale verdeling. De meeste leerlingen hebben een startniveau tussen de 30,00 en 50,00. Figuur 3 toont dat er sprake is van 9 outliers. De 5 leerlingen met een opvallend laag startniveau hebben allen ouders met een migratieachtergrond, komen niet uit een eenoudergezin en hebben niet gedoubleerd. De 4 leerlingen met een opvallend hoog startniveau komen ook niet uit een eenoudergezin en hebben niet gedoubleerd. Bovendien hebben zij allen Nederlandse ouders, van wie minstens één gemiddeld opgeleid is. Bij 31,2% van de steekproef mist het startniveau. Binnen deze groep leerlingen is het percentage dat stagneert (68,7%) significant lager dan binnen de groep bij wie het startniveau bekend is (91,5%), $\chi^2(1) = 90,412; p <,001$. Ook geldt dat zij gemiddeld significant minder afnames hebben gehad ($M=4,08$) ten opzichte van leerlingen

van wie het startniveau bekend is ($M=5,94$), $t(1061) = 33,666$; $p <,001$. Of het startniveau mist, hangt ook significant samen met de school. Van 8 van de 51 scholen in de steekproef ontbreekt het startniveau bij iedere leerling. Op de overige variabelen verschillen leerlingen bij wie het startniveau wel en niet ontbreekt niet significant.

Figuur 3

Steekproefverdeling startniveau (N=731)



Tabel 3 toont hoe de categorische predictoren in de steekproef samenhangen met stagnatie. In verhouding zijn er binnen de groep met migratieachtergrond significant minder leerlingen die stagneren dan in de groep zonder. Ook geldt dat in verhouding minder vaak stagnatie optreedt bij leerlingen met ouders met een zeer laag en/of laag opleidingsniveau, maar het verschil met een gemiddeld opleidingsniveau is niet significant. Geslacht, doublure, eenouder en geboortemaand hangen niet significant samen met stagnatie.

Tabel 3*Kruistabellen en Chi-kwadraattoetsen stagnatie en categorische variabelen*

		Stagnatie n(%)			χ^2 - waarde	df	p-waarde
		Wel	Niet	Totaal			
Eenouder	Wel	104(80,0)	26(20,0)	130	2,128	1	,145
	Niet	791(85,0)	140(15,0)	931			
	Totaal	897(84,4)	166(15,6)	1061			
Migratie	Wel	78(69,6)	34(30,4)	112	20,644	1	<,001
	Niet	819(86,1)	132(13,9)	951			
	Totaal	897(84,4)	166(15,6)	1063			
Opleidingsniveau ouders	Zeer laag	34(75,6)	11(24,4)	45	3,465	2	,177
	Laag	57(80,3)	14(19,7)	71			
	Gemiddeld	765(84,7)	138(15,3)	903			
	Totaal	856(84,0)	13(16,0)	1019			
Geslacht	Jongen	455(84,9)	81(15,1)	536	0,209	1	,648
	Meisje	442(83,9)	85(16,1)	527			
	Totaal	897(84,4)	166(15,6)	1063			
Doublure	Wel	81(89,0)	10(11,0)	91	1,578	1	,209
	Niet	815(84,0)	155(16,0)	970			
	Totaal	896(84,4)	165(15,6)	1061			
Geboortemaand	Jan-Sept	677(84,5)	124(15,5)	801	0,008	11	,930
	Okt-Dec	220(84,3)	41(15,7)	261			
	Totaal	897(84,5)	165(15,5)	1062			

Tabel 4 toont hoe de continue variabelen samenhangen met stagnatie. De variabele stagnatie hangt niet significant samen met leeftijd $t(1060) = -0,363$; $p = ,717$. De assumptie van normaliteit wordt echter geschonden. Wel vinden er significant meer afnames plaats bij leerlingen die wel stagneren dan bij leerlingen die niet stagneren $t(1061) = -14,648$; $p <,001$ (Figuur 4). Ook geldt dat leerlingen bij wie minstens 1 stagnatie opgetreden is, gemiddeld een hoger startniveau hebben ($M=41,89$) dan leerlingen bij wie geen stagnatie opgetreden is ($M=25,08$), $t(729) = -7,169$; $p <,001$ (Figuur 5). Het startniveau hangt niet significant samen met het aantal afnames ($r = ,009$).

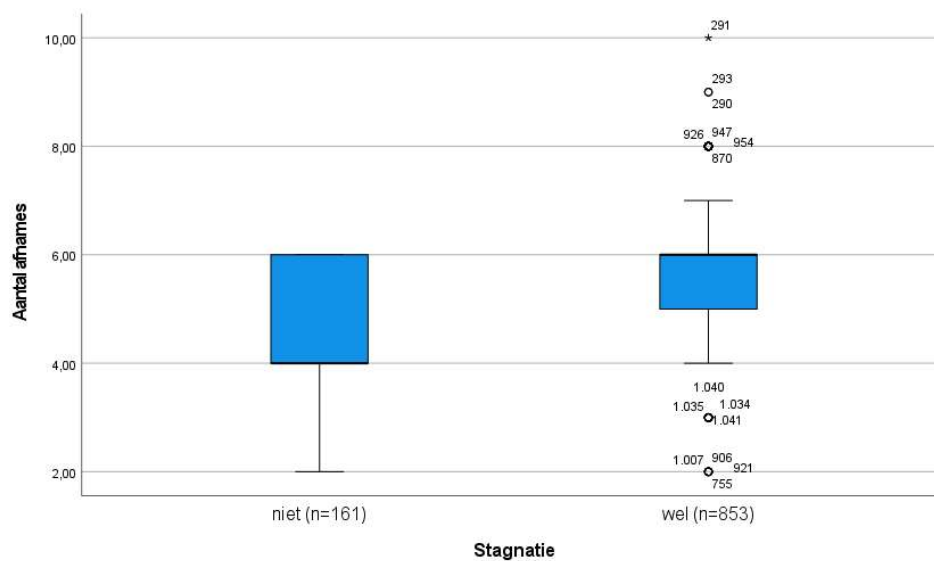
Tabel 4

Gemiddelde(SD) per continue variabele van leerlingen die wel en niet stagneren

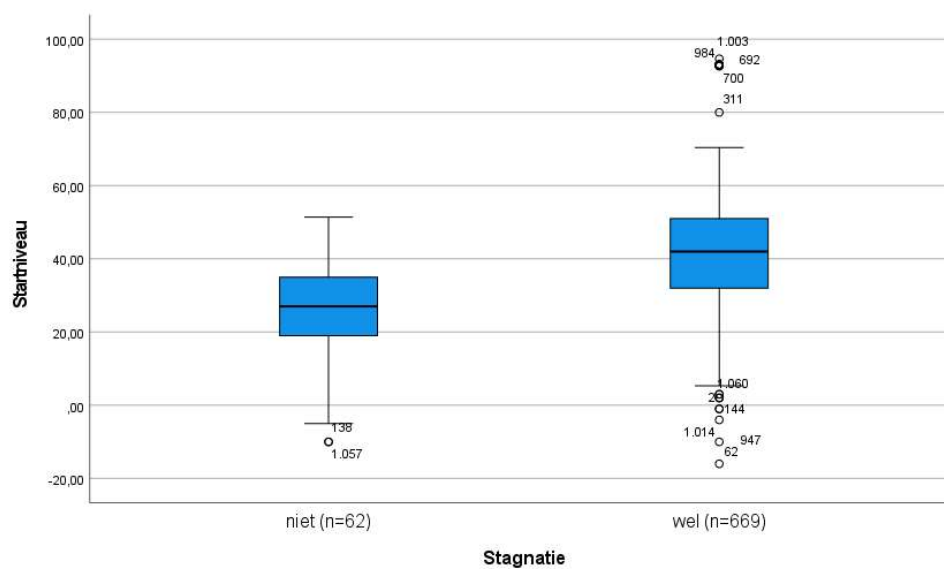
		Leeftijd	Startniveau	Afnames
Stagnatie	Wel	9,83(0,67)	41,89(17,94)	5,57(1,05)
	Niet	9,81(0,72)	25,08(14,26)	4,22(1,30)

Figuur 4

Samenhang aantal afnames en stagnatie

**Figuur 5**

Samenhang startniveau en stagnatie



Uit de bivariate relaties resulteerde dat startniveau, aantal afnames en migratie significant samenhangen met stagnatie. Startniveau is als eerst in het logistische regressiemodel opgenomen (Tabel 5 model 1). De regressiecoëfficiënt bleek significant ($p < ,001$), dus startniveau werd definitief opgenomen. Na toevoeging van het aantal afnames zijn beide regressiecoëfficiënten significant (Tabel 5 model 2). De proportie verklaarde variantie is met 9,0% toegenomen ten opzichte van model 1 (van 15,9% naar 24,9%). Vervolgens is de derde predictor (migratie) met een significante relatie met stagnatie toegevoegd. De regressiecoëfficiënt bleek niet significant ($b=0,242$; $SE=0,415$; $p=,559$) en de proportie verklaarde variantie nam met slechts 0,1% toe. Migratie werd vanwege de minimale toegevoegde waarde niet in het model opgenomen. Vanwege het ontbreken van significante regressiecoëfficiënten na toevoeging van de variabelen eenoudergezin ($b=-0,222$; $SE=0,374$; $p=,552$), opleidingsniveau van de ouders ($b=0,537$; $SE=0,452$; $p=,236$), leeftijd ($b=0,312$; $SE=0,224$; $p=,164$), geslacht ($b=0,041$; $SE=0,291$; $p=,887$), doublures ($b=18,834$; $SE=4486,856$; $p=,997$) en geboortemaand ($b=-0,182$; $SE=0,328$; $p=,579$), zijn deze ook niet van toegevoegde waarde in het verklaren van de variantie in stagnatie. Dit blijkt ook uit het gebrek aan of de minimale toename van de proportie verklaarde variantie na toevoeging van deze variabelen aan het model. Model 2 is het definitieve regressiemodel.

Tabel 5

Modelbouw logistische regressie

	<i>b</i>	<i>SE</i>	<i>Exp(b)</i>	<i>df</i>	<i>p</i> -waarde	Nagelkerke <i>R</i> ²
Model 1						
Constante	0,402	0,281	1,495	1	,152	,159
Startniveau	0,059	0,009	1,061	1	<,001	
Model 2						
Constante	-4,538	0,930	0,011	1	<,001	,249
Startniveau	0,059	0,009	1,061	1	<,001	
Afnames	0,863	0,155	2,371	1	<,001	

Om het definitieve model nader te interpreteren, is gebruik gemaakt van de odds van de predictoren en de proportie juiste voorspellingen op basis van het model. De odds van startniveau is gelijk aan 1,06. Bij een gemiddeld startniveau ($M=40,46$; $SD=18,26$) en een standaard aantal afnames (6), is de kans op stagnatie 95,4%. Een leerling met een gelijk aantal

afnames maar een startniveau van een standaarddeviatie hoger, heeft 98,4% kans op stagnatie. Een leerling met een gemiddeld startniveau en 7 afnames (1 afname meer dan het standaard aantal afnames), heeft 98,0% kans op stagnatie. Tabel 6 laat zien dat bij 92,2% van de steekproef op basis van het model correct voorspeld werd of stagnatie heeft plaatsgevonden. Opvallend is dat een aanzienlijk hoger percentage stagneerders correct voorspeld wordt, dan niet-stagneerders.

Tabel 6

Vergelijking geobserveerde en voorspelde stagnatie in de steekproef

		Voorspelde stagnatie ^a		Percentage correct voorspeld
		Wel	Niet	
Geobserveerde stagnatie	Wel	662	7	99,0
	Niet	57	5	8,1
	Totaal			91,2 ^b

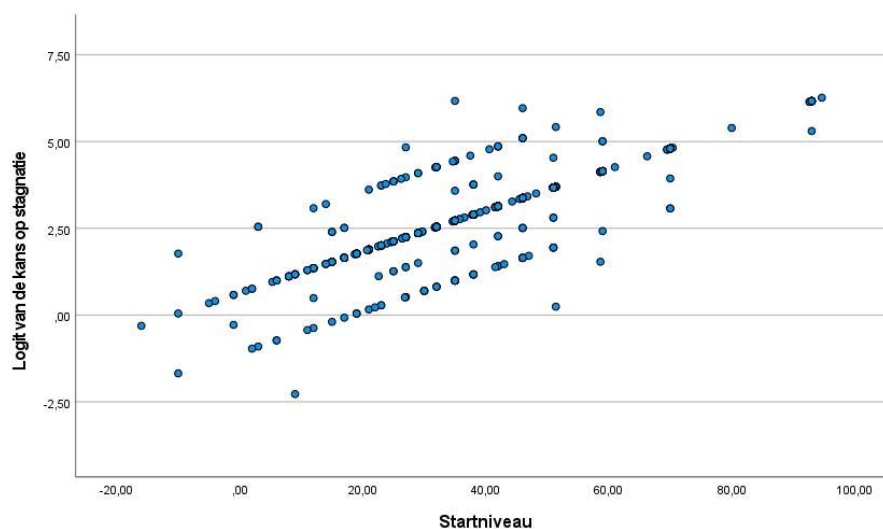
^a Cut-value ,5

^b 91,5 in het model zonder leerlingkenmerken

De voorspellers binnen het model hangen niet met elkaar samen, aangezien VIF=1,00. Cook's distance ligt per waarneming tussen de 0,00 en 0,08. Er is dus geen sprake van invloedrijke waarden en eventuele outliers zullen weinig invloed hebben gehad op de modelschattingen. De assumptie van lineariteit wordt niet geschonden (Figuur 6 en 7).

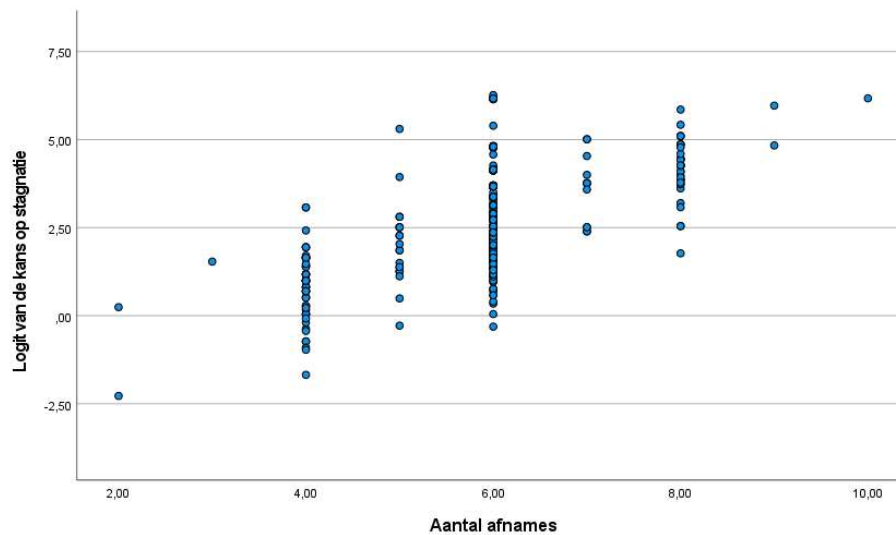
Figuur 6

Verband tussen logit van de kans op stagnatie en startniveau (N=731)



Figuur 7

Verband tussen logit van de kans op stagnatie en aantal afnames (N=731)



Discussie

Het doel van dit onderzoek is om na te gaan in hoeverre stagnatie in groei van woordenschatvaardigheid vaker voorkomt bij leerlingen met bepaalde kenmerken en in hoeverre stagnatie op basis daarvan voorspeld kan worden. Uit de resultaten bleek dat stagnatie frequent voorkomt: bij 84,4% van de steekproef. Bivariate relaties laten zien dat de kans op stagnatie significant groter is voor leerlingen met een hoger startniveau, meer toetsafnames en leerlingen zonder migratieachtergrond. Het effect van migratie verdwijnt echter bijna volledig in het model met startniveau en aantal afnames. Leerlingen zonder migratieachtergrond hebben een hoger gemiddeld startniveau en aantal afnames, wat samenhangt met meer kans op stagnatie. Startniveau en aantal afnames spelen een significante rol in het voorspellen van stagnatie. Sterke voorspellers zijn het niet, aangezien de regressiecoëfficiënten en standaardfouten van beide variabelen relatief laag zijn en het model met deze kenmerken van een lager percentage leerlingen correct voorspelt of ze stagneren (91,2%) dan een model zonder leerlingkenmerken (91,5%). De overige onderzochte leerlingkenmerken hangen niet significant samen met stagnatie.

Op basis van eerder onderzoek werd verwacht dat stagnatie in vaardigheidsgroei van woordenschat voorspeld kon worden op basis van specifieke kenmerken waar stagnatie systematisch mee zou samenhangen. De hypothese dat stagnatie samenhangt met het startniveau en aantal afnames, wordt bevestigd. Een hoger startniveau hangt samen met meer kans op stagnatie. Dit spreekt eerder onderzoek van Huttenlocher et al. (1991) tegen, waarin

leerlingen met een grotere woordenschat minder kans hebben op stagnatie. Dat onderzoek is echter uitgevoerd bij een jongere doelgroep en mogelijk is een meer betrouwbare toets gebruikt. Een grotere kans op stagnatie bij een hoger startniveau in dit onderzoek duidt op het meten van ruis, aangezien het statistisch verklaard kan worden door ‘regression toward the mean’: hoe verder boven of onder gemiddeld een leerling scoort, hoe groter de kans dat de daaropvolgende score dichterbij het gemiddelde ligt en dus respectievelijk lager of hoger uitvalt. (Agresti, 2018). Een verklaring voor het resultaat dat een hoger aantal afnames samenhangt met meer kans op stagnatie, is dat bij meer meetmomenten de kans op het vaststellen van stagnatie groter is. Ook is mogelijk dat bij stagnerende leerlingen meer toetsen worden afgenomen om hun ontwikkeling beter te monitoren. Als dit echter de verklaring zou zijn, is de verwachting dat leerlingen die doubleren vaker stagneren. We zien echter dat bij leerlingen die doubleren wel significant meer toetsafnames hebben plaatsgevonden, maar ze niet significant vaker stagneren. De kans op stagneren hangt dus samen met het toetsbeleid dat door scholen wordt gehanteerd.

De hypothese kan verworpen worden wat betreft de overige relaties. Het door Keuleers (2015) gevonden effect dat de kans op stagnatie met de leeftijd toeneemt, is in dit onderzoek niet vastgesteld. Mogelijk is de leeftijdsrange van dit onderzoek hier te smal voor. Het verbreden van deze range is aanbevolen om hier wel zicht op te krijgen. Bovendien wordt bij de analyses met leeftijd de assumptie van normaliteit geschonden, wat de betrouwbaarheid schaadt. Verschillende onderzoeken hebben aangetoond dat een lager opleidingsniveau van ouders samenhangt met meer kans op stagnatie (Huttenlocker et al., 1991; Jones & Rowland, 2017; Prevoo et al., 2014). In dit onderzoek bleken leerlingen met ouders met een laag tot zeer laag opleidingsniveau ten opzichte van een gemiddeld opleidingsniveau juist een kleinere kans op stagnatie te hebben, al was het verschil niet significant. Uit onderzoek van Dagevos et al. (2003) volgt dat basisschoolleerlingen met migratieachtergrond een taalachterstand hebben ten opzichte van autochtone leerlingen. Huidig onderzoek laat het tegenovergestelde zien, al verdwijnt dit effect wanneer het startniveau in acht genomen wordt. Mogelijk profiteren leerlingen van ouders met een lager opleidingsniveau en/of een migratieachtergrond zodanig van de taalinput op school, dat zij minder stagneren. Ook kan een rol hebben gespeeld dat het startniveau in eerder onderzoek niet meegenomen is, terwijl dit een invloedrijke predictor blijkt.

De steekproef is wat betreft startniveau en aantal afnames representatief voor de populatie, aangezien het gemiddeld startniveau ($M=40,46$; $SD=18,26$) niet ver afwijkt van het gemiddelde gebaseerd op normeringsonderzoek ($M=39,60$; $SD=14,95$) en er bij de meeste

leerlingen het te verwachten aantal afnames heeft plaatsgevonden (Berkel et al., 2010). Ook voor het percentage doublures, leerlingen in een eenoudergezin en het opleidingsniveau van de ouders is de steekproef representatief (CBS, 2021; Ministerie van Onderwijs, Cultuur en Wetenschap, 2016; Van Vuuren & Van der Wiel, 2015). De populatie bevat momenteel procentueel wel meer leerlingen met migratieachtergrond (30%) dan in de steekproef (10,5%) (CBS, 2021). Wanneer de steekproef ook hierin representatief zou zijn, bevordert dit de validiteit. Een sterk punt van dit onderzoek is dat data van een groot aantal leerlingen van verschillende scholen is geanalyseerd. Dit verhoogt de externe validiteit. Een beperking is dat leerlingen bij wie het startniveau mist, significant minder vaak stagneren in de steekproef. Hierdoor kan het percentage stagneerders zijn overschat, aangezien een groep leerlingen met een kleinere kans op stagnatie niet in het model is opgenomen. Echter kan deze kleinere kans op stagnatie mogelijk (deels) verklaard worden door minder afnames, aangezien de eerste meting mist.

Sterk aan dit onderzoek is dat de assumpties van lineariteit en multicollineariteit niet geschonden zijn. De assumptie van onafhankelijkheid is wel geschonden, aangezien sommige leerlingen van dezelfde school en/of uit dezelfde klas komen. De verhouding leerlingen die wel en niet stagneert verschilt tussen scholen, wat wijst op afhankelijkheid binnen de data waardoor resultaten eerder significantie bereiken. De meetfout wordt kleiner geschat dan hij werkelijk is. Door de verwerpingsregel strenger te maken, zou hier voor gecorrigeerd kunnen worden. Een betere oplossing is het gebruik van multilevel analyse: een techniek die rekening kan houden met afhankelijkheid en geclusterde data op verschillende niveaus (Hox et al., 2018). Een sterk punt van het onderzoek is ook dat meerdere theoretisch relevante variabelen zijn meegenomen. De interne validiteit wordt echter wel bedreigd doordat variabelen die niet in het onderzoek zijn meegenomen, bijvoorbeeld op schoolniveau, alternatieve verklaringen zouden kunnen bieden voor de gevonden relaties. De kwaliteit van de leerkracht kan een mediërende rol hebben gespeeld. Een meer gestandaardiseerde onderzoeksopzet zou de interne validiteit kunnen verhogen. Echter is dit ethisch onverantwoord en moeilijk te verwezenlijken binnen de onderwijscontext. Wel is het mogelijk om in vervolgonderzoek data over onder andere de kwaliteit van de leerkracht te verzamelen en hiervoor te controleren. Hier kan de afhankelijkheid ook verder mee in kaart worden gebracht.

Op basis van de resultaten van dit onderzoek is niet precies vast te stellen in hoeverre er sprake is van ontwikkelingsproblemen en/of meetfouten, en wat onderliggend is aan eventuele meetfouten. Mogelijk zouden een aanvullend kwalitatief onderzoek en meerdere informatiebronnen meer inzicht opleveren.

Een aanbeveling voor de praktijk is om bewust te zijn van eventuele statistische verklaringen voor een vastgestelde stagnatie bij leerlingen op basis van woordenschattoetsen. Een hoger startniveau (mogelijk veroorzaakt door een meetfout) verhoogt de kans op stagnatie, als gevolg van 'regression toward the mean'. Een hoger aantal afnames, maakt de kans op stagnatie groter. Door hier bewust van te zijn, kan worden voorkomen dat stagnatie onterecht wordt toegeschreven aan leerlingkenmerken en dat remediëring plaatsvindt wanneer geen sprake is van ontwikkelingsproblemen. Het is onwaarschijnlijk dat de 84,4% stagneerders in de steekproef allemaal ontwikkelingsproblemen hebben. Het is belangrijk om in de praktijk niet blindelings op toetsscores te vertrouwen, maar om ook kwalitatieve data over de ontwikkeling te verzamelen, zodat hulp de leerlingen toekomt die dit nodig hebben en werkelijke problemen niet gebagatelliseerd worden.

Stagnatie in vaardigheidsgroei van woordenschat lijkt dus het best voorspeld te kunnen worden op basis van het startniveau en aantal afnames. Vanwege het ontbreken van statistisch onverklaarbare systematische samenhang van de overige leerlingkenmerken met stagnatie en vanwege de onwaarschijnlijkheid dat bij 84,4% sprake is van ontwikkelingsproblemen, zijn de Cito Woordenschat toetsen niet betrouwbaar genoeg om stagnaties in individuele ontwikkeling mee vast te stellen. Het percentage stagnatie is zodanig hoog; dit zouden we verwachten wanneer we puur naar ruis kijken. Als gevolg van meetfouten over- of onderschat de toetsscore het woordenschatniveau, waardoor ook bij leerlingen zonder ontwikkelingsproblemen stagnatie wordt vastgesteld. Vervolgonderzoek zou moeten uitwijzen in hoeverre een verminderde toetsbaarheid van deze jonge leerlingen hierin een rol speelt en/of welke andere mechanismen hier van kracht zijn.

Literatuurlijst

- Agresti, A. (2018). *Statistical Methods for the Social Sciences* (5de editie). Pearson.
- Bauer, D. J., Goldfield, B. A., & Reznick, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics*, 23(3), 313–35. <http://doi.org/10.1017.S0142716402003016>
- Berkel, S. van, Hilte, M., Engelen, R., Kamphuis, F., Kleintjes, F. & Krom, R. (2010). *Leerling- en onderwijsvolgsysteem*. Woordenschat Groep 3 t/m 5. Wetenschappelijke verantwoording. Arnhem: Cito
- CBS. (2021, 1 november). *Jongeren in Nederland - Jaarrapport Landelijke Jeugdmonitor 2021*. Jongeren in Nederland - Jaarrapport Landelijke Jeugdmonitor 2021 | CBS. <https://longreads.cbs.nl/jeugdmonitor-2021/jongeren-in-nederland/>
- Cito. (z.d.). *Leerling in beeld - nieuw! : Maakt elke leerling zichtbaar*. www.cito.nl. Geraadpleegd op 4 mei 2023, van <https://www.cito.nl/onderwijs/primair-onderwijs/lvs-leerling-in-beeld>
- Dagevos, J. Gijbers, M. & Van Praag, C. (2003). *Rapportage minderheden: Onderwijs, arbeid en sociale integratie*. Sociaal en Cultureel Planbureau. Geraadpleegd op 25 mei 2023, van https://www.researchgate.net/publication/306359531_Rapportage_Minderheden_2003_Onderwijs_arbeid_en_social-culturele_integratie
- Driessen, G., Mulder, L., Leest, B. & Verrijt, T. (2014). Zittenblijven in Nederland: Een probleem? *Tijdschrift voor Orthopedagogiek*, 53, 297-311. https://www.researchgate.net/publication/263320118_Zittenblijven_in_Nederland_Een_probleem
- Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, 56(8), 848–856. <https://doi.org/10.1111/jcpp.12378>
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). COTAN Beoordelingssysteem voor de kwaliteit van tests. In *cotandocumentatie.nl*. https://www.researchgate.net/publication/254836210_COTAN_Beoordelingssysteem_voor_de_kwaliteit_van_tests
- Frans, N. (2019). *A Captivating Snapshot of Standardized Testing in Early Childhood: on the stability and utility of the Cito preschool/kindergarten tests* [MA thesis]. 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://research.rug.nl/en/publications/identification-with-the-cito-preschoolkindergarten-tests-below-av>
- Gelderblom, G., Schildkamp, K., Pieters, J., & Ehren, M. (2016). Data-based decision making for instructional improvement in primary education. *International Journal of Educational Research*, 80, 1–14. <https://doi.org/10.1016/j.ijer.2016.07.004>
- Hox, J. J., Moerbeek, M., & Van De Schoot, R. (2018). *Multilevel Analysis: Techniques and Applications, Third Edition* (3de editie) [E-book]. Routledge. https://books.google.nl/books?hl=nl&lr=&id=iLD_DwAAQBAJ&oi=fnd&pg=PP1&q=multilevel+analysis&ots=a4oX_74tXd&sig=k7JcLrShLCzC7sXD8fleCT3uX-4#v=onepage&q=multilevel%20analysis&f=false
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(2), 236–248. <https://doi-org.proxy-ub.rug.nl/10.1037/0012-1649.27.2.236>
- Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, 98, 1–21. <https://doi.org/10.1016/j.cogpsych.2017.07.002>
- Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*, 68(8), 1665–1692. <https://doi.org/10.1080/17470218.2015.1022560>
- Mayo, A., & Leseman, P. P. (2008). Off to a good start? Vocabulary development and differences in early family and classroom experiences of children from native-dutch and immigrant families in the Netherlands. *Educational and Child Psychology*, 25(3), 70–82. <https://doi.org/10.53841/bpsecp.2008.25.3.70>
- Ministerie van Onderwijs, Cultuur en Wetenschap. (2016, 13 augustus). *Differentiatie van het opleidingsniveau van ouders bij het beoordelen van opbrengsten van de basisschool*. Publicatie | Inspectie van het onderwijs. <https://www.onderwijsinspectie.nl/documenten/publicaties/2013/01/08/differentiatie-opleidingsniveau-ouders-bij-beoordelen-opbrengsten-basisschool>

- Prevo, M. J., Malda, M., Mesman, J., Emmen, R. A. G., Yeniad, N., Van IJzendoorn, M., & Linting, M. (2014). Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: Different pathways for host and ethnic language. *Journal of Child Language*, *41*(5), 963-984.
<http://doi.org/10.1017/S0305000913000299>
- Schrooten, W., & Vermeer, A. (1994). *Woorden in het basisonderwijs: 15.000 woorden aangeboden aan leerlingen*. Tilburg University Press.
<https://annevermeer.github.io/wrdlstin.html>
- Sénéchal, M., Ouelette, G., & Rodney, D. (2006). The Misunderstood Giant: On the Predictive Role of Early Vocabulary to Future Reading [E-book]. In *Handbook of Early Literacy Research* (Vol. 2, pp. 173–182). Guilford Press.
<https://ebookcentral.proquest.com/lib/rug/detail.action?docID=362557>
- Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42. <https://doi.org/10.12691/ajams-8-2-1>
- Van Vuuren, D., & Van der Wiel, K. (2015, 17 januari). *Zittenblijven in het primair en voortgezet onderwijs: Een inventarisatie van de voor- en nadelen*. Publicatie | Centraal Planbureau. <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-policy-brief-2015-01-zittenblijven-het-primair-en-voortgezet-onderwijs.pdf>
- Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*(3), 407–423.
<https://doi.org/10.1002/acp.1414>
- Vlug, K.F.M. (1997). Because every pupil counts: the success of the pupil monitoring system in the netherlands. *Education and Information Technologies*, *2*(4), 287–306.
<https://doi.org/10.1023/A:1018629701040>
- Waber, D. P. (1976). Sex differences in cognition: a function of maturation rate? *Science*, *192*(4239), 572–574. <http://www.jstor.org/stable/1741573>
- Werkgroep IBO. (2017). Onderwijsachterstandenbeleid, een duwtje in de rug? In <https://archieff.rijksbegroting.nl/>. Ministerie van Financiën. Geraadpleegd op 20 maart 2023, van <https://archieff.rijksbegroting.nl/system/files/12/ibo-onderwijsachterstandenbeleid-eindrapport-een-duwtje-de-rug.pdf>