

**Het verschil in de mate van stabiliteit van de leerkrachtverwachtingen tussen het  
leerjaar en gender van leerlingen**

Milou van der Maat

S5717892

Onderwijsinnovatie - Rijksuniversiteit Groningen

Master's Thesis - PAMAOW05

Begeleider - Dr. A. C. (Anneke) Timmermans

31 mei 2024

## Samenvatting

Eerdere onderzoeken hebben aangetoond dat resultaten van leerlingen beïnvloed kunnen worden door leerkrachtverwachtingen. Veelal gebruikten deze onderzoeken één meting van de leerkrachtverwachtingen aan het begin van het schooljaar. Hierdoor is een gebrek aan onderzoek waar de stabiliteit van de leerkrachtverwachtingen gedurende een langere tijdsperiode wordt gemeten. Om deze reden is in dit onderzoek de stabiliteit van de leerkrachtverwachtingen gedurende één schooljaar, met behulp van drie meetmomenten bepaald. De verwachtingen van 88 leerkrachten over 2461 leerlingen in twaalf Nieuw-Zeelandse basis- en voortgezet openbare scholen werden onderzocht voor de vakken wiskunde en lezen. Hierbij is gekeken in hoeverre de stabiliteit van de leerkrachtverwachtingen van individuele leerlingen gedurende één schooljaar stabiel bleef. Ook is onderzocht in hoeverre de mate van stabiliteit van de leerkrachtverwachtingen van individuele leerlingen verschilt tussen de verschillende leerjaren of het gender van leerlingen. Als laatste is het verschil in de mate van de stabiliteit van de leerkrachtverwachtingen tussen het gender van leerlingen uitgesplitst in verschillende leerjaren onderzocht. De resultaten van het onderzoek laten zien dat de stabiliteit van de leerkrachtverwachtingen aan het eind van het schooljaar sterker is dan aan het begin van het schooljaar. Daarnaast blijkt dat er geen duidelijk verband zichtbaar is tussen het leerjaar van leerlingen en de stabiliteit van de leerkrachtverwachtingen. Ook laat de stabiliteit van de leerkrachtverwachtingen geen significant verschil zien tussen jongens en meisjes. Als laatste laat de vergelijking van de correlaties tussen jongens en meisjes in verschillende leerjaren enkele significante verschillen zien, zonder duidelijk patroon.

*Sleutelwoorden:* leerkrachtverwachtingen, stabiliteit leerkrachtverwachtingen, leerjaar, gender.

### **Abstract**

Previous studies have shown that student outcomes can be influenced by teacher expectations. Often, these studies used one measurement for teacher expectations at the beginning of the school year. As a result, there is a lack of research where the stability of teacher expectations is measured over a longer period. For this reason, this study determined the stability of teacher expectations over one school year, using three measurement moments. The expectations of 88 teachers across 2461 students in twelve New Zealand primary and secondary public schools were examined for the subjects of mathematics and reading. This involved examining the extent to which the stability of teacher expectations of individual students remained stable over one school year. The extent to which the degree of stability of individual pupils' teacher expectations differed across grade levels or gender of pupils was also examined. Lastly, the difference in the degree of stability of teacher expectations between the gender of pupils disaggregated into different grade levels was examined. The results of the study shows that the stability of teacher expectations is stronger at the end of the school year than at the beginning of the school year. It also shows that no clear relationship is visible between pupils' grade levels and the stability of teacher expectations. Also, the stability of teacher expectations shows no significant difference between boys and girls. Finally, the comparison of correlations between boys and girls in different grade levels shows some significant differences, with no clear pattern.

**Keywords:** teacher expectations, stability teacher expectations, grade, gender.

## Inhoudsopgave

Samenvatting .....	2
Abstract .....	3
Inhoudsopgave .....	4
Inleiding .....	6
Stabiliteit van de Leerkrachtverwachtingen .....	7
Stabiliteit van de Leerkrachtverwachtingen Bij het Leerjaar van Leerlingen .....	11
Stabiliteit van de Leerkrachtverwachtingen Bij het Gender van Leerlingen.....	12
De Huidige Studie .....	13
Methode.....	14
Onderzoeksdesign.....	14
Context.....	14
Participanten .....	15
Instrumenten .....	15
Leerkrachtverwachtingen.....	15
Leerling-prestaties.....	16
Procedure .....	18
Analyse .....	18
Missende Waarden .....	18
Onderzoeksvragen.....	19
Resultaten .....	21
Regressiemodellen.....	21
Stabiliteit van de Leerkrachtverwachtingen Wiskunde en Lezen Gedurende Eén Schooljaar .....	22
Vershil Mate van Stabiliteit Leerkrachtverwachtingen Tussen het Leerjaar van Leerlingen .....	23
Wiskunde.....	23
Lezen.....	25
Vershil Mate van Stabiliteit Leerkrachtverwachtingen Wiskunde en Lezen Tussen het Gender van Leerlingen .....	27
Vershil Mate van Stabiliteit Leerkrachtverwachtingen Tussen het Gender van Leerlingen Uitgesplitst in Verschillende Leerjaren.....	28
Wiskunde.....	28
Lezen.....	29
Conclusie en Discussie.....	31

Sterke Punten en Limitaties .....	34
Referenties.....	37
Bijlage 1 .....	44

## Inleiding

Het onderzoek naar leerkrachtverwachtingen begon bij het boek *'Pygmalion in the classroom'* van Rosenthal en Jacobson (1968). Dit onderzoek wordt gezien als de basis van het onderzoek naar leerkrachtverwachtingen in de daaropvolgende periode (Good et al., 2018). Leerkrachtverwachtingen worden door Good en Brophy (1997) gedefinieerd als: “conclusies die leerkrachten trekken over het toekomstige gedrag of de toekomstige schoolprestaties van hun leerlingen, op basis van wat ze nu over deze leerlingen weten” (p. 79, eigen vertaling). Leerkrachten baseren hun verwachtingen van individuele leerlingen op de academische prestaties van leerlingen en de betrokkenheid en het gedrag van leerlingen in de klas (Dusek & Joseph, 1983; Wang et al., 2018). Ook worden de verwachtingen van leerkrachten gebaseerd op basis van verschillende demografische kenmerken van leerlingen zoals; etniciteit (Tenenbaum & Ruck, 2007), gender (Robinson-Cimpian et al., 2014) en sociaal economische status (Westphal et al., 2016).

Bij onderzoek naar leerkrachtverwachtingen worden vaak onderzoeksopzetten gebruikt met één meetmoment van verwachtingen vroeg in het schooljaar (Rubie-Davies et al., 2018). Dit wordt gedaan aangezien de hypothese is dat de verwachtingen van leerkrachten over de prestaties van leerlingen gedurende een schooljaar stabiel blijven. Om deze reden zijn er maar enkele onderzoeken beschikbaar waarbij gebruik wordt gemaakt van meer dan één meetmoment en waarbij wordt gekeken of de verwachtingen van docenten voor individuele leerlingen in de loop van de tijd stabiel zijn of dat ze hun eerste verwachtingen bijstellen wanneer er nieuwe informatie beschikbaar komt (Wang et al., 2020). Over de resultaten van deze onderzoeken kan nog geen eenduidige conclusie getrokken worden ten aanzien van de stabiliteit van de leerkrachtverwachtingen, doordat deze verschillende resultaten laten zien (Timmermans et al., 2021; Wang et al., 2018).

Over het algemeen wordt er in het beperkte stabiliteitsonderzoek nog niet gekeken naar de eventuele verschillen in de stabiliteit van de leerkrachtverwachtingen tussen groepen

leerlingen, bijvoorbeeld naar verschillen omtrent de etniciteit, het gender of het leerjaar van leerlingen. Dit is wel van belang om zo te kunnen begrijpen hoe leerkrachtverwachtingen zich ontwikkelen en om te kunnen achterhalen welke groepen mogelijk zorgen voor verschillen in de mate van de stabiliteit van de leerkrachtverwachtingen. Aangezien het over- of onderschatten van leerlingen door leerkrachten invloed kan hebben op hoe leerlingen presteren, zelfs op de lange termijn, is het van belang om als leerkracht correcte verwachtingen te hebben van alle groepen leerlingen (Hattie, 2008; Jussim & Harber, 2005; Rubie-Davies et al., 2018). Op het moment dat een leerkracht blijft geloven dat een bepaalde verwachting waar is, kan het handelen van een leerkracht ervoor zorgen dat de verwachting uitkomt, ook wel een self-fulfilling prophecy genoemd (Merton, 1948; Timmons, 2019). Om je verwachtingen als leerkracht aan te kunnen passen, als de verwachting onjuist bleek te zijn, is een zekere mate van instabiliteit van belang. Hierbij worden eerdere verwachtingen door de leerkracht bijgesteld aan de hand van veranderingen in het gedrag of academische prestaties van leerlingen (Timmermans et al., 2021). Op het moment dat leerkrachten hun verwachtingen aanpassen en nauwkeuriger worden, is de kans kleiner dat er door de leerkrachtverwachtingen een self-fulfilling prophecy ontstaat (Jussim & Harber, 2005). Door te achterhalen of leerkrachtverwachtingen voor bepaalde groepen wel of niet stabiel blijven, kunnen leerkrachten een stabiele of niet stabiele verwachting aanpassen op het moment dat dit een positief verschil kan maken ten aanzien van de leerling-prestaties.

In dit onderzoek zal dan ook worden onderzocht in hoeverre de verwachtingen van de leerkrachten voor wiskunde en lezen stabiel zijn gedurende één schooljaar en in hoeverre het leerjaar van een leerling, het gender van een leerling en de combinatie van deze twee factoren verschillen laten zien in de mate van de stabiliteit van de leerkrachtverwachtingen.

### **Stabiliteit van de Leerkrachtverwachtingen**

Stabiliteit van de leerkrachtverwachtingen wordt in dit onderzoek gedefinieerd als: “of leerkrachten wel of niet vasthouden aan bestaande verwachtingen bij het verkrijgen van tegenstrijdige informatie of dat ze deze verwachtingen aanpassen naarmate de leerlingen vorderen gedurende het jaar” (Rubie-Davies et al., 2018, p. 223, eigen vertaling). Zeven studies hebben onderzoek gedaan naar de stabiliteit van de leerkrachtverwachtingen over een langere periode. Beginnend met drie studies die de stabiliteit van de leerkrachtverwachtingen onderzochten waarbij geen rekening gehouden werd met de academische prestaties van de leerlingen. De studies van Good et al. (1980) en Martinek (1980) hebben beide onderzoek gedaan naar de stabiliteit van de leerkrachtverwachtingen in het basisonderwijs. De studie van Good et al. (1980) heeft dit bij zestien leerkrachten en 192 leerlingen onderzocht in relatie tot de domeinen ‘algemeen academisch potentieel’ en ‘succes bij het voltooien van verbale academische taken’. De studie van Martinek (1980) heeft dit bij zes leerkrachten van *grade 2*, 4 en 6 (7 tot 11 jaar) en 179 leerlingen in relatie tot het domein ‘vaardigheden lichamelijke opvoeding’ onderzocht. Bij het onderzoek van Good et al. (1980) waren de correlaties van de drie meetmomenten allemaal boven de 0.75 en bij Martinek (1980) varieerde de correlaties van de twee meetmomenten tussen de 0.68 en 0.96. De correlaties uit deze twee onderzoeken tonen een sterke stabiliteit van de leerkrachtverwachtingen. Dit impliceert dat de leerlingen van wie de leerkrachten op een bepaald moment hoge verwachtingen hadden, waarschijnlijk dezelfde leerlingen waren van wie de leerkrachten bij een later meetmoment ook hoge verwachtingen hadden (Timmermans et al., 2021). Deze twee onderzoeken hebben ook limitaties. Allereest is de steekproefgrootte van zowel de leerkrachten ( $N = 16$  en  $N = 6$ ) als de leerlingen ( $N = 192$  en  $N = 179$ ) vrij klein. Daarnaast is de tijd die tussen de meetmomenten zit, respectievelijk, drie á vier maanden (Good et al., 1980) en acht weken (Martinek, 1980), vrij lang aangezien leerkrachten op dagelijks/wekelijkse basis informatie verzamelen over hun leerlingen. Het onderzoek van Kuklinski en Weinstein (2000), waarbij ook geen rekening



gehouden werd met de academische prestaties van de leerlingen, gebruikte nog een grotere tijdsperiode, van zes maanden, tussen de twee meetmomenten. In de eerste deel van deze studie hadden 48 leerkrachten van *grade* 1, 3 en 5 (6 tot 10 jaar), 464 leerlingen beoordeeld van 'slecht' tot 'uitstekend', gaande over de leerkrachtverwachtingen ten aanzien van de leesprestaties van leerlingen. In het tweede deel van de studie rangschikten twaalf leerkrachten van *grade* 1, 3 en 5 (6 tot 10 jaar), 138 leerlingen. Hierbij kwam naar voren dat de leerkrachtverwachtingen bij het geven van een beoordeling (en niet bij het rangschikken) meer stabiel waren voor leerlingen in *grade* 5 dan in *grade* 1. Kuklinski en Weinstein (2000) gaven als mogelijke verklaring hiervoor, dat dit komt door het gebruik van informatie van leerkrachten uit voorgaande leerjaren. Wat betekent dat leerkrachten het idee hadden dat eerder verzamelde informatie over leerlingen een juiste weergave was van het kunnen van leerlingen. Ook zouden leerlingen zich in hogere leerjaren consequenter kunnen gedragen naar de verwachtingen van de leerkracht. Voor beide delen van de studie bleek dat de leerkrachtverwachtingen een middelmatige tot sterke stabiliteit hadden waarbij het gemiddelde hiervan varieerde van 0.65 tot 0.69.

Ondanks dat in de onderzoeken van Good et al. (1980), Martinek (1980) en Kuklinski en Weinstein (2000) de stabiliteit van de leerkrachtverwachtingen is onderzocht zonder rekening te houden met de academische prestaties van de leerlingen, is het wel van belang om deze variabele mee te nemen. Dit aangezien dan bij het interpreteren van de academische resultaten niet beargumenteerd zou kunnen worden dat de leerkrachtverwachtingen stabiel zijn door stabiliteit bij de academische resultaten van leerlingen (Wang et al., 2020). Onderzoeken waarbij de academische prestaties van de leerlingen wel werden meegenomen en waarbij grotere steekproefgroottes zijn gebruikt, zijn: Rubie-Davies et al. (2018), Wang et al. (2020), Timmermans et al. (2021) en Hao et al. (2022).

In het onderzoek van Rubie-Davies et al. (2018) werden de verwachtingen van 94 leerkrachten van de basisschool en de onderbouw van het voortgezet onderwijs onderzocht waarbij gefocust werd op de stabiliteit van de leerkrachtverwachtingen op groepsniveau en niet op leerlingniveau. Er waren zeven meetmomenten verspreid over drie schooljaren. De resultaten lieten zien dat de leerkrachtverwachtingen op groepsniveau stabiel bleven binnen elk van de drie schooljaren. Er waren weinig variaties van jaar tot jaar en de leerkrachtverwachtingen waren voor alle drie de schooljaren vergelijkbaar, ondanks veranderingen in leerling cohorten elk schooljaar.

In de onderzoeken van Wang et al. (2020), Timmermans et al. (2021) en Hao et al. (2022) is gekeken naar de stabiliteit van de leerkrachtverwachtingen op leerlingniveau. Bij het onderzoek van Wang et al. (2020) is gebruik gemaakt van drie meetmomenten in één schooljaar bij Chinese leerlingen (12-15 jaar). Het onderzoek werd uitgevoerd onder vijftig leerkrachten van de vakken wiskunde en engels en bij 1.199 leerlingen. Nadat de academische resultaten van de leerlingen werden meegenomen, waren de correlaties tussen de leerkrachtverwachtingen veel lager, er was een matig tot sterke stabiliteit voor 18-22% van de gevallen. Het onderzoek van Timmermans et al. (2021) onderzocht onder 89 leerkrachten en 2.536 leerlingen de stabiliteit van de leerkrachtverwachtingen van de wiskunde binnen één schooljaar met drie meetmomenten met een tijdsinterval van vier maanden. Er werd een sterke stabiliteit van de leerkrachtverwachtingen gevonden '*between-student level*', wat blijkt uit de correlaties die varieerden van 0.736 tot 0.946. Deze bevinding gaf aan dat de rangorde van leerlingen op basis van de verwachtingen van de leerkrachten vrij stabiel bleef gedurende het schooljaar. Uit de resultaten '*within-student level*' in dit onderzoek naar voren dat leerkrachten hun verwachtingen voor individuele leerlingen tijdens de eerste schoolmaanden aanpasten en hun aangepaste verwachtingen later in het jaar handhaafden. Ook bleek dat leerkrachten hun verwachtingen aanpasten om op één lijn te komen met de prestaties van de

leerlingen en niet andersom. De meest recente studie van Hao et al. (2022) sluit aan bij de resultaten uit het onderzoek van Timmermans et al. (2021) waarin de verschillen ‘*within-student level*’ wordt onderzocht. Hier bleek ook dat leerkrachten in de eerste paar maanden hun verwachtingsvooroordelen aanpasten, maar deze aangepaste verwachtingen vervolgens in stand hielden. De studie van Hao et al. (2022) onderzocht dit binnen twee schooljaren met vier meetmomenten onder vijftig leerkrachten met hoge, gemiddelde en lage verwachtingen en 567 Chinese middelbare scholieren voor de vakgebieden wiskunde, chinees en engels. Het aanpassen van de leerkrachtverwachtingen in de eerste maanden zou erop kunnen wijzen dat leerkrachten zich bewust worden van onnauwkeurige verwachtingen over de werkelijke capaciteiten van de leerlingen op het moment dat ze hun leerlingen beter leren kennen (Hao et al., 2022). De vooringenomen verwachtingen van leerkrachten aan het begin van het schooljaar, zouden mogelijk gebaseerd kunnen zijn op ervaringen van leerkrachten die de leerlingen eerder in de klas hebben gehad (Ready & Wright, 2011). Mogelijk is deze vooringenomenheid bijgestuurd door nieuwe en betere informatie die leerkrachten hebben verzameld over de leerlingen (Brophy, 1983).

### **Stabiliteit van de Leerkrachtverwachtingen Bij het Leerjaar van Leerlingen**

Naar de mogelijke verschillen van het leerjaar waar leerlingen zitten op de stabiliteit van de leerkrachtverwachtingen is maar in één onderzoek gekeken. Het onderzoek van Kuklinski en Weinstein (2000) laat zien dat de leerkrachtverwachtingen bij lezen van leerlingen uit *grade 1* (6 jaar), minder stabiel en meer flexibel waren dan in de hogere leerjaren *grade 3* (8 jaar) en *grade 5* (10 jaar). Stabieler leerkrachtverwachtingen in een hoger leerjaar zouden verklaard kunnen worden door de gedachte dat leerkrachten in hogere leerjaren meer waarde hechten aan informatie uit de schoolgegevens van leerlingen en doordat leerlingen zich consequenter zouden kunnen gedragen naar de verwachtingen van de leerkracht (Kuklinski & Weinstein, 2000). Er zal, naast het enige onderzoek van Kuklinski en

Weinstein (2000), die alleen kijkt naar het vakgebied lezen, meer onderzoek gedaan moeten worden naar verschillen in de stabiliteit van de leerkrachtverwachtingen in verschillende leerjaren. Dit is van belang om meer te weten te komen over de stabiliteit van leerkrachtverwachtingen in de afzonderlijke leerjaren en voor verschillende vakgebieden in het onderwijs. Risico's ten aanzien van de prestaties van leerlingen, die door een mogelijke self-fulfilling prophecy van leerkrachten ontstaan, kunnen dan verkleind worden doordat leerkrachten hun verwachtingen van leerlingen van een bepaald leerjaar kunnen aanpassen als dat nodig blijkt uit de resultaten (Jussim & Harber, 2005).

### **Stabiliteit van de Leerkrachtverwachtingen Bij het Gender van Leerlingen**

Waarschijnlijk worden de verwachtingen van leerkrachten deels gevormd door de stereotypering die leerkrachten hebben over het karakter en demografische kenmerken van een gender van leerlingen (Muntoni & Retelsdorf, 2018; Robinson-Cimpian et al., 2014). In dit onderzoek zal gender dan ook worden gezien als een sociaal-cultureel gevormd verschil, vanwege de verschillende leerkrachtverwachtingen over jongens en meisjes, die ontstaan door stereotypingen over groepen (Muehlenhard & Peterson, 2011).

Over de mogelijke verschillen van het gender van leerlingen op de leerkrachtverwachtingen is, volgens de reviewstudie van Wang et al. (2018), momenteel nog geen eenduidige conclusie te trekken door de tegenstrijdige resultaten uit verschillende onderzoeken. Onderzoeken zonder een specifieke focus op een vakgebied in het onderwijs laten zien dat leerkrachten meestal hogere verwachtingen hebben voor meisjes dan voor jongens (e.g. De Boer et al., 2010; Timmermans et al., 2015, 2016). Maar er zijn ook onderzoeken die geen verschillen van gender van leerlingen op de vorming van leerkrachtverwachtingen vinden (e.g. Fitzpatrick et al., 2016; Van Den Bergh et al., 2010). Voor specifieke vakgebieden is er bewijs uit onderzoek dat leerkrachten hogere verwachtingen hebben voor jongens bij wiskunde (e.g. Holder & Kessels, 2017; Rubie-Davies

& Peterson, 2016) en voor meisjes bij lezen (e.g. Hinnant et al., 2009; Hornstra et al., 2010; Meissel et al., 2017). Dit verschil zou verklaard kunnen worden doordat het beeld van leerkrachten over de motivatie en de gewoonten van wiskunde bij jongens vaak positiever is dan bij meisjes, wat waarschijnlijk zorgt voor hogere leerkrachtverwachtingen van wiskunde bij jongens (Robinson-Cimpian et al., 2014). Daarentegen hebben leerkrachten bij meisjes een positiever beeld over de motivatie en de gewoonten bij lezen, wat waarschijnlijk zorgt voor hogere leerkrachtverwachtingen van lezen bij meisjes (Gentrup, et al., 2018; Timmermans et al., 2016).

Onderzoek naar het verschil in de stabiliteit van de leerkrachtverwachtingen tussen het gender van leerlingen is nog niet eerder onderzocht. Dit is wel van belang aangezien alle leerlingen gelijke kansen zouden moeten krijgen die niet beïnvloed zouden mogen worden door de self-fulfilling prophecy die een leerkracht van een bepaald gender heeft (Rosenthal & Jacobson, 1968).

### **De Huidige Studie**

Aangezien er nauwelijks onderzoek is gedaan of de stabiliteit van de leerkrachtverwachtingen gedurende één schooljaar verschillend is voor verschillende leerjaren of voor het gender van de leerlingen en waarbij ook de academische prestaties van de leerlingen worden meegenomen, zijn de volgende onderzoeksvragen opgesteld:

1. In hoeverre blijven de leerkrachtverwachtingen over individuele leerlingen op het gebied van de wiskunde- of leesprestaties stabiel gedurende één schooljaar?
2. In hoeverre verschilt de mate van stabiliteit van de leerkrachtverwachtingen over individuele leerlingen gedurende één schooljaar ten aanzien van de toekomstige wiskunde- of leesprestaties **tussen verschillende leerjaren?**

3. In hoeverre verschilt de mate van stabiliteit van de leerkrachtverwachtingen over individuele leerlingen gedurende één schooljaar ten aanzien van de toekomstige wiskunde- of leesprestaties **tussen het gender van leerlingen?**
4. In hoeverre verschilt de mate van stabiliteit van de leerkrachtverwachtingen over individuele leerlingen gedurende één schooljaar ten aanzien van de toekomstige wiskunde- of leesprestaties **tussen het gender van leerlingen in verschillende leerjaren.**

## **Methode**

### **Onderzoeksdesign**

In dit onderzoek werd gebruik gemaakt van een bestaande dataset (Rubie-Davies et al., 2018; Timmermans et al., 2021). De data komen voort uit een drie jaar durend longitudinaal panelonderzoek, welke is uitgevoerd in Nieuw-Zeeland (Baumgartner, 2020). In dit onderzoek zal alleen gebruik worden gemaakt van de data uit het eerste jaar, die zijn verzameld in 2014. In het panelonderzoek zijn de leerkrachtverwachtingen en de academische resultaten van leerlingen uit leerjaar 3 tot en met 8 (6-13 jaar), voor zowel wiskunde als lezen, drie keer op dezelfde wijze gedurende elk schooljaar gemeten. Deze metingen zijn uitgevoerd in de maanden februari (begin van het schooljaar), juni (medio het schooljaar) en oktober (richting einde van het schooljaar).

### **Context**

In Nieuw-Zeelandse onderwijs (Ministry of Education., z.d.) gaan leerlingen naar de basisschool (*primary school*) voor leerjaar 1 tot en met 6 (5-10 jaar) en vervolgens naar de onderbouw van het voortgezet onderwijs (*middle school*) voor leerjaar 7 en 8 (11-12 jaar). Deze acht schooljaren zijn voor de leerlingen verplicht om te volgen. Daarna kunnen de leerlingen voor leerjaar 9 tot en met 13 naar de bovenbouw van het voortgezet onderwijs (*high school*), hier zitten de leerlingen met verschillende niveaus bij elkaar in de klas. Deze schooljaren zijn van belang voor toelating tot het hoger onderwijs.

De meeste leerlingen in Nieuw-Zeeland gaan naar de plaatselijke school. Door de overheid worden de scholen gerangschikt op een schaal van 1-10. In deze schaal staat '1' voor een school in een zeer arm gebied en '10' voor een school in een welvarend gebied. Deze rangschikking wordt onder andere gebaseerd op sociaaleconomische informatie van het geografische gebied waar een school staat, maar ook op basis van gegevens over het percentage gezinnen uit minderheidsgroepen in het gebied.

## **Participanten**

De scholen uit de steekproef die binnen één geografisch gebied vielen, zijn onderverdeeld in lage, midden en hoge sociaaleconomische scholen. Vervolgens zijn de scholen binnen deze drie sociaaleconomische categorieën willekeurig geselecteerd en uitgenodigd om deel te nemen aan het onderzoek. Op deze manier zijn leerlingen uit elke sociaaleconomische categorie verkregen en is de steekproef op dit gebied representatief (Timmermans et al., 2021). De leerkrachten van de twaalf participerende scholen boden zich vervolgens vrijwillig aan voor deelname aan het onderzoek.

Aan het onderzoek namen 2461 leerlingen (51,4% jongens) deel. De gemiddelde leeftijd van de leerlingen was 9.62 jaar ( $SD = 1,65$ ). Op de basisscholen zaten leerlingen in leerjaar 3 tot en met 6 (6-11 jaar). In de onderbouw van het voortgezet onderwijs zaten leerlingen in leerjaar 7 en 8 (10-13 jaar). Van de deelnemende leerlingen had 49,4% een Europees/Nieuw-Zeelandse achtergrond, 17,3% was Maori (de inheemse groep), 15,6% was Pasifica (afkomstig van een van de eilanden in de Stille Oceaan), 14,5% was Aziatisch en de resterende groep (overig) bestaat uit 3,3% van de leerlingen. Daarnaast namen 88 leerkrachten (28,0% man) deel aan het onderzoek. De ervaring van de leerkrachten varieerde van 1 tot 5 jaar (32,3%) tot meer dan 25 jaar (18,1%).

## **Instrumenten**

### ***Leerkrachtverwachtingen***

De verwachtingen van de leerkrachten op het gebied van wiskunde en lezen werden afzonderlijk van elkaar in de eerste maand van het schooljaar (februari) gemeten. De leerkrachten hadden hiervoor toegang tot informatie over de leerlingen van voorgaande leerkrachten, onder andere over de prestatieniveaus. Zelf hadden de leerkrachten nog geen gestandaardiseerde testen in hun groep uitgevoerd. De leerkrachten gebruikten hiervoor de eerdere wiskunde- en leesresultaten van de leerlingen.

De leerkrachten kregen een lijst van de leerlingen uit hun klas en beantwoorde voor elke leerling, voor zowel wiskunde als lezen, de vraag: ‘Welk niveau verwacht je dat deze leerling bereikt aan het eind van het schooljaar?’ Hierop konden de leerkrachten de leerlingen scoren van 1 = heel erg onder het gemiddelde tot 7 = heel erg boven het gemiddelde. Ditzelfde instrument werd ook afgenomen in het midden van het schooljaar (juni) en tegen het einde van het schooljaar (oktober). Tegen het einde van het schooljaar voorspelden de leerkrachten het niveau dat hun leerlingen zouden bereiken tegen het einde van het daaropvolgende schooljaar, wanneer de leerlingen een nieuwe leerkracht zouden krijgen. De leerkrachten gebruikten over het algemeen het gehele bereik van de Likert-schaal om hun verwachtingen voor hun individuele leerlingen aan te geven (Rubie-Davies et al., 2018; Timmermans et al., 2021).

Aangezien de leerkrachtverwachtingen werden gemeten met één vraag is er geen informatie omtrent de psychometrische kwaliteit van de metingen. Het gebruik van een 1-item instrument in onderzoek naar leerkrachtverwachtingen is vrij gebruikelijk (e.g. Rubie-Davies et al., 2018; Timmermans et al., 2021; Turner et al., 2015). Mogelijk is dit gebruikelijk vanwege pragmatische overwegingen om de belasting van leerkrachten beperkt te houden, aangezien ze de leerkrachtverwachtingen voor alle individuele leerlingen, afzonderlijk voor de vakken wiskunde en lezen, in moesten vullen.

### ***Leerling-prestaties***



Met behulp van de e-asTTle (Electronic Assessment Tools for Teaching and Learning) werden de prestaties van leerlingen op het gebied van wiskunde en lezen op dezelfde drie meetmomenten als de leerkrachtverwachtingen gemeten (E-asTTle, z.d.). In veel Nieuw-Zeelandse scholen worden deze gestandaardiseerde testen gebruikt. Met behulp van dit systeem konden de docenten en onderzoekers toetsen op het gebied van lezen en wiskunde maken door zowel de lengte, het niveau en de specifieke onderdelen van het curriculum te kiezen. De getoetste curriculum onderdelen bij wiskunde waren getalkennis, getalbegrip en algebra en bij lezen waren het processen en strategieën, ideeën en taalkenmerken (Rubie-Davies et al., 2018; Timmermans et al., 2021).

De toetsen in het onderzoek waren allemaal veertig minuten lang en de leerkrachten kozen het meest geschikte niveau voor verschillende leerlingen in de klas. Om ervoor te zorgen dat leerlingen een vergelijkbare score op hun toets behalen, ongeacht welke toets hun leerkracht heeft geselecteerd, werd de Item Response Theory (IRT) scoringsprocedures gebruikt om de items over toetsen met een verschillende moeilijkheid te kalibreren (Embretson & Reise, 2000). Op deze manier konden de scores van de leerlingen vergeleken worden tussen verschillende klassen, jaarniveaus en scholen.

De scores van de wiskundeprestaties lagen tussen de waarden 1194 en 1765. De scores van de leesprestaties lagen tussen de waarden 1165 en 1780. Hierbij geldt, hoe hoger de score van een leerling hoe beter de prestatie. Echter hebben deze scores in de verschillende leerjaren een hele andere betekenis. Een score van bijvoorbeeld 1460 zou voor een leerling uit leerjaar 4 gemiddeld zijn, terwijl dezelfde score in leerjaar 8 aangeeft dat er ver onder het gemiddelde gescoord is voor dat leerjaar. Om de scores tussen verschillende leerjaren te kunnen vergelijken zijn de nationale normscores, voor elk leerjaar en het betreffende semester, van de gestandaardiseerde scores van de leerlingen afgetrokken (Rubie-Davies et al., 2018; Timmermans et al., 2021).

## **Procedure**

Allereerst is er ethische goedkeuring van het onderzoek door de institutionele commissie gegeven en daarna heeft er een actieve consent procedure plaatsgevonden voor zowel leerkrachten als ouders. Vervolgens hebben de leerkrachten in februari, juni en oktober hun schattingen van de prestaties van hun leerlingen gedaan en rapporteerden ze ook het gender en het leerjaar van elke leerling. Tegelijkertijd maakten de leerlingen de e-asTTle toetsen. Nadat de toetsen bij de leerlingen waren afgenomen, werden ze teruggestuurd naar de onderzoekers en beoordeeld door de onderzoekers.

## **Analyse**

De data-analyse methodes van dit kwantitatieve onderzoek werden uitgevoerd in IBM SPSS Statistics 29.0 (IBM Corp, 2022). Allereerst is gekeken naar de missende waarden van de data en vervolgens is de data-analyse bij elke onderzoeksvraag beschreven.

### ***Missende Waarden***

Bij het analyseren van de data is rekening gehouden met missende waarden. Allereerst is gekeken of de data behoorde tot fase 1 van het onderzoek, alleen deze data werden gebruikt en meegenomen in de data-analyse. De data horend bij fase 2 ( $n = 77$ ) zijn om deze reden verwijderd. Daarnaast is één case verwijderd aangezien niet bekend was bij welke fase deze hoorde. Daarna is gekeken naar de ontbrekende gegevens van de leerkrachtverwachtingen van wiskunde en lezen voor alle meetmomenten. Van één leerkracht ontbrak al deze data, om deze reden zijn alle bijbehorende cases ( $n = 21$ ) verwijderd.

Vervolgens is er gekeken naar data van de academische resultaten van de leerlingen van zowel wiskunde als lezen voor alle meetmomenten. Van de in totaal 2.461 cases ontbrak een groot deel van de data, variërend van  $n = 566$  (wiskunde meetmoment 1) tot  $n = 965$  (wiskunde meetmoment 3). Op het moment dat alleen volledige cases geanalyseerd zouden worden, zou er maar 33,9% van alle cases overblijven. Voor het behoud van de data, is ervoor

gekozen de cases met missende waarden bij de academische resultaten te behouden. Dit betekent dat bij verschillende analyses de steekproefgrootte kan verschillen afhankelijk van de meetmomenten waarover wordt gesproken. Als laatste is er gekeken naar het leerjaar en gender van de leerlingen. Bij deze variabelen waren geen missende waarden.

### ***Onderzoeksvragen***

Om de bij de eerste onderzoeksvraag te onderzoeken of de leerkrachtverwachtingen van individuele leerlingen op het gebied van de wiskunde- of leesprestaties stabiel blijven gedurende een schooljaar, zijn de data-analyseprocedures van Hinnant et al. (2009) en Wang et al. (2020) gebruikt. Om de stabiliteit van de leerkrachtverwachtingen bij wiskunde en lezen te berekenen, is een afzonderlijke regressie-analyse uitgevoerd voor elk van de drie meetmomenten, en afzonderlijk voor wiskunde en lezen. Hierbij zijn de leerkrachtverwachtingen meegenomen als afhankelijke variabele en de academische resultaten van de leerlingen op hetzelfde meetmoment als onafhankelijke variabele. De zes regressieanalyses creëerde zes ongestandaardiseerde residuen voor elke leerling. Een residu kan worden geïnterpreteerd als de mate waarin leerkrachten op een specifiek moment een leerling onderschat (negatief residu), overschat (positief residu), of nauwkeurig voorspelt (residu dicht bij nul).

Bij het checken van de assumpties bleek dat de residuen continu waren. Daarnaast was er een lineair verband tussen de leerkrachtverwachtingen en de academische prestaties van de leerlingen. Ook kan er vanwege de grote hoeveelheid cases vanuit gegaan worden dat de data normaal verdeeld is. Dit wordt ook wel de centrale limiet theorie genoemd (Pólya, 1920). Als laatste is er een sensitiviteitsanalyse gedaan door de uitschieters te checken. Het aantal uitschieters varieerde van  $n = 1$  (wiskunde meetmoment 3) tot  $n = 12$  (lezen meetmoment 3). Er waren geen grote verschillen te zien tussen de correlaties wanneer de uitschieters wel of

niet behouden werden in de dataset (zie bijlage 1). Om deze reden zijn de uitschieters niet verwijderd.

Aangezien geconcludeerd kan worden dat de assumpties niet zijn geschonden, zijn er meerdere Pearson-correlatieanalyse uitgevoerd met de residuen van de leerkrachtverwachtingen van de drie meetmomenten voor zowel wiskunde als lezen. Op deze manier kan de consistentie van de residuen over de drie meetmomenten binnen één schooljaar bepaald worden. De correlatiecoëfficiënten kunnen op verschillende manieren geïnterpreteerd worden (Schober et al., 2018). Een positieve correlatie laat zien dat hoge waarden op de ene variabele samengaan met hoge waarden op de andere variabele. Voor de interpretatie van de Pearson-correlatiecoëfficiënt wordt het volgende aangehouden: een sterke stabiliteit ( $r \geq 0.7$ ), een middelmatige stabiliteit ( $0.7 > r \geq 0.5$ ), een zwakke stabiliteit ( $0.5 > r \geq 0.3$ ) en nauwelijks tot geen stabiliteit ( $0.3 > r \geq -0.3$ ) (Wang et al., 2020).

Om bij de tweede onderzoeksvraag te bekijken in hoeverre de stabiliteit van de leerkrachtverwachtingen gedurende een schooljaar voor individuele leerlingen ten aanzien van de toekomstige wiskunde- of leesprestaties tussen de leerjaren verschilt, zal eveneens meerdere Pearson-correlatieanalyse worden uitgevoerd. Hierbij zal gekeken worden naar de correlatie tussen de drie verschillende meetmomenten zowel voor wiskunde als lezen, die is uitgesplitst in de verschillende leerjaren. Om de correlaties van de verschillende leerjaren met elkaar te kunnen vergelijken zullen meerdere Fisher's Z-testen worden uitgevoerd (Lenhard & Lenhard, 2014). Deze zorgen ervoor dat de correlaties met verschillende steekproefgroottes met elkaar vergeleken kunnen worden. De verkregen p-waardes laten vervolgens zien of er significante verschillen zijn tussen de correlaties van de verschillende meetmomenten.

Bij onderzoeksvragen drie en vier zal dezelfde manier van het analyseren van de data worden gebruikt als bij de tweede onderzoeksvraag. Bij onderzoeksvraag drie zal leerjaar

vervangen worden door gender. En bij de laatste onderzoekvraag zal de correlatie worden uitgesplitst in verschillende leerjaren en daarnaast ook het gender van de leerlingen.

## Resultaten

### Regressiemodellen

Zes regressieanalyses zijn gebruikt om de leerkrachtverwachtingen te voorspellen op basis van de academische resultaten van de leerlingen. Hieruit blijkt dat de academische resultaten van de leerlingen, de leerkrachtverwachtingen voor alle meetmomenten van wiskunde en lezen, significant voorspelde (zie Tabel 1). De verbanden tussen deze twee variabelen, voor alle meetmomenten van wiskunde en lezen, zijn positief. De sterkte van de verbanden varieerden van nauwelijks tot geen sterkte, bijvoorbeeld bij wiskunde meetmoment drie  $R^2 = .090$ ,  $F(1,1494) = 148.14$ ,  $p < .001$  tot een zwakke sterkte, bijvoorbeeld bij lezen meetmoment drie  $R^2 = .452$ ,  $F(1,1502) = 1230.45$ ,  $p < .001$ .

**Tabel 1**

*Regressieanalyse Voorspelling Leerkrachtverwachtingen op basis van Academische Resultaten Leerlingen*

Meetmomenten	<i>B</i>	<i>SE</i>	95% CI		$R^2$
			<i>LL</i>	<i>UL</i>	
Wiskunde	1	.009*	.000	.010	.375
	2	.011*	.000	.010	.276
	3	.005*	.000	.004	.005
Lezen	1	.008*	.000	.008	.259
	2	.010*	.000	.010	.402
	3	.011*	.000	.011	.012

*Noot.* \* $p < .001$ . *B* = ongestandaardiseerde coëfficiënten horend bij de voorspeller academische resultaten.

## Stabiliteit van de Leerkrachtverwachtingen Wiskunde en Lezen Gedurende Eén

### Schooljaar

De residuen die zijn gecreëerd bij de regressieanalyse zijn gebruikt om correlaties tussen de leerkrachtverwachtingen van de drie meetmomenten te berekenen. In Tabel 2 is te zien dat de leerkrachtverwachtingen van leerlingen op het gebied van wiskunde zwak tot sterk gecorreleerd waren. Deze varieerden van  $r(1594) = .446, p < .001$  tussen meetmomenten één en twee tot  $r(1333) = .709, p < .001$  tussen meetmomenten twee en drie. Op het gebied van lezen waren de leerkrachtverwachtingen middelmatig tot sterk gecorreleerd. Deze varieerden van  $r(1431) = .544, p < .001$  tussen meetmomenten één en twee tot  $r(1253) = .801, p < .001$  tussen meetmomenten twee en drie.

Bij de vergelijking van de correlaties tussen de verschillende meetmomenten, met behulp van de Fisher's Z-testen (Tabel 2), valt op dat, van zowel wiskunde als lezen, de correlaties tussen meetmomenten twee en drie significant hoger zijn dan tussen meetmomenten één en twee. Dit duidt op meer instabiliteit van de leerkrachtverwachtingen tussen meetmomenten één en twee. Daarnaast laat Tabel 2 zien dat alleen de correlatie bij lezen ( $z = -0.895, p = .185$ ) tussen meetmomenten één en twee en meetmomenten één en drie niet significant van elkaar verschillen.

**Tabel 2**

*Correlaties Leerkrachtverwachtingen voor Wiskunde en Lezen en Z-scores*

	Meetmomenten	<i>r</i>	<i>n</i>	Meetmomenten	<i>z</i>	<i>p</i>
Wiskunde	1, 2	.446*	1596	1, 3	-3.387	.000
				2, 3	-10.921	.000
	1, 3	.542*	1278	2, 3	-7.100	.000
	2, 3	.709*	1335			
Lezen	1, 2	.544*	1433	1, 3	-0.895	.185
				2, 3	-12.701	.000
	1, 3	.568*	1239	2, 3	-11.393	.000

---

2, 3	.801*	1255
------	-------	------

---

*Noot.* \* $p < .001$ . *P*-waardes dikgedrukt van correlaties die significant van elkaar verschillen.

## **Verschil Mate van Stabiliteit Leerkrachtverwachtingen Tussen het Leerjaar van Leerlingen**

### ***Wiskunde***

Een overzicht van de correlaties tussen de leerkrachtverwachtingen van de drie meetmomenten van wiskunde, uitgesplitst in het leerjaar van leerlingen, is te vinden in Tabel 3. Alleen de correlatie van leerjaar 3 tussen meetmomenten één en twee was niet significant verschillend van nul  $r(64) = .218, p = .078$ . Verder is te zien dat de leerkrachtverwachtingen van leerlingen nauwelijks tot sterk gecorreleerd waren. Deze varieerden van  $r(64) = .218, p < .001$  tussen meetmomenten één en twee tot  $r(218) = .770, p < .001$  tussen meetmomenten twee en drie.

De vergelijking van de correlaties tussen de verschillende leerjaren, met behulp van de Fisher's Z-testen, duidt op enkele significante verschillen tussen de correlaties van de verschillende meetmomenten. Tabel 3 laat zien dat de correlaties tussen meetmomenten één en twee van alle leerjaren, in combinatie met leerjaar 8, significant zijn. Het significantieniveau varieerde  $z = -3.154, p = .001$  tot  $z = -1.778, p = .038$ . Dit duidt erop dat de stabiliteit van de leerkrachtverwachtingen in leerjaar 8, tussen meetmomenten één en twee, significant hoger is dan de stabiliteit van de leerkrachtverwachtingen uit de andere leerjaren. Dit patroon is niet te zien als je de stabiliteit van de leerkrachtverwachtingen van meetmomenten één en twee van andere leerjaren met elkaar vergelijkt. Daarentegen is er bij de correlaties tussen meetmomenten twee en drie helemaal geen patroon tussen de leerjaren te vinden. Dit duidt erop dat de stabiliteit van de leerkrachtverwachtingen in groep 8 tussen

meetmomenten twee en drie niet meer significant verschilt met de leerkrachtverwachtingen uit de andere leerjaren.

**Tabel 3**

*Correlaties Leerkrachtverwachtingen Wiskunde uitgesplitst in Leerjaar en Z-scores*

Meetmomenten	Leerjaar	<i>r</i>	<i>n</i>	leerjaar	<i>z</i>	<i>p</i>
1,2	3	.218	66	4	-1.778	<b>.038</b>
				5	-1.495	.067
				6	-1.376	.084
				7	-2.062	<b>.020</b>
				8	-2.950	<b>.002</b>
	4	.438*	279	5	0.498	.309
				6	0.727	.234
				7	-0.461	.322
				8	-2.286	<b>.011</b>
	5	.404*	308	6	0.225	.411
				7	-0.974	.165
				8	-2.862	<b>.002</b>
	6	.389*	337	7	-1.215	.112
				8	-3.154	<b>.001</b>
	7	.469*	286	8	-1.824	<b>.034</b>
	8	.577*	320			
1,3	3	.356**	67	4	-1.986	<b>.024</b>
				5	-1.117	.132
				6	-1.216	.112
				7	-2.427	<b>.008</b>
				8	-2.775	<b>.003</b>
	4	.574*	229	5	1.376	.084
				6	1.245	.107
				7	-0.718	.236
				8	-1.168	.121
	5	.484*	262	6	-0.151	.440
				7	-2.035	<b>.021</b>
				8	-2.634	<b>.004</b>
	6	.494*	274	7	-1.912	<b>.028</b>
				8	-2.506	<b>.006</b>
	7	.620*	185	8	-0.360	.359
	8	.641*	261			
2,3	3	.633*	66	4	-0.070	.472
				5	-1.279	.101
				6	-0.769	.221
				7	-1.914	<b>.028</b>
				8	-1.374	.085
	4	.639*	220	5	-1.835	<b>.033</b>



Meetmomenten	Leerjaar	<i>r</i>	<i>n</i>	leerjaar	<i>z</i>	<i>p</i>
				6	-1.074	.141
				7	-2.748	<b>.003</b>
				8	-2.015	<b>.022</b>
	5	.729*	254	6	0.838	.201
				7	-1.011	.156
				8	-0.125	.450
	6	.693*	282	7	-1.841	<b>.033</b>
				8	-0.997	.159
	7	.770*	220	8	0.925	.178
	8	.734*	293			

*Noot.* \* $p < .001$ . \*\* $p < .01$ . *P*-waardes dikgedrukt van correlaties die significant van elkaar verschillen.

### **Lezen**

Een overzicht van de correlaties tussen de leerkrachtverwachtingen van de drie meetmomenten van lezen, uitgesplitst in het leerjaar van leerlingen, is te vinden in Tabel 4. Hier is te zien dat de leerkrachtverwachtingen van leerlingen zwak tot sterk gecorreleerd waren en varieerden van  $r(54) = .404$ ,  $p = .002$  tussen meetmomenten één en twee tot  $r(279) = .811$ ,  $p < .001$  tussen meetmomenten twee en drie.

De vergelijking van de correlaties tussen de verschillende leerjaren, met behulp van de Fisher's Z-testen (Tabel 4), duidt op enkele significante verschillen tussen de correlaties van de verschillende meetmomenten waarin geen duidelijk patroon te vinden is. Opvallend is dat alle correlaties tussen meetmomenten twee en drie, op leerjaar 3 in combinatie met leerjaar 6 na ( $z = -1.652$ ,  $p = .049$ ), niet significant van elkaar verschillen. Dit duidt erop dat de stabiliteit van de leerkrachtverwachtingen tussen meetmomenten twee en drie in alle leerjaren nagenoeg gelijk zijn.

### **Tabel 4**

*Correlaties Leerkrachtverwachtingen Lezen uitgesplitst in Leerjaar en Z-scores*

Meetmomenten	Leerjaar	<i>r</i>	<i>n</i>	leerjaar	<i>z</i>	<i>p</i>
1,2	3	.404**	56	4	-1.012	.156
				5	-1.391	.082
				6	-1.452	.073
				7	-0.311	.378
				8	-1.658	<b>.049</b>
	4	.524*	223	5	-2.015	<b>.022</b>
				6	-2.168	<b>.015</b>
				7	-0.249	.402
				8	-2.499	<b>.006</b>
	5	.564*	254	6	-0.069	.473
				7	1.915	<b>.028</b>
				8	-0.434	.332
	6	.568*	306	7	2.084	<b>.019</b>
				8	-0.384	.351
7	.442*	302	8	-2.187	<b>.014</b>	
8	.589*	292				
1,3	3	.478*	54	4	-0.931	.176
				5	-1.734	<b>.041</b>
				6	-1.336	.091
				7	0.108	.457
				8	-0.385	.350
	4	.582*	216	5	-1.296	.097
				6	-0.648	.259
				7	1.681	<b>.046</b>
				8	0.947	.172
	5	.658*	229	6	0.674	.250
				7	3.012	<b>.001</b>
				8	2.341	<b>.010</b>
	6	.621*	240	7	2.379	.009
				8	1.665	<b>.048</b>
7	.465*	222	8	-0.833	.203	
8	.522*	278				
2,3	3	.716*	66	4	-0.743	.229
				5	-1.297	.097
				6	-1.652	<b>.049</b>
				7	-0.846	.199
				8	-1.610	.054
	4	.765*	183	5	-0.763	.223
				6	-1.273	.102
				7	-0.121	.452
				8	-1.211	.113
	5	.795*	222	6	-0.499	.309
				7	0.676	.249
				8	-0.434	.332
	6	.811*	281	7	1.215	.112
				8	0.069	.473
7	.770*	223	8	-1.149	.125	

Meetmomenten	Leerjaar	<i>r</i>	<i>n</i>	leerjaar	<i>z</i>	<i>p</i>
	8	.809*	280			

*Noot.* \* $p < .001$ . \*\* $p < .01$ . *P*-waarden dikgedrukt van correlaties die significant van elkaar verschillen.

### **Verschil Mate van Stabiliteit Leerkrachtverwachtingen Wiskunde en Lezen Tussen het Gender van Leerlingen**

Een overzicht van de correlaties tussen de leerkrachtverwachtingen van de drie meetmomenten, uitgesplitst in het gender van leerlingen, is te vinden in Tabel 5. De leerkrachtverwachtingen van de jongens op het gebied van wiskunde waren zwak tot middelmatig gecorreleerd. Deze varieerden van  $r(791) = .439, p < .001$  tussen meetmomenten één en twee tot  $r(660) = .696, p < .001$  tussen meetmomenten twee en drie. Op het gebied van wiskunde voor de meisjes waren de correlaties zwak tot sterk gecorreleerd. Deze varieerden van  $r(801) = .454, p < .001$  tussen meetmomenten één en twee tot  $r(671) = .724, p < .001$  tussen meetmomenten twee en drie.

De leerkrachtverwachtingen op het gebied van lezen waren voor zowel de jongens als de meisjes middelmatig tot sterk gecorreleerd. Voor de jongens variërend van  $r(690) = .535, p < .001$  tussen meetmomenten één en twee tot  $r(613) = .792, p < .001$  tussen meetmomenten twee en drie. En voor de meisjes variërend van  $r(739) = .552, p < .001$  tussen meetmomenten één en twee tot  $r(638) = .812, p < .001$  tussen meetmomenten twee en drie.

De vergelijking van de correlaties tussen jongens en meisjes, met behulp van de Fisher's *Z*-testen, duidt erop dat er geen significante verschillen zijn gevonden tussen de correlaties bij alle verschillende meetmomenten. Dit geldt zowel voor wiskunde als lezen, wat te zien is aan de *p*-waarden die varieerde van  $z = -1.035, p = .150$  tot  $z = -0.374, p = .354$  (Tabel 5). Dit duidt erop dat de stabiliteit van de leerkrachtverwachtingen bij alle meetmomenten, geen significant verschil laat zien tussen jongens en meisjes.

**Tabel 5**

*Correlaties Leerkrachtverwachtingen Wiskunde en Lezen uitgesplitst in Gender en Z-scores*

Meetmomenten	Jongens		Meisjes		<i>z</i>	<i>p</i>	
	<i>r</i>	<i>n</i>	<i>r</i>	<i>n</i>			
Wiskunde	1, 2	.439*	793	.454*	803	-0.374	.354
	1, 3	.523*	628	.561*	650	-0.960	.169
	2, 3	.696*	662	.724*	673	-1.030	.151
Lezen	1, 2	.535*	692	.552*	741	-0.455	.324
	1, 3	.547*	594	.587*	645	-1.035	.150
	2, 3	.792*	615	.812*	640	-0.991	.161

*Noot.* \* $p < .001$ .

### **Verschil Mate van Stabiliteit Leerkrachtverwachtingen Tussen het Gender van Leerlingen Uitgesplitst in Verschillende Leerjaren**

#### ***Wiskunde***

Een overzicht van de correlaties tussen de leerkrachtverwachtingen van de drie meetmomenten van wiskunde, tussen het gender van leerlingen uitgesplitst in verschillende leerjaren, is te vinden in Tabel 6. Hier is te zien dat de leerkrachtverwachtingen van leerlingen nauwelijks tot sterk gecorreleerd waren en varieerden van  $r(31) = .157, p = .383$  tussen meetmomenten één en twee tot  $r(147) = .745, p < .001$  tussen meetmomenten twee en drie.

De vergelijking van de correlaties tussen jongens en meisjes in verschillende leerjaren, met behulp van de Fisher's Z-testen (Tabel 6), duidt op significante verschillen tussen de correlaties van jongens en meisjes van leerjaar 6 ( $z = -2.642, p = .004$ ) en 7 ( $z = -1.768, p = .039$ ) tussen meetmomenten één en twee. Ook is er een significant verschil gevonden tussen de correlaties van jongens en meisjes van leerjaar 6 ( $z = -1.928, p = .027$ ) tussen meetmomenten twee en drie.

**Tabel 6**

*Correlaties Leerkrachtverwachtingen Wiskunde uitgesplitst in Gender en Leerjaar en Z-scores*

Meetmomenten	Leerjaar	Jongens		Meisjes		z	p
		r	n	r	n		
1,2	3	.157****	33	.256****	33	-0.401	.344
	4	.510*	136	.358*	143	1.554	.060
	5	.416*	160	.394*	148	0.229	.410
	6	.262*	167	.507*	170	-2.642	<b>.004</b>
	7	.428*	136	.515*	150	-1.768	<b>.039</b>
	8	.582*	161	.572*	159	0.133	.447
1,3	3	.281****	40	.454****	27	-0.767	.222
	4	.598*	114	.540*	115	0.641	.261
	5	.460*	126	.507*	136	-0.491	.312
	6	.435*	133	.551*	141	-1.258	.104
	7	.558*	88	.673*	113	-1.290	.099
	8	.669*	127	.612*	134	0.773	.220
2,3	3	.739*	36	.516 **	30	1.454	.073
	4	.595*	107	.681*	113	-1.065	.144
	5	.711*	126	.747*	128	0.446	.328
	6	.631*	137	.751*	145	-1.928	<b>.027</b>
	7	.741*	107	.800*	113	-1.067	.143
	8	.745*	149	.722*	144	0.422	.337

*Noot.* \* $p < .001$ . \*\* $p < .01$ . \*\*\* $p < .05$ . \*\*\*\* $p > .05$ . *P*-waardes dikgedrukt van correlaties die significant van elkaar verschillen.

### **Lezen**

Een overzicht van de correlaties tussen de leerkrachtverwachtingen van de drie meetmomenten van lezen, tussen het gender van leerlingen uitgesplitst in verschillende leerjaren, is te vinden in Tabel 7. Hier is te zien dat de leerkrachtverwachtingen van leerlingen nauwelijks tot sterk gecorreleerd waren en varieerden van  $r(26) = .281$ ,  $p = .147$  tussen meetmomenten één en twee tot  $r(107) = .833$ ,  $p < .001$  tussen meetmomenten twee en drie.

De vergelijking van de correlaties tussen jongens en meisjes in verschillende leerjaren, met behulp van de Fisher's Z-testen (Tabel 7), duidt op significante verschillen tussen enkele de correlaties. Er is een significant verschil tussen meetmomenten één en twee bij leerjaar 5 ( $z = 1.921, p = .027$ ) en leerjaar 7 ( $z = -2.196, p = .014$ ). Ook is er een significant verschil tussen meetmomenten één en drie bij leerjaar 5 ( $z = 1.885, p = .030$ ) en leerjaar 7 ( $z = -2.402, p = .008$ ). Als laatste is er een significant verschil gevonden tussen de correlaties van jongens en meisjes tussen meetmomenten twee en drie van leerjaar 4 ( $z = -1.946, p = .026$ ).

**Tabel 7**

*Correlaties Leerkrachtverwachtingen Lezen uitgesplitst in Gender en Leerjaar en Z-scores*

Meetmomenten	Leerjaar	Jongens		Meisjes		z	p
		r	n	r	n		
1,2	3	.281****	28	.570**	28	-1.268	.102
	4	.580*	100	.471*	123	1.107	.134
	5	.636*	128	.468*	126	1.921	<b>.027</b>
	6	.511*	149	.623*	157	-1.435	.076
	7	.339*	146	.543*	156	-2.196	<b>.014</b>
	8	.589*	141	.576*	151	0.166	.434
1,3	3	.662*	30	.414***	24	1.223	.111
	4	.590*	104	.558*	112	0.346	.365
	5	.728*	108	.586*	121	1.885	<b>.030</b>
	6	.556*	112	.667*	128	-1.361	.087
	7	.341*	108	.593*	114	-2.402	<b>.008</b>
	8	.475*	132	.549*	146	-0.827	.204
2,3	3	.672*	36	.784*	30	-0.930	.176
	4	.711*	93	.828*	90	-1.946	<b>.026</b>
	5	.833*	109	.756*	113	1.551	.060
	6	.809*	132	.813*	149	-0.097	.461
	7	.736*	111	.807*	112	-1.301	.097
	8	.815*	134	.803*	146	0.287	.387

*Noot.* \* $p < .001$ . \*\* $p < .01$ . \*\*\* $p < .05$ . \*\*\*\* $p > .05$ . P-waardes dikgedrukt van correlaties die significant van elkaar verschillen.

## Conclusie en Discussie

In dit onderzoek is onderzocht in hoeverre de verwachtingen van de leerkrachten voor wiskunde en lezen stabiel blijven gedurende één schooljaar door de correlaties tussen drie verschillende meetmomenten in één schooljaar met elkaar te vergelijken. Ook is voor wiskunde en lezen onderzocht in hoeverre er een verschil is in de mate van de stabiliteit van de leerkrachtverwachtingen tussen het leerjaar van een leerling, het gender van een leerling en de combinatie van deze twee factoren. Hierbij is gebruik gemaakt van Fisher's Z-testen waardoor correlaties uit verschillende meetmomenten en groepen met elkaar vergeleken konden worden.

Bij de eerste onderzoeksvraag is onderzocht in hoeverre de leerkrachtverwachtingen over individuele leerlingen op het gebied van wiskunde- of leesprestaties stabiel bleven gedurende één schooljaar. De resultaten uit dit onderzoek lieten zien dat de leerkrachtverwachtingen over individuele leerlingen gedurende één schooljaar voor wiskunde een zwakke tot sterke stabiliteit hadden. Voor lezen is er een middelmatige tot sterke stabiliteit van de leerkrachtverwachtingen over individuele leerlingen gedurende één schooljaar gevonden. Hierbij viel op dat de stabiliteit van de leerkrachtverwachtingen aan het einde van het schooljaar, voor zowel wiskunde als lezen, sterker was dan aan het begin van het schooljaar. Dit duidt op meer instabiliteit van de leerkrachtverwachtingen aan het begin van het schooljaar. Deze bevinding lijkt sterk op de resultaten uit de studies van Kuklinski en Weinstein (2000), Wang et al. (2020), Timmermans et al. (2021) en Hao et al., (2022) waarbij in het tweede deel van het schooljaar de veranderingen van de leerkrachtverwachtingen kleiner waren dan in het eerste deel van het schooljaar. De verwachtingen aan begin van het schooljaar, zouden mogelijk gebaseerd kunnen zijn op ervaringen van leerkrachten die de leerlingen eerder in de klas hebben gehad (Ready & Wright, 2011). Aan het eind van het schooljaar zijn de leerkrachtverwachtingen waarschijnlijk bijgestuurd door nieuwe en betere

informatie over de leerlingen (Brophy, 1983). Dit doordat leerkrachten zich waarschijnlijk bewust zijn geworden van eerdere onnauwkeurige verwachtingen over de werkelijke capaciteiten van hun leerlingen op het moment dat ze hun leerlingen beter leerden kennen (Hao et al., 2022).

Bij de tweede onderzoeksvraag is onderzocht in hoeverre de mate van stabiliteit van de leerkrachtverwachtingen over individuele leerlingen gedurende één schooljaar ten aanzien van de toekomstige wiskunde- of leesprestaties verschilt tussen verschillende leerjaren. Uit de resultaten van dit onderzoek blijkt dat er geen duidelijk verband zichtbaar is tussen het leerjaar van leerlingen en de stabiliteit van de leerkrachtverwachtingen. Toch is opvallend dat de stabiliteit van de leerkrachtverwachtingen van wiskunde in leerjaar 8, aan het begin van het schooljaar, significant hoger zijn dan de stabiliteit van de leerkrachtverwachtingen uit de andere leerjaren. Deze bevinding sluit aan bij het onderzoek van Kuklinski en Weinstein (2000) waarbij de leerkrachtverwachtingen van leerlingen stabiel werden in hogere leerjaren. Echter is het patroon van een hogere stabiliteit van de leerkrachtverwachtingen in dit onderzoek niet meer aanwezig bij wiskunde aan het eind van het schooljaar en helemaal niet bij het vak lezen. Niet alle resultaten uit dit onderzoek lieten dus zien dat de leerkrachtverwachtingen stabiel worden in hogere leerjaren, hier kan dan ook geen algemene conclusie over getrokken worden. De stabiele leerkrachtverwachtingen bij wiskunde aan het begin van het schooljaar in leerjaar 8 zouden verklaard kunnen worden door de gedachte dat leerkrachten gebruik kunnen maken van meer schoolgegevens die eerder zijn verzameld dan in lagere leerjaren (Kuklinski & Weinstein, 2000). Ook zouden leerlingen zich in hogere leerjaren consequenter kunnen gedragen naar de verwachtingen van de leerkracht (Kuklinski & Weinstein, 2000). Vervolgonderzoek, waarbij onderzocht wordt wat de onderliggende reden is van mogelijke stabielere leerkrachtverwachtingen in hogere leerjaren, is van belang om leerkrachten aan de hand van de verkregen resultaten bewust te maken



waarom ze verwachtingen van bepaalde leerjaren wel of niet aanpassen. Een kwalitatieve studie waarbij onderzocht wordt waarom leerkrachten hun verwachtingen veranderen of behouden op basis van informatie over de leerlingen kan helpend zijn om de kwantitatieve resultaten uit deze studie te begrijpen. Deze kennis over de onderliggende reden kan leerkrachten vervolgens helpen om de verwachtingen over leerlingen nauwkeuriger te maken, hierdoor kan de kans kleiner worden dat er een self-fulfilling prophecy voor een bepaald leerjaar ontstaat.

Bij de derde onderzoeksvraag is onderzocht in hoeverre de mate van stabiliteit van de leerkrachtverwachtingen over individuele leerlingen gedurende één schooljaar ten aanzien van de toekomstige wiskunde- of leesprestaties verschilt tussen het gender van leerlingen. De resultaten van dit onderzoek lieten zien dat er geen significante verschillen zijn gevonden in de mate van de stabiliteit van de leerkrachtverwachtingen tussen het gender van leerlingen. Bij dit onderzoek naar de stabiliteit van de leerkrachtverwachtingen worden verschillen in de verwachtingen die leerkrachten kunnen hebben over verschillende genders van leerlingen, op basis van de stereotypering over een bepaald gender, niet zichtbaar (Muntoni & Retelsdorf, 2018; Robinson-Cimpian et al., 2014).

Bij de laatste onderzoeksvraag is onderzocht in hoeverre de mate van stabiliteit van de leerkrachtverwachtingen over individuele leerlingen gedurende één schooljaar ten aanzien van de toekomstige wiskunde- of leesprestaties verschilt tussen het gender van leerlingen in verschillende leerjaren. De resultaten uit dit onderzoek lieten, in tegenstelling tot de onderzoeksvraag waarbij het gender niet uitgesplitst wordt in verschillende leerjaren, zien dat er zowel bij wiskunde als lezen enkele verschillen waren in de mate van stabiliteit van de leerkrachtverwachtingen tussen het gender van leerlingen in verschillende leerjaren. Doordat er geen duidelijk patroon zichtbaar is tussen deze verschillen, kan niet geconcludeerd worden

of er in bepaalde leerjaren verschillen zijn in de mate van de stabiliteit van de leerkrachtverwachtingen tussen het gender van leerlingen.

### **Sterke Punten en Limitaties**

Dit onderzoek is een toevoeging aan het geringe onderzoek over de stabiliteit van leerkrachtverwachtingen dat er op dit moment is. Dit aangezien er gekeken is naar de twee vakgebieden (wiskunde en lezen) en naar de verschillen in de stabiliteit van de leerkrachtverwachtingen voor twee groepen (leerjaar en gender). Een ander sterk punt is dat er bij het onderzoeken van de stabiliteit van de leerkrachtverwachtingen rekening is gehouden met de academische resultaten van de leerlingen (Wang et al., 2020). De resultaten van dit onderzoek zijn om deze reden een toevoeging aan de vier onderzoeken waarbij deze variabele ook is meegenomen (Hao et al., 2022; Rubie-Davies et al., 2018; Timmermans et al., 2021; Wang et al., 2020). Ook maakt dit onderzoek gebruik van grotere steekproefgroottes van studenten en leerkrachten dan drie van de zeven eerdere uitgevoerde studies over de stabiliteit van de leerkrachtverwachtingen: Good et al. (1980), Martinek (1980) en Kuklinski en Weinstein (2000).

Ondanks dat dit onderzoek een toevoeging is aan het al bestaande onderzoek over de stabiliteit van de leerkrachtverwachtingen, is er nog veel meer onderzoek nodig. Onderzoek naar de stabiliteit van de leerkrachtverwachtingen waarbij gebruik wordt gemaakt van verschillende contexten, zoals verschillende leeftijden, landen en vakgebieden. Na het doen van meer onderzoek naar de stabiliteit van de leerkrachtverwachtingen kunnen de resultaten uiteindelijk gegeneraliseerd worden en in meerdere verschillende contexten ingezet worden in de praktijk. Het is van belang om de resultaten in te zetten in de praktijk aangezien het over- of onderschatten van leerlingen door leerkrachten invloed kan hebben op hoe leerlingen presteren, zelfs op de lange termijn (Hattie, 2008; Jussim & Harber, 2005; Rubie-Davies et al., 2018). Waar mogelijk kunnen leerkrachten aan de hand van de resultaten naar de stabiliteit

van de leerkrachtverwachtingen zich ook bewust worden (van de de vorming en impact van de verwachtingen) om de verwachtingen waar nodig aan te passen en zo de vooroordelen over (bepaalde groepen) leerlingen te verminderen en de leerling resultaten te verhogen (Yeager & Walton, 2011; Rubie-Davies, 2014)

In dit onderzoek werd gebruik gemaakt van een bestaande dataset waarbij ongeveer een derde deel van de data van de academische prestaties van de leerlingen miste. In vervolgonderzoek zal geprobeerd moeten worden om bij het verzamelen van de data meer volledige resultaten te verkrijgen. Dit door als onderzoekers nauw in contact te staan met de leerkrachten die de data aanleveren en ze erop te attenderen op het moment dat er data mist, zodat ze deze data alsnog kunnen aanleveren voor het onderzoek.

Ook is er tijdens het verzamelen van de data gebruik gemaakt van een relatief lange tijd (vier maanden) tussen de meetmomenten. Dit is net als in eerdere onderzoeken naar de stabiliteit van leerkrachtverwachtingen, waar acht weken tot zes maanden tussen de verschillende meetmomenten zit (Good et al., 1980; Hao et al., 2022; Kuklinski & Weinstein, 2000; Martinek, 1980; Rubie-Davies et al., 2018; Timmermans et al., 2021; Wang et al., 2020). Leerkrachtverwachtingen worden waarschijnlijk al in de eerste twee weken van het schooljaar gevormd en worden in de loop van het jaar stabiel (Raudenbush, 1984). Mogelijke aanpassingen van leerkrachtverwachtingen in de eerste weken van het schooljaar zijn mogelijk niet gedetecteerd vanwege de grote tijdspanne van vier maanden tussen de meetmomenten. Om deze mogelijke aanpassingen in de eerste weken van het schooljaar te kunnen onderzoeken is voor vervolgonderzoek een kortere tijdsspanne met meer meetmomenten tussen de verschillende metingen van de stabiliteit van de leerkrachten aan te raden (Hao et al., 2022).

Ook is er tijdens het verzamelen van de data voor deze studie gebruik gemaakt van één item om de leerkrachtverwachtingen te meten. Echter zorgt het gebruik van dezelfde vraag er

niet automatisch voor dat bij elke meting hetzelfde construct gemeten wordt (Seddig & Leitgöb, 2018). In de loop van de tijd kunnen leerkrachten dezelfde vraag op een andere manieren interpreteren en beantwoorden, wat impliceert dat het begrip is veranderd (e.g. Ferrer et al., 2008; Widaman et al., 2010). Het is van belang om na te gaan of leerkrachten de vraag die werd gesteld nog hetzelfde interpreterden voor alle verschillende meetmomenten. Vervolgonderzoek zou de één-item schaal kunnen valideren door gebruik te maken van een korte gestandaardiseerde schaal. Er kan mogelijk ook een andere manier van het meten van de leerkrachtverwachtingen gebruikt worden, waarbij niet één, maar meerdere vragen aan de leerkrachten worden gesteld om de leerkrachtverwachtingen op de verschillende meetmomenten te achterhalen. Bij het opstellen van meerdere vragen zal wel rekening gehouden moeten worden of leerkrachten de hoeveelheid vragen aankunnen om het onderzoek af te ronden voor alle leerlingen in de klas.

Al met al laat dit onderzoek zien dat de leerkrachtverwachtingen van individuele leerlingen voor zowel wiskunde als lezen stabiel blijven gedurende één schooljaar. Ook laat dit onderzoek zien dat de stabiliteit van de leerkrachtverwachtingen aan het einde van het schooljaar sterker waren dan aan het begin van het schooljaar. Ook is gebleken dat de mate van stabiliteit van de leerkrachtverwachtingen over individuele leerlingen gedurende één schooljaar ten aanzien van de toekomstige wiskunde- of leesprestaties niet verschilt tussen het leerjaar van leerlingen, het gender van leerlingen en de combinatie van deze twee factoren. Verder kwalitatief onderzoek, waarbij bijvoorbeeld inzicht verkregen wordt in de contextuele en ecologische factoren die de verwachtingen van leerkrachten beïnvloeden, zal gedaan kunnen worden om de achterliggende redenen te achterhalen waarom deze kwantitatieve studie deze resultaten laat zien.

## Referenties

- Baumgartner, S. E. (2020). Longitudinal data and panel data: Methods and analysis. *The International Encyclopedia of Media Psychology*, 1–5.  
<https://doi.org/10.1002/9781119011071.iemp0021>
- Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, 75(5), 631–661. <https://doi.org/10.1037/0022-0663.75.5.631>
- De Boer, H., Bosker, R., & Van Der Werf, G. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, 102(1), 168–179. <https://doi.org/10.1037/a0017289>
- Dusek, J. B., & Joseph, G. E. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology*, 75(3), 327–346. <https://doi.org/10.1037/0022-0663.75.3.327>
- E-asTTLe. (z.d.). *About E-asTTLe*. Geraadpleegd op 21 mei 2024, van <https://e-asttle.tki.org.nz/About-e-asTTLe>
- Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Erlbaum.
- Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology*, 4(1), 22–36.  
<https://doi.org/10.1027/1614-2241.4.1.22>
- Fitzpatrick, C., Côté-Lussier, C., & Blair, C. (2016). Dressed and groomed for success in elementary school: Student appearance and academic adjustment. *Elementary School Journal*, 117(1), 30–45. <https://doi.org/10.1086/687753>
- Gentrup, S., Rjosk, C., Stanat, P., & Lorenz, G. (2018). Einschätzungen der schulischen motivation und des arbeitsverhaltens durch grundschullehrkräfte und deren bedeutung für verzerrungen in leistungserwartungen [Beoordelingen van schoolmotivatie en

- werkgedrag door leerkrachten in het basisonderwijs en hun betekenis voor verstoringen in prestatieverwachtingen]. *Zeitschrift für Erziehungswissenschaft*, 21(4), 867–891. <https://doi.org/10.1007/s11618-018-0806-2>
- Good, T. L., & Brophy, J. E. (1997). *Looking in classrooms* (7de editie). Longman
- Good, T. L., Cooper, H., & Blakey, S. L. (1980). Classroom interaction as a function of teacher expectations, student sex, and time of year. *Journal of Educational Psychology*, 72(3), 378–385. <https://doi.org/10.1037/0022-0663.72.3.378>
- Good, T. L., Sterzinger, N. K., & Lavigne, A. L. (2018). Expectation effects: Pygmalion and the initial 20 years of research. *Educational Research and Evaluation*, 24(3–5), 99–123. <https://doi.org/10.1080/13803611.2018.1548817>
- Hao, L., Rubie-Davies, C. M., & Watson, P. (2022). Do teachers maintain their expectation bias for students? A longitudinal investigation. *Social Psychology of Education*, 25(4), 719–744. <https://doi.org/10.1007/s11218-022-09714-6>
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, 101(3), 662–670. <https://doi.org/10.1037/a0014306>
- Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social Psychology of Education*, 20(3), 471–490. <https://doi.org/10.1007/s11218-017-9384-z>
- Hornstra, L., Denessen, E., Bakker, J., Van Den Bergh, L., & Voeten, M. (2010). Teacher attitudes toward dyslexia: Effects on teacher expectations and the academic achievement of students with dyslexia. *Journal of Learning Disabilities*, 43(6), 515–529. <https://doi.org/10.1177/0022219409355479>

- IBM Corp. (2022). *IBM SPSS statistics for Windows* (Versie 29.0) [Computer programma].
- Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131–155. [https://doi.org/10.1207/s15327957pspr0902\\_3](https://doi.org/10.1207/s15327957pspr0902_3)
- Kuklinski, M., & Weinstein, R. (2000). Classroom and grade level differences in the stability of teacher expectations and perceived differential teacher treatment. *Learning Environments Research*, 3, 1-34. <https://doi.org/10.1023/A:1009904718353>
- Lenhard, W. & Lenhard, A. (2014). *Hypothesis tests for comparing correlations*. *Psychometrica*. <https://www.psychometrica.de/correlation.html>.
- Martinek, T. J. (1980). Stability of teachers' expectations for elementary school aged children. *Perceptual and Motor Skills*, 51(suppl2), 1269–1270. <https://doi.org/10.2466/pms.1980.51.3f.1269>
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgements: Exploring student characteristics that influence teacher judgements of student ability. *Teaching and Teacher Education*, 65, 48–60. <https://doi.org/10.1016/j.tate.2017.02.021>
- Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review*, 8(2), 193. <https://doi.org/10.2307/4609267>
- Ministry of Education. (z.d.). *Education in New Zealand*. Geraadpleegd op 21 mei 2024, <https://www.education.govt.nz/our-work/our-role-and-our-people/education-in-nz/#Early>
- Muehlenhard, C. L., & Peterson, Z. D. (2011). Distinguishing between sex and gender: History, current conceptualizations, and implications. *Sex Roles*, 64(11–12), 791–803. <https://doi.org/10.1007/s11199-011-9932-5>

- Muntoni, F., & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading: The role of teachers' gender stereotypes. *Contemporary Educational Psychology, 54*, 212–220. <https://doi.org/10.1016/j.cedpsych.2018.06.012>
- Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should ALWAYS check for them). *Practical Assessment, Research and Evaluation, 9*(1), 1–8. <https://doi.org/10.7275/qf69-7k43>
- Pólya, G. (1920). Über den zentralen grenzwertsatz der wahrscheinlichkeitsrechnung und das momentenproblem [De centrale limietstelling van de kansrekening en het momentenprobleem]. *Mathematische Zeitschrift, 8*(3–4), 171–181. <https://doi.org/10.1007/bf01206525>
- Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76*(1), 85–97. <https://doi.org/10.1037/0022-0663.76.1.85>
- Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities. *American Educational Research Journal, 48*(2), 335–360. <https://doi.org/10.3102/0002831210374874>
- Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Çopur-Gençtürk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology, 50*(4), 1262–1281. <https://doi.org/10.1037/a0035073>
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. Holt, Rinehart & Winston.
- Rubie-Davies, C. (2014). *Becoming a high expectation teacher: Raising the bar*. Routledge. <https://doi.org/10.4324/9781315761251>



- Rubie-Davies, C. M., & Peterson, E. R. (2016). Relations between teachers' achievement, over- and underestimation, and students' beliefs for Māori and Pākehā students. *Contemporary Educational Psychology, 47*, 72–83.  
<https://doi.org/10.1016/j.cedpsych.2016.01.001>
- Rubie-Davies, C. M., Watson, P., Flint, A., Garrett, L., & McDonald, L. (2018). Viewing students consistently: How stable are teachers' expectations? *Educational Research and Evaluation, 24*(3–5), 221–240. <https://doi.org/10.1080/13803611.2018.1550836>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia, 126*(5), 1763–1768.  
<https://doi.org/10.1213/ane.0000000000002864>
- Seddig, D., & Leitgöb, H. (2018). Approximate measurement invariance and longitudinal confirmatory factor analysis: concept and application with panel data. *Directory of Open Access Journals*. <https://doi.org/10.18148/srm/2018.v12i1.7210>
- Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*(2), 253–273. <https://doi.org/10.1037/0022-0663.99.2.253>
- Timmermans, A. C., de Boer, H., & van der Werf, M. P. C. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education, 19*, 217–240.  
<https://doi.org/10.1007/s11218-015-9326-6>
- Timmermans, A. C., Kuyper, H., & van der Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology, 85*, 459–478.  
<https://doi.org/10.1111/bjep.12087>

- Timmermans, A., Rubie-Davies, C. M., & Wang, S. (2021). Adjusting expectations or maintaining first impressions? The stability of teachers' expectations of students' mathematics achievement. *Learning and Instruction, 75*, 101483. <https://doi.org/10.1016/j.learninstruc.2021.101483>
- Timmons, K. (2019). Kindergarten Expectations and Outcomes: Understanding the influence of educator and child expectations on children's self-regulation, early reading, and vocabulary outcomes. *Journal of Research in Childhood Education, 33*(3), 471–489. <https://doi.org/10.1080/02568543.2019.1609144>
- Turner, H., Rubie-Davies, C. M., & Webber, M. (2015). Teacher expectations, ethnicity and the achievement gap. *New Zealand Journal of Educational Studies, 50*(1), 55–69. <https://doi.org/10.1007/s40841-015-0004-1>
- Van Den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers. *American Educational Research Journal, 47*(2), 497–527. <https://doi.org/10.3102/0002831209353594>
- Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation, 24*(3–5), 124–179. <https://doi.org/10.1080/13803611.2018.1548798>
- Wang, S., Rubie-Davies, C. M., & Meissel, K. (2020). The stability and trajectories of teacher expectations: Student achievement level as a moderator. *Learning and Individual Differences, 78*, 101819. <https://doi.org/10.1016/j.lindif.2019.101819>
- Westphal, A., Becker, M., Vock, M., Maaz, K., Neumann, M., & McElvany, N. (2016). The link between teacher-assigned grades and classroom socioeconomic composition: The role of classroom behavior, motivation, and teacher characteristics. *Contemporary Educational Psychology, 46*, 218–227. <https://doi.org/10.1016/j.cedpsych.2016.06.004>

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child*

*Development Perspectives*, 4(1), 10–18. <https://doi.org/10.1111/j.1750->

8606.2009.00110.x

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education:

They're not magic. *Review of Educational Research*, 81(2), 267-301.

<https://doi.org/10.3102/0034654311405999>

### Bijlage 1

Om te kijken of de uitschieters in de dataset invloed hadden op de correlaties van de leerkrachtverwachtingen van wiskunde en lezen, zijn de correlaties opnieuw berekend nadat de uitschieters zijn verwijderd uit de dataset. Voor het verwijderen van de uitschieters is de algemene regel aangehouden om de Z-waardes kleiner dan -3.0 en groter dan 3.0 te verwijderen uit de dataset (Osborne & Overbay, 2004). In de Tabel 8 is te zien dat het verschil tussen de correlaties met en zonder uitschieters niet groot is. Het grootste verschil zit tussen de correlatie van lezen tussen meetmoment 1 en 2. Zonder uitschieters is deze correlatie  $r(1422) = .528, p < .001$  en met uitschieters is deze correlatie  $r(1431) = .544, p < .001$ . Om deze reden is besloten om de uitschieters in de dataset te behouden.

**Tabel 8**

*Correlaties Leerkrachtverwachtingen Wiskunde en Lezen Zonder en Met Uitschieters*

Meetmomenten	Wiskunde		Lezen	
	1	2	1	2
Zonder uitschieters	1	1	1	
	2	.449* (1588)	.528* (1424)	1
	3	.540* (1174)	.707* (1329)	.567* (1228)
Met uitschieters	1	1	1	
	2	.446* (1596)	.544* (1433)	1
	3	.542* (1278)	.709* (1335)	.568* (1239)

*Noot.* \* $p < .001$ . De steekproefgrootte van de correlaties staan tussen haakjes.