

De relatie tussen het leerkrachtoordeel over en het niveau van de voorbereidende rekenvaardigheid van kinderen uit groep 2 en 3.

Thesis Pedagogische Wetenschappen
(200600042)
Universiteit Utrecht
2016-2017

Sanne Nuijten (3698176)

Werkgroep 17

Begeleider: Bernadette van de Rijt

Datum: 24 juni 2016

Abstract

Background: According to research, there is a positive relation between early numeracy and later success in mathematics. Therefore, teachers need to discover delays in the child's early numeracy to make early intervention possible. For this, teachers need to form a judgment either through observing in the classroom and analyzing tests or through observing the child in general. No consensus has been reached by researchers in literature concerning the value of teacher judgment. **Aim:** The aim of this study is to examine whether there is correlation between early numeracy and teacher judgment. Given the fact that the Utrechtse Getalbegrip Toets- Revised (UGT-R) offers a valid measurement of this early numeracy, this instrument is used to assess the children's level of early numeracy. The value of teacher judgment is reviewed with a correlation-test and regression analysis. The research is executed with 92 children, aged 5-7, from years 2 and 3 from several Dutch primary schools. **Results:** Results of the present study show significant correlations between results on the UGT-R and teacher judgment for years 2 and 3, and ages 5-7. Tests have also shown a positive correlation between teacher judgment and results on two categories of the UGT-R (i.e. telvaardigheden en Piagetiaanse vaardigheden) for year 3 and on all categories for year 2. Additionally, a regression analysis has shown a significant influence of children's early numeracy, as measured with the UGT-R, on teacher judgment of that early numeracy. **Conclusion:** According to this study there is a significant correlation between teacher judgment and early numeracy. The level of early numeracy is also a significant predictor of teacher judgment. This thesis implies teacher judgment should be used more, instead of testing. Further research with a longitudinal design should be done on the correlation between teacher judgment and early numeracy through the years.

Keywords: Early Numeracy, UGT-R, Teacher Judgment, Age

Inleiding

In het Nederlandse rekenen-wiskunde onderwijs wordt gestreefd naar het aanleren van vaardigheden die bijdragen aan het behalen van 11 kerndoelen, die de kinderen voldoende bagage meegeven voor actieve participatie in de maatschappij (SLO, 2009). In groep 3 komen de kinderen voor het eerst in aanraking met het formele rekenonderwijs, dat gericht is op het behalen van deze doelen aan het eind van groep 8. Hoewel alle kinderen dezelfde overstap maken van spelenderwijs leren in de kleuterklas naar systematisch leren in groep 3, is er een groep leerlingen die een grote achterstand oploopt bij rekenen ten opzichte van haar medeleerlingen (SLO, 2009). Het verschil in deze rekenvaardigheid tussen leerlingen is te verklaren door verschillende factoren. Uit eerder onderzoek van Aubrey, Dahl en Godfrey (2006) bleek dat factoren als kindmotivatie, leerkrachtinstructie en opleidingsniveau van ouders van invloed zijn op de rekenprestaties van kinderen. De voorbereidende rekenvaardigheid is volgens verschillende onderzoekers echter de grootste voorspeller van rekensucces (Aubrey, Dahl, & Godfrey, 2006; Aunio & Niemivirta, 2010). Het is van belang dat leerkrachten de ontwikkeling van de voorbereidende rekenvaardigheid beoordelen, zodat vroegtijdige signalering mogelijk is. Het leerkrachtoordeel betreft een oordeel over de vaardigheden van een kind (Urhahne, 2015) waarop hij zijn verwachtingen en begeleiding aanpast. Hieronder zullen eerst de voorbereidende rekenvaardigheid en de transitie naar het formele rekenonderwijs uiteengezet worden. Daarna wordt ingezoomd op het leerkrachtoordeel en de relatie tussen dit oordeel en het niveau van voorbereidende rekenvaardigheid van het kind.

Voorbereidende rekenvaardigheid

In recent onderzoek worden onder voorbereidende rekenvaardigheid steeds twee, onderling gerelateerde, deelgebieden aangehaald: getalbegrip en relationele vaardigheden (Aunio, Hautamäki, Heiskari, & Van Luit, 2006; Aunio, Hautamäki, Sajaniemi, & Van Luit, 2009). Het eerste deelgebied, getalbegrip, betreft telvaardigheden, kennis van getallen en getaltransformaties (Aunio & Niemivirta, 2010; Jordan et al., 2006; Jordan, Kaplan, Locuniak, & Ramineni, 2007). Getalbegrip vereist kennis, gevoel en vaardigheden betreffende de mentale getallenlijn (Berch, 2005). Het tweede deelgebied bestaat uit de relationele vaardigheden. Piaget (zoals geciteerd in Aunio et al., 2005) geeft aan dat de ontwikkeling van rekenkundig denken is gerelateerd aan het begrip van relationele uitspraken. Het gaat hierbij om vaardigheden als vergelijken, classificeren, één-op-één corresponderen en het seriëren van getallen.

De voorbereidende rekenvaardigheid heeft een groot aandeel in het rekensucces (Aubrey et al., 2011). Kinderen kunnen namelijk pas strategieën ontwikkelen om complexere wiskundige problemen op te lossen, als deze vaardigheden voldoende beheerst worden (Aunio & Niemivirta, 2010; Berch, 2005; Gersten, Jordan, & Flojo,

2005). Geconcludeerd kan worden dat beide gebieden voldoende moeten zijn ontwikkeld bij de transitie naar het formele rekenonderwijs in groep 3. Leerkrachten in groep 1/2 moeten de leerlingen hier voldoende in ondersteunen (Inspectie van Onderwijs, 2011).

Transitie naar formeel rekenonderwijs

Vanaf het moment dat kinderen starten in groep 1 is er al een bepaalde voorbereidende rekenvaardigheid aanwezig. De ontwikkeling van deze vaardigheid verloopt volgens zowel Aunola, Leskinen, Lerkkanen en Nurmi (2004), Jordan, Kaplan, Olah en Locuniak (2006) als Aunio, Heiskari, Van Luit en Vuori (2015) ongeveer even snel bij alle leerlingen. Het niveau waarmee ze binnenkomen, maakt echter het verschil tussen zwakke en sterke rekenaars (Aunio et al., 2015). Hoewel Aunio en Niemivirta (2010) aangeven dat kinderen met een zwakke voorbereidende rekenvaardigheid moeite blijven hebben met rekenen, blijkt dat deze kinderen tot het einde van de kleuterjaren nog veel kunnen inlopen (Aunio et al., 2015). Om dit mogelijk te maken, is het van belang dat leerlingen vroeg gesignaleerd worden door de leerkracht en dat interventies worden ingezet die de leerling begeleiden bij het verbeteren van de voorbereidende rekenvaardigheid (Bodovsky & Farkas, 2007; Bryant et al., 2011).

Leerkrachtoordeel

Leerkrachten maken gebruik van zowel methodeafhankelijke als -onafhankelijke toetsen om de rekenvaardigheid van kinderen te analyseren (Cito, 2017). Naast het analyseren van toetsgegevens, observeren leerkrachten om te kijken hoe het kind het doet. Volgens Cito (2013) dient het toetsen als hulpmiddel om leerkrachten te ondersteunen bij het nemen van beslissingen, zodat ze kinderen een passend onderwijsaanbod kunnen bieden. Uit onderzoek van Cornwell, Mustard en Van Parys (2013) is gebleken dat leerkrachten bij het beoordelen van leerlingen niet alleen rekening houden met de toetsresultaten, maar een totaalbeeld schetsen van de leerling. Aangezien ontwikkelingsprocessen bij jonge kinderen niet ordelijk en geleidelijk verlopen, raadt Goorhuis (2014) dit ook aan. Op deze manier kunnen leerkrachten beter onderscheid maken tussen sterke, gemiddelde en zwakke leerlingen en hier hun onderwijs op aansluiten (Kilday, Kinzie, Mashburn, & Whittaker, 2012). Het leerkrachtoordeel wordt beïnvloed door verschillende factoren, zoals de norm waarmee de leerkracht vergelijkt (Südkamp, Kaiser, & Möller, 2012), de kennis van de leerkracht betreffende voorbereidende rekenvaardigheid (Kilday et al., 2012) en kindkenmerken, zoals concentratie (Aunio & Niemivirta, 2010). In het onderzoek van Taylor, Anselmo, Foreman, Schatschneider en Angelopoulos (2000) komt de kracht van het leerkrachtoordeel in de kleuterklas naar voren. Leerkrachten blijken goed in staat om leerproblemen vroeg te signaleren. Volgens Aunio en anderen (2015) hebben leerkrachten echter onvoldoende competenties om kinderen in een vroeg stadium te signaleren als rekenzwakke leerlingen.

Vanuit deze tegenstrijdigheid volgt de vraag: 'Is er een relatie tussen het leerkrachtoordeel over het niveau van de voorbereidende rekenvaardigheid van kinderen uit groep 2 en 3 en hun voorbereidende rekenvaardigheid?'. Verwacht wordt dat er een relatie is tussen het leerkrachtoordeel en het niveau van de leerling. Indien een samenhang wordt gevonden, volgt de vraag of er sprake is van causaliteit: kan het leerkrachtoordeel voorspeld worden aan de hand van de voorbereidende rekenvaardigheid? In andere woorden: 'Is het niveau van voorbereidende rekenvaardigheid van het kind van invloed op het oordeel van de leerkracht en hoe sterk is dit effect?'.

Methode

Participanten

De populatie van het onderzoek betreft leerlingen uit de groepen 2 en 3 van de basisschool. Voor de steekproef zijn 56 leerlingen geselecteerd met behulp van een quotasteekproef. Dit is een niet-willekeurige steekproef waarbij van tevoren is vastgesteld hoeveel deelnemers van elke categorie geselecteerd worden (Neuman, 2014). Aan de hand van een korte beoordelingslijst (bijlage 1) is een overzicht gemaakt van het niveau van de leerlingen volgens de leerkracht. Uit deze beoordelingen zijn de leerlingen geselecteerd. Er is een gelijke verdeling van zowel sterke als zwakke rekenaars, jongens en meisjes, en de groepen 2 en 3 gemaakt. Vervolgens is het databestand aangevuld met gegevens van 36 leerlingen, die voldeden aan gestelde eisen en zijn getest door studenten van de opleiding pedagogische wetenschappen. De totale steekproef bestaat uit $n=92$ kinderen.

Procedure

Voor het onderzoek zijn verschillende scholen telefonisch en via email benaderd. De scholen die hebben meegewerkt, waren bekend bij de onderzoekers. 28 leerlingen van groep 2 zijn geselecteerd op een school in Breda en 28 leerlingen van groep 3 op een school in Utrecht. De leerlingen zijn op de eigen school getest in een aparte kamer. Hiermee is geprobeerd om zoveel mogelijk prikkels bij de leerling te ontnemen. De leerlingen zijn in één sessie van 20-40 minuten getest met minimale onderbreking. De versies A en B van de Utrechtse Getalbegrip Toets-Revised (Van Luit & Van de Rijt, 2009) zijn evenredig verdeeld over de verschillende groepen, leerkrachtoordelen en seksen.

Wetenschappelijke verantwoording

De wetenschappelijke relevantie heeft betrekking op de samenhang tussen het leerkrachtoordeel en de resultaten op de UGT-R. Zoals eerder vermeld, lopen de meningen betreffende de noodzaak van het toetsen van kleuters uiteen (Cito, 2014). Met dit onderzoek kan ondersteuning worden gevonden voor de hypothese dat het leerkrachtoordeel overeenkomt met het niveau van voorbereidende rekenvaardigheid van de leerlingen. Kleutertoetsen zijn ontwikkeld om leerkrachten te ondersteunen. Wanneer

blijkt dat er sprake is van een significant positief verband tussen het oordeel en het niveau, zou de noodzaak van toetsen heroverwogen kunnen worden. Als er echter sprake is van een grote discrepantie, zal de hypothese worden verworpen.

In dit onderzoek wordt rekening gehouden met ethische richtlijnen. In het algemeen geldt dat hoe groter de kans op schade is bij de deelnemer, hoe belangrijker het is dat de informed consent volledig is (Neuman, 2014). Omdat dit onderzoek geen risico's met zich meedraagt, is enkel met een e-mail naar ouders om toestemming gevraagd. Wanneer ouders bezwaar hadden tegen het onderzoek, konden ze dit aangeven bij de leerkracht. In dit geval zijn hun kinderen niet in de selectie opgenomen. Aan het begin van elke test is de leerling mondeling om toestemming gevraagd en ook gedurende de afname is rekening gehouden met het welbevinden van het kind. Na afloop van het onderzoek zijn de resultaten met de leerkrachten besproken. Met het oog op privacy zijn de namen van de leerlingen gefingeerd in leerlingnummers. Hierdoor is de anonimiteit gegarandeerd.

Instrumenten en variabelen

In dit onderzoek is sprake van zowel verkennend als beschrijvend onderzoek. Deze onderzoeken richten zich respectievelijk op het ontwikkelen van een algemeen begrip en op het presenteren van een kwantitatief beeld van de situatie (Neuman, 2014). Aangezien het leerkrachtoordeel niet eerder is meegenomen in onderzoek naar voorbereidende rekenvaardigheid, is dit onderzoek verkennend. Omdat het onderzoek zich daarnaast op een kwantitatieve manier richt op de samenhang tussen twee variabelen wordt gesproken van beschrijvend onderzoek. De afhankelijke variabele is het niveau van voorbereidende rekenvaardigheid en de onafhankelijke het leerkrachtoordeel. De operationalisatie in instrumenten en de validiteit en betrouwbaarheid worden hierna uiteengezet.

Voorbereidende rekenvaardigheid

De voorbereidende rekenvaardigheid betreft het begrijpen van en het ontwikkelen van strategieën in de omgang met getallen (Butterworth, 2005). Om het niveau van voorbereidende rekenvaardigheid te meten, is de Utrechtse Getalbegrip Toets - Revised (UGT-R) ingezet (Van Luit & Van de Rijt, 2009). De toets bestaat uit telvaardigheden, Piagetiaanse vaardigheden en schatten. Deze drie onderdelen zijn opgedeeld in totaal negen categorieën, te weten: vergelijken, hoeveelheden koppelen, een op een correspondentie, ordenen, telwoorden gebruiken, synchroon en verkort tellen, resultaatief tellen, toepassen van kennis van getallen en schatten (Van Luit & Van de Rijt, 2009). Voor iedere categorie zijn vijf opdrachten ontworpen, waarmee in totaal 45 punten kunnen worden gehaald. Een hoge vaardigheidsscore op de UGT-R betekent een goede voorbereidende rekenvaardigheid en andersom (Van Luit & Van de Rijt, 2009). Binnen het onderzoek wordt gewerkt met de totaalscores en de vaardigheidsscores van de

leerlingen op de UGT-R. Deze scores zijn discreet. De totaalscores zijn van ratio meetniveau en de vaardigheidsscores van interval meetniveau.

In de handleiding van de UGT-R wordt geconcludeerd dat de UGT-R een valide instrument is (Van Luit & Van de Rijt, 2009b). De begripsvaliditeit en de criteriumvaliditeit worden door de Commissie Testaangelegenheden Nederland [COTAN] beoordeeld als onvoldoende, omdat er nog te weinig onderzoek is gedaan (Evers et al., 2010). De UGT-R wordt gezien als een betrouwbaar instrument om voorbereidende rekenvaardigheid te meten (Van Luit & Van de Rijt, 2009). De Cronbach's alfa van de UGT-R is gemiddeld .93. De COTAN beoordeelt de betrouwbaarheid van de UGT-R als voldoende (Evers et al., 2010). In dit onderzoek wordt de toets afgenomen door studenten Orthopedagogiek. De UGT-R scoort voldoende op betrouwbaarheid en daarom wordt verwacht dat, mits de handleiding is gevolgd, betrouwbare gegevens zijn vergaard.

Leerkrachtoordeel

Het leerkrachtoordeel kan gebaseerd zijn op zowel de resultaten van methode- en niet-methodetoetsen als op andere kenmerken van het kind, de ouders en de verhoudingen in de groep (Aunio & Niemivirta, 2010; Cito, 2013; Kilday et al., 2012; Südkamp, Kaiser, & Möller, 2012). Voor dit onderzoek is het leerkrachtoordeel geoperationaliseerd in een beoordelingslijst (bijlage 1). In deze lijst vult de leerkracht in hoe een kind presteert op het gebied van rekenen in vergelijking met zijn klasgenoten. Hierbij wordt nadrukkelijk verzocht om geen rekening te houden met toetsresultaten. Aangezien alleen de leerlingen met een beoordeling 'sterk' of 'zwak' worden betrokken bij de selectie, is het leerkrachtoordeel een dichotome variabele.

Het leerkrachtoordeel is een subjectief oordeel. De mening van leerkrachten is persoonlijk, veranderlijk en beïnvloedbaar. Omdat er verwacht wordt dat het leerkrachtoordeel kwalitatief is, zullen de leerkrachtoordelen over dezelfde leerlingen niet veel verschillen. Dit resulteert in een hoge interbeoordelaarsbetrouwbaarheid. Daarnaast wordt verwacht dat de leerkrachtoordelen bij het herhaaldelijk invullen van het instrument niet veel zullen wijzigen. Het instrument beoogt te meten wat de leerkracht ziet. De interpretatie van de leerkracht beïnvloedt de validiteit. Wanneer deze mening wordt beïnvloed door andere factoren dan de rekenvaardigheid, heeft dat een negatief effect op de validiteit. Het beoordelingsinstrument beoogt de betrouwbaarheid van het onderzoek te vergroten. De lijst maakt de selectie van leerlingen transparant en herhaalbaar.

Data-analyse

Eerst zal met beschrijvende statistieken en een t-toets voor onafhankelijke groepen een algemeen beeld geschetst worden van de verdeling in resultaten. Daarna zal met behulp van een punt-biseriële correlatietoets een antwoord worden gezocht op de onderzoeksvraag. Hierbij worden de totaalscores (ratio) van de leerlingen vergeleken

met het leerkrachtoordeel (dichotoom). De samenhang zal individueel worden getoetst voor de groepen 2 en 3. Daarnaast zal gekeken worden naar de samenhang tussen de variabelen voor de verschillende leeftijdscategorieën. Hiervoor is gekozen vanwege een overlap in leeftijd tussen de groepen 2 en 3. Tot slot wordt de samenhang getoetst tussen het leerkrachtoordeel en de drie categorieën op de UGT-R (Piagetiaanse vaardigheden, telvaardigheden en schatten). Hierbij worden groep 2 en 3 wederom opgesplitst. Verwacht wordt dat er een positieve relatie is tussen het leerkrachtoordeel en het niveau van de leerlingen op de UGT-R. Deze hypothese wordt aangenomen als er een significante positieve samenhang wordt geconstateerd tussen de twee variabelen. Er wordt eenzijdig getoetst met de kritische waarde $\alpha = .05$.

Als een significante samenhang wordt geconstateerd met de punt-biseriële correlatietoets, is er de mogelijkheid om met behulp van een regressieanalyse een uitspraak te doen over de invloed van de voorbereidende rekenvaardigheid op het leerkrachtoordeel. Omdat er in deze situatie sprake is van een dichotome afhankelijke variabele wordt de invloed van de score op de UGT-R geanalyseerd aan de hand van een logistische regressieanalyse. Omdat er sprake is van een redelijk kleine sample ($n=92$) wordt gekozen voor een kritische waarde van $\alpha = .01$ op de Wald test. Voorafgaand aan de toets zijn de assumpties steeds gecontroleerd. Hoewel niet aan alle assumpties is voldaan, wordt hiermee, door de beperkingen van dit onderzoek, toch verder gewerkt.

Resultaten

Met behulp van statistische analyses is de volgende onderzoeksvraag onderzocht: 'Is er een relatie tussen het leerkrachtoordeel over het niveau van de voorbereidende rekenvaardigheid van kinderen uit groep 2 en 3 en hun voorbereidende rekenvaardigheid?'. Een tweede onderzoeksvraag luidt: 'Is het niveau van voorbereidende rekenvaardigheid van invloed op het oordeel van de leerkracht?'. De resultaten van de analyses worden hieronder weergegeven.

Relatie UGT-R en leerkrachtoordeel

Samenhang per groep

De verdeling van de leerlingen en hun scores op de UGT-R is af te lezen in tabel 1. Hierin is aangegeven hoeveel leerlingen, volgens het leerkrachtoordeel, een goede voorbereidende rekenvaardigheid (sterk) en een slechte voorbereidende rekenvaardigheid (zwak) hebben. Met een t-toets voor onafhankelijke groepen zijn de verschillen in gemiddeldes getoetst. Hieruit komt naar voren dat het verschil tussen groep 2 en 3 ($t(90)=-3.62, p<.01$) en tussen zwak en sterk ($t(65.48)=-9.45, p<.01$) significant is. Dit is ook het geval per groep; de resultaten van de zwakke en sterke leerlingen verschillen significant van elkaar. Groep 2 met $t(40)=-8.51, p<.01$ en groep 3 met $t(48)=-8.45, p<.01$.

Tabel 1

Beschrijvende Statistieken voor de Sterke en Zwakke Leerlingen voor Groep 2 en 3

		<i>n</i>	Minimum	Maximum	<i>M</i>	<i>SD</i>
Groep 2	Zwak	18	17	33	25.00	4.88
	Sterk	24	29	40	35.21	2.86
	Totaal	42	17	40	30.83	6.37
Groep 3	Zwak	23	21	35	30.83	2.80
	Sterk	27	33	44	39.00	4.01
	Totaal	50	21	44	35.24	5.32

In tabel 2 zijn de resultaten van de punt-biseriële correlatietoets weergegeven. Hierin komt naar voren dat er sprake is van een positief verband tussen leerkrachtoordeel en score op de UGT-R voor zowel groep 2 als 3. De correlatie is voor beide groepen significant en wordt als sterk beoordeeld.

Tabel 2

Resultaten van de Punt-Biseriële Correlatietoets voor de Samenhang tussen het Leerkrachtoordeel en de Totaalscore op de UGT-R voor de Groepen 2 en 3

		Totaalscore UGT-R		
		<i>n</i>	<i>r</i>	<i>p</i>
Leerkrachtoordeel	Groep 2	42	.80	<.01
	Groep 3	50	.77	<.01

Samenhang per leeftijdscategorie

Tussen de leeftijden van de kinderen uit groep 2 (63 -79 maanden) en 3 (75-94 maanden) is sprake van overlap. Om deze reden is de variabele 'leeftijdscategorie' aangemaakt. Met de correlatietoets wordt de samenhang tussen het leerkrachtoordeel en de score op de UGT-R per leeftijdsgroep geanalyseerd. In tabel 3 zijn de resultaten op de correlatietoets opgenomen. Hierin komt naar voren dat de samenhang tussen leerkrachtoordeel en totaalscore UGT-R voor alle leeftijdscategorieën significant is. Het verband wordt als sterk beoordeeld.

Tabel 3

Resultaten van de Punt-Biseriële Correlatietoets voor de Samenhang tussen het Leerkrachtoordeel en de Totaalscore op de UGT-R per Leeftijdscategorie

	Leeftijdscategorie	Totaalscore UGT-R		
		<i>n</i>	<i>r</i>	<i>p</i>
	5	21	.74	<.01
Leerkrachtoordeel	6	50	.69	<.01
	7	21	.74	<.01

Samenhang per categorie van de UGT-R

De UGT-R is op te delen in drie categorieën, te weten: Piagetiaanse vaardigheden, telvaardigheden en schatten. Onder de eerste twee categorieën vallen beide 20 opdrachten en de categorie schatten bestaat uit vijf opdrachten. In tabel 4 zijn de beschrijvende statistieken voor de groepen 2 en 3 opgenomen per onderdeel. Ook zijn de gemiddelden voor de sterke en zwakke leerlingen opgenomen.

Tabel 4

Gemiddelde Scores UGT-R per Onderdeel

		<i>n</i>	Min.	Max.	<i>M</i>	<i>SD</i>	<i>M</i> _{zwak}	<i>M</i> _{sterk}
Groep 2	PT	42	9	19	15.52	2.56	13.56	17.00
	TV	42	3	19	13.07	3.78	9.83	15.50
	Schatten	42	0	4	2.24	1.21	1.61	2.71
Groep 3	PT	50	11	20	16.84	2.19	15.22	18.22
	TV	50	6	20	15.10	3.177	12.57	17.26
	Schatten	50	1	5	3.30	1.11	3.04	3.52

Noot: PT = Piagetiaanse vaardigheden, TV = telvaardigheden

In tabel 5 zijn de resultaten uit de punt-biseriële correlatietoets opgenomen. Daarin komt naar voren dat het leerkrachtoordeel significant, positief samenhangt met de Piagetiaanse- en telvaardigheden in zowel groep 2 als 3. Deze verbanden worden beoordeeld als sterk. De categorie 'schatten' hangt alleen in groep 2 significant samen met het leerkrachtoordeel. Dit verband wordt beoordeeld als matig.

Tabel 5

Resultaten van de Punt-Biseriële Correlatietoets voor de Samenhang tussen het Leerkrachtoordeel en de Scores UGT-R per Categorie

	Categorieën binnen de UGT-R	Leerkrachtoordeel		
		<i>r</i>	<i>p</i>	<i>n</i>
Groep 2	PV	.67	.00	42
	TV	.75	.00	42
	Schatten	.46	.00	42
Groep 3	PV	.69	.00	50
	TV	.74	.00	50
	Schatten	.22	.13	50

Noot: PV=Piagetiaanse vaardigheden, TV= telvaardigheden.

Het leerkrachtoordeel voorspellen

Aangezien de relatie tussen het leerkrachtoordeel en de score op de UGT-R significant is, kan met de logistische regressieanalyse worden berekend wat de kans is dat een leerling het leerkrachtoordeel 'sterk' of 'zwak' krijgt. Met verschillende testen is de fit van het model bepaald. In tabel 6 is aangetoond dat het model met de score van de UGT-R beter past bij de data dan een model zonder deze variabele. De fit van het model is daarnaast bepaald met behulp van de Nagelkerke r^2 , een pseudo r^2 -maat die vergelijkbaar is met de r^2 uit de lineaire regressieanalyse. Het model scoorde op deze toets $r^2 = .72$.

Tabel 6

Resultaten op de Omnibus Test voor Model Coëfficiënten

	χ^2	<i>df</i>	<i>p</i>
Model	71.33	1	.00

Met de Hosmer en Lemeshow Goodness of Fit Test (zoals geciteerd in Sieben & Linssen, 2009) zijn de individuen geordend in tien ongeveer even grote groepjes naar oplopende voorspelde slaagkans. Voor elke groep wordt gekeken hoeveel individuen een 0 (zwak) of 1 (sterk) hebben gescoord en hoeveel op basis van het model verwacht zou worden (tabel 7). Met de χ^2 -toets is aangetoond dat de verschillen niet significant zijn $\chi^2(8)=5.69$, $p=.68$. Dit wil zeggen dat het model goed bij de data past.

Tabel 7

Afhankelijkheidstabel voor de Hosmer en Lemeshow Test Betreffende Leerkrachtoordeel

	Zwak		Sterk		Totaal
	Geobserveerd	Verwacht	Geobserveerd	Verwacht	
1	8	8.00	0	0.00	8
2	9	8.89	0	0.11	9
3	7	7.42	1	0.58	8
4	6	6.58	3	2.42	9
5	3	4.17	5	3.83	8
6	6	3.25	3	5.75	9
7	2	1.98	10	10.02	12
8	0	0.50	9	8.50	9
9	0	0.2	13	12.80	13
10	0	0.02	7	6.98	7

Gezien het feit dat de fit van het model bepaald is, kan het effect van de variabele 'totaalscore op de UGT-R' berekend worden. Met behulp van de Wald-statistiek is berekend dat de score op de UGT-R een significante invloed heeft op het leerkrachtoordeel sterk/zwak $Wald(1)=19.39, p=.00$. De $Exp(B)=1.92$ toont aan dat het effect van de score op de UGT-R positief is.

Conclusie en discussie

Conclusie

Het huidige onderzoek is gericht op het onderzoeken van de correlatie tussen de voorbereidende rekenvaardigheid van leerlingen, zoals gemeten met de UGT-R, en het leerkrachtoordeel over deze rekenvaardigheid voor leerlingen van groep 2 en 3. Hierbij is onderscheid gemaakt tussen de groepen 2 en 3, en de leeftijden van de leerlingen. Uit de resultaten kan geconcludeerd worden dat er sprake is van een positief, significant verband tussen het leerkrachtoordeel en de voorbereidende rekenvaardigheid. Dit wil zeggen dat wanneer een leerling een goede voorbereidende rekenvaardigheid heeft, de kans groot is dat de leerkracht de leerling als sterk beoordeelt en wanneer de leerling een slechte voorbereidende rekenvaardigheid heeft, als zwak. Dit is zowel het geval voor de groepen 2 en 3 als voor de drie leeftijdscategorieën (5, 6 en 7 jaar oud). Het positieve verband is daarnaast significant voor zowel groep 2 als 3 op de onderdelen Piagetiaanse vaardigheden en telvaardigheden, maar alleen voor groep 2 op het onderdeel schatten. Binnen de resultaten is geen significant verschil aangetoond tussen de correlaties van de resultaten op de UGT-R van de groepen 2 en 3 met het leerkrachtoordeel. De resultaten uit het huidige onderzoek komen overeen met de verwachtingen vanuit het literatuuronderzoek.

De tweede onderzoeksvraag richt zich op de invloed van de voorbereidende rekenvaardigheid op het leerkrachtoordeel. Vanuit de literatuur werd verwacht dat de rekenvaardigheid van leerlingen een positieve invloed heeft op het leerkrachtoordeel. Dit wordt in het huidige onderzoek bevestigd. De fit van het model is aangetoond aan de hand van verschillende statistische analyses. Hieruit komt met $r^2 = .72$ naar voren welk deel van de variabiliteit in de scores verklaard kan worden door de relatie tussen de scores op de UGT-R en het leerkrachtoordeel. Met de Wald-toets is aangetoond dat deze invloed significant en positief is. Het leerkrachtoordeel kan voorspeld worden aan de hand van de voorbereidende rekenvaardigheid van de leerling.

Discussie

De voorbereidende rekenvaardigheid is volgens verschillende onderzoekers de belangrijkste voorspeller van rekensucces (Aubrey, Dahl, & Godfrey, 2006; Aunio & Niemivirta, 2010). Daarom is het belangrijk dat een achterstand vroegtijdig wordt gesignaleerd (Bodovsky & Farkas, 2007; Bryant et al., 2011). Leerkrachten maken gebruik van verschillende toetsen, opdrachten en observaties om de rekenvaardigheid van kinderen te analyseren (Cito, 2017). Daaruit vormen zij een oordeel. In literatuuronderzoek kwamen tegenstrijdige resultaten betreffende de kwaliteit van het oordeel naar voren. In dit onderzoek is echter aangetoond dat leerkrachten een passend oordeel vellen over de voorbereidende rekenvaardigheid van hun leerlingen. Aangezien onderzoek naar het leerkrachtoordeel schaars is, wordt aanvullend onderzoek aanbevolen.

In literatuuronderzoek is daarnaast naar voren gekomen dat verschillende factoren betrokken worden bij het vormen van een leerkrachtoordeel (o.a. Cornwell et al., 2013). In dit onderzoek is enkel gekeken naar de invloed van het niveau van voorbereidende rekenvaardigheid op het oordeel. De invloed is zowel sterk als positief wat impliceert dat de variantie in leerkrachtoordeel voor een groot deel verklaard kan worden door het daadwerkelijke niveau van voorbereidende rekenvaardigheid van de leerling. Factoren als concentratievermogen, kennis van de leerkracht en de norm waarmee de leerkracht vergelijkt hebben dus slechts een bescheiden invloed.

Beperkingen

Dit onderzoek wijst uit dat het leerkrachtoordeel samenhangt met de voorbereidende rekenvaardigheid, zoals gemeten met de UGT-R. Dit onderzoek kent echter een aantal beperkingen, waardoor de resultaten met voorzichtigheid moeten worden geïnterpreteerd. Ten eerste is er een beperkte diversiteit van participanten, doordat de participanten zijn geselecteerd met een quotasteekproef op een gering aantal scholen. Dit maakt, samen met de geringe omvang van de steekproef ($n=92$), dat de resultaten niet generaliseerbaar zijn voor leerlingen van groep 2 en 3 in het Nederlandse basisonderwijs. De steekproef is echter wel representatief voor de populatie van de

scholen die hebben geparticipeerd aan het onderzoek. De onderzoeksbevindingen zijn bruikbaar op deze scholen.

Ten tweede zijn voorafgaand aan het onderzoek verschillende assumpties geschonden. Zo was er binnen groep 2 en de leeftijdscategorie '6 jaar' geen sprake van gelijkheid in variantie tussen sterke en zwakke leerlingen. Dit is enerzijds mogelijk te verklaren door het langer kleuteren van leerlingen in groep 2. Hierdoor worden de verschillen binnen de groep groter. Immers, des te ouder het kind, des te beter de voorbereidende rekenvaardigheid (Aunio, Aubrey, Godfey, Pan & Liu, 2008). Anderzijds valt de transitie tussen groep 2 en 3 in de leeftijdscategorie '6 jaar'. Dit kan negatieve invloed hebben op de validiteit van de meting. Kinderen krijgen in beide groepen namelijk op een geheel andere wijze onderwijs.

Ook zijn de assumpties geschonden door een uitschieter binnen de resultaten van de zwakke leerlingen uit groep 3. Hierdoor waren de resultaten niet normaal verdeeld. In een vervolgonderzoek wordt aanbevolen om gebruik te maken van een grotere steekproef met meer diversiteit in leerlingen. Mogelijk representeert deze steekproef onvoldoende de variëteit in scores onder de Nederlandse leerlingen van groep 3, waardoor resultaten van de leerlingen rooskleuriger zijn dan in de populatie het geval is.

Tevens zijn de assumpties gecontroleerd voor de drie categorieën van de UGT-R. Opvallend was dat vrijwel binnen alle categorieën geen sprake was van normaliteit en gelijkheid in variantie. Dit is mogelijk te verklaren door de geringe hoeveelheid opdrachten binnen de categorieën, met name in de categorie 'schatten'.

Als derde moet rekening worden gehouden met de mogelijke beperking in validiteit en betrouwbaarheid van het toetsingsinstrument voor het leerkrachtoordeel. Hoewel aan de leerkrachten expliciet is gevraagd om de scores op de CITO niet mee te nemen in het oordeel, is niet uitgesloten of dit indirect toch invloed heeft gehad op het leerkrachtoordeel. Wel zijn de resultaten op de CITO door de onderzoeker bekeken en zijn daarin enkele opvallende verschillen gevonden tussen leerkrachtoordeel en resultaat op de CITO. Omdat deze verschillen niet zijn getoetst op significantie, kunnen hier geen uitspraken over worden gedaan. Betreffende de betrouwbaarheid van het instrument wordt aangeraden om leerkrachten, die samen op de groep staan, hetzelfde instrument in te laten vullen. Op deze manier kan worden onderzocht of het oordeel van de kinderen overeenkomt op het instrument. Hiermee kan de interbeoordelaarsbetrouwbaarheid van het instrument worden ondersteund.

Aanbeveling

Hoewel uit het huidige onderzoek naar voren komt dat het leerkrachtoordeel significant samenhangt met de voorbereidende rekenvaardigheid, kan door de beperkingen van het onderzoek nog onvoldoende worden gezegd over de kwaliteit van het leerkrachtoordeel over het niveau van voorbereidende rekenvaardigheid in groep 2

en 3 in het algemeen. Met het oog op de uitblijvende consensus betreffende het afnemen van toetsen in groep 2 en 3 lijkt een vervolgonderzoek met een grotere, aselechte steekproef noodzakelijk.

Aangezien de categorie 'schatten' slechts bestaat uit 5 opdrachten, moeten vooral met deze resultaten voorzichtig worden omgegaan. Als in vervolgonderzoek meer vragen worden toegevoegd binnen het onderdeel 'schatten', kan een beter beeld geschetst worden over de vaardigheden van de leerlingen. Het is daarnaast belangrijk dat er wordt onderzocht waarom de leerlingen op dit specifieke onderdeel slecht scoren. Mogelijk wordt het nog onvoldoende aangeboden, waardoor resultaten achterblijven.

Tot slot impliceert dit onderzoek dat het leerkrachtoordeel voorspeld wordt door de voorbereidende rekenvaardigheid. Het leerkrachtoordeel is accuraat. In vervolgonderzoek zou onderzocht kunnen worden hoe de kwaliteit van het leerkrachtoordeel is op lange termijn. Aangezien de voorbereidende rekenvaardigheid een voorspeller is voor rekensucces en de leerkracht een goede inschatting hiervan kan maken, zouden leerlingen op lange termijn gevolgd moeten worden. Op die manier kan de voorspellende kracht van het leerkrachtoordeel getoetst worden.

Referenties

- Aunio, P., Aubrey, C., Godfrey, R., Pan, Y., & Liu, Y. (2008). Children's early numeracy in England, Finland and People's Republic of China. *International Journal of Early Years Education, 16*(3), 203-221.
- Aubrey, C., Dahl, S., & Godfrey, R. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal, 18*(1), 27-46.
- Aunio, P., Hautamäki, J., Heiskari, P., & Van Luit, J. E. H. (2006). The Early Numeracy Test in Finnish - Children's norms. *Scandinavian Journal of Psychology, 47*(5), 369-378.
- Aunio, P., Hautamäki, J., Sajaniemi, N., & Van Luit, J. E. H. (2009). Early numeracy in low-performing young children. *British Educational Research Journal, 35*(1), 25-46.
- Aunio, P., Heiskari, P., Van Luit, J. E., & Vuorio, J. M. (2015). The development of early numeracy skills in kindergarten in low-, average-and high-performance groups. *Journal of Early Childhood Research, 13*(1), 3-16.
- Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and individual differences, 20*(5), 427-435.
- Aunola, K., Leskinen, E., Lerkkanen, M. K., & Nurmi, J. E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96*(4), 699.

- Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38*(4), 333-339.
- Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal, 108*(2), 115-130.
- Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children, 78*(1), 7-23.
- Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry, 46*, 3-18. doi:10.1111/j.1469-7610.2005.00374.x
- Cito. (2013). *Toetsen van jonge kinderen. Deel 1*. Arnhem: Cito.
- Cito. (2017). *Veelgestelde vragen over methode en methodeonafhankelijke toetsen in het basisonderwijs*. Geraadpleegd op 13-03-2017, van http://www.cito.nl/~media/cito_nl/files/onderzoek%20en%20wetenschap/cito_veelgestelde_vragen_methode_methodeonafhankelijke_toetsen_2013.ashx.
- Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources, 48*, 236-264. doi:10.3368/jhr.48.1.236
- Evers, A., Egberink, I. J. L., Braak, M. S. L., Frima, R. M., Vermeulen, C. S. M., & Van Vliet-Mulder, J. C. (2010). *COTAN Documentatie*. Amsterdam: Boom test uitgevers.
- Gersten R., Jordan N. C., & Flojo J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 293-304.
- Goorhuis, S. (2014). *Peuters en kleuters zijn niet toetsbaar*. Geraadpleegd op 11-06-2017, van <http://wij-leren.nl/kleuters-toetsen.php>.
- Inspectie van Onderwijs. (2011). *Automatiseren bij rekenen-wiskunde*. Utrecht: Inspectie van het Onderwijs.
- Jordan, N. C., Kaplan, D., Locuniak M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice, 22*(1), 36-46.
- Jordan, N. C., Kaplan, D., Olah, L., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*(1), 153-175.
- Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2012). Accuracy of teacher judgments of preschoolers' math skills. *Journal of Psychoeducational Assessment, 30*(2), 148-159.
- Neuman, W. L. (2014). *Understanding Research*. Harlow: Pearson Education.

- Sieben, I., & Linssen, L. (2009). *Logistische regressie analyse: Een handleiding*. Nijmegen: Radboud Universiteit.
- SLO. (2009). *Fundamentele doelen Rekenen-Wiskunde: Uitwerking van het Fundamenteel niveau 1F voor einde basisonderwijs, versie 1.2*. Nationaal Expertisecentrum Leerplanontwikkeling: Enschede.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743-762.
- Taylor, H. G., Anselmo, M., Foreman, A. L., Schatschneider, C., & Angelopoulos, J. (2000). Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities, 33*(2), 200-210.
- Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education, 45*, 73-82.
- Van Luit, J. E. H., & Van de Rijt, B. A. M. (2009). *Utrechtse getalbegrip toets-Revised [Early numeracy test-Revised]*. Doetinchem: Graviant.

Bijlage 1

Beste leerkracht,

Wij willen u vragen om voor elke leerling de volgende vraag te beantwoorden: "Hoe is het begripsniveau van voorbereidend rekenen van de leerling in vergelijking met het gemiddelde begripsniveau van zijn of haar klasgenoten?" U mag het getal dat van toepassing is voor de desbetreffende leerling omcirkelen.

Wij zullen deze informatie gebruiken om onze onderzoeksvraag "Is er een relatie tussen het leerkrachtoordeel van kinderen uit groep 2 en 3 en hun prestaties op de Utrechtse Getalbegrip Toets – Revised?" te beantwoorden. Het kost ongeveer 15 minuten tijd om dit in te vullen voor alle leerlingen uit de klas.

Hartelijk bedankt voor de medewerking.

Naam leerling 1. Lager 2. Beneden gemiddeld 3. Gemiddeld 4. Iets beter 5. Hoger	Naam leerling 1. Lager 2. Beneden gemiddeld 3. Gemiddeld 4. Iets beter 5. Hoger
Naam leerling 1. Lager 2. Beneden gemiddeld 3. Gemiddeld 4. Iets beter 5. Hoger	Naam leerling 1. Lager 2. Beneden gemiddeld 3. Gemiddeld 4. Iets beter 5. Hoger