Differentiatie in de Rekenles: Zelfinschatting en Observatie van Leerkrachtgedrag

Masterthesis

Universiteit Utrecht
Masteropleiding Pedagogische Wetenschappen
Masterprogramma Orthopedagogiek

Door : A. (Andries) Kamminga (3973034)
Begeleider: Emilie Prast
Tweede beoordelaar: Hans van Luit
Datum: 20 juni 2014

Voorwoord

Dit onderzoek vormt de masterthesis van mijn studie orthopedagogiek aan de Universiteit Utrecht en maakt deel uit van het project ‘Ieder kind heeft recht op gedifferentieerd rekenonderwijs’ (GROW). Het oorspronkelijke onderzoeksplan heb ik samen met Susan Gaasbeek geschreven, waarna wij ieder verschillende onderzoeksvragen betreffende het leerkrachtgedrag onafhankelijk van elkaar hebben uitgewerkt. Het observatieinstrument waar wij gebruik van maakten is in de onderzoeksperiode vele malen gescoord en daarvoor ben ik mijn medestudenten Elizabeth, Henriëtte, Sarah, Susan en Tessa zeer erkentelijk. Een bijzonder woord van dank gaat uit naar mijn thesisbegeleider Emilie Prast en ook naar Hans van Luit, die haar in de beginfase korte tijd heeft vervangen. Door hun kritische blik en aanmoediging kon ik deze scriptie voortdurend verbeteren en voltooien. Tenslotte wil ik Rita bedanken voor de vele ongezellige uren die ze zonder mij heeft moeten doorbrengen, terwijl ik op zaterdag- en zondagochtend ‘filmpjes scoorde’ en op bijna alle andere vrije momenten anderszins aan deze scriptie werkte. (Ik beloof je dat we nu weer leuke dingen gaan doen!) Met dit onderzoek hoop ik mijn bijdrage te hebben geleverd aan de wetenschap en ik wens dat deze bijdrage op de een of andere manier zijn weg zal vinden naar de praktijk van alledag. Deze uitdaging laat ik graag aan de lezer!

Samenvatting

Leerkrachten in het primair onderwijs staan voor de taak alle leerlingen beter te leren rekenen. Om in de rekenles beter te leren differentiëren naar de onderwijsbehoeften van de leerlingen namen 430 leerkrachten van de groepen 3 tot en met 8 van 32 basisscholen in Nederland deel aan het driejarige project 'Gedifferentieerd RekenOnderwijs' (GROW). In het jaar 2012-2013 kreeg het eerste cohort van leerkrachten de GROW-training aangeboden en de cohorten die in de daarop volgende jaren zouden worden getraind dienden als controlegroep. Deze studie meet het effect van de training op het leerkrachtgedrag met behulp van een leerkrachtvragenlijst en deze meting wordt vervolgens vergeleken met het geobserveerde leerkrachtgedrag. De gebruikte analysetechnieken zijn MANOVA en Pearson correlatie. Volgens de inschatting van de leerkrachten zelf is er een klein significant positief effect van de GROW-training op het differentiatiegedrag van leerkrachten ten aanzien van het vaststellen van en het stellen van gedifferentieerde doelen. Op schaalniveau is echter geen samenhang gevonden tussen het zelfingeschatte en het geobserveerde differentiatiegedrag. Vervolgonderzoek bij de volgende trainingsgroepen en follow-up onderzoek moet uitwijzen of de training inderdaad het beoogde effect heeft op het differentiatiegedrag van leerkrachten, of dit effect blijvend is en hoe dit effect kan worden vergroot.

Kernwoorden: differentiatie, rekenonderwijs, zelfinschatting, observatie

Abstract

One of the responsibilities of teachers in primary education is to enhance all pupils' mathematical skills. Four hundred and thirty teachers teaching grades 3 to 8 in 32 primary schools in the Netherlands aimed to improve the differentiation in their maths lessons to meet the educational needs of their pupils and therefore participated in the three year project 'Differentiated Math Education'(GROW). In the academic year 2012-2013 the first group of teachers was offered the GROW-training, while a second and third group, to be trained subsequently, served as a control. This study measures the effect of the training on teachers' behaviour by the use of a self-assessment questionnaire and results will be compared to scores obtained by observation. The analysis was carried out using MANOVA and Pearson's correlation. According to the teachers' own assessment the GROW-training has had a small significant positive effect on teachers' differentiation behaviour with respect to assessing pupils' educational needs (R^2 =.034) and setting differentiated goals (R^2 =.025). However, at scale level no correlation was found between self-assessed and observed behaviour. Further research and follow-up studies will have to reveal firstly whether the training will bring about

the intended change in teachers' behaviour, secondly whether this change will be sustainable and finally how this effect can be increased.
Keywords: differentiation, maths education, self-assessment, observation

Differentiatie in de Rekenles: Zelfinschatting en Observatie van Leerkrachtgedrag

Het Nederlandse reken- en wiskundeonderwijs heeft internationaal een goede naam, maar volgens het *Programme for International Student Assessment* (PISA) van de Organisatie voor Economische Samenwerking en Ontwikkeling (OESO) is Nederland inmiddels zijn plek in de top 10 van deelnemende landen kwijtgeraakt. Nederland zakte in drie jaar tijd van de vijfde naar de elfde plaats in 2009. Er is voldoende aandacht voor zwakke rekenaars, maar er lijkt in Nederland weinig aandacht voor de excellente rekenaars (OECD, 2010). De commissie 'Rekenonderwijs op de basisschool' van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) onderzocht de relatie tussen rekendidactiek en rekenvaardigheid in het Nederlandse basisonderwijs en kwam tot de conclusie dat de kwaliteit van de leerkracht doorslaggevend is voor de rekenprestaties van de leerlingen en dat de sleutel voor verbetering ligt in de opleiding en nascholing van de leerkracht (KNAW, 2009). Onderzoek laat inderdaad zien dat voor een verbetering van de rekenprestaties van de kinderen de leerkracht het verschil maakt: onderwijstijd, didactische vaardigheden en interactie tussen leerkracht en leerling blijken belangrijker dan bijvoorbeeld de rekenmethode (Gelderblom, 2010; Nye, Konstantopoulus, & Hedges, 2004). Bij succesvol rekenonderwijs, dat wil zeggen rekenonderwijs waardoor alle leerlingen beter gaan presteren, gaat de deskundigheid van de leerkrachten en de kwaliteit van de interactie tussen leraar en leerling boven de didactiek en het curriculum (Gelderblom, 2007; Slavin & Lake, 2008). Onderwijsmedewerkers geven zelf in een peiling van onderzoeksbureau TNS-NIPO aan bij- en nascholing belangrijk te vinden om de prestaties van de leerlingen te verbeteren, maar er is één maatregel die ze nog belangrijker achten, namelijk meer aandacht voor de persoonlijke ontwikkeling van de leerling (Bulsink, 2010). Het een spreekt het ander niet tegen en de combinatie van een betere scholing van leerkrachten en een betere aansluiting van het onderwijsaanbod bij de onderwijsbehoeften van de individuele leerling is dan ook het onderwerp van deze studie

Het internationale onderzoek *Trends in International Mathematics and Science Study* (TIMSS) uit 2007 vergeleek de rekenvaardigheid van leerlingen in groep 6 van Nederlandse basisscholen met die van leeftijdsgenoten in andere landen (Mullis et al., 2008). De Nederlandse kinderen bezetten ondanks een significante daling ten opzichte van 1995 een

verdienstelijke negende plaats. De Nederlandse leerkracht voelde zich gemiddeld meer dan voldoende toegerust om les te geven in rekenen en wiskunde en bleek in vergelijking tot leerkrachten in andere landen het minst aan nascholing in rekenen te doen. Opmerkelijk is dat de Nederlandse leerkracht het laagst scoorde op 'het uitleggen van antwoorden', maar juist het hoogst op 'zelfstandig werken aan opgaven' (KNAW, 2009). Mogelijk wordt het belang van instructie in Nederland onderschat. Een andere verklaring voor de geringe animo voor nascholing en de terugloop van de instructietijd wordt door Valli en Buese (2007) gezocht in de veranderende rol van de leerkracht in wat Nederlanders waarschijnlijk een 'afrekencultuur' zouden noemen. Gedifferentieerde instructie, volgens Tomlinson (2005) het proactief, flexibel en multi-mediaal bieden van (reken-)instructie voor verschillende niveaus in de klas, blijkt een forse verzwaring van de taak voor leerkrachten. Zeker bij meer includerend onderwijs en wanneer er veel tijd en energie gaat zitten in bijvoorbeeld het afnemen van toetsen, is het voor veel leerkrachten moeilijk om dat alles op eigen kracht te doen (Valli & Buese, 2007). Differentiëren vraagt voorbereiding en voorbereidingstijd (Janson, 2012). In het eerder genoemde rapport van de KNAW wordt dan ook niet voor niks aanbevolen om nascholing in het rekenonderwijs krachtig te stimuleren door begeleiding te bieden op de werkvloer en nascholing onder schooltijd te laten plaatsvinden (KNAW, 2009).

In het Delphi onderzoek naar gedifferentieerd primair rekenonderwijs (Van de Weijer-Bergsma & Prast, 2013) gaven rekenexperts aan welke kennis en vaardigheden leerkrachten nodig hebben om beter te kunnen differentiëren in de rekenles, waarbij aangenomen is dat leerlingen meer zullen profiteren van de rekenles als de instructie en verwerking nauwer aansluiten bij de onderwijsbehoeften van de leerlingen. Deze aanname lijkt voor de hand te liggen voor zwakke rekenaars, maar ook begaafde leerlingen blijken gebaat te zijn bij instructie op maat (Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004). De uitkomsten van dit Delphi onderzoek vormden de basis van de tiendaagse training 'Ieder kind heeft recht op gedifferentieerd rekenonderwijs' (GROW), waarin leerkrachten kennis en vaardigheden opdoen met betrekking tot convergente differentiatie. Bij convergente differentiatie werken alle leerlingen gedurende de rekenles in hun eigen groep aan de rekenstof die de methode voorschrijft, waarbij zwakke, gemiddelde en sterke rekenaars werken aan verschillende doelstellingen, variërend van fundamenteel tot uitdagend (www.gedifferentieerdrekenonderwijs.nl). Zoals door de KNAW werd aanbevolen is de GROW-training aan hele schoolteams aangeboden, zodat die onder schooltijd op studiedagen kon worden gegeven en er zijn per school coaches opgeleid om de leerkrachten op de werkvloer te begeleiden. De GROW-training beoogt een zevental vaardigheden, 'bouwstenen'

genoemd, bij de leerkrachten te bevorderen, namelijk onderwijsbehoefte vaststellen, doelen stellen, differentiatie in instructie, differentiatie in verwerking, evaluatie, organisatie en voorwaarden voor differentiatie. Er zijn doelen geformuleerd voor leerkrachten en ook voor gevorderden en experts. Inmiddels heeft een eerste groep leerkrachten deze training gevolgd en in dit artikel wordt verslag gedaan van de eerste resultaten in termen van leerkrachtgedrag.

Om het gedrag van individuele leerkrachten te onderzoeken worden de meetinstrumenten zelfrapportage, interview en observatie het vaakst toegepast, waarbij elk instrument zowel voordelen als nadelen heeft (Desimone, 2009). In deze studie vinden de metingen plaats middels zelfrapportage en observatie. Zelfrapportage is een krachtig middel om het eigen functioneren van de leerkracht te verbeteren, maar wordt krachtiger in combinatie met observatie en feedback door collega's en externe deskundigen, aangezien er altijd een gevaar schuilt in onder- en overschatting van de eigen kennis en vaardigheid (Ross & Bruce, 2007). Deze onder- of overschatting kan onbewust plaatsvinden, maar in zelfrapportagelijsten kunnen respondenten ook bewust sociaal wenselijke antwoorden geven, zeker bij gevoelige onderwerpen (Tourangeau & Yan, 1995). Observatie is een andere mogelijkheid om het differentiatiegedrag van de leerkracht te onderzoeken. Voorbeelden van observatieinstrumenten voor getrainde observatoren met een goede interbeoordelaarsbetrouwbaarheid zijn de Classroom Instructional Practices Scale (CIPS) en de Classroom Observation Scale Revised (COS-R), maar deze instrumenten richten zich vooral op differentiatie ten aanzien van begaafde leerlingen en zijn niet specifiek voor de rekenles ontwikkeld (Johnsen, Haensley, Ryser, & Ford, 2002; Van Tassel-Baska, Quek, & Feng, 2007). Studies waarin leerkrachtgedrag met verschillende instrumenten wordt gemeten zijn niet eensluidend over de samenhang tussen de instrumenten (Desimone, 2009). Onderzoekers uit de jaren '60 en '70 vonden heel weinig samenhang tussen zelfrapportage en klassenobservaties (Hook & Rosenshine, 1979). Later werden er echter wel matige tot sterke correlaties gevonden, zeker wanneer er meerdere observaties werden gedaan en vooral werd gekeken naar feitelijk kwantificeerbaar gedrag (Mayer, 1999; Ross, McDougall, Hogaboam-Gray, & LeSage, 2003). Mogelijk heeft dit te maken met de voordelen die het observeren met behulp van de camera biedt (Jacobs, Kawanaka, & Stigler, 1999; Borko, Jacobs, Eiteljorg, & Pittmann, 2008). In een recent onderzoek met betrekking tot het integreren van informatietechnologie in het onderwijs, bleek overigens dat de geobserveerde praktijk minder positief uitpakte dan op grond van lesplan en zelfrapportage werd verondersteld (Ageyi & Voogt, 2011).

Om het effect van de GROW-training op het leerkrachtgedrag te meten zijn twee nieuwe meetinstrumenten ontwikkeld, gericht op differentiatie in de rekenles, namelijk de Differentiatie Zelfinschattingsvragenlijst [Differentiation Self-Assessment Questionnaire, (DSAQ)] en de Observatieschaal Differentiatie in Mathematische Instructie [Observation scale Differentiated Mathematical Instruction (DMI)] (Prast, Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2013a, 2013b). Beide instrumenten zullen in de methode sectie worden toegelicht. In dit artikel worden de opbrengsten van de GROW-training beschreven in termen van verandering in differentiatiegedrag in de rekenles, zoals gemeten met de vragenlijst DSAQ. Daarnaast zullen de scores op één van de schalen van de DSAQ worden vergeleken met één van de schalen van het observatieinstrument DMI, waarbij de gekozen schalen min of meer hetzelfde gedrag beschrijven. Het zal duidelijk zijn dat het gebruik van een vragenlijst vele malen eenvoudiger en goedkoper is om te meten of een leerkracht van de training geprofiteerd heeft dan het inzetten van het observatieinstrument (Desimone, 2009). Als beide instrumenten een vergelijkbare verandering aangeven, kan dit een overweging zijn om in vervolgonderzoek alleen de vragenlijst in te zetten. De onderzoeksvragen zijn: ‘Wat is het effect van de GROW-training op de zelfingeschatte *verandering* in differentiatiegedrag van de leerkrachten in de rekenles?’ en ‘Wat is de samenhang tussen de *zelfingeschatte* en de *geobserveerde* verandering in differentiatiegedrag van de leerkrachten in de rekenles?’ Aangezien het bij de GROW-training om een doelbewuste interventie gaat om het differentiatiegedrag van de leerkrachten in de rekenles te bevorderen is de verwachting dat de verandering in het gedrag een positieve richting heeft en dat dit met beide onderzoeksinstrumenten kan worden gemeten. De onderzoekshypothesen luiden dan als volgt: ‘De zelfingeschatte verandering in differentiatiegedrag van de trainingsgroep laat een positief significant verschil zien ten opzichte van de zelfingeschatte verandering in differentiatiegedrag van de controlegroep’ en ‘De zelfingeschatte verandering in differentiatiegedrag van de trainingsgroep hangt positief samen met de geobserveerde verandering in differentiatiegedrag’.

Methode

Participanten

De te onderzoeken populatie bestaat uit leerkrachten van de groepen 3 tot en met 8 in het Nederlands primair onderwijs. De steekproef bestaat uit alle leerkrachten die in het schooljaar 2012-2013 deelnamen aan de GROW-training (experimentele groep, cohort 1) en uit die welke in schooljaar 2013-2014 of 2014-2015 deelnemen aan de training (controlegroep, cohort 2 en 3). De scholen hebben zich vrijwillig voor de training aangemeld,

maar zijn vervolgens aselect over de cohorten verdeeld, zodat er een experimentele en een controlegroep zijn gevormd. De deelnemende scholen staan verspreid over het land en kennen uiteenlopende karakteristieken: er zijn grote en kleine scholen, dorpsscholen en stadsscholen, openbare scholen en scholen met een specifieke geloofsachtergrond, scholen met en zonder combinatiegroepen. Het gaat om volledige schoolteams en alle leerkrachten van de groepen 3 tot en met 8 van een geselecteerde school doen per definitie mee, zodat de representativiteit van de steekproef niet nadelig wordt beïnvloed door de bereidwilligheid van de individuele leerkracht. Mogelijk is er wel sprake van enige bias doordat de leerkrachten in de controlegroep weten dat ze de training nog gaan volgen. In 2012-2013 is de vragenlijst DSAQ door 277 leerkrachten ingevuld en in 2013-2014 door 258 leerkrachten. Omdat het in dit onderzoek om verschilscores gaat zijn voor de analyses alleen die participanten meegenomen bij wie twee metingen zijn verricht en bleken er 176 geschikt voor analyse. In 2012-2013 is het observatieinstrument DMI minimaal eenmaal afgenomen bij 83 leerkrachten, waarvan er nog 51 geschikt bleken voor analyse, aangezien bij deze participanten minimaal één voormeting en één nameting plaatsvond. Voor de correlatievraag ten slotte konden de gegevens van 33 participanten worden geanalyseerd bij wie zowel een verschilscore op de vragenlijst als een verschilscore op het observatieinstrument kon worden berekend.

Meetinstrumenten

Het gaat hier om een quasi-experimenteel onderzoeksdesign met een experimentele groep en een controlegroep, een voormeting (VM) en een nameting (NM). Het zelfingeschatte leerkrachtgedrag wordt gemeten met de DSAQ en het geobserveerde leerkrachtgedrag met de DMI.

De DSAQ bevindt zich nog in een experimenteel stadium en bestaat uit een algemeen descriptief deel en dertien kwantificerende schalen, waarvan vijf zelfinschattingsschalen met betrekking tot differentiatiegedrag (Prast, Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2013a). Bij deze vijf zelfinschattingsschalen gaat het om het inschatten van het eigen differentiatiegedrag ten aanzien van Onderwijsbehoefte, Doelen stellen, Instructie, Verwerking en Evaluatie. Er is voor deze schalen gekozen omdat in dit onderzoek het gedrag van de leerkracht centraal staat en niet diens attitudes, kennis of self-efficacy. De items van de schalen worden gescoord op een vijfpunts Likertschaal, van ‘helemaal niet van toepassing op mij’ tot en met ‘helemaal van toepassing op mij’. Een voorbeeld van een item uit de schaal Zelfinschatting Differentiatiegedrag Instructie (ZI) is ‘Ik stel bewust vragen van verschillende moeilijkheidsgraad tijdens de klassikale instructie’. De betrouwbaarheid van de zelfinschattingsschalen is goed (Cronbach’s α = .85).

Ook het observatieinstrument DMI dat zich richt op de differentiatie in de rekenles is nieuw (Prast, Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2013b). Het observatieinstrument bestaat naast een descriptief deel uit een viertal kwantificerende schalen, namelijk Breed Aanbod (BA), Klassenmanagement, Differentiatie voor Zwakke Rekenaars en Differentiatie voor Sterke Rekenaars. Elke schaal is opgebouwd uit een aantal items die op een vijfpuntsschaal worden gescoord, waarbij elk item weer gebaseerd is het voorkomen van *good practice* voorbeelden, gescoord op een driepuntsschaal. Eén van de vijf BA items is: 'De leerkracht brengt de stof op verschillende manieren over' en één van de vier *good practice* voorbeelden bij dit item is 'De leerkracht biedt dezelfde stof aan op verschillende handelingsniveaus'. De interbeoordelaarsbetrouwbaarheid is voor alle schalen laag en dat geldt ook voor de schaal Breed Aanbod (Cronbach's α =.55). De interne consistentie van de schalen is goed, ook voor de schaal Breed Aanbod (Cronbach's α = .88) (Field, 2009). De validiteit van beide instrumenten is nog niet onderzocht.

Voor de vergelijking tussen de DSAQ en de DMI wordt gebruik gemaakt van de schaal Zelfinschatting Differentiatiegedrag Instructie (ZI) uit de DSAQ en de schaal Breed Aanbod (BA) uit de DMI, omdat beide schalen min of meer hetzelfde construct meten, namelijk de input van de leerkracht in de klassikale instructie. Zie voor een overzicht van de items van deze schalen Bijlage A en voor een vergelijking van beide schalen Bijlage B

Procedure

In het schooljaar 2011-2012 is de GROW-training ontwikkeld en op een viertal scholen in verkorte vorm als pilot uitgevoerd. Voor het hoofdonderzoek en de gehele training konden schoolteams zich inschrijven voor deelname. Alleen de scholen die zich met het gehele team inschreven kwamen voor de training in aanmerking. In verband met een beperkte capaciteit zijn alleen de eerste 32 scholen die zich aanmeldden toegelaten tot het onderzoek. In het schooljaar 2012-2013 hebben in alle deelnemende scholen metingen plaatsgevonden met de DSAQ en de DMI. In het begin van het jaar vond een voormeting plaats, waarbij alle deelnemende leerkrachten de DSAQ invulden en er bij elke deelnemende leerkracht binnen enkele weken tijd twee maal een rekenles werd gefilmd om de DMI te scoren. Vervolgens volgden de leerkrachten van het eerste cohort, de experimentele groep, het nascholingstraject GROW. Het nascholingstraject bestaat voor de leerkrachten uit tien bijeenkomsten van 3 of 4 uur, waarvan vier onder leiding van een projectcoach op de eigen school. De leerkrachten van cohort 2 en 3, de controlegroep, volgden deze training niet. Aan het einde van schooljaar 2012-2013 werd als nameting bij alle deelnemende leerkrachten opnieuw tweemaal een rekenles gefilmd en hebben de deelnemende leerkrachten aan het begin van schooljaar 2013-

2014 opnieuw de DSAQ ingevuld. In het schooljaar 2013-2014 zijn observanten getraind in het gebruik van de DMI en is de interbeoordelaarsbetrouwbaarheid bepaald. Tenslotte zijn alle videopnamen met behulp van de DMI gescoord en ingevoerd in het statistiekprogramma SPSS, evenals de scores op de DSAQ, waarna de data konden worden geanalyseerd. Om de data voor het doel van dit onderzoek te prepareren werden per leerkracht de verschilscores tussen voor- en nameting berekend, zowel voor de zelfinschattingsvragen van de DSAQ als voor de schaal Breed Aanbod van de DMI. Voor dit laatste was het nodig om de data verkregen uit de twee video-opnamen van de voormeting te aggregeren en datzelfde te doen voor de twee opnamen van de nameting.

Data analyse

Om de verschilscores op de zelfinschattingsschalen differentiatiegedrag van de DSAQ te vergelijken tussen de experimentele groep en de controlegroep is gebruik gemaakt van *Multivariate Analysis of Variance* (MANOVA). Hiermee kan worden aangetoond op welke schalen de leerkrachten in de experimentele groep een groter verschil in het eigen differentiatiegedrag rapporteren dan de leerkrachten in de controlegroep. Om de samenhang tussen de zelfingeschatte verandering in leerkrachtgedrag en de geobserveerde verandering in leerkrachtgedrag te onderzoeken is gebruik gemaakt van de Pearson correlatie.

Resultaten

De beschrijvende statistieken voor de verschilscores op de zelfinschattingsschalen van de DSAQ zijn weergegeven in Tabel 1 en die voor de verschilscores op de schaal Breed Aanbod van de DMI in Tabel 2.

Tabel 1

Gemiddelden en Standaardafwijkingen van Verschilscores (nameting-voormeting) op de Zelfinschattingsschalen van de DSAQ.

	Experimentele groep			Controlegroep		
	n	*M*	*SD*	*n*	*M*	*SD*
Onderwijsbehoefte	68	0.16	0.45	108	-0.03	0.45
Doelen	68	0.24	0.46	108	0.07	0.48
Instructie	67	0.06	0.44	107	0.07	0.36
Verwerking	67	0.14	0.50	106	0.05	0.45
Evaluatie	67	0.19	0.45	106	0.12	0.49

Tabel 2

Gemiddelden en Standaardafwijkingen van Verschilscores (nameting-voormeting) op de schaal Breed Aanbod van de Observatieschaal Differentiatie in Mathematische Instructie

	Experimentele groep			Controlegroep		
	n	*M*	*SD*	*n*	*M*	*SD*
Breed Aanbod	24	-.08	.40	27	.05	.28

Voor de eerste onderzoeksvraag betreffende de zelfingeschatte verandering in differentiatiegedrag van de leerkrachten in de rekenles is een MANOVA uitgevoerd, waarvan de resultaten zijn weergegeven in Tabel 3. Aan de assumpties voor een MANOVA is voldaan. Er is een significant effect van de GROW-training op de zelfingeschatte verandering in differentiatiegedrag ten aanzien van het vaststellen van de onderwijsbehoeften van de leerlingen, $F(1,171) = 7.07$, $p < .01$, $R^2 = .034$. Er is eveneens een significant effect van de GROW-training op de zelfingeschatte verandering in differentiatiegedrag ten aanzien van het stellen van (gedifferentieerde) doelen, $F(1,171) = 5.44$, $p < .05$, $R^2 = .025$. In beide gevallen gaat het om een klein effect (Field, 2009). Er is geen significant effect gemeten van de GROW-training op de zelfingeschatte verandering in differentiatiegedrag ten aanzien van instructie, verwerking en evaluatie.

Tabel 3

Zelfingeschatte Verandering in Differentiatiegedrag

	MS	*F*	*p*	R^2
Onderwijsbehoefte	1.46	7.07*	<.01	.034
Doelen	1.22	5.44**	.02	.025
Instructie	<0.01	0.04	.83	-.006
Verwerking	0.36	1.61	.21	.004
Evaluatie	0.19	0.86	.36	-.001

*p<.01, **p<.05

De tweede onderzoeksvraag betreffende de samenhang tussen het zelfingeschatte en het geobserveerde differentiatiegedrag van de leerkrachten in de rekenles is geanalyseerd met behulp van de Pearson correlatie. Hiervoor zijn de scores op de observatieschaal BA en de zelfinschattingsschaal ZI gebruikt, zowel bij de voormeting als bij de nameting, alsmede de verschilscores op beide schalen. De resultaten zijn weergegeven in Tabel 4. Er is op schaalniveau geen significante correlatie gevonden. Dit geldt voor de voormeting, de

nameting en de verschilscores. De spreidingsdiagrammen van de verschilscores op de genoemde schalen in de experimentele, controle en totale groep en van de schaalscores op de voormeting bevestigen dit beeld.

Tabel 4

Correlatie tussen Zelfingeschat(te) en Geobserveerd(e) (verandering in) Differentiatiegedrag (schaalniveau)

	Experimentele groep			Controlegroep			Totaal		
	n	*p*	*r*	*n*	*p*	*r*	*n*	*p*	*r*
Schaalscores Voormeting									
DSAQ ZI en DMI BA	28	.17	.27	22	.78	.07	50	.18	.19
Schaalscores Nameting									
DSAQ ZI en DMI BA	21	.18	-.31	27	.41	-.16	48	.16	-.21
Verschilscores									
DSAQ ZI en DMI BA	19	.47	.18	14	.92	.03	33	.53	.12

Op itemniveau is er alleen samenhang gevonden voor de controlegroep, zoals weergegeven in Tabel 5, maar niet voor de experimentele of de totale groep. Ook hiervoor is gebruik gemaakt van de Pearson correlatie. Voor de inhoud van de items zie Bijlage 1. In de controlegroep is een significante correlatie gevonden tussen items die de geobserveerde verandering in differentiatiegedrag meten en items die de zelfingeschatte verandering in differentiatiegedrag meten voor de items DSAQ-ZI 1 en DMI-BA 2 (r = -.64, n = 14, p < .05, tweezijdig), voor de items DSAQ-ZI 1 en DMI-BA 3 (r = -.54, n = 14, p < .05, tweezijdig) voor de items DSAQ-ZI 3 en DMI-BA 4 (r = -.71, n = 14, p < .01, tweezijdig), voor de items DSAQ-ZI 4 en DMI-BA 2 (r = .55, n = 14, p < .05, tweezijdig) en voor de items DSAQ-ZI 5 en DMI-BA 2 (r = .62, n = 14, p < .05, tweezijdig). In alle gevallen gaat het om een sterke correlatie, r > .5. Er is behalve naar de correlatie van verschilscores ook gekeken naar de correlatie van itemscores in de voormeting. Er is een significante correlatie gevonden tussen het geobserveerde en zelfingeschatte differentiatiegedrag voor de items DMI-BA 1 en DSAQ-ZI 3 (r = .29, n = 51, p < .05, tweezijdig) en voor de items DMI-BA 1 en DSAQ-ZI 5 (r = .38, n = 51, p < .01, tweezijdig) bij de voormeting in de totale groep (experimentele en controlegroep samen).

Tabel 5

Correlatie tussen Zelfingeschatte en Geobserveerde Verandering in Differentiatiegedrag in de Controlegroep (itemniveau)

DMI Breed Aanbod		Item 1	Item 2	Item 3	Item 4	Item 5
DSAQ Instructie		*n*=15	*n*=14	*n*=14	*n*=14	*n*=15
Item 1	*r*	.40	-.64*	-.54*	.27	.22
	p	.14	.01	.05	.34	.43
Item 2	*r*	.02	-.18	-.06	.21	-.02
	p	.93	.55	.85	.47	.94
Item 3	*r*	.06	-.05	.01	-.71**	.14
	p	.82	.87	.97	.00	.61
Item 4	*r*	.18	.55*	.27	.05	.10
	p	.52	.04	.35	.86	.73
Item 5	*r*	-.08	.62*	.17	-.09	-.01
	p	.78	.02	.55	.76	.96
Item 6	*r*	.09	-.19	-.06	.05	.22
	p	.76	.51	.84	.86	.43
Item 7	*r*	-.49	.18	-.17	.15	-.36
	p	.06	.54	.55	.62	.18

*$p < .05$, ** $p < .01$

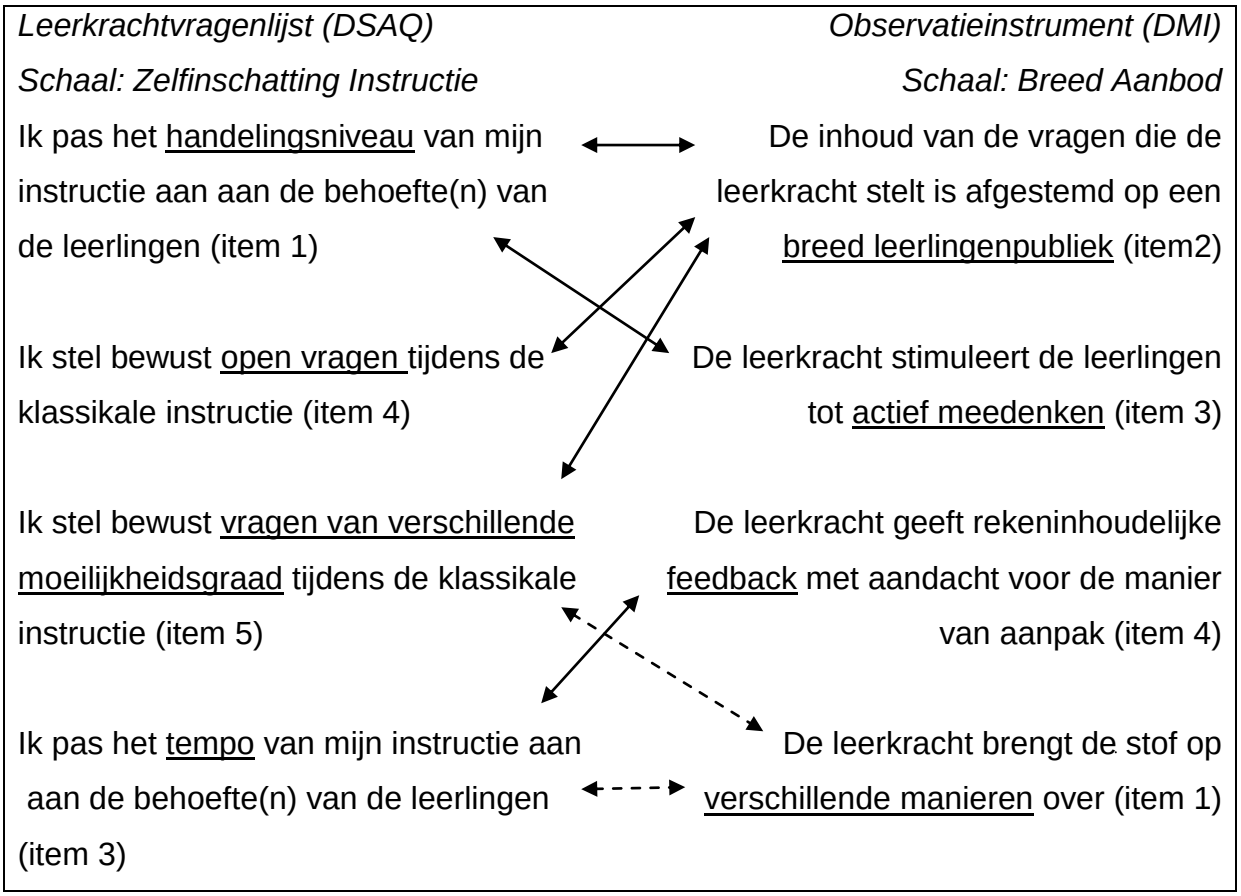
Discussie en conclusie

De GROW-training lijkt gemeten met de DSAQ op twee van de vijf onderzochte schalen een klein positief effect te hebben op het differentiatiegedrag van de leerkrachten, namelijk op de schalen Onderwijsbehoefte en Doelen stellen. Dit betekent dat de getrainde leerkrachten naar eigen zeggen na de training beter in staat zijn om rekentoetsen en –taken te analyseren en de onderwijsbehoefte van specifieke leerlingen door observatie of diagnostisch gesprek in te schatten. Het betekent ook dat de getrainde leerkrachten volgens hun eigen inschatting beter in staat zijn om verschillende doelen te hanteren voor leerlingen van verschillende rekenniveaus, dat ze de mogelijkheden van de rekenmethode hiervoor beter kennen en deze ook beter benutten. Op de andere schalen, Instructie, Verwerking en Evaluatie, lijken de verschilscores weliswaar positief, maar niet significant groter dan die van de leerkrachten die de training nog niet gevolgd hebben.

Dat de getrainde leerkrachten juist op deze twee schalen een positieve verandering in het eigen differentiatiegedrag rapporteren, heeft er mogelijk mee te maken dat beide ‘bouwstenen’ als eerste in de training aan bod komen. De leerkrachten hebben dus langer de tijd gehad om zich deze vaardigheden eigen te maken. Een andere verklaring kan zijn dat bij de voormeting de experimentele groep op beide schalen gemiddeld lager scoorde dan de controlegroep, in tegenstelling tot de andere drie schalen. Daarmee was er voor de leerkrachten in de experimentele groep meer ‘groeiruimte’ ten opzichte van de leerkrachten in de controlegroep. Bij de nameting blijkt de getrainde groep de achterstand inderdaad te hebben ingehaald en zijn de scores op alle schalen hoger dan die van de niet-getrainde groep. Beide schalen laten zich echter moeilijker observeren in de groep dan bijvoorbeeld de schalen Instructie en Verwerking. In de wetenschap dat de leerkrachten ook geobserveerd worden is het voor de trainingsgroep wellicht ‘veiliger’ om op die onderdelen vooruitgang te rapporteren die minder gemakkelijk door observatie gecrossvalideerd kunnen worden. Hier speelt de bias van sociale wenselijkheid wellicht een rol (Tourangeau & Yan, 1995). De getrainde leerkrachten zullen namelijk wensen dat ze baat hebben gehad bij de training, aangezien ze er zelf tijd en energie in hebben gestopt. Bovendien zullen ze graag de projectleiders, coaches en directies tevreden willen stellen.

De schaal Instructie lijkt de schaal te zijn waarin de leerkracht het meest uitdrukkelijk direct observeerbaar gedrag laat zien. Deze schaal van de DSAQ is dan ook gekozen om de tweede onderzoeksvraag, de vraag naar de samenhang tussen zelfingeschat en geobserveerd gedrag, te beantwoorden. Op schaalniveau is er echter geen samenhang gevonden tussen de observatieschaal Breed Aanbod en de zelfinschattingsschaal Instructie. Dit is niet in lijn met de verwachtingen, aangezien beide schalen ongeveer hetzelfde construct lijken te meten. Op itemniveau zijn wel enkele correlaties gevonden. Deze itemcorrelaties zijn weergeven in Figuur1. Voor de verschilscores werden alleen in de controlegroep correlaties gevonden en niet in de experimentele groep of in de gehele groep. Deze controlegroep is klein (n = 15) waardoor deze correlaties weinig zeggingskracht hebben. Wel komen er twee van de vier op voorhand verwachte correlaties naar voren, zoals vermeld in Bijlage B. De meeste waarde lijkt te kunnen worden gehecht aan de items van de twee instrumenten die bij de voormeting in de gehele groep (n = 51) blijken te correleren, namelijk ‘De leerkracht brengt de stof op verschillende manieren over’ uit de DMI met ‘Ik stel bewust vragen van verschillende moeilijkheidsgraad tijdens de klassikale instructie’ en ‘Ik pas het tempo van mijn instructie aan aan de behoefte(n) van de leerlingen’ uit de DSAQ. Deze items vertonen inhoudelijk een gelijkenis, want onder ‘verschillende manieren van het overbrengen van lesstof’ kan

inderdaad zowel ‘vragen stellen van verschillende moeilijkheidsgraad’ als ‘het tempo van de instructie aanpassen’ worden verstaan.



Figuur 1. Correlaties op itemniveau tussen DMI (schaal Breed Aanbod) en DSAQ (schaal Zelfinschatting Instructie): verschilscores in de controlegroep (stippellijn) en itemscores van de totale groep in de voormeting (doorgetrokken lijn).

Toch is de samenhang tussen de instrumenten hiermee niet duidelijk aangetoond. Een sterke samenhang tussen de beide schalen zou aannemelijk hebben gemaakt dat het effect van de interventie op het leerkrachtgedrag zowel door het ene als door het andere instrument kan worden gemeten, zodat in de toekomst ook voor slechts één van de meetinstrumenten gekozen kan worden. Echter nu deze samenhang niet of nauwelijks gevonden is moet voorlopig worden aangenomen dat de gekozen schalen van beide instrumenten ieder een ander construct meten. Het is dan ook niet de vraag welk van de twee instrumenten een beter beeld geeft van de (verbeterde) differentiatievaardigheden van de leerkrachten, maar meer welk construct de bedenkers van de GROW-training precies willen meten, het gedrag van leerkrachten in de perceptie van de observator of in de perceptie van de leerkrachten zelf. De instrumenten leveren niet dezelfde informatie op, maar kunnen beide waardevol zijn indien ze goed geconstrueerd zijn, zoals Desimone (2009) betoogt. Simultaan met deze studie loopt een

onderzoek van de Universiteit Utrecht waarin nader wordt ingegaan op de metingen verricht met het observatieinstrument DMI. Mogelijk wordt in dat onderzoek duidelijk of er, net als met de DSAQ, positieve effecten van de training gemeten worden of dat de leerkracht zichzelf overschat, zoals in het onderzoek van Ageyi en Voogt (2011).

Een mogelijke verklaring voor de geringe samenhang is voor wat betreft de verschilscores gelegen in het moment waarop de DSAQ als nameting is afgenomen, namelijk aan het begin van het volgende schooljaar. Dit in tegenstelling tot de DMI, waarvan de voor- en nameting in hetzelfde schooljaar plaatsvonden. Hoewel de leerkrachten dezelfde gebleven zijn is de context voor veel leerkrachten na de zomervakantie waarschijnlijk een andere dan voor de zomervakantie, bijvoorbeeld doordat de groep kinderen een andere is. Het is te verwachten dat leerkrachten in juni hun klas beter kennen dan in september en beter kunnen inschatten welke leerlingen welke instructie behoeven. Bovendien moet een nieuwe groep ook aan de onderwijsstijl van de leerkracht wennen en moeten routines worden ingeslepen. Mogelijk is aan het begin van een schooljaar aan de voorwaarden voor differentiatie, zoals die genoemd worden in de GROW-training, minder goed voldaan dan aan het einde van een schooljaar. Hierbij valt te denken aan een goed pedagogisch klimaat en klassenmanagement. Soms geeft een leerkracht in het tweede schooljaar les aan een andere jaargroep en heeft hij of zij minder goed zicht op de leerlijnen van de betreffende jaargroep. Met betrekking tot de DSAQ is er daarom eerder sprake van een nieuwe voormeting dan van een betrouwbare nameting, terwijl de nameting met de DMI in dit opzicht zuiverder is. Een andere verklaring betreft de validiteit van beide instrumenten, waarover nog weinig bekend is. Meten de instrumenten wat ze pretenderen te meten, namelijk ‘breed aanbod van de basisstof’ en ‘differentiatie van instructie’? De interbeoordelaarsbetrouwbaarheid van de DMI is evenmin boven iedere twijfel verheven. Het betreft nieuwe, experimentele instrumenten die hun bruikbaarheid in de toekomst nog moeten bewijzen. In deze studie is gepoogd twee schalen van de twee instrumenten met elkaar in verband te brengen. Deze schalen zijn gekozen omdat ze op elkaar lijken, maar ze zijn niet met dit doel geconstrueerd. Wanneer de DSAQ en de DMI op eenzelfde leest geschoeid zouden zijn en er in de DSAQ expliciet naar datgene gevraagd wordt wat in de DMI wordt geobserveerd is het mogelijk dat er meer samenhang wordt gevonden of kan met meer zekerheid worden vastgesteld dat zelfrapportage juist andere resultaten oplevert dan observatie. Vraagtekens kunnen worden geplaatst bij de begripsvaliditeit van beide typen instrumenten en de fundamentele vraag dringt zich op in hoeverre zelfinschattingsvragenlijsten minder of meer geschikt zijn voor het meten van gedrag dan observatieinstrumenten. Waar Hook en Rosenshine (1979) nog betoogden dat

observatie te verkiezen is boven zelfrapportage, blijkt uit het werk van Howard en collega's dat zelfrapportage evengoed superieur kan worden bevonden (Howard, 1994; Howard, Conway, & Maxwell, 1985; Howard, Maxwell, Wiener, Boynton, & Rooney, 1980).

Bij een volgend correlatie-onderzoek tussen de DMI en de DSAQ verdient het aanbeveling de nameting op beide instrumenten in dezelfde fase van het schooljaar af te nemen. Dan is het in de toekomst wellicht mogelijk om de DSAQ te crossvalideren met de DMI. Op grond van dit onderzoek kan niet overtuigend worden aangetoond dat voor het meten van specifiek leerkrachtgedrag zelfinschatting even goed voldoet als observatie. Op grond van zelfinschatting kan voor deze training wel voorzichtig worden geconcludeerd dat de leerkrachten in de GROW-training geleerd hebben de onderwijsbehoefte van de leerlingen beter vast te stellen en ook dat ze gedifferentieerder doelen kunnen stellen.

Referenties

Agyei, D., & Voogt, J. (2011). Determining teachers' TPACK through observations and self-report data. In M. Koehler & P. Mishra (Eds.), *Proceedings of society for information technology & teacher education international conference 2011* (pp. 2314-2319). Chesapeake, VA: AACE.

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education, 24*, 417-436. doi:10.1016/j.tate.2006.11.012

Bulsink, M. (2011). *Verhogen prestaties leerlingen basisonderwijs. Professionalisering enpersoonlijke aandacht vormen de sleutel volgens onderwijsmedewerkers.* (TNS NIPO rapport G 1345). Retrieved from: http://www.onderwijsraad.nl/upload/publicaties/627/documenten/verhogen-prestaties-leerlingen-basisonderwijs.pdf

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181-199. doi:10.3102/0013189X08331140

Field, A. (2009). *Discovering statistics using SPSS* (pp. 676-681). London: Sage.

Gelderblom, G. (2007). *Effectief omgaan met verschillen in het rekenonderwijs.* Amersfoort: CPS.

Gelderblom, G. (2010). Effectieve instructie is hart van rekenonderwijs. *Didactief*, 40(7), 6-7.

Hook, C. M., & Rosenshine, B. V. (1979). Accuracy of teacher reports of their classroom behavior. *Review of Educational Research, 49,* 1-12. doi:10.3102/00346543049001001

Howard, G. S. (1994). Why do people say nasty things about self-reports? *Journal of Organizational Behavior, 15,* 399-404. doi:10.1002/job.4030150505

Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology, 77,* 187-196. doi:10.1037/0022-0663.77.2.187

Howard, G. S., Maxwell, S. E., Wiener, R. L., Boynton, K. S., & Rooneym, W. M. (1980). Is a behavioral measure the best estimate of behavioral parameters? Perhaps not. *Applied Pscyhological Measurement, 4,* 293-311. doi:10.1177/014662168000400302

Jacobs, J. K., Kawanaka, T., & Stigler, J. W. (1999). Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. *International Journal of Educational Research, 31,* 717-724. doi:10.1016/S0883-0355(99)00036-1

Janson, D. (2012). Differentiatie vraagt voorbereiding: Een groepsplan in de rekenles. *Volgens Bartjens, 31*(5), 10-13.

Johnsen, S. K., Haensly, P. A., Ryser, G. R., & Ford, R. F. (2002). Changing general education classroom practices to adapt for gifted students. *Gifted Child Quarterly, 46,* 45-63. doi: 10.1177/001698620204600105

Koninklijke Nederlandse Akademie van Wetenschappen (KNAW), (2009). *Rekenonderwijs op de basisschool. Analyse en sleutels tot verbetering.* Amsterdam: KNAW.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21,* 29-45 . doi:10.3102/01623737021001029

Mullis, I. V. S., Martin, M. O., Olson, J. F., Berger, D. R., Milne, D., & Stanco, G. M. (2008). *TIMSS 2007 Encyclopedia. A guide to mathematics and science education around the World. Part 2,* Boston College, TIMSS & PIRLS International Study Center.

Nye, B., Konstantopoulus, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation Policy Analysis, 26,* 237-257. doi: 10.3102/01623737026003237

Organisation for Economic Co-operation and Development (OECD) (2010). *PISA 2009 results: Executive summary.* Retrieved from: http://www.oecd.org/pisa/pisaproducts/46619703.pdf

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E., & Van Luit, J. E. H. (2013a). *Teacher self-assessment of differentiation strategies recommended by experts.* Paper presented at the EARLI 2013, 27-31 August 2013, Munich, Germany.

Prast, E., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2013b). Observatieschaal Differentiatie in Mathematische Instructie (DMI). [Observation scale Differentiated Mathematical Instruction (DMI)]. Universiteit Utrecht: ongepubliceerd intern document.

Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education, 23,* 146–159. doi: 10.1016/j.tate.2006.04.035

Ross, J. A., McDougall, D., Hogaboam-Gray, A., & LeSage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics Education, 34,* 344-363. doi:10.2307/30034787

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78,* 427-515. doi: 10.3102/0034654308317473

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133,* 859-883. doi:10.1037/0033-2909.133.5.859

Valli, L. & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal, 44,* 519-558. doi: 10.3102/0002831207306859

Van de Weijer-Bergsma, E., & Prast, E. J. (2013). Gedifferentieerd primair rekenonderwijs volgens experts: De resultaten uit een Delphi-onderzoek. *Orthopedagogiek: Onderzoek en Praktijk, 52,* 336-349

Van Tassel-Baska, J., Quek, C., & Feng, A. X. (2007). The development and use of a structured teacher observation scale to assess differentiated best practice. *Roeper Review, 29,* 84-92. doi: 10.1080/02783190709554391

Ysseldyke, J., Tardrew, S., Betts, J., Thill, T., & Hannigan, E. (2004). Use of an instructional management system to enhance math instruction of gifted and talented students. *Journal for the Education of the Gifted, 27,* 293-310. doi: 10.4219/jeg-2004-319

Bijlage A

Samenstelling van de schalen Zelfinschatting Instructie (DSAQ) en Breed Aanbod (DMI)

Leerkrachtvragenlijst (DSAQ) *Schaal: Zelfinschatting Instructie*	*Observatieinstrument (DMI)* *Schaal: Breed Aanbod*
1. Ik pas het handelingsniveau van mijn instructie aan aan de behoefte(n) van de leerlingen. 2. Ik pas de modaliteit van mijn instructie (visueel, verbaal, handelend) aan aan de behoefte(n) van de leerlingen. 3. Ik pas het tempo van mijn instructie aan aan de behoefte(n) van de leerlingen. 4. Ik stel bewust open vragen tijdens de klassikale instructie. 5. Ik stel bewust vragen van verschillende moeilijkheidsgraad tijdens de klassikale instructie. 6. Ik geef regelmatig extra instructie (verlengde instructie, preteaching) aan rekenzwakke kinderen. 7. Ik geef rekensterke kinderen regelmatig instructie of begeleiding op hun niveau, in groepsverband of individueel.	1. De leerkracht brengt de stof op verschillende manieren over. 2. De inhoud van de vragen die de leerkracht stelt is afgestemd op een breed leerlingenpubliek. 3. De leerkracht stimuleert de leerlingen tot actief meedenken. 4. De leerkracht geeft rekeninhoudelijke feedback met aandacht voor de manier van aanpak. 5. De leerkracht biedt de leerlingen keuzemogelijkheden waarmee zij aspecten van instructie en verwerking beter kunnen aan kunnen laten sluiten bij hun onderwijsbehoefte.

Bijlage B

Overeenkomsten tussen de schalen Zelfinschatting Instructie (DSAQ) en Breed Aanbod (DMI).

Zelfinschatting Instructie (DSAQ)	*Breed Aanbod (DMI)*
Ik pas het handelingsniveau van mijn instructie aan aan de behoefte(n) van de leerlingen (1) Ik pas de modaliteit van mijn instructie (visueel, verbaal, handelend) aan aan de behoefte(n) van de leerlingen (2)	De leerkracht brengt de stof op verschillende manieren over (1).
Ik stel bewust vragen van verschillende moeilijkheidsgraad tijdens de klassikale instructie (5).	De inhoud van de vragen die de leerkracht stelt is afgestemd op een breed leerlingenpubliek (2).
Ik stel bewust open vragen tijdens de klassikale instructie (4).	De leerkracht stimuleert de leerlingen tot actief meedenken (3).