

De Predictieve Validiteit van de UGT-R

*Het voorspellen van rekenprestaties
op basis van testcores op de UGT-R*

Masterthesis 2008-2009

Door: Ingrid van Dongen

Studentnummer: 3208176

Thesisbegeleider: Dhr. Van Luit

Datum: 30-06-2009



Universiteit Utrecht

Voorwoord

Met enige onzekerheid, maar genoeg motivatie, ben ik aan het uitvoeren van dit onderzoek begonnen. De enige echte problemen die ik gedurende het proces ben tegen gekomen hebben betrekking gehad op het vinden van basisscholen die dit schooljaar (2008/ 2009) nogmaals hun medewerking zouden willen verlenen aan het onderzoek. Door telefonisch contact op te blijven nemen met verschillende scholen in de provincie Zuid-Holland en tevens het belang van het huidige onderzoek voor de schoolpraktijk te blijven benadrukken hebben uiteindelijk toch twee scholen zich hiervoor geleend.

Graag zou ik dan ook de directie, de intern begeleiders en natuurlijk de leerkrachten en leerlingen van de Henri Dunantschool te Sliedrecht, alsmede de Van Rijckevorselschool te Hoek van Holland willen bedanken voor hun inzet gedurende mijn dataverzameling.

Ook wil ik graag mijn begeleider Hans van Luit, maar ook mijn 'back-up' begeleidster Bernadette van de Rijt, willen bedanken voor hun ondersteuning op de momenten dat ik deze nodig had. Als ware statistiek 'leek' heb ik onder andere veel waarde gehecht aan de bevestigende woorden van beide, welke voor mij net dat duwtje in de rug vormden om met vertrouwen in eigen kunnen het onderzoek voort te kunnen zetten en af te kunnen ronden.

Tot slot wil ik alle mensen uit mijn directe omgeving (Helmer, Brigitte, mijn ouders en Fenneke) bedanken voor hun steun, maar ook het geduld dat zij voor mij hebben kunnen opbrengen in deze voor mij toch wel stressvolle periode.

Abstract

Background: A lot of research has been performed regarding early numeracy and its relevancy towards mathematical competence and performance. It has been shown that insufficient development of early numeracy can lead towards low arithmetic performance and difficulties later in life. Researchers have constructed the Early Numeracy Test-Revised (ENT-R) in order to measure the comprehension of early numeracy. **Aim:** The aim of this study is to investigate whether the ENT-R test scores of children in kindergarten and first grade are predictive for these children's mathematical performance one year later. Differences between ENT-R comprehension and level scores are examined, as well as the stability and correlation between these two. Also the correlation between the ENT-R and the M3/ M4 Cito mathematical performance scales has been investigated. **Method:** Out of 2137 young pupils, aged 4 to 7 years old, 100 pupils were selected on a basis of availability and a total of 89 pupils actually participated in the present study. Children in kindergarten and grade 1 were tested with the ENT-R during three test cycles to examine their early mathematical competence. Children in kindergarten 2 and grade 1 were also tested with the M3/ M4 Cito mathematical performance scales in order to examine their mathematical competence. **Results:** Significant differences were found between the ENT-R comprehension and level scores during the three testing cycles. The scores were found to be stable within the year groups, but did lose stability under the influence of formal arithmetic education. ENT-R scores found during the first testing cycle were indeed predictive of the following ENT-R scores. Correlation between the ENT-R and the M3/ M4 mathematical performance scales was found to be much stronger for the M3 than the M4 scale. ENT-R comprehension and level scores measured during the first test cycle were both well able to predict mathematical performance measured at the mid-point of grade 1. Predicting mathematical performance in grade 2 could be performed less precisely. **Conclusion:** ENT-R test scores of children in kindergarten 1 and 2 are indeed predictive of these children's mathematical performance one year later, whereas the mathematical performance of children in grade 2 could not be that accurately predicted on behalf of the ENT-R test scores measured in grade 1. Pupils with extremely low test scores should therefore be carefully monitored during their mathematical learning process in order to perform early intervention and prevent the development of mathematical difficulties.

Keywords: *Early numeracy, mathematical performance, predictive validity, ENT-R, 4-7 year-olds.*

Inleiding

In toenemende mate wordt het belang van vroege en voorschoolse educatie voor latere schoolprestaties erkend. Dit heeft ertoe geleid dat onderzoek hiernaar aanzienlijk is toegenomen en dat onder andere de relatie tussen vroege educatie en latere rekenprestaties meer onder de aandacht is gebracht (Aubrey & Godfrey, 2003). Hierbij neemt het concept 'getalbegrip' een belangrijke positie in.

Getalbegrip en de ontwikkeling ervan

Voor het definiëren van getalbegrip zijn vooral de wetenschappelijke bevindingen van cognitieve ontwikkelingspsychologen van belang geweest. Het is een relatief lastig te definiëren construct dat verwijst naar de gemakkelijke en flexibele omgang van kinderen met getallen, het besef van hoeveelheden, het geven van betekenis aan getallen, het besef dat getallen meerdere betekenissen en functies kunnen hebben, het begrijpen van deze betekenissen en functies om later rekenkundige oplossingsstrategieën te ontwikkelen, het leggen van verbanden tussen de betekenissen en het vermogen om uiteindelijk een rekentaak tot een goed einde te brengen (Berch, 2005; Gersten & Chard, 1999; Treffers & De Moor, 1990; Van de Rijt, 1996).

Een belangrijke voorwaarde voor het ontstaan van getalbegrip is het kunnen conserveren van aantallen. Dit houdt in dat kinderen begrijpen dat een bepaalde hoeveelheid objecten in een groep niet kan veranderen, tenzij er objecten aan worden toegevoegd of uit worden verwijderd. Kinderen worden hiertoe in staat gesteld wanneer zij zich het cognitieve concept van getalvolgorde (ordering van de telrij) eigen hebben gemaakt en derhalve kunnen tellen. Tellen vormt dus een belangrijke ondersteunende factor in de ontwikkeling van getalbegrip en is op zichzelf een ontwikkelend proces welke achtereenvolgens bestaat uit akoestisch, (a)synchroon, resultaatief en verkort tellen. Het heeft betrekking op het opnoemen van telwoorden en het achtereenvolgens koppelen van objecten aan deze telwoorden, waarna een precies aantal objecten kan worden vastgesteld. (Kidd, Paskas, Gadzichowski, Ferral-Like & Gallington, 2008; Ruijsenaars, Van Luit & Van Lieshout, 2004; Treffers & De Moor, 1990; Van de Rijt & Van Luit, 1998).

Van de Rijt (1996) definieert getalbegrip als een construct dat uit een aantal voorbereidende deelvaardigheden bestaat, welke onder meer betrekking hebben op zojuist genoemde onderdelen. Dit betreffen de vaardigheden vergelijken, classificeren, het leggen van één-op-één relaties (correspondentie), seriëren, telwoorden gebruiken, synchroon tellen, verkort tellen, resultaatief tellen en het toepassen van kennis van getallen (Van de Rijt, 1996).

Daarnaast constateren Siegler en Booth (2004) bijvoorbeeld dat numeriek schatten eveneens een belangrijke component vormt binnen de voorbereidende rekenvaardigheid.

Vanuit de cognitieve ontwikkelingspsychologie gezien vangt de ontwikkeling van getalbegrip al ver voordat kinderen werkelijk aan het formele rekenonderwijs op de basisschool deelnemen aan. Het blijkt een proces te zijn waarbij de oorsprong ligt in een aangeboren gevoeligheid voor hoeveelheden en getallen. Deze aangeboren gevoeligheid wordt beïnvloed door de hoeveelheid en variatie aan ervaringen die kinderen met getallen en kwantitatieve situaties, zowel binnen als buiten de schoolcontext, opdoen. Binnen dit ontwikkelingsproces worden vervolgens de verschillende deelvaardigheden en kennis verworven die uiteindelijk nodig zijn voor vroege rekenvaardigheid, waarbij dus een belangrijke rol is weggelegd voor de omgeving van het kind (Berch, 2005; Braams & Denis, 2003; Dehaene, 1997; Gersten & Chard, 1999; Torbeyns et al., 2002; Van de Rijt & Van Luit, 1998).

Belang van getalbegrip voor rekenvaardigheid en verband met latere rekenprestaties

Voor het proces van leren rekenen is het van belang dat kinderen over een voldoende mate van beheersing van getalbegrip beschikken. Het wordt als een voorwaarde beschouwd om in groep 3 op adequate manier aan het rekenonderwijs te kunnen deelnemen. Verminderd getalbegrip kan het verwerven van rekenvaardigheden namelijk bemoeilijken en uiteindelijk zelfs leiden tot verminderde rekenprestaties en rekenproblemen (Braams & Denis, 2003; Gersten & Chard, 1999; Van Luit, 2000; Van Luit & Van de Rijt, 2009; Van de Rijt & Van Luit, 1998; Van de Rijt, Van Luit & Pennings, 1994).

Een aantal onderzoeken hebben dit belang van getalbegrip voor rekenvaardigheid inmiddels onderzocht. Hierbij is ook aandacht besteed aan het verband dat bestaat tussen getalbegrip en rekenprestaties (alle te gebruiken bronnen).

Zo is het belang van de voorbereidende deelvaardigheden seriëren, conserveren en vergelijken voor de ontwikkeling van (latere) rekenvaardigheid aangetoond door Kidd, Pasnak, Gadzichowski, Ferral-Like en Gallington (2008). Kinderen die niet over dergelijke vaardigheden beschikken zouden het moeilijk krijgen tijdens hun pogingen om zich de abstracte rekenkundige concepten en vaardigheden eigen te maken en daardoor risico lopen op het ontwikkelen van rekenproblemen.

Het belang van de deelvaardigheid tellen voor rekenvaardigheid is onder andere aangetoond in het onderzoek dat door Passolunghi, Vercelloni en Schadee (2007) bij kinderen uit groep 3 is uitgevoerd. Numerieke kennis, in deze studie onderverdeeld in getalherkenning

en -productie enerzijds, en het vermogen om te kunnen tellen anderzijds, blijkt een belangrijke factor te vormen voor de ontwikkeling van rekenvaardigheid. Hierbij houden getalherkenning en -productie volgens de onderzoekers geen direct verband met vroege rekenvaardigheid, maar is er wel een directe relatie gevonden tussen het vermogen om te tellen en rekenvaardigheid. De onderzoekers geven aan dat het derhalve mogelijk is om op basis hiervan te voorspellen hoe het rekenleerproces zal verlopen, waardoor eventuele rekenproblemen vroeg gesignaleerd, dan wel voorkomen kunnen worden.

Dat telvaardigheid een belangrijke voorspeller is voor rekenprestaties blijkt ook uit onderzoek naar de ontwikkelingsdynamiek van rekenprestaties bij kinderen tussen de 5 en 6 jaar. Dit onderzoek wijst uit dat kinderen die in de kleuterklas over een hoger niveau van telvaardigheid beschikken, een hoger niveau op rekenprestaties laten zien in groep 3 en hier ook een grotere groei in vertonen (Aunola, Leskinen, Lerkkanen & Nurmi, 2004).

Ook het onderzoek van Jordan, Kaplan, Locuniak en Ramineni (2007) heeft een dergelijke relatie tussen getalbegrip en latere rekenprestaties aangetoond. Getalbegrip bij kinderen uit de kleuterklas zou een betrouwbare en krachtige voorspeller zijn van rekenprestaties bij dezelfde kinderen aan het einde van groep 3. Kinderen die aan het begin van de kleuterklas een lagere score behaalden op getalbegriptaken en tijdens een tweede en derde meting geen vooruitgang hierin lieten zien, behaalden aan het einde van groep 3 lagere rekenprestaties dan kinderen die een dergelijke lage score behaalden, maar vervolgens tijdens de tweede en derde meting vooruitgang boekten. Het screenen van getalbegrip in de kleuterklas zou nuttig zijn voor het signaleren van kinderen die mogelijk te maken zullen krijgen met problemen op het gebied van rekenen om daarmee rekenproblemen te voorkomen (Jordan et al., 2007).

Deze bevindingen worden ook gevonden in een ander onderzoek dat is uitgevoerd naar de relatie tussen getalbegrip van kleuters en rekenvaardigheid in groep 4. De onderzoekers hebben geconstateerd dat verminderd getalbegrip een krachtige voorspeller is van rekenvaardigheid en zelfs rekenproblemen, waardoor vroegtijdige screening en signalering van belang is (Locuniak & Jordan, 2008).

Naast getalbegrip blijken echter ook andere, al dan niet aan getalbegrip gerelateerde, factoren van invloed te zijn op de ontwikkeling van rekenvaardigheid en rekenprestaties. Zo is getalbegrip voor het leren rekenen wel noodzakelijk, maar blijkt het beschikken over voldoende getalbegrip niet vanzelfsprekend te leiden tot betere rekenvaardigheid en betere rekenprestaties. Onder meer het leren automatiseren van basale rekenkennis en toepassen van adaptieve rekeninstructies zouden hier mede van invloed op zijn (Berch, 2005; Gersten &

Chard, 1999). Ook Torbeyns en collega's (2002) constateren dat de instructie welke geboden wordt binnen rekenonderwijs, alsmede de organisatie ervan en de algemene visie op leerdoelen met betrekking tot rekenonderwijs, mogelijk kunnen leiden tot verschillen in rekenvaardigheid en –prestaties en derhalve beïnvloedende factoren zijn. Onderzoek wijst verder bijvoorbeeld uit dat getalbegrip bij Taiwanese leerlingen minder goed ontwikkeld is vanwege het kwalitatief minder goede voorbereidende rekenonderwijs, waardoor lagere scores worden behaald op rekentoetsen (Yang, Li & Lin, 2007).

De ontwikkeling van rekenvaardigheid blijkt dus enerzijds afhankelijk te zijn van de vroege ervaringen die het kind opdoet met rekenkundige activiteiten binnen de directe leefomgeving en waarmee het een bepaalde mate van getalbegrip ontwikkelt. Anderzijds dient het formele rekenonderwijs hier verder vorm aan te geven (Varol & Farran, 2006).

Het huidige onderzoek

Gezien het belang van voldoende mate van beheersing van getalbegrip voor deelname aan het formele rekenonderwijs, alsmede latere rekenvaardigheid en –prestaties, lijkt het nuttig om met behulp van een betrouwbaar en valide meetinstrument, van tevoren te kunnen voorspellen welke kinderen een mogelijk risicogroep zullen vormen met betrekking tot het vertonen van uitval op rekenprestaties.

De Utrechtse Getalbegrip Toets - Revised (UGT-R, 2009) is een meetinstrument dat tot doel heeft om het beheersingsniveau van getalbegrip te meten bij kinderen in de leeftijd van 4,0 jaar tot 7,6 jaar (groep 1, 2 en 3 basisonderwijs), en is gebaseerd op de door Van de Rijt benadrukte voorbereidende deelvaardigheden (Van Luit & Van de Rijt, 2009; Van de Rijt, Van Luit & Pennings, 1994). In het huidige onderzoek wordt nagegaan wat de predictieve validiteit van dit meetinstrument is. Onderzocht wordt of rekenprestaties van kinderen uit groep 2, 3 en 4 voorspeld kunnen worden op basis van de eerder door hen behaalde testcores op de UGT-R. Hierbij zal enerzijds gekeken worden naar de relatieve stabiliteit van de door de leerlingen uit groep 1 en 2 behaalde vaardigheids- en niveauscores op de UGT-R, alsmede naar de mate waarin deze scores met elkaar samenhangen en te voorspellen zijn. Ook wordt nagegaan of de scores die de leerlingen in groep 1 en 2 op de UGT-R hebben behaald verschillen van de door hen behaalde scores een jaar later. Er wordt verwacht dat er een positieve samenhang zal bestaan tussen de scores over de meetmomenten heen genomen, dat de vaardigheidsscores significant van elkaar zullen verschillen, maar dat de niveauscores gelijk zullen blijven.

Anderzijds zal gekeken worden naar het verband tussen de UGT-R en de methodeonafhankelijke rekenvaardigheidstoetsen van het Cito-leerlingvolgsysteem, om derhalve uitspraken te kunnen doen over de mate waarin het mogelijk is om rekenprestaties te kunnen voorspellen. Er wordt verwacht dat er een positieve samenhang zal bestaan tussen de scores op de UGT-R en de M3/ M4 Rekenen-Wiskunde toetsen.

In een onderzoek, uitgevoerd door Aubrey en Godfrey (2003), is namelijk met behulp van de Utrecht Early Numeracy Test, een Engelse vertaling van de inmiddels herziene UGT, nagegaan of de totaalscore van 5-jarige Engelse kinderen op deze toets hun rekenprestaties van een jaar later kon voorspellen. Uit dit onderzoek komt naar voren dat de totaalscore die wordt behaald tijdens een eerste meetmoment (halverwege het schooljaar), inderdaad de rekenprestaties op 6-jarige leeftijd voorspelt. Ook onderzoek van Kavkler, Tancig en Magajna (2003) wijst uit dat 5-jarige onderpresteerders op de UGT op latere leeftijd een lager rekenvaardigheidniveau behalen dan hun leeftijdgenoten die ruim voldoende tot goed op deze test scoren. Ten aanzien van het huidige onderzoek wordt dan ook verwacht dat de predictieve validiteit van de UGT-R voldoende zal zijn.

Methode

Participanten

Uit 2137 basisschoolkinderen, waarbij in het schooljaar van 2007/ 2008 de Utrechtse Getalbegrip Toets-Revised op twee momenten ten behoeve van hernormingsonderzoek is afgenomen, zijn 100 leerlingen uit drie jaargroepen (groep 2, 3 en 4) op basis van beschikbaarheid geselecteerd om deel te nemen aan het huidige valideringsonderzoek.

Aan het eerder uitgevoerde normeringsonderzoek hebben in totaal 45 scholen, die op aselecte wijze uit 12 provincies in de steekproef zijn opgenomen, meegewerkt. Praktische overwegingen hebben ertoe geleid dat voor het huidige onderzoek alleen de zes scholen uit Zuid-Holland telefonisch benaderd zijn. Van zes scholen hebben twee reguliere openbare basisscholen hun medewerking aan het valideringsonderzoek verleend.

De steekproeftrekking van dit onderzoek betreft dus een selecte toevallige steekproef en kan vanwege de verdeling van de populatie in deelgroepen als quotasteekproef worden aangeduid. Het is derhalve niet mogelijk om vergaande generaliserende uitspraken over de onderzoeksuitkomsten te doen.

Wegens verhuizing, ziekte of andere redenen hebben een aantal leerlingen die vorig schooljaar wel hebben deelgenomen aan het hernormingsonderzoek niet kunnen participeren aan het huidige onderzoek. Dit reduceert de totale populatie tot 89 leerlingen,

bestaande uit 43 jongens en 46 meisjes, met een gemiddelde leeftijd van 6 jaar en 9 maanden. In Tabel 1 kunnen de beschrijvende statistieken van de onderzoekspopulatie worden teruggevonden.

Tabel 1

Beschrijvende statistieken onderzoekspopulatie

	Geslacht		Leeftijd (in maanden)	
	Jongens	Meisjes	M	SD
Groep 2	13	12	70,7	4,1
Groep 3	12	17	82,5	3,4
Groep 4	18	17	93,1	4,9
Totaal (N=89)	43	46	83,4	10,1

Meetinstrumenten

Getalbegrip

De Utrechtse Getalbegrip Toets-Revised (UGT-R) is een methodeonafhankelijke taakgerichte toets en is afgenomen om getalbegrip bij de leerlingen uit groep 2 en 3 van dit schooljaar te meten. De toets bestaat uit negen verschillende onderdelen welke betrekking hebben op vergelijken, classificeren, correspondentie leggen, seriëren, telwoorden gebruiken, synchroon en verkort tellen, resultaatief tellen, toepassen van kennis van getallen en schatten. Scores op deze onderdelen geven een beeld van de sterke en zwakke kanten van het voorbereidende rekenvaardigheidsniveau (niveau A t/m E, vergelijkbaar met de Cito vaardigheidsscores). Ook kan worden nagegaan welke onderdelen al wel, of nog niet (voldoende) worden beheerst (Van Luit & Van de Rijt, 2009).

In dit onderzoek is gebruik gemaakt van een pilotversie van de UGT-R. Deze bestaat uit zes verschillende testboekjes, ieder voorzien van 45 items. Dit betekent dat er per toetsonderdeel vijf items aan iedere leerling zijn aangeboden. De leerlingen hebben gedurende het onderzoek uit dezelfde boekjes gewerkt als die hen tijdens het hernormeringsonderzoek zijn aangeboden. Uit onderzoek is gebleken dat de Crombach's alpha en betrouwbaarheidscoëfficiënt van de gebruikte boekjes hoog is. Het gemiddelde van de zes gevonden betrouwbaarheidscoëfficiënten is .94. Verschillen in getalbegrip kunnen derhalve betrouwbaar gemeten worden (Van Luit & Van de Rijt, 2009).

Rekenprestaties

Met behulp van de LVS Rekenen-Wiskunde M3 en M4 toetsen van het Cito, beiden methodeonafhankelijke toetsen, zijn de rekenprestaties van de kinderen uit groep 3 en 4 van dit schooljaar, uitgedrukt in niveauscores, gemeten. De niveauscores geven de positie van de kinderen ten opzichte van een normgroep aan. Hierbij worden de niveau's A, B, C, D en E onderscheiden. Niveau A en B liggen boven het landelijk gemiddelde (ieder 25%), C, D en E eronder (respectievelijk 25%, 15% en 10%). Voor de kinderen uit groep 2 worden rekenprestaties gebaseerd op scores van de UGT-R, welke dezelfde vaardigheidniveau's omvatten. De Cotan heeft de betrouwbaarheid en begripsvaliditeit van de M3 en M4 toets als goed beoordeeld (Commissie Testaangelegenheden Nederland, 1999; Jansen & Engelen, 2002).

Onderzoeksdesign en procedure

Voor het huidige valideringsonderzoek, welke in principe aan de kenmerken van een 'follow-up' onderzoek voldoet, is gebruik gemaakt van zowel bestaande als nieuw verzamelde onderzoeksgegevens. Deze gegevens zijn op drie verschillende meetmomenten verzameld. De eerste twee meetmomenten hebben plaatsgevonden in januari/ februari (T1) en mei/ juni 2008 (T2), in het kader van hernormeringsonderzoek van de UGT-R. Het derde meetmoment (T3) heeft in januari/ februari 2009 plaatsgevonden, in het kader van het huidige onderzoek naar de predictieve validiteit van de UGT-R.

Bij de leerlingen die in het schooljaar 2007/2008 in groep 1 en 2 de UGT-R op de twee eerder genoemde meetmomenten hebben gemaakt, is in januari/ februari 2009 opnieuw de UGT-R individueel afgenomen. Hierbij werd gebruik gemaakt van de werkboekjes waaruit de leerlingen tijdens de eerder genoemde meetmomenten ook hadden gewerkt. De toetsduur bedroeg ongeveer 30 minuten per leerling. Leerlingen waarbij vorig schooljaar in groep 2 en 3 de UGT-R is afgenomen, hebben in januari/ februari 2009 van dit schooljaar de M3, respectievelijk M4 Rekenen-Wiskunde toets klassikaal gemaakt. Deze toetsafname heeft ongeveer drie kwartier per groep in beslag genomen en werd uitgevoerd door de groepsleerkracht.

Resultaten

De gemiddelde vaardigheidsscores op de UGT-R en de verschillen hiertussen over de drie meetmomenten genomen zijn per basisschoolgroep weergegeven in Tabel 2 en 3.

Met behulp van de Paired-samples T-test is nagegaan of de UGT-R vaardigheidsscores van kinderen uit groep 1 en 2 in 2008 en de door hen behaalde testcores in groep 2 en 3 een jaar later verschillend zijn.

Resultaten wijzen uit dat de gemiddelde vaardigheidsscores van leerlingen uit alle drie de groepen tussen T1 en T2, respectievelijk T3 toenemen, zie Tabel 3. Ook van T2 naar T3 blijken de scores van leerlingen uit groep 1 en groep 2 toe te nemen (groep 1 $t(24) = 12,67$, $p < .001$ en groep 2 $t(28) = 10,14$, $p < .001$). De verschillen in vaardigheidsscores blijken volgens verwachting significant te zijn.

Tabel 2

Gemiddelde vaardigheidsscores op de UGT-R per groep over de drie verschillende meetmomenten gemeten

	Voormeting (T1)		Nameting (T2)		Uitgestelde nameting (T3)	
	M	SD	M	SD	M	SD
Groep 1	41.28	8.08	46.30	9.19	57.60	9.47
Groep 2	52.48	6.81	56.88	6.36	62.29	5.15

Tabel 3

Vershil in vaardigheidsscores op de UGT-R per groep over de drie verschillende meetmomenten gemeten

	Vershil T1-T2		Vershil T1-T3		Vershil T2-T3	
	t	df	t	df	t	df
Groep 1	5.78	24	8.84	24	12.67	24
Groep 2	5.24	28	6.26	28	10.14	28

Alle verschillen zijn significant bij $p < .01$

De relatieve stabiliteit van de vaardigheidsscores is met behulp van de Pearson's productmomentcorrelatie nagegaan. Binnen groep 1 blijkt sprake te zijn van een positieve en sterke mate van samenhang tussen de verschillende scores van meetmoment naar meetmoment, zie Tabel 4. De vaardigheidsscores van de leerlingen zijn hiermee in groep 1, alsook gedurende het eerste halfjaar groep 2, relatief vrij stabiel te noemen.

Enkelvoudige lineaire regressie wijst bovendien uit dat 78% van de variantie in vaardigheidsscores op T2 kan worden verklaard door de scores op T1 ($F(1) = 80.00$, $p < .001$), en dat op basis van de scores op T1 55% van de variantie in scores op T3 kan worden verklaard ($F(1) = 28.08$, $p < .001$). Op basis van de scores op T2 kan 59% van de variantie in scores op T3 worden verklaard ($F(1) = 32.55$, $p < .001$). De vaardigheidsscores van groep 2

leerlingen in het huidige schooljaar kunnen in redelijk hoge mate op basis van hun eerder behaalde vaardigheidsscores worden voorspeld.

Binnen groep 2 is de relatieve stabiliteit van de vaardigheidsscores van T1 naar T2 nog steeds hoog te noemen ($r = .77$). De proportie verklaarde variantie van 59% wijst erop dat vaardigheidsscores hierbij goed kunnen worden voorspeld ($F(1) = 38.19, p < .001$). Bij de overgang van groep 2 naar groep 3 (T2-T3) is de stabiliteit eveneens matig te noemen ($r = .69$). Op basis van de vaardigheidsscores op T2 kan 48% van de variantie in scores op T3 worden verklaard ($F(1) = 24.72, p < .001$), hetgeen betekent dat ook deze scores nog goed, maar onder invloed van formeel rekenonderwijs iets minder nauwkeurig kunnen worden voorspeld.

Tabel 4

Samenhang en proportie verklaarde variantie tussen de vaardigheidsscores over de drie meetmomenten gemeten

	Correlatie T1-T2	R^2	Correlatie T1-T3	R^2	Correlatie T2- T3	R^2
Groep 1	.88	.78	.74	.55	.77	.59
Groep 2	.77	.59	.65	.42	.69	.48

Alle correlaties zijn significant bij $p < .01$

Om te kunnen bepalen of de UGT-R niveauscores van kinderen uit groep 1 en 2 in het jaar 2008 verschillen van de door hen behaalde niveauscores in groep 2 en 3 van het schooljaar 2009 is analyse met behulp van de Wilcoxon Signed-ranktoets uitgevoerd.

Binnen groep 1 blijkt een verschil ($Z = -3.22, p = .001$) te bestaan tussen de niveauscores die op het eerste (T1) en tweede meetmoment (T2) zijn gemeten. Van de 25 leerlingen blijken er 14 een hogere score op T1 te hebben dan op T2. Slechts eenmaal blijkt er een hogere score op T2 te zijn behaald. In totaal 10 leerlingen behouden dezelfde score. Van T1 naar T3 blijken de niveauscores, in tegenstelling tot de verwachting, eveneens te verschillen ($Z = -2.35, p = .019$). Er blijken 13 leerlingen een hogere score op T1 dan op T3 te hebben. Slechts 3 leerlingen behalen een hogere score op T3 en 9 leerlingen behouden dezelfde score. Van T2 naar T3 blijken de niveauscores niet significant te verschillen ($Z = -0.29, p = .77$).

De relatieve stabiliteit van de niveauscores is met behulp van de Spearman's rho nagegaan. Er blijkt een hoge mate van samenhang te bestaan tussen de scores op T1 en de scores op T2 ($r_s = .81, p < .001$). Van T2 naar T3, en van T1 naar T3 blijken de scores matig te correleren ($r_s = .70, p < .001$, respectievelijk $r_s = .59, p = .002$). De scores zijn hiermee binnen groep 1 (T1-T2)

relatief vrij stabiel te noemen. Bij overgang naar de volgende jaargroep (T2-T3) neemt de stabiliteit enigszins af.

Binnen groep 2 bestaat er eveneens een verschil tussen de niveauscores op T1 en T2 ($Z = -3.63, p < .001$). Van de 29 leerlingen blijken 17 leerlingen een hogere score te hebben behaald op T1 dan op T2. Slechts 1 leerling behaalt een hogere score op T2 en 11 leerlingen behouden dezelfde score. Van T1 naar T3 blijken er binnen deze groep volgens de verwachting geen significante verschillen te bestaan in niveauscores ($Z = -0.73, p = .466$). Van T2 naar T3 zijn de verschillen in niveauscores wel significant ($Z = -3.49, p < .001$). Een hogere score op T3 wordt door 18 leerlingen behaald en slechts 3 leerlingen behalen een hogere score op T2. In totaal behouden 8 leerlingen dezelfde scores. Uit correlatieanalyse blijkt dat er sprake is van een sterke positieve samenhang tussen niveauscores op T1 en T2 gemeten ($r_s = .73, p < .001$) en van matige positieve samenhang tussen de niveauscores op T1 en T3, respectievelijk T2 en T3 gemeten, zie Tabel 5. Ook binnen groep 2 (T1-T2) blijken de niveauscores relatief vrij stabiel te zijn, maar wordt de stabiliteit onder invloed van het formele rekenonderwijs (T2-T3) aanzienlijk geringer.

Tabel 5

Correlatie tussen de niveauscores op de UGT-R binnen de verschillende groepen over drie meetmomenten genomen

	Correlatie T1-T2	Correlatie T1-T3	Correlatie T2- T3
Groep 1	.81	.59	.70
Groep 2	.73	.50	.55

Alle correlaties zijn significant bij $p < .01$

De samenhang tussen de UGT-R en de M3 en M4 Rekenen-Wiskunde toetsen van het Cito leerlingvolgsysteem op basis van ruwe scores is met behulp van de Pearson's correlatie nagegaan. De ruwe scores op de UGT-R die tijdens T1 en T2 zijn gemeten binnen groep 2 blijken volgens verwachting een sterke positieve samenhang te vertonen met de ruwe scores op de M3 toets ($r = .84, p < .001$ respectievelijk $r = .76, p < .001$). Enkelvoudige lineaire regressie wijst uit dat bij een proportie verklaarde variantie van 71% ($F(1) = 63.76, p < .001$) de ruwe scores die tijdens T1 zijn gemeten op de UGT-R hierbij de scores op de M3 toets nauwkeuriger blijken te kunnen voorspellen dan de ruwe scores op de UGT-R die tijdens T2 zijn gemeten. De proportie verklaarde variantie betreft hier namelijk 58% ($F(1) = 35.28, p < .001$).

Binnen groep 3 blijkt er een zwakke positieve samenhang te bestaan tussen de ruwe scores op de UGT-R, zowel op T1 als T2 gemeten, en de M4 toets ($r = .48, p < .001$, respectievelijk $r = .40, p = .02$). Bij een proportie verklaarde variantie van 23% voor T1 ($F(1) = 9.42, p = .004$) en 16% voor T2 ($F(1) = 6.01, p = .020$) blijkt het voorspellen van de ruwe scores op de M4 toets op basis van tijdens T1 en T2 gemeten ruwe scores op de UGT-R, in vergelijking met het voorspellen van ruwe scores op de M3 toets, veel minder nauwkeurig te kunnen geschieden.

Voor de niveauscores is de samenhang tussen de UGT-R en de M3/ M4 toetsen met behulp van de Spearman's rho berekend, zie Tabel 6. Er blijkt volgens verwachting een sterk positief verband te bestaan tussen de niveauscores die leerlingen halverwege groep 2 (T1) behalen op de UGT-R en de door hen behaalde niveauscores op de M3 een jaar later ($rs = .71, p < .001$). De gevonden positieve samenhang tussen de UGT-R niveauscores op T2 met de M3 niveauscores blijkt matig te zijn ($rs = .60, p = .001$). Een matige positieve samenhang is eveneens gevonden voor de behaalde niveauscores op de UGT-R halverwege groep 3 (T1) en de niveauscores die een jaar later op de M4 toets zijn behaald ($rs = .53, p = .001$). De gevonden positieve samenhang tussen de UGT-R niveauscores aan het einde van groep 3 en de halverwege groep 4 behaalde M4 niveauscores blijkt zwak te zijn ($rs = .49, p = .003$).

Conclusie en discussie

In dit onderzoek is nagegaan of rekenprestaties van kinderen uit groep 2, 3 en 4 voorspeld kunnen worden op basis van de eerder door hen behaalde testcores op de UGT-R in groep 1, 2 en 3.

Allereerst is nagegaan of de eerder behaalde UGT-R vaardigheids- en niveauscores van kinderen uit groep 1 en 2 significant verschillen van de door hen behaalde UGT-R scores in groep 2 en 3. Daarnaast is de relatieve stabiliteit van en de samenhang tussen deze scores nagegaan. Resultaten wijzen uit dat de UGT-R vaardigheidsscores van de leerlingen uit groep 1 en 2, zoals verwacht significant verschillen van de door hen behaalde vaardigheidsscores in groep 2, respectievelijk 3. In tegenstelling tot de verwachting dat de niveauscores met oog op de predictieve validiteit gelijk zouden blijven, blijken deze scores evenwel van elkaar te verschillen. Binnen groep 1 is dit over alle drie de meetmomenten heen te zien. Ook blijken niveauscores voornamelijk ten opzichte van het eerste meetmoment af te nemen. Voor groep 2 is een dergelijk verschil alleen gevonden voor het tweede meetmoment. De niveauscores die halverwege groep 2 zijn behaald blijken namelijk niet significant te verschillen van de behaalde niveauscores halverwege groep 3. Doordat er in de ontwikkeling van getalbegrip

gedurende de kleutertijd grote individuele verschillen bestaan en kinderen verschillend kunnen profiteren van het informele rekenonderwijs dat aangeboden wordt (Pennings, Van de Rijt & Van Luit, 1994; Van de Rijt, 1996), kan dit er bij het huidige onderzoek toe hebben geleid dat vaardigheidsscores, gemiddeld genomen, in een jaar tijd wel significant toenamen, maar dat toename hierin voor sommige kinderen niet voldoende bleek om eenzelfde vaardigheidsniveau te behalen. Dit zou een verklaring kunnen bieden voor de gevonden verschillen in niveauscores.

Wat betreft de stabiliteit van de UGT-R scores over drie verschillende meetmomenten heen genomen is uit het onderzoek gebleken dat zowel de vaardigheidsscores als de niveauscores binnen beide jaargroepen zelf relatief vrij stabiel zijn, maar met name onder invloed van formeel rekenonderwijs aan stabiliteit inboeten. Dit komt overeen met de bevindingen van Van de Rijt (1996) in haar onderzoek naar voorbereidende rekenvaardigheden bij kleuters. Zij heeft eveneens geconstateerd dat de relatieve positie van de leerlingen gezien hun scores in de periode dat zij formeel onderwijs kregen sterker in beweging waren dan daarvoor.

Het huidige onderzoek heeft tevens uitgewezen dat zowel de vaardigheidsscores als de niveauscores van de UGT-R positief met elkaar samen te hangen. Dit betekent dat een tijdens een eerste meetmoment behaalde hogere, dan wel lagere vaardigheid- cq. niveauscore op de UGT-R zal leiden tot een overeenkomstig hogere, dan wel lagere UGT-R score op volgende meetmomenten. Met name de vaardigheidsscores blijken op basis van eerder behaalde scores goed te kunnen voorspellen wat de scores op volgende meetmomenten zullen zijn. Tijdens de overgang van groep 2 naar groep 3, dus op het moment dat de leerlingen gaan participeren aan het formele rekenonderwijs, blijkt een nauwkeurige voorspelling echter minder goed mogelijk en is enige voorzichtigheid hierbij geboden. Een mogelijke verklaring hiervoor zou kunnen liggen in het feit dat de leerlingen in groep 3 meer gestructureerde rekeninstructies krijgen, waardoor zowel het getalbegrip, alsmede de rekenvaardigheden extra worden beïnvloed (Torbeyns et al., 2002).

Ten tweede is het verband tussen de UGT-R en de M3/ M4 Rekenen-Wiskunde toetsen van het Cito leerlingvolgsysteem nagegaan. Hierbij is gekeken naar de samenhang tussen zowel de ruwe totaalscores, alsook de niveauscores op beide toetsen. Het gevonden verband tussen de toetsen blijkt sterker te zijn voor de M3 toets dan voor de M4 toets. Zowel de ruwe totaalscores als de niveauscores die halverwege het schooljaar in groep 2 met de UGT-R gemeten worden blijken goed te kunnen voorspellen wat de leerlingen uit deze groep een jaar later zullen scoren op de M3 Rekenen-Wiskunde toets. Voor beide scores geldt dat

hogere, dan wel lagere scores die halverwege het schooljaar zijn behaald op de UGT-R zullen leiden tot overeenkomstig hogere, dan wel lagere scores op de M3 toets. Dit komt overeen met de bevindingen uit ander onderzoek (Aubrey & Godfrey, 2003; Jordan et al., 2007; Kavkler, Tancig & Magajna, 2003). Het voorspellen van ruwe totaal- en niveauscores op de M4 toets, op basis van de eerder behaalde scores op de UGT-R halverwege groep 3, kan echter een stuk minder nauwkeurig geschieden. Mogelijk hangt dit samen met de toegenomen invloed van onderwijsgerelateerde factoren waaraan de ontwikkeling van getalbegrip en rekenvaardigheid binnen deze groep onderhevig is (Berch, 2005; Gersten & Chard, 1999; Varol & Farran, 2006). Voorzichtigheid is hierbij dan ook geboden.

De UGT-R blijkt op basis van de onderzoeksresultaten een instrument te zijn dat over het algemeen in redelijk sterke mate in staat is om voorspellende uitspraken te doen over latere rekenprestaties. De predictieve validiteit van de UGT-R is hiermee ruim voldoende tot goed te noemen. Dit betekent dat middels afname van de UGT-R betrouwbaar en valide gesignaleerd kan worden welke kinderen een verhoogde kans hebben op het vertonen van uitval in rekenprestaties of zelfs het ontwikkelen van rekenproblemen. Kinderen uit groep 1 en 2 die extreem lage scores op de UGT-R behalen dienen om die reden in hun rekenvaardigheidproces extra goed in de gaten te worden gehouden. Dit maakt het namelijk mogelijk om al vroeg in basisschooltijd rekeninterventies uit te voeren, zodat eventuele ontwikkeling van rekenproblemen voorkomen dan wel verminderd kan worden. Het belang van dergelijke vroegtijdige screening is eveneens in onderzoek van Locuniak en Jordan (2008) en Jordan en collega's (2007) aangetoond.

Het huidige onderzoek kent echter wel een aantal beperkingen waar tijdens de interpretatie en generalisatie van de gevonden resultaten rekening mee moet worden gehouden. Allereerst heeft het onderzoek betrekking gehad op een selecte steekproef. Dit heeft tot gevolg dat de conclusie die is getrokken ten aanzien van de predictieve validiteit van de UGT-R vooralsnog alleen van toepassing is op de populatie van het huidige onderzoek en derhalve niet generaliseerbaar is naar de Nederlandse populatie. Nader onderzoek naar de predictieve validiteit van de UGT-R waaraan meerdere scholen, die eerder aan het hernomeringsonderzoek hebben deelgenomen, participeren en die uit verschillende Nederlandse provincies afkomstig zijn, is met het oog op generaliseerbaarheid van de bevindingen dan ook gewenst.

Tevens betrof het huidige onderzoek een relatief kleine steekproef. Het gevolg daarvan was dat een aantal wenselijke analyses ten aanzien van de ontwikkeling van rekenprestaties niet uitgevoerd konden worden. Het bleek onder andere niet mogelijk om op

basis van niveauscores op de UGT-R, met behulp van de Chi-kwadraat toets na te gaan of de gevonden niveauscores op de M3 en M4 toetsen afwijken van, dan wel gelijk blijven aan, de verwachte scores op deze toetsen. Op basis hiervan hadden uitspraken gedaan kunnen worden over de waarschijnlijkheid waarmee leerlingen met A/B of D/E niveauscores op de UGT-R ook A/B, respectievelijk D/E scores op de Rekenen-Wiskunde toetsen zouden behalen. Deze hadden de gevonden onderzoeksresultaten sterker kunnen onderbouwen en zouden wellicht in grootschaliger onderzoek naar de predictieve validiteit van de UGT-R nog uitgevoerd kunnen worden.

Referenties

- Aubrey, C., & Godfrey, R. (2003). The development of children's early numeracy through Key Stage 1. *British Educational Research Journal*, 29, 821-840.
- Aunola, K., Leskinen, E. Lerkkanen, M. K., & Nurmi, J. E. (2004). Developmental dynamics of mathematical performance from preschool to grade 2. *Journal of Educational Psychology*, 96, 699-713.
- Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, 38, 333-339.
- Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, 79, 1016-1031.
- Braams, T., & Denis, D. (2003). Getalbegrip: Een noodzakelijke voorwaarde voor het leren rekenen. *Tijdschrift voor Remedial Teaching*, 2, 16-20.
- Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education*, 33, 18-28.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Janssen, J., & Engelen, R. (2002). Verantwoording van de toetsen Rekenen-Wiskunde 2002. Arnhem: Citogroep. Gedownload op 14/05/2009 van <http://toetswijzer.kennisnet.nl>.
- Jordan, N. C., Kaplan, D., Olah, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, 77, 153-175.
- 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, 36-46.

- Kavkler, M., Tancig, S., & Magajna, L. (2003). Longitudinal study of children with very low mathematical competence in preschool years. Paper presented at the EARLI2003, Padova, Italy, 25–30 August.
- Kidd, J. K., Pasnak, R., Gadzichowski, M., Ferral-Like, M., & Gallington, D. (2008). Enhancing early numeracy by promoting the abstract thought involved in the oddity principle, seriation and conservation. *Journal of Advanced Academics*, *19*, 164-200.
- Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities*, *41*, 451-459.
- Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development*, *22*, 165-184.
- Ruijsenaars, A. J. J. M., Van Luit, J. E. H., & Van Lieshout, E. C. D. M. (2004). *Rekenproblemen en dyscalculie: Theorie, onderzoek, diagnostiek en behandeling*. Rotterdam: Lemniscaat.
- Torbeyns, J., Van den Noortgate, W., Ghesquiere, P., Verschaffel, L., Van de Rijt, B. A. M., & Van Luit, J. E. H. (2002). Development of early numeracy in 5- to 7-year old children: A comparison between Flanders and The Netherlands. *Educational Research and Evaluation*, *8*, 249-275.
- Treffers, A., & De Moor, E. (1990). *Proeve van een nationaal programma voor het rekenwiskundeonderwijs op de basisschool. Deel 2: Basisvaardigheden en cijferen*. Tilburg: Zwijssen.
- Van de Rijt, B. A. M. (1996). *Voorbereidende rekenvaardigheid bij kleuters. De ontwikkeling van rekenvaardigheidsschalen en een onderzoek naar de invloed van een programma*. Doetinchem: Graviant.
- Van de Rijt, B. A. M., & Van Luit, J. E. H. (1998). Effectiveness of the additional early mathematics program for teaching children early mathematics. *Instructional Science*, *26*, 337-358.
- Van Luit, J. E. H., & Schopman, E. A. M. (2000). Improving early numeracy of young children with special Educational Needs. *Remedial & Special Education*, *21*, 27-41.
- Van Luit, J. E. H., & Van de Rijt, B. A. M. (2009). *Utrechtse Getalbegrip Toets – Revised*. Doetinchem: Graviant.
- Van Luit, J. E. H., Van de Rijt, B. A. M., & Pennings, A. H. (1994). *Utrechtse Getalbegrip Toets*. Doetinchem: Graviant.

- Varol, F., & Farran, C. (2006). Early mathematical growth: How to support young children's mathematical development. *Early Childhood Education Journal*, 33, 381-387.
- Yang, D. C., Li, M. N., & Lin, C. I. (2007). A study of the performance of 5th graders in number sense and its relationship to achievement in mathematics. *International Journal of Science and Mathematics Education*, 6, 789-807.