

Is de cloze-toets een betrouwbare en valide maat voor begrijpend lezen?

M. M. L. Muijselaar, E. H. de Bree, E. G. Steenbeek-Planting en P. F. de Jong

Samenvatting

De cloze-toets (een toets waarbij kinderen gaten in de tekst moeten opvullen) is een korte leestoets die gebruikt wordt om begrijpend lezen te meten. Maar over de kwaliteit van de cloze-toets als maat voor begrijpend lezen bestaat twijfel. In dit onderzoek is in twee studies de betrouwbaarheid en de validiteit van twee typen meerkeuze cloze-toetsen onderzocht: een cloze-toets met systematisch geplaatste gaten (SG), gaten die om het 11e woord zijn geplaatst, en een cloze-toets met variabel geplaatste gaten (VG), waarbij de gaten zo zijn geplaatst dat tekstbegrip wordt gevraagd. De eerste studie is uitgevoerd bij 312 kinderen uit groep 7, de tweede bij 100 kinderen uit de groepen 5 en 6. De resultaten van beide studies laten zien dat de VG cloze-toets hoger samenhangt met woordenschat dan de SG cloze-toets, terwijl de SG cloze-toets juist hoger samenhangt met technisch lezen dan de VG cloze-toets. Daarnaast hangt de accuratesse waarmee de cloze-toets wordt gemaakt hoger samen met begrijpend lezen dan de snelheidscore. Geconcludeerd wordt dat de accuratesse van de VG cloze-toets de meest valide cloze-toets is om begrijpend lezen te meten. De betrouwbaarheid van deze toets is echter onvoldoende. Daarom wordt geconcludeerd dat geen van de onderzochte varianten van de cloze-toets daadwerkelijk geschikt lijkt om begrijpend lezen te meten.

Kernwoorden: begrijpend lezen; cloze-toets; validiteit; betrouwbaarheid; technisch lezen; woordenschat

schoolsucces en het is een vereiste voor een leven lang leren (zie bijvoorbeeld Spörer & Brunstein, 2009). Ondanks het feit dat het kunnen lezen en begrijpen van teksten noodzakelijk is om in de huidige maatschappij te functioneren, heeft een substantieel deel van de kinderen (zo'n 10 tot 15%) ernstige problemen met het begrijpen van teksten (Kordes, Bolsinova, Limpens, & Stolwijk, 2013). Om zulke problemen snel te kunnen signaleren, is het belangrijk om de ontwikkeling van deze vaardigheid nauwkeurig te volgen. Voor het meten van begrijpend lezen wordt soms een cloze-toets gebruikt. Bij deze toets wordt een tekst gepresenteerd waarin woorden zijn weggelaten, een zogenaamde tekst met gaten. De opdracht is om de tekst zo snel mogelijk te lezen. Daarbij kan de vraag zijn om in te vullen welk woord zou horen bij het gat (Janzen & Boersma, 2013; Kamalski, 2007; Kamalski, Sanders, Lentz, & Van den Bergh, 2005; Kraf & Pander Maat, 2009; Van Dooren, Van den Bergh, & Evers-Vermeul, 2012). Een andere mogelijkheid is om bij elk gat aan te geven welk woord uit een keuze van drie past in de tekst (Espin, Wallace, Lembke, Campbell, & Long, 2010; Gellert & Elbro, 2013). De snelheid en accuratesse waarmee dit wordt gedaan, zou een maat zijn voor begrijpend lezen.

De cloze-toets is aantrekkelijk omdat de afname maar een paar minuten duurt. De afname van standaardtoetsen voor begrijpend lezen, zoals de toetsen begrijpend lezen van het Leerling- en Onderwijs Volgstelsel (Cito, 2008), is vele malen tijdrovender (ongeveer anderhalf uur). Deze toetsen worden in de bovenbouw van de basisschool dan ook slechts één keer per schooljaar afgenomen. Eén keer per jaar toetsen is te weinig om te signaleren of kinderen problemen hebben met begrijpend lezen. Het tussendoor onderzoeken van begrijpend lezen zou wenselijk zijn. Cloze-toetsen zijn ontwikkeld om

1 Inleiding

Begrijpend lezen is een van de belangrijkste vaardigheden die kinderen moeten verwerven in de bovenbouw van de basisschool: het is de meest belangrijke voorspeller voor later

de (technische) leesvaardigheid van kinderen te volgen (zie bijvoorbeeld Espin, McMaster, Rose, & Wayman, 2012; Espin et al., 2010). Door deze toets elke zes weken af te nemen, kan de ontwikkeling van de algemene leesvaardigheid goed in kaart worden gebracht. Bij de cloze-toetsen zijn echter nog geen normscores voorhanden, zoals dat bij Cito wel is.

Een andere vraag is of de cloze-toets een goede maat is voor begrijpend lezen in termen van betrouwbaarheid en validiteit. Daarover bestaat twijfel (zie bijvoorbeeld Kamalski, 2007; Gellert & Elbro, 2013; Muijselaar, Kendeou, De Jong, & Van den Broek, 2017). In deze studie worden de validiteit en betrouwbaarheid van verschillende typen cloze-toetsen voor het meten van begrijpend lezen nader onderzocht. De betrouwbaarheid zegt iets over hoe goed vragen van een toets met elkaar samenhangen en over de herhaalbaarheid van de toetsscore; in deze studie wordt daarvoor de paralleltestbetrouwbaarheid onderzocht. Wat betreft validiteit worden de convergente en de begripsvaliditeit nader bekeken. De convergente validiteit betreft hoe goed de ontwikkelde toets lijkt op soortgenoten. De begripsvaliditeit verwijst naar hoe goed de ontwikkelde toets het onderliggende construct begrijpend lezen meet. De convergente validiteit is onderdeel van de begripsvaliditeit.

Tekstbegrip is een complex cognitief proces waarbij een coherent mentaal model van waar de tekst over gaat wordt geconstrueerd. Dit product van begrijpend lezen wordt ook wel een situatiemodel genoemd (Kintsch & Van Dijk, 1978). Aangezien in de meeste teksten niet alle informatie staat die nodig is om zo'n coherent situatiemodel te construeren, moet de lezer inferenties maken om nabije stukjes informatie in de tekst met elkaar te verbinden (lokale inferenties) en om de tekst aan te vullen met informatie die niet letterlijk in de tekst staat (globale inferenties) (Currie & Cain, 2015; Cain & Oakhill, 1999; Long & Chong, 2001). Het is daarom belangrijk dat toetsen voor begrijpend lezen zowel lokaal begrip meten (op woord- en zinsniveau), als ook globaal begrip op tekstniveau (Gellert & Elbro, 2013).

Begrijpend lezen steunt op twee belangrijke deelvaardigheden: het efficiënt decoderen van woorden (technisch lezen), en het weten of kunnen achterhalen van de betekenis van die woorden in de tekst (woordenschat; zie bijvoorbeeld Hoover & Gough, 1990; Verhoeven & Perfetti, 2008). Een toets voor begrijpend lezen zal dus een beroep moeten doen op zowel technisch lezen als woordenschat. Deze vaardigheden faciliteren ook het maken van inferenties en zijn dus nodig om een coherent situatiemodel van de tekst te creëren. Van de standaardtoetsen voor begrijpend lezen is bekend dat ze samenhangen met zowel technisch lezen als woordenschat (zie bijvoorbeeld Muijselaar & De Jong, 2015). Uit een aantal internationale studies blijkt echter dat cloze-toetsen vaak een sterker beroep doen op technisch lezen dan standaardtoetsen voor begrijpend lezen (zie bijvoorbeeld Keenan, Betjemann, & Olson, 2008; Kendeou, Papadopoulos, & Spanoudis, 2012; Muijselaar, Kendeou et al., 2017; Nation & Snowling, 1997) terwijl de rol van woordenschat juist minder groot is bij cloze-toetsen (Muijselaar, Kendeou et al., 2017). Dit soort resultaten suggereert dat de cloze-toets vooral een maat voor lokaal tekstbegrip en technisch lezen is en dus minder geschikt is voor het meten van begrijpend lezen. Om een goede maat voor begrijpend lezen te zijn, met een goede begripsvaliditeit, moet de cloze-toets zowel lokaal als globaal tekstbegrip meten en dus vergelijkbaar samenhangen met technisch lezen en woordenschat als standaardtoetsen voor begrijpend lezen.

Uit Nederlandse studies over de cloze-toets (zie bijvoorbeeld Jansen & Boersma, 2013; Kamalski et al., 2005; Sanders, Lentz & Van den Bergh, 2005) blijkt dat het standaard cloze-format is dat woorden systematisch worden weggelaten uit de tekst en worden vervangen door een streepje. De opdracht is om het juiste woord in te vullen op de plaats van het streepje. In deze studies geldt geen tijdslimiet voor de cloze-toets. In internationale literatuur is het standaard cloze-format daarentegen dat in een tekst op de plek van elk zevende woord drie meerkeuze alternatieven worden geplaatst, te weten het goede antwoord en twee afleiders (zie bijvoorbeeld

Fuchs & Fuchs, 1992). Bijvoorbeeld: 'De jongen is niet op zoek (koop / naar / ster) een nieuwe telefoon, een flatscreen of (een / hok / jam) tablet.' (CED-groep, 2013). Bij de internationale studies geldt wel een tijdslijm. Een belangrijk verschil tussen de cloze-toets in Nederland en die in internationale studies is dus het cloze-format (open vragen vs. meerkeuze vragen). Omdat in deze studie een vergelijking wordt gemaakt met eerder internationaal onderzoek (Gellert & Elbro, 2013), is ervoor gekozen om het format met meerkeuze vragen te hanteren (hieronder meer informatie over het onderzoek van Gellert en Elbro, 2013). Er wordt op verschillende manieren gescoord. Zo wordt het aantal goed gemaakte vragen binnen de tijd gehanteerd als score (vanaf nu somscore genoemd), maar ook het aantal goed gemaakte vragen gedeeld door het aantal vragen gemaakt binnen de tijd (vanaf nu proportiescore genoemd). De vraag is welke scoringsmethode van de cloze-toets het meest geschikt is om begrijpend lezen te meten.

Gellert en Elbro (2013) stellen dat eerder ontwikkelde cloze-toetsen veelal een beroep doen op lokaal tekstbegrip, dus tussen frasen en zinnen, en daardoor minder geschikt zijn om het construct begrijpend lezen te meten. Gellert en Elbro ontwikkelden daarom een cloze-toets die meer globaal tekstbegrip vereist, en dus een betere maat zou zijn voor begrijpend lezen. In de cloze-toets waren de gaten variabel geplaatst in plaats van systematisch. In plaats van elk n -de woord te vervangen door drie alternatieven, waarbij meer lokaal begrip gevraagd wordt, werden woorden weggelaten die begrip vergen van meerdere zinnen, de gehele paragraaf of zelfs de hele tekst. Daardoor, zo veronderstelden Gellert en Elbro, is meer globaal begrip noodzakelijk en is een coherente mentale representatie van de tekst nodig om de test goed te kunnen maken. De woorden die werden weggelaten waren verbindingswoorden (voegwoorden, verwijswaarden en synoniemen) of woorden die vragen om inductie of causale inferentie. De cloze-toets van Gellert en Elbro had een afdameduur van tien minuten.

In het onderzoek van Gellert en Elbro (2013) werd de taak afgenomen bij Deense

(jong)volwassenen (gemiddelde leeftijd 34.1; $SD = 12.8$) die volwassenonderwijs volgden vanwege hun zwak ontwikkelde leesvaardigheden. Uit de resultaten bleek dat de taak sterk samenhangt met een standaard meerkeuze toets begrijpend lezen ($r = .84$), wat een aanwijzing is voor de (convergente) validiteit van de cloze toets (Evers, Lucassen, Meijer, & Sijtsma, 2009). De taak hing ook vergelijkbaar sterk samen met technisch lezen ($r_{cloze} = .73$; $r_{standaard} = .63$), woordenschat ($r_{cloze} = .57$; $r_{standaard} = .60$) en een standaard toets begrijpend lezen, wat eveneens pleit voor een goede begripsvaliditeit. Het enige verschil was dat de cloze-toets een iets sterker beroep leek te doen op technisch lezen dan de standaard toets begrijpend lezen. Dit verschil verdween wanneer de proportiescore in plaats van de somscore werd gebruikt ($r_{cloze} = .45$; $r_{standaard} = .43$). Bij de somscore is zowel de accuratesse als de snelheid waarmee de toets gemaakt wordt van belang voor de score, en deze is daardoor ook meer afhankelijk van de leessnelheid. Bij de proportiescore bepaalt alleen de accuratesse hoe hoog gescoord wordt. Deze score leunt (mogelijk) meer op inhoud en dus begrip van de tekst. De resultaten van Gellert en Elbro suggereren dat de proportiescore geschikter is voor het meten van begrijpend lezen dan de somscore.

1.1 Het huidige onderzoek

In dit onderzoek werd met twee deelstudies de betrouwbaarheid en de validiteit van de cloze-toets onderzocht. In lijn met de studie van Gellert en Elbro (2013) werd een cloze-toets ontwikkeld die beoogt meer globaal tekstbegrip te meten dan de eerder ontwikkelde cloze-toetsen. Deze cloze-toets zal verder cloze-toets met variabel geplaatste gaten (VG) genoemd worden. Om te kunnen onderzoeken of deze VG cloze-toets een ander beroep doet op begrijpend lezen en onderliggende vaardigheden werd deze vergeleken met een cloze-toets met systematisch geplaatste gaten (SG). Een vergelijking tussen een VG en een SG cloze-toets werd in de studie van Gellert en Elbro (2013) niet gemaakt. Daarnaast waren de deelnemers in de studie van Gellert & Elbro (2013) (jong) volwassenen met zwak ontwikkelde leesvaar-

digheden. Aan de huidige studies werd deelgenomen door basisschoolleerlingen met een ruime variatie in leesvaardigheid. Het doel van deze studies was om de betrouwbaarheid en validiteit van de cloze-toets te onderzoeken in de Nederlandse onderwijssetting.

De betrouwbaarheid van de cloze-toetsen werd in kaart gebracht door het bepalen van de paralleltest betrouwbaarheid. Voor de validiteit werd eerst onderzocht hoe de SG cloze-toets en VG cloze-toets samenhangen met standaardtoetsen voor begrijpend lezen. Bij de VG cloze-toetsen werden, net als in de studie van Gellert en Elbro (2013), verbindingswoorden weggelaten (voegwoorden, verwijswaarden en synoniemen) of woorden die inductie of causale inferentie vereisten. Bij de SG cloze-toetsen werden de antwoordalternatieven zo gekozen dat de woordlengte van alle alternatieven vergelijkbaar was. Een hoge samenhang tussen de cloze-toetsen en toetsen voor begrijpend lezen is een aanwijzing voor een goede convergente validiteit van de toets (Evers et al., 2009). Ten tweede werd onderzocht of het uitmaakt of een snelheidsbenadering (somscore) of een accuratessebenadering (proportiescore) wordt gebruikt bij beide cloze-toetsen. Ten derde is onderzocht in welke mate deze cloze-toetsen een beroep doen op technisch lezen en woordenschat. Als dit vergelijkbaar is met standaardtoetsen voor begrijpend lezen, is dit ook een aanwijzing voor een goede begripsvaliditeit. Ten vierde is onderzocht of de tijd die kinderen krijgen om de VG cloze-toets te maken van invloed is op hoe goed deze toets begrijpend lezen meet.

Om deze vragen te beantwoorden werd in Studie 1 een cloze-toets met een standaard cloze-format, dus een toets met systematisch geplaatste gaten (SG1 cloze-toets), vergeleken met een toets met variabel geplaatste gaten (VG1 cloze-toets). Beide toetsen bestonden uit twee teksten, waarvoor de kinderen elk twee minuten de tijd krijgen. Er wordt verwacht dat de VG1 cloze-toets sterker samenhangt met begrijpend lezen dan de SG1 cloze-toets. Daarnaast werd voor beide cloze-toetsen het verschil onderzocht in het gebruik van een snelheidsbenadering (somscore), waarbij accuratesse een zeer kleine

rol speelt, of een pure accuratessebenadering (proportiescore), waarbij verwacht wordt dat de relatie met begrijpend lezen hoger is dan bij de snelheidsbenadering. In lijn met de bevindingen van Gellert en Elbro (2013) wordt verwacht dat de VG1 cloze-toets sterker samenhangt met woordenschat en minder sterk samenhangt met technisch lezen dan de SG1 cloze-toets, aangezien de VG1 cloze-toets waarschijnlijk zowel lokaal als globaal tekstbegrip meet, terwijl de SG1 cloze-toets voornamelijk lokaal begrip zal meten.

In Studie 2 werd eveneens een toets met systematisch geplaatste gaten (SG2 cloze-toets) vergeleken met een toets met variabel geplaatste gaten (VG2 cloze-toets). Het belangrijkste verschil tussen Studie 1 en Studie 2 was dat kinderen voor de VG1 cloze-toets tweemaal twee minuten de tijd krijgen en voor de VG2 cloze-toets eenmaal 10 minuten. Verwacht wordt dat de VG2 cloze-toets nog meer lijkt op de standaardtoetsen voor begrijpend lezen dan de VG1 cloze-toets, omdat het bij de kortere toets mogelijk zo is dat sommige kinderen maar een paar vragen af krijgen, waardoor er minder een beroep wordt gedaan op het situatiemodel. Dat wil zeggen dat verwacht wordt dat de VG2 cloze-toets een vergelijkbare samenhang heeft met technisch lezen en woordenschat als standaardtoetsen voor begrijpend lezen.

2 Methoden Studie 1

2.1 Deelnemers

De data van dit onderzoek waren onderdeel van een groter longitudinaal en experimenteel onderzoek naar begrijpend lezen (zie Muijselaar, Swart et al., 2017; Swart et al., 2016). Aan het huidige onderzoek namen 312 kinderen deel uit groep 7 met een gemiddelde leeftijd van 10 jaar en 10 maanden ($SD = 5.51$), van wie ongeveer de helft jongen was (jongen: $N = 157$, meisje: $N = 155$). De kinderen waren afkomstig uit 13 klassen van 12 basisscholen in Nederland. Van deze kinderen was 4.7% niet in Nederland geboren en spreekt 2% met beide ouders een andere taal dan het Nederlands.

2.2 Onderzoeksinstrumenten

Om de verschillen tussen de twee typen cloze-toetsen te onderzoeken, werden twee cloze-toetsen ontworpen. Daarnaast werden toetsen voor begrijpend lezen, technisch lezen en woordenschat geselecteerd.

Cloze-toetsen

Voor dit onderzoek werden twee verschillende cloze-toetsen ontwikkeld. Beide cloze-toetsen bestaan uit dezelfde twee teksten met 67 gaten (vragen). De teksten waren informatief en bevatten respectievelijk 373 en 410 woorden. De teksten waren afkomstig van de methode Nieuwsbegrip en hebben niveau B (geschikt voor kinderen uit groep 7 en 8). De teksten werden achter elkaar gezet in een toetsboekje waar de vragen met antwoordopties in verwerkt zijn. Elke vraag had drie antwoordmogelijkheden; het kind diende het antwoord te omcirkelen dat het beste op de plek van het gat, de lege plek in de tekst, paste. De eerste cloze-toets bevatte systematisch geplaatste gaten (SG1 cloze-toets) en de tweede cloze-toets variabel geplaatste gaten (VG1 cloze-toets). Bij de SG1 cloze-toets werd, na de eerste zin die intact bleef, steeds elk elfde woord vervangen door drie antwoordmogelijkheden. Deze antwoordmogelijkheden werden aselekt gekozen, waarbij de woordlengtes van de opties gelijk zijn gehouden. Een item is bijvoorbeeld: "Een beter leven. Unilever wil kippen gaan gebruiken die een (omdat / zagen / beter) leven hebben gehad." Bij de VG1 cloze-toets werden de gaten variabel gekozen, waarbij in totaal evenveel gaten zijn gecreëerd als bij de SG1 cloze-toets. Hierbij werd onderscheid gemaakt tussen verbindingswoorden (voegwoorden, verwijswwoorden en synoniemen) en inductie of causale inferenties. De verhouding tussen verbindingswoorden en inductie of causale inferenties was gelijk gehouden. De antwoordalternatieven waren altijd van dezelfde woordsoort als het weggelaten woord, ofwel het juiste antwoord. Een deel van de VG1 cloze-toets was bijvoorbeeld: "Een andere vereiste is dat de kippen snel kunnen groeien: dan hebben de fokkers zelf niet zo veel (winst / kosten / inkomen), maar kunnen ze (toch / nooit / bijvoorbeeld) een

goede prijs voor hun kippen vragen." Hierbij is het eerste gat een causale inferentie en het tweede gat een verbindingswoord. Voor de afname van de cloze-toets zijn vijf voorbeeldvragen besproken en zijn de goede antwoorden toegelicht. De kinderen kregen voor elke tekst twee minuten de tijd om zoveel mogelijk vragen goed te maken. De teksten werden direct achter elkaar gemaakt.

Voor beide cloze-toetsen werd zowel een somscore als een proportiescore berekend. De somscore is het aantal goed gemaakte vragen, dit is dus een maat voor snelheid en accuratesse. De proportiescore is het aantal goed gemaakte vragen gedeeld door het totaal aantal gemaakte vragen, waarbij dus alleen accuratesse een rol speelt.

Begrijpend lezen

Voor begrijpend lezen werd de toets begrijpend lezen voor groep 6, 7 en 8 afgenomen van Aarnoutse en Kapinga (2006). Deze toets bestond uit een boekje met zeven korte verhalende en informatieve teksten (122 tot 288 woorden). Na elke tekst waren er 5 tot 7 meerkeuzevragen en de toets bevatte 44 vragen in het totaal. De vragen hadden twee (waar/niet waar) of vier antwoordmogelijkheden. Voor het begin van de test werd een voorbeeldtekst samen met de kinderen gemaakt en werden de goede antwoorden op de bijbehorende vragen toegelicht. De teksten uit de toets bleven beschikbaar tijdens het maken van de test. De testafname kostte ongeveer 50 minuten. De score was het aantal goed beantwoorde vragen. De betrouwbaarheid was goed (Cronbachs alfa = .83).

Technisch lezen

Technisch lezen werd gemeten met de Eén-MinuuT-Test (Brus & Voeten, 1979). Aan de kinderen werd gevraagd om zoveel mogelijk woorden correct te lezen van een lijst die bestaat uit 116 woorden met één tot vijf lettergrepen. De score was het aantal woorden dat een kind juist had gelezen in één minuut. De gerapporteerde betrouwbaarheid was goed (gemiddelde parallel-test relatie is $r = .90$, Van den Bos, Lutje Spelberg, Scheepstra, & de Vries, 1994).

Woordenschat

De Nederlandse klassikale versie van de Peabody Picture Vocabulary Test werd gebruikt om woordenschat te meten (Schlichting, 2005). Er werd een aangepaste versie gebruikt waarin sets 8 tot en met 13 zijn afgenomen, zodat de afname minder tijd in beslag nam. De toets bestond in het totaal uit 72 items met elk vier antwoordmogelijkheden. Door deze aanpassing kon de test klassikaal worden afgenomen. Kinderen werden geïnstrueerd om het plaatje te onderstrepen dat hoorde bij het woord dat de onderzoeksassistent uitsprak. Twee oefenitems werden afgenomen voor het begin van de test. De afname van de test nam ongeveer 30 minuten in beslag. De score was het aantal juiste antwoorden. De betrouwbaarheid was voldoende (Cronbachs alfa was .73).

2.3 Procedure

De cloze-toetsen, de toets begrijpend lezen en de woordenschattoets werden in de klas afgenomen in twee ochtenden met voldoende pauzes. De kinderen werden binnen klassen aselekt toegewezen aan één van beide cloze-toetsen: de twee cloze-toetsen werden om en om uitgedeeld aan de kinderen. De ene groep kinderen maakte de SG1 cloze-toets, de andere groep de VG1 cloze-toets. Op de eerste testdag werd de woordenschattoets afgenomen. Op de tweede testdag werd eerst de toets begrijpend lezen afgenomen en vervolgens de cloze-toetsen. De test voor technisch lezen werd individueel afgenomen.

3 Resultaten Studie 1

3.1 Controle van de data en beschrijvende statistieken

Voor het uitvoeren van de analyses, werden alle variabelen gecontroleerd op uitbijters ($-3 < z < 3$). Voor de 7 uitbijters werden de scores vervangen door de score die hoort bij $z = -/+ 3$. De data werd ook gecontroleerd op ontbrekende scores. Ongeveer 4% van de scores ontbrak vanwege bijvoorbeeld ziekte van de kinderen. De beschrijvende statistieken van de verschillende variabelen zijn vermeld in Tabel 1. Alle variabelen waren normaal verdeeld, met uitzondering van de

proportiescore van de SG1 cloze-toets. De correlaties tussen de variabelen staan beschreven in Tabel 2.

3.2 Betrouwbaarheid van de cloze-toetsen

Voor de somscores van de SG1 cloze-toets bleek de paralleltest betrouwbaarheid van de twee teksten goed ($r = .79$), van de proportiescores van de SG1 cloze-toets was deze onvoldoende ($r = .46$). De betrouwbaarheid van de VG1 cloze-toets was voldoende: $r = .61$ voor de somscores, en onvoldoende: $r = .43$ voor de proportiescores.

3.3 Validiteit van de cloze-toetsen

Allereerst werd de samenhang tussen de cloze-toetsen met begrijpend lezen onderzocht. De cloze-toetsen (zowel somscore als proportiescore) hingen matig tot goed samen met begrijpend lezen ($r = .33 - r = .65$, zie Tabel 2). Ten tweede werden de verschillen tussen de correlaties van de twee scoringswijzen van de cloze-toetsen met begrijpend lezen getoetst. Hiervoor werd de Fisher's z -test gebruikt. De relatie tussen de proportiescores van beide cloze-toetsen en begrijpend lezen bleek hoger te zijn dan deze relatie tussen de somscores van beide cloze-toetsen en begrijpend lezen (SG1: $z = 3.21$, $p < .001$; VG1: $z = 2.22$, $p = .026$, zie Tabel 2 voor de correlaties). Alhoewel het leek dat de SG1 cloze-toets minder hoog correleerde met begrijpend lezen dan de VG1 cloze-toets, was dit verschil niet significant.

Ten derde werd onderzocht of de samenhang van de SG1 cloze-toets met technisch lezen, woordenschat en begrijpend lezen anders was dan van de VG1 cloze-toets. Hiervoor werd een hiërarchische regressieanalyse uitgevoerd met de score op de cloze-toetsen als afhankelijke variabele. De onafhankelijke variabelen waren type toets (SG1 vs. VG1), technisch lezen, woordenschat, begrijpend lezen, de interactie van type toets en technisch lezen, de interactie van type toets en woordenschat, en de interactie van type toets en begrijpend lezen. Om de interpretatie van de interactietermen te vergemakkelijken, werden de variabelen technisch lezen, woordenschat en begrijpend lezen gecentreerd (Echambadi & Hess, 2007).

Tabel 1

Beschrijvende Statistieken voor de Cloze-toetsen, Begrijpend Lezen, Technisch Lezen en Woordenschat in Studie 1

	Maximum	M	SD	Scheefheid	Gepiektheid
SG1 cloze-toets som	57	28.42	9.48	0.59	0.38
SG1 cloze-toets proportie	1	.91	.10	-1.57	1.83
VG1 cloze-toets som	40	19.15	7.21	0.47	-0.30
VG1 cloze-toets proportie	1	.68	.15	-0.34	-0.53
Begrijpend lezen	43	30.63	6.60	-0.42	-0.71
Technisch lezen	112	73.61	14.49	0.05	0.00
Woordenschat	57	42.15	6.54	-0.10	-0.15

Noot. SG1 cloze-toets: cloze-toets met systematisch geplaatste gaten van Studie 1; VG1 cloze-toets: cloze-toets met variabel geplaatste gaten van Studie 1.

Tabel 2

Correlaties tussen de Cloze-toetsen, Begrijpend Lezen, Technisch Lezen en Woordenschat in Studie 1

	1	2	3	4	5	6	7
1 SG1 cloze-toets som	1						
2 SG1 cloze-toets proportie	.29**	1					
3 VG1 cloze-toets som	-	-	1				
4 VG1 cloze-toets proportie	-	-	.46**	1			
5 Begrijpend lezen	.33**	.62**	.47**	.65**	1		
6 Technisch lezen	.63**	.34**	.47**	.25**	.37**	1	
7 Woordenschat	.30**	.36**	.50**	.58**	.60**	.19**	1

Noot. Aangezien de kinderen maar één cloze-toets maakten, is de correlatie tussen deze typen cloze-toetsen niet te berekenen.

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

In Tabel 3 zijn de resultaten van de hiërarchische regressieanalyses met de somscores van de cloze-toetsen als afhankelijke variabele weergegeven. Het negatieve effect van type toets op de somscores van de cloze-toetsen geeft aan dat op de SG1 cloze-toets een hogere score werd behaald dan op de VG1 cloze-toets. De somscore op beide cloze-toetsen werd voor een groot deel bepaald door technisch lezen, en voor een klein deel door woordenschat. Begrijpend lezen speelde geen significante rol bij de somscore van de cloze-toetsen als gecontroleerd was voor technisch lezen en woordenschat. De significante negatieve interactieterm van type toets met technisch lezen geeft aan dat, zoals verwacht, de

rol van technisch lezen groter was bij de SG1 cloze-toets dan bij de VG1 cloze-toets. De interactietermen van type toets met woordenschat en met begrijpend lezen waren niet significant, dus de rol van woordenschat en begrijpend lezen verschilde niet voor de twee cloze-toetsen.

De resultaten van de hiërarchische regressieanalyse met de proportiescores van de cloze-toetsen als afhankelijke variabele zijn ook weergegeven in Tabel 3. Hieruit bleek dat er een negatief effect was van type toets op de proportiescores van beide cloze-toetsen, wat betekent dat kinderen een hogere score behalen op de SG1 cloze-toets dan op de VG1 cloze-toets. De proportiescore van de cloze-

Tabel 3

Hiërarchische Regressieanalyse voor de Verschillen tussen de Somscore en de Proportiescore van de SG1 en VG1 Cloze-toetsen

Somscore van cloze-toetsen		
	ΔR^2	B
1. Type toets	.24**	-.48**
2. Technisch lezen	.25**	.56**
3. Woordenschat	.06**	.16*
4. Begrijpend lezen	.00	.04
5. Type toets x Technisch lezen	.01**	-.18**
6. Type toets x Woordenschat	.00	.08
7. Type toets x Begrijpend lezen	.00	.02
Proportiescore van cloze-toetsen		
	ΔR^2	B
1. Type toets	.44	-.65
2. Technisch lezen	.04	.07
3. Woordenschat	.11	-.00
4. Begrijpend lezen	.07	.33
5. Type toets x Technisch lezen	.00	-.04
6. Type toets x Woordenschat	.02	.18
7. Type toets x Begrijpend lezen	.00	.06

Noot. De gestandaardiseerde regressiecoëfficiënten zijn gepresenteerd voor het model met alle variabelen.

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

toets werd bepaald door technisch lezen, woordenschat en begrijpend lezen. De interactietermen van type toets met technisch lezen en met begrijpend lezen waren niet significant. De positieve, significante interactieterm van type toets met woordenschat wijst uit dat de VG1 cloze-toets een sterker beroep deed op woordenschat dan de SG1 cloze-toets.

4 Conclusie Studie 1

In Studie 1 werden de betrouwbaarheid en validiteit van twee verschillende cloze-toetsen onderzocht, een cloze-toets met systematisch geplaatste gaten (SG1) en een cloze-toets met variabel geplaatste gaten (VG1). De betrouwbaarheid van de proportiescores bleek voldoende tot goed, maar de somscores waren onvoldoende betrouwbaar. Wat

betreft de validiteit werd eerst onderzocht hoe de verschillende cloze-toetsen samenhangen met begrijpend lezen. Deze samenhang was matig tot goed, wat betekent dat, afhankelijk van het type toets, de convergente validiteit matig tot goed is. Ten tweede werd onderzocht of het uitmaakt of somscores of proportiescores werden gebruikt in de analyses. De proportiescores van beide cloze-toetsen hingen sterker samen met begrijpend lezen dan de somscores. Op basis daarvan kan geconcludeerd worden dat de convergente validiteit van de pure accuratesbenadering (gebruik van proportiescores) hoger is dan van de snelheidsbenadering met accuratesse (gebruik van somscores). Ten derde werd onderzocht in welke mate de cloze-toetsen een beroep deden op technisch lezen, woordenschat en begrijpend lezen. De regressieanalyses toonden aan dat technisch lezen een grotere rol speelt bij de somscore

van de SG1 cloze-toets dan bij de somscore van de VG1 cloze-toets. Daarnaast bleek woordenschat belangrijker te zijn bij de proportiescore van de VG1 cloze-toets dan bij de proportiescore van de SG1 cloze-toets. Ook bleek dat begrijpend lezen, na controle voor technisch lezen en woordenschat, geen rol speelt bij de somscores van de cloze-toetsen, maar wel bij de proportiescores. Uit deze analyses bleek de accuratessebenadering (proportiescores) van de VG1 cloze-toets het beste te zijn als toets voor begrijpend lezen; de VG1 cloze-toets heeft de beste begripsvaliditeit. Opvallend was ook dat wanneer de regressiecoëfficiënten van beide regressieanalyses vergeleken worden, de rol van technisch lezen groter lijkt voor de somscores dan voor de proportiescores. De snelheidsbenadering leek dus meer gebaseerd te zijn op technisch lezen dan de accuratessebenadering. Samenvattend kan gesteld worden dat de accuratessebenadering van de VG1 cloze-toets een meer valide manier is om begrijpend lezen te meten, maar daarbij moet direct opgemerkt worden dat de betrouwbaarheid van de accuratessescore onvoldoende is, wat de toets in de huidige vorm ongeschikt maakt als maat voor begrijpend lezen.

De vierde onderzoeksvraag, of de tijd die kinderen kregen voor het maken van de VG cloze-toets invloed heeft op de validiteit van de toets, kan beantwoord worden door de resultaten van Studie 1, waarbij een korte VG cloze-toets is gebruikt, te vergelijken met de resultaten van Studie 2, waarbij kinderen meer tijd kregen voor de VG cloze-toets. Daarnaast werd in Studie 2 onderzocht of de resultaten van Studie 1 gerepliceerd kunnen worden bij iets jongere kinderen (groep 5 en 6 in Studie 2 versus groep 7 in Studie 1). In Studie 2 is getracht het onderscheid tussen de verschillende cloze-toetsen te vergroten door de VG2 cloze-toets te verlengen, en door de SG2 cloze-toets nog meer een maat voor technisch lezen te laten zijn, waar nauwelijks leesbegrip voor nodig is. Dit is gedaan door voor de antwoordopties van de SG2 cloze-toets niet langer te kiezen voor drie heel verschillende woorden, maar voor drie woorden die maximaal twee letters van elkaar verschillen.

5 Methoden Studie 2

5.1 Deelnemers

De data van deze studie maakten deel uit van een groter cross-sectioneel onderzoek naar begrijpend luisteren. Aan de huidige studie namen 100 kinderen deel uit groep 5 en 6 van gemiddeld 9 jaar en 6 maanden ($SD = 8.39$), van wie ongeveer de helft jongen was (jongen: $N = 46$, meisje: $N = 54$). De kinderen zijn afkomstig uit 6 klassen van 3 basisscholen in Nederland. Van de kinderen spreekt 94% het liefste Nederlands.

5.2 Onderzoeksinstrumenten

In Studie 2 werden twee andere cloze-toetsen ontwikkeld dan in Studie 1. Daarnaast werden toetsen voor begrijpend lezen, technisch lezen en woordenschat geselecteerd.

Cloze-toetsen

Voor deze studie werden twee verschillende cloze-toetsen ontwikkeld, die vergelijkbaar zijn met de cloze-toetsen uit Studie 1. De cloze-toets met systematisch geplaatste gaten uit Studie 2 (SG2 cloze-toets) bevatte dezelfde 67 systematisch geplaatste gaten en dezelfde twee teksten als de SG1 cloze-toets. De antwoordmogelijkheden bij de SG2 cloze-toets waren echter anders. In de SG1 cloze-toets werden willekeurig drie opties gegeven. In de SG2 cloze-toets werd van het juiste antwoord steeds één of twee letters veranderd om de andere antwoordopties te maken. Een voorbeeld van een item uit de SG2 cloze-toets is: "Het bedrijf wil kippen gaan gebruiken die een beter leven (hekken / hebben / hellen) gehad." De antwoordopties waren allemaal bestaande Nederlandse woorden. Voor de afname van de SG2 cloze-toets werden twee voorbeeldvragen besproken. De kinderen kregen voor beide teksten twee minuten de tijd om zoveel mogelijk vragen goed te maken, en beide teksten werden direct na elkaar afgenomen.

Bij de VG2 cloze-toetsen werden de gaten, net als bij de VG1 cloze-toets, variabel gekozen. Hierbij was weer onderscheid gemaakt tussen verbindingswoorden (voegwoorden, verwijfswoorden en synoniemen) en inductie of causale inferenties. De verhouding tussen

verbindingswoorden en inductie of causale inferenties was gelijk gehouden. Een deel van de VG2 cloze-toets met eerst een verbindingswoord en vervolgens een causale inferentie is: “Mensen maken zich zorgen over die bevingen, (want / zoals / tenzij) door de (bodem / gassen / trillingen) raken huizen beschadigd.” De toets bestond uit vijf informatieve teksten van tussen de 304 en 373 woorden. De teksten waren afkomstig van de methode Nieuwsbegrip en hebben niveau A (geschikt voor kinderen uit groep 5 en 6). De teksten werden achter elkaar gezet in een toetsboekje waar de vragen met antwoordopties in verwerkt zijn. De verhouding van cloze-items versus woorden in de tekst was 1:20, met in totaal 83 cloze-items. Voor de afname van de VG2 cloze-toets werden drie voorbeeldvragen besproken. De kinderen kregen tien minuten de tijd om zoveel mogelijk vragen goed te maken van de vijf teksten.

Voor zowel de SG2 als de VG2 cloze-toetsen werd een somscore en een proportiescore berekend. De somscore is het aantal goed gemaakte vragen. De proportiescore is het aantal goed gemaakte vragen gedeeld door het totaal aantal gemaakte vragen.

Begrijpend lezen

Om begrijpend lezen te meten werd gebruik gemaakt van de toetsen voor groep 5 en 6 van het Leerling- en Onderwijs Volgstelsel (LOVS; Cito, 2008). Deze testen zijn deels adaptief aan het niveau van begrijpend lezen van het kind. Dat wil zeggen dat alle kinderen het eerste deel van de toets maken, en dat de score op het eerste deel bepaalt of ze het makkelijkere of moeilijker tweede deel moeten maken. Elk deel bestond uit 25 meerkeuzevragen. De toetsen voor groep 5 en groep 6 bestonden uit verschillende verhalende en informatieve teksten. Zowel het eerste als tweede deel van de toets nam ongeveer 45 minuten in beslag. Voor de analyses werd de ruwe score omgezet in een vaardigheidscore. De betrouwbaarheid van de LOVS toetsen was goed (Cronbachs alfa = .75, Muijselaar & De Jong, 2015).

Technisch lezen

Technisch lezen werd, net als in Studie 1,

gemeten met de Eén-Minuu-Test (Brus & Voeten, 1979).

Woordenschat

De Nederlandse versie van de Peabody Picture Vocabulary Test werd gebruikt om woordenschat te meten (Schlichting, 2005). In tegenstelling tot Studie 1, werd in Studie 2 wel de adaptieve individuele test afgenomen. Deze test bestaat uit 208 items, welke zijn verdeeld over 18 sets. Het instapniveau werd bepaald door de leeftijd van elk kind. Als van de instapset vijf of meer items onjuist werden beantwoord, werd een makkelijkere set afgenomen totdat een set was afgenomen waarin maximaal vier items fout waren. De toets werd afgebroken als het kind 9 of meer items fout had. Kinderen waren geïnstrueerd om het plaatje aan te wijzen dat hoorde bij het woord dat de onderzoeksassistent uitsprak. Voor het begin van de test werden twee oefenitems afgenomen. De afname van de test kostte ongeveer 15 minuten. De ruwe score werd omgezet in de Woordbegripquotiënt, welke werd gebruikt in de analyses. De betrouwbaarheid van de test was goed (Cronbachs alfa = .95; Luk & Bialystok, 2013).

5.3 Procedure

De data van de toets begrijpend lezen is verkregen via de scholen. De cloze-toetsen werden afgenomen in twee klassikale testsessies. Op alle scholen werd eerst klassikaal de VG2 cloze-toets afgenomen en enkele dagen later de SG2 cloze-toets. De toetsen voor woordenschat en technisch lezen werden in een individuele testsessie afgenomen.

6 Resultaten Studie 2

6.1 Controle van de data en beschrijvende statistieken

Voor het uitvoeren van de analyses werd de data gecontroleerd op uitbijters ($z < -3$ of $z > 3$). Uitbijterscores (6) werden vervangen door de score die hoort bij $z = -/+ 3$. De data werd ook gecontroleerd op ontbrekende scores. Ongeveer 1% van de scores ontbrak vanwege bijvoorbeeld ziekte van de kinderen. De beschrijvende statistieken van de verschil-

Tabel 4

Beschrijvende Statistieken voor de Cloze-toetsen, Begrijpend Lezen, Technisch Lezen en Woordenschat in Studie 2

	Maximum	<i>M</i>	<i>SD</i>	Scheefheid	Gepiektheid
SG2 cloze-toets som	45	22.29	7.62	0.35	0.24
SG2 cloze-toets proportie	1	.91	.11	-2.24	6.70
VG2 cloze-toets som	56	25.13	10.45	0.44	0.01
VG2 cloze-toets proportie	1	.75	.14	-0.41	-0.54
Begrijpend lezen	54	26.11	14.05	-0.44	0.23
Technisch lezen	91	61.02	13.57	-0.04	0.10
Woordenschat	141	101.93	12.61	0.22	0.39

Tabel 5

Partiële Correlaties tussen de Cloze-toetsen, Begrijpend Lezen, Technisch Lezen en Woordenschat in Studie 2, gecontroleerd voor Groep

	1	2	3	4	5	6	7
1 SG2 cloze-toets som	1						
2 SG2 cloze-toets proportie	.28**	1					
3 VG2 cloze-toets som	.63**	.26*	1				
4 VG2 cloze-toets proportie	.16	.46**	.42**	1			
5 Begrijpend lezen	.30**	.43**	.38**	.53**	1		
6 Technisch lezen	.53**	.35**	.55**	.27**	.41**	1	
7 Woordenschat	.19	.30**	.31**	.50**	.57**	.27**	1

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

lende variabelen zijn vermeld in Tabel 4. Alle variabelen waren normaal verdeeld, met uitzondering van de SG2 cloze-toets. Ondanks dat is de SG2 cloze-toets, net als in Studie 1, wel geanalyseerd met parametrische toetsen. De partiële correlaties tussen de variabelen, waarbij gecontroleerd werd voor groep (groep 5 of groep 6), staan beschreven in Tabel 5.

6.2 Betrouwbaarheid van de cloze-toetsen

De paralleltest betrouwbaarheid van de somscores van de SG2 cloze-toets was voldoende ($r = .73$), maar voor de proportiescores was deze matig ($r = .56$). Voor de VG2 cloze-toets kon geen betrouwbaarheid berekend worden, aangezien deze toets in één keer was afgenomen en niet bestond uit meerdere paralleltests.

6.3 Validiteit van de cloze-toetsen

Ten eerste werd onderzocht hoe de verschillende cloze-toetsen samenhangen met standaardtoetsen voor begrijpend lezen: deze samenhang blijkt matig tot goed te zijn ($r = .30$ tot $.53$, zie Tabel 5). Ten tweede werden de verschillen onderzocht tussen de correlaties van de cloze-toetsen en begrijpend lezen. Om verschillen tussen correlaties te toetsen, werd gebruik gemaakt van een toets voor verschillen tussen correlaties in twee afhankelijke steekproeven (Lee & Preacher, 2013). Ondanks dat het leek alsof de proportiescore van de VG2 cloze-toets hoger correleerde met begrijpend lezen dan alle andere scores op beide cloze-toetsen, bleek geen van de verschillen significant te zijn.

Ten derde werd er een hiërarchische regressieanalyse uitgevoerd om te toetsen of de samenhang van de SG2 cloze-toets met technisch lezen, woordenschat en begrijpende lezen anders was dan voor de VG2 cloze-toets. Hiervoor zijn de scores van alle kinderen op de ene toets in SPSS onder de scores op de andere toets gezet, zodat een 'long' format van de data werd verkregen. Dit maakte het mogelijk om het effect van de onafhankelijke variabelen type toets (SG2 vs. VG2), technisch lezen, woordenschat, begrijpend lezen, de interactie van type toets en technisch lezen, de interactie van type toets en woordenschat, en de interactie van type toets en begrijpend lezen op de cloze-toets te onderzoeken. In deze analyse werd ook gecontroleerd voor de groep waarin kinderen zaten. De variabelen technisch lezen, woordenschat en begrijpend lezen werden gecentreerd om de interpretatie van de interactietermen te vereenvoudigen (Echambadi & Hess, 2007).

In Tabel 6 zijn de resultaten van de hiërarchische regressieanalyse met de somscores van de cloze-toetsen als afhankelijke variabele weergegeven. Het effect van type toets was significant en positief. Dit is echter niet betekenisvol aangezien de toetsen niet te vergelijken zijn als het gaat om aantal items en de tijd die de kinderen kregen voor de toets. Het significante effect van groep toonde dat kinderen uit groep 6 hoger scoorden op de cloze-toetsen dan kinderen uit groep 5. De resultaten lieten ook een significant positief effect van technisch lezen op de somscores van de cloze-toetsen zien, wat betekent dat technisch lezen een belangrijke rol speelt bij deze toetsen. Er was geen effect van woordenschat en begrijpend lezen, en ook niet van de interactietermen op de somscores van de cloze-toetsen.

De resultaten van de hiërarchische regressieanalyse met de proportiescores van de cloze-toetsen als afhankelijke variabele zijn ook weergegeven in Tabel 6. Hieruit blijkt een negatief effect van type toets op de proportiescores van beide cloze-toetsen. Dat betekent dat kinderen een hogere score behalen op de SG2 cloze-toets dan op de VG2 cloze-toets. Het effect van groep was niet sig-

nificant, wat betekent dat kinderen uit groep 5 en 6 vergelijkbaar scoren op de cloze-toetsen. De effecten van technisch lezen en begrijpend lezen waren beide significant, wat betekent dat beide variabelen een onafhankelijke rol spelen in de proportiescore van de cloze-toetsen. De interactietermen van type toets met technisch lezen, woordenschat en begrijpend lezen waren niet significant. De niet-significante verbanden wezen echter wel in de verwachte richting: de proportiescore van de VG2 cloze-toets lijkt een sterker beroep te doen op woordenschat en begrijpend lezen, en een minder sterk beroep op technisch lezen dan de proportiescore van de SG2 cloze-toets.

7 Conclusie Studie 2

In Studie 2 werden, net als in Studie 1, de betrouwbaarheid en validiteit van twee verschillende cloze-toetsen onderzocht, een cloze-toets met systematisch geplaatste gaten (SG2) en een cloze-toets met variabel geplaatste gaten (VG2). In deze studie kregen de kinderen een langere cloze-toets en meer tijd om die te maken dan in Studie 1. De betrouwbaarheid van de VG2 cloze-toets kon niet berekend worden, aangezien deze toets niet bestond uit twee deelversies zoals bij Studie 1. De betrouwbaarheid van de somscore van de SG2 cloze-toets was voldoende, die van de proportiescore was matig. Met betrekking tot de validiteit werd allereerst onderzocht hoe de cloze-toetsen samenhangen met begrijpend lezen. De samenhang tussen de cloze-toetsen en begrijpend lezen was middelmatig tot hoog. Ten tweede werd onderzocht of het uitmaakt of somscores of proportiescores werden gebruikt. Er werden geen verschillen gevonden in de relaties van de somscores en de proportiescores met begrijpend lezen, alhoewel de proportiescores van de VG2 cloze-toets het sterkste lijken samen te hangen met begrijpend lezen. Als derde werd onderzocht in welke mate de cloze-toetsen een beroep deden op technisch lezen, woordenschat en begrijpend lezen. Uit de regressieanalyses kon opgemaakt worden dat de SG2 en VG2 cloze-toetsen een verge-

Tabel 6

Hiërarchische Regressieanalyse voor de Verschillen tussen de Somscore en de Proportiescore van de SG2 en VG2 Cloze-toetsen

Somscore van cloze-toetsen		
	ΔR^2	B
1. Type toets	.03*	.16**
2. Groep	.09**	.17**
3. Technisch lezen	.29**	.41**
4. Woordenschat	.01*	.02
5. Begrijpend lezen	.00	.07
6. Type toets x Technisch lezen	.01	.10
7. Type toets x Woordenschat	.01	.12
8. Type toets x Begrijpend lezen	.00	.02
Proportiescore van cloze-toetsen		
	ΔR^2	B
1. Type toets	.30**	-.54**
2. Groep	.04**	.06
3. Technisch lezen	.07**	.17*
4. Woordenschat	.07**	.05
5. Begrijpend lezen	.05**	.25*
6. Type toets x Technisch lezen	.00	-.10
7. Type toets x Woordenschat	.01*	.14
8. Type toets x Begrijpend lezen	.00	.07

Noot. De gestandaardiseerde regressiecoëfficiënten zijn gepresenteerd voor het model met alle variabelen.

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

lijikbaar beroep doen op technisch lezen. Wel bleek dat woordenschat belangrijker was voor de somscore van de VG2 cloze-toets dan voor de somscore van de SG2 cloze-toets. Dit verschil werd ook gezien bij de proportiescores, maar was hier niet significant. Ten slotte bleek begrijpend lezen wel een rol te spelen bij de proportiescore, maar niet bij de somscore van de SG2 en VG2 cloze-toetsen. Op basis van deze resultaten zou geconcludeerd kunnen worden dat bij het gebruik van de pure accuratessebenadering (proportiescores) de VG2 cloze-toets meer lijkt op een standaard toets begrijpend lezen dan wanneer de snelheidsbenadering (somscores) wordt gebruikt. Ten vierde was onderzocht of de tijd die kinderen kregen om de VG cloze-toets te maken uitmaakte voor hoe goed deze toets begrijpend lezen meet. De VG cloze-toets uit Studie 2, met een langere afnametijd, leidde

niet tot een sterker beroep op woordenschat en begrijpend lezen dan de VG cloze-toets uit Studie 1. Op basis van deze studie kan dus niet geconcludeerd worden dat een langere afname de VG cloze-toets meer een toets voor begrijpend lezen maakt.

8 Discussie

In dit onderzoek werden de betrouwbaarheid en de validiteit van verschillende soorten meerkeuze cloze-toetsen voor begrijpend lezen onderzocht. Hiertoe zijn toetsen met systematisch geplaatste gaten (SG) vergeleken met toetsen met variabel geplaatste gaten (VG). De betrouwbaarheid van de snelheidsbenadering van de cloze-toetsen blijkt voldoende tot goed, maar de accuratessebenadering van de cloze-toetsen, die de voorkeur verdient als maat voor begrijpend lezen, blijkt

Tabel 7
Samenvatting van de Resultaten van Studie 1 en 2

	Parallelttest betrouwbaarheid	Correlatie met begrijpend lezen	Correlatie met technisch lezen	Correlatie met woordenschat
SG1 cloze-toets som	.79**	.33**	.63**	.30**
SG1 cloze-toets proportie	.46**	.62**	.34**	.36**
VG1 cloze-toets som	.61**	.47**	.47**	.50**
VG1 cloze-toets proportie	.43**	.65**	.25**	.58**
SG2 cloze-toets som	.73**	.30**	.53**	.19
SG2 cloze-toets proportie	.56**	.43**	.35**	.30**
VG2 cloze-toets som	-	.38**	.55**	.31**
VG2 cloze-toets proportie	-	.53**	.27**	.50**

** $p < .01$.

onvoldoende. Zie ook Tabel 7 met samenvattende gegevens betreffende beide studies.

Wat betreft de validiteit betrof de eerste onderzoeksvraag de samenhang tussen de SG cloze-toets en de VG cloze-toets met standaardtoetsen voor begrijpend lezen. De relaties tussen de somscores en de proportiescores van de cloze-toetsen en begrijpend lezen zijn middelmatig tot hoog. Dit betekent dat, afhankelijk van het type cloze toets, de convergente validiteit matig tot goed is.

Als tweede werd onderzocht of het uitmaakt of een snelheidsbenadering met accuratesse (gebruik van somscore) of een pure accuratessebenadering (gebruik van proportiescore) werd gebruikt bij het scoren van de toetsen. Uit de resultaten van Studie 1 kan opgemaakt worden dat proportiescores van cloze-toetsen sterker samenhangen met begrijpend lezen dan somscores. Dit verschil in samenhang werd niet gevonden in Studie 2. In beide studies werd echter wel gevonden dat begrijpend lezen een unieke bijdrage levert aan de proportiescores van beide type cloze-toetsen en niet aan de somscores. De accuratessebenadering is dus een meer valide manier om begrijpend lezen te meten. Hierbij moet wel de kanttekening gemaakt worden dat de proportiescores van de SG cloze-toetsen niet normaal verdeeld zijn, doordat veel leerlingen bijna alle vragen goed hebben op deze toets. Dit betekent dat een groot deel van de leerlingen een bijna maximale score behaalt, waardoor deze combinatie van sco-

ren en toetsen niet geschikt is om begrijpend lezen te meten.

Ten derde werd onderzocht of de verschillende cloze-toetsen een verschillend beroep doen op technisch lezen, woordenschat en begrijpend lezen. De regressieanalyses van Studie 1 laten zien dat als wordt gekeken naar de somscores, de SG1 cloze-toets een sterker beroep doet op technisch lezen dan de VG1 cloze-toets. Uit de analyses met proportiescores, waar in feite wordt gecorrigeerd voor leessnelheid, blijkt dat woordenschat een grotere rol speelt bij de VG1 cloze-toets dan bij de SG1 cloze-toetsen. Alhoewel vergelijkbare verbanden werden gevonden in Studie 2, waren ze in Studie 2 niet significant. Op basis van Studie 1 kan dus geconcludeerd worden dat de VG cloze-toetsen een groter beroep doen op woordenschat en een minder groot beroep op technisch lezen dan de SG cloze-toetsen. Dit ondersteunt de hypothese dat de VG cloze-toetsen meer valide zijn om begrijpend lezen te meten dan de SG cloze-toetsen.

Als vierde is onderzocht of het verlengen van de afnametijd uitmaakt voor de mate waarin de cloze-toets begrijpend lezen meet. Als de relaties van de VG1 cloze-toets en begrijpend lezen vergeleken worden met de relaties van de VG2 cloze-toets en begrijpend lezen kan geconcludeerd worden dat op basis van deze twee studies geen verschillen worden gevonden tussen cloze-toetsen met een verschillende afnametijd. Het feit dat geen verschillen worden gevonden in de afname-

tijd zou echter verklaard kunnen worden door het verschil in leeftijd van de kinderen. Het is mogelijk dat als de VG1 en VG2 cloze-toets bij kinderen van dezelfde leeftijd worden afgenomen, er dan wel verschillen worden gevonden. Een andere verklaring kan zijn dat de VG1 cloze-toets in twee keer is afgenomen, terwijl de VG2 cloze-toets slechts één afnamemoment bevatte. Dit maakt dat de VG1 en de VG2 cloze-toets dus niet alleen verschillen in afnameduur, waardoor het effect hiervan niet goed onderzocht kan worden.

Op basis van de uitkomsten van deze studies kan geconcludeerd worden dat bij gebruik van een cloze-toets om begrijpend lezen te meten, het de aanbeveling verdient om een toets te nemen met variabel geplaatste gaten; bij de VG cloze-toetsen blijkt de relatie met woordenschat namelijk hoger dan bij de SG cloze-toetsen. Daarnaast verdient de accuratessebenadering (proportiescore) de voorkeur boven de snelheidsbenadering (somscore). Bij gebruik van somscores doet de cloze-toets een sterker beroep op vloeiend lezen (snel en accuraat) en een minder groot beroep op begrijpend lezen. Kinderen die de toets slecht maken lijken dus vooral niet zo vlot te kunnen lezen (zie bijvoorbeeld Keenan et al., 2014). Vooral leerlingen met dyslexie, die doorgaans een lager leestempo hebben (Steenbeek-Planting, 2012), worden met de snelheidsbenadering als het ware gestraft voor een lager leestempo, terwijl zo niet de beoogde onderliggende vaardigheid begrijpend lezen wordt gemeten. Met betrekking tot de testduur van de VG cloze-toetsen blijkt dat er in de relaties met een andere toets begrijpend lezen, technisch lezen en woordenschat geen duidelijke verschillen worden gevonden tussen de VG1 en de VG2 cloze-toets. Ondanks dat wordt aanbevolen om de cloze-toets langer te maken dan twee minuten, aangezien bij een korte toets sommige kinderen maar heel weinig vragen kunnen beantwoorden, wat ervoor zorgt dat de toets minder betrouwbaar is.

Omdat de betrouwbaarheid van de gebruikte VG cloze-toetsen in dit onderzoek onvoldoende is, kunnen de cloze-toetsen niet in de huidige vorm gebruikt worden voor

begrijpend lezen. Deze lage betrouwbaarheden werden niet gevonden in eerdere Nederlandse studies, maar in die studies werd een cloze-toets met open vragen in plaats van multiple choice vragen gebruikt (zie bijvoorbeeld Kamalski, 2007). Ook zijn de gevonden relaties tussen de cloze-toets en standaard toetsen begrijpend lezen ($r = .65$ voor VG1; $r = .53$ voor VG2) niet zo hoog als in het onderzoek van Gellert en Elbro (2013) ($r = .84$). Nader onderzoek naar het gebruik van de cloze-toets met variabel geplaatste gaten als maat voor begrijpend lezen wordt daarom aanbevolen. Het is interessant om deze manipulaties op de teksten ook uit te voeren bij oudere kinderen. Tevens is in het huidige onderzoek slechts gebruik gemaakt van één toets voor begrijpend lezen om de cloze-toetsen aan te relateren. Dit is te weinig om te stellen dat de cloze-toets een goede maat is voor begrijpend lezen (zie bijvoorbeeld Kamalski, 2007). Daarnaast is het relevant om de betrouwbaarheid van de taak nader te onderzoeken door gebruik te maken van meerdere afnamemomenten waardoor de testhertest betrouwbaarheid berekend kan worden.

Verschillen tussen cloze-toetsen zijn nog nooit eerder systematisch getoetst. De huidige studies leveren een belangrijke bijdrage aan de discussie of verschillende typen cloze-toetsen gebruikt kunnen worden als maat voor begrijpend lezen. Op basis van deze studie kan geconcludeerd worden dat de cloze-toets zoals deze op dit moment vaak gebruikt wordt, met systematisch geplaatste gaten en het gebruik van de snelheidsscore, geen beroep doet op begrijpend lezen en meer een maat is voor technisch lezen. De accuratessebenadering van de cloze-toets met variabel geplaatste gaten blijkt nog de beste maat te zijn voor begrijpend lezen, maar ook bij deze benadering is de toets nog onvoldoende betrouwbaar, waardoor geconcludeerd moet worden dat de cloze-toetsen in dit onderzoek niet geschikt zijn om begrijpend lezen te meten. Nader onderzoek naar de validiteit en betrouwbaarheid van deze toets is dan ook noodzakelijk. Vooralnog verdienen toetsen begrijpend lezen met meerdere teksten en verschillende soorten vragen de voorkeur

boven cloze-toetsen.

Noot

Wij willen graag M. M. A. de Haan, M. B. C. Tool en E. M. Tessensohn bedanken voor de input die zij met hun bachelor-onderzoek hebben geleverd aan dit onderzoek. We willen José van der Hoeven bedanken voor het verzamelen van de Cito toets-scores van Studie 2. Daarnaast willen we CED-groep bedanken voor het gebruik van de Nieuwsbegrip-teksten voor de ontwikkeling van de cloze-toetsen. Tot slot willen we de reviewers bedanken voor hun feedback/review.

Literatuur

- Aarnoutse, C., & Kapinga, T. (2006). *Begrijpend lezen*. Ridderkerk, Nederland: Onderwijs Adviesing.
- Brus, B. Th., & Voeten, M. J. M. (1979). *Een-Minuu-Test, vorm A en B: Verantwoording en handleiding*. Lisse, Nederland: Swets & Zeitlinger.
- Cain, K., & Oakhill, J. V. (1999). Inference making and its relation to comprehension failure. *Reading and Writing. An Interdisciplinary Journal*, 11, 489-503. doi:10.1023/A:1008084120205
- CED-groep (2013). Handleiding sneltoets informatieve teksten. Rotterdam, Nederland: CED-groep.
- Cito (2008). *LOVS begrijpend lezen groep 6*. Arnhem, Nederland: Cito.
- Currie, N. K., & Cain, K. (2015). Children's inference generation: the role of vocabulary and working memory. *Journal of Experimental Child Psychology*, 137, 57-75. doi:10.1016/j.jecp.2015.03.005
- Echambadi, R., & Hess, J. D. (2007). Mean-centering does not alleviate collinearity problems in moderated multiple regression models. *Marketing Science*, 26, 438-445. doi:10.1287/mksc.1060.0263
- Espin C. A., McMaster K., Rose S. & Wayman M. (Red.) (2012). *A measure of success: The influence of curriculum-based measurement on education*. Minneapolis, MN: University of Minnesota Press.
- Espin, C. A., Wallace, T., Lembke, E., Campbell, H., & Long, J. D. (2010). Creating a progress measurement system in reading for middle-school students: Monitoring progress towards meeting high stakes standards. *Learning Disabilities Research and Practice*, 25, 60-75. doi:10.1111/j.1540-5826.2010.00304.x
- Evers, A., Lucassen, W., Meijer, R. R., & Sijtsma, K. (2009). *COTAN Beoordelingssysteem voor de kwaliteit van tests, geheel herziene versie*. Amsterdam: NIP/COTAN.
- Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, 21(1), 45-59.
- Gellert, A. S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, 31, 16-28. doi:10.1177/0734282912451971
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127-160. doi:10.1007/BF00401799
- Jansen, C., & Boersma, N. (2013). Meten is weten? Over de waarde van de leesbaarheidsvoorspellingen van drie geautomatiseerde Nederlandse meetinstrumenten. *Tijdschrift voor Taalbeheersing*, 35(1), 47-62.
- Kamalski, J. (2007). *Coherence marking, comprehension and persuasion. On the processing and representation of discourse*. Dissertatie. Utrecht: Universiteit Utrecht.
- Kamalski, J., Sanders, T., Lentz, L., & Van den Bergh, H. (2005). Hoe kun je het beste meten of eenleerling een tekst begrijpt? Een vergelijkend onderzoek naar vier methoden. *Levende Talen Tijdschrift*, 6(4), 3-9.
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12, 281-300. doi:10.1080/10888430802132279
- Keenan, J. M., Hua, A. N., Meenan, C. E., Pennington, B. F., Willcutt, E., & Olson, R. K. (2014). Issues in identifying poor comprehenders. *L'Année Psychologique*, 114, 753-777. doi:10.4074/2FS0003503314004072
- Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing demands of reading com-

- prehension tests in young readers. *Learning and Instruction*, 22, 354-367. doi:10.1016/j.learninstruc.2012.02.001
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(1), 363-394.
- Kordes, J., Bolsinova, M., Limpens, G., & Stolwijk, R. (2013). *Resultaten PISA-2012. Praktische kennis en vaardigheden van 15-jarigen. Nederlandse uitkomsten van het Programme for International Student Assessment (PISA) op het gebied van wiskunde, natuurwetenschappen en leesvaardigheid in het jaar 2012*. Arnhem, Nederland: CITO.
- Kraf, R., & Pander Maat, H. (2009). Leesbaarheids-onderzoek: oude problemen, nieuwe kansen. *Tijdschrift voor Taalbeheersing*, 31(2), 97-123.
- Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Beschikbaar via <http://quantpsy.org>
- Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology*, 27, 1424-1429. doi:10.1037/0278-7393.27.6.1424
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25, 605-621. doi:10.1080/20445911.2013.795574
- Muijselaar, M. M. L., & De Jong, P. F. (2015). The effects of updating ability and knowledge reading strategies on reading comprehension. *Learning and Individual Differences*, 43, 111-117. doi:10.1016/j.lindif.2015.08.011
- Muijselaar, M. M. L., Kendeou, P., De Jong, P. F., & Van den Broek, P. W. (2017). What does the CBM-Maze test measure? *Scientific Studies of Reading*. doi:10.1080/10888438.2016.1263994
- Muijselaar, M. M. L., Swart, N. M., Steenbeek-Planting, E. G., Droop, M., Verhoeven, L., & De Jong, P. F. (2017). The dimensions of reading comprehension: Is differentiation by text and question type necessary? *Journal of Educational Psychology*, 109, 70-83. doi:10.1037/edu0000120
- Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67, 359-370. doi:10.1111/j.2044-8279.1997.tb01250.x
- Schlichting, L. (2005). *Peabody picture vocabulary test-III-NL*. Amsterdam, Nederland: Harcourt Test Publisher.
- Spörer, N., & Brunstein, J. C. (2009). Fostering the reading comprehension of secondary school students through peer-assisted learning: Effects on strategy knowledge, strategy use, and task performance. *Contemporary Educational Psychology*, 34, 289-297. doi:10.1016/j.cedpsych.2009.06.004
- Steenbeek-Planting, E. G. (2012). *Should reading errors be taken seriously in assessment and intervention?* Dissertatie. Radboud Universiteit Nijmegen.
- Swart, N. M., Muijselaar, M. M. L., Steenbeek-Planting, E. G., Droop, M., De Jong, P. F., & Verhoeven, L. (2016). Differential lexical predictors of reading comprehension in fourth graders. *Reading and Writing: An Interdisciplinary Journal*, 30, 489-507. doi:10.1007/s11145-016-9686-0
- Van den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J. M., & de Vries, J. R. (1994). *De Klepel: Een test voor de leesvaardigheid van pseudowoorden*. Lisse, Nederland: Swets and Zeitlinger.
- Van Dooren, W., Van den Bergh, H., & Evers-Vermeul, J. (2012). Leesbare teksten? Over de invloed van structuurmarkeringen op het tekstbegrip en de tekstwaardering van zwakke en sterke lezers. *Levende Talen Tijdschrift*, 13(4), 31-38.
- Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology*, 22, 293-301. doi:10.1002/acp.1417

Auteurs

Marloes Muijselaar is Postdoc bij het Behavioral Science Institute aan de Radboud Universiteit. **Elise de Bree** is Universitair Docent bij de afdeling Pedagogische en Onderwijswetenschappen aan de Universiteit van Amsterdam. **Esther Steenbeek-Planting** is Postdoc bij het Behavioral Science Institute aan de Radboud Universiteit. **Peter de Jong** is Hoogleraar Ortho-

pedagogiek bij de afdeling Pedagogische en Onderwijswetenschappen aan de Universiteit van Amsterdam.

Correspondentieadres: M. M. L. Muijselaar, postdoctoral researcher, Radboud Universiteit, Postbus 9104, 6500 HE Nijmegen, M.M.L.Muijselaar@gmail.com

Abstract

Is the cloze test a reliable and valid measure for reading comprehension?

The cloze test (short gap-filling text) is currently used to measure reading comprehension. The quality of this measure for reading comprehension, however, has been subject of debate. Two different studies were conducted to examine the reliability and validity of two types of multiple-choice cloze-tests; one with fixed-ratio gaps, every 11th word is replaced by three answer options, and one with variably-spaced gaps, gaps that are selected so that they require inferencing. Participants were 312 fifth graders in the first study and 100 third and fourth graders in the second study. The results showed that the variably-spaced cloze test had a stronger correlation with vocabulary than the fixed-ratio cloze test, whereas the fixed-ratio cloze test had a higher correlation with reading fluency than the variably-spaced cloze test. Furthermore, the accuracy approach correlated stronger with reading comprehension than the speed approach. The accuracy approach of the variably-spaced cloze test is thus the most valid measure for reading comprehension. The reliability of this test, however, was poor. Thus none of the cloze-tests used in the current study can be used as a measure for reading comprehension.

keywords: comprehension, cloze test, validity, reliability, reading fluency, vocabulary