

# **Curriculum Based Measurement in leesvaardigheid: betrouwbaarheid en validiteit van de SIT maze- taken.**



C.E.M. Giesen

Masterscriptie Education and Child Studies

Faculteit Sociale Wetenschappen

Universiteit Leiden

1° beoordelaar: Prof. Dr. C.A. Espin

2° beoordelaar: Dr. M.J. van Dijken

Juni/ juli 2014

## **Abstract**

Het doel van deze studie is om te onderzoeken of de scores van de SIT maze-taken betrouwbaar, valide en sensitief voor groei voor leesvaardigheid zijn. De steekproef bestaat uit leerlingen (N = 856) van klas 1 tot en met 3 van het middelbaar onderwijs in Nederland. De leerlingen hebben op vier meetmomenten SIT maze-taken gemaakt en ook een begrijpend lezentoets als pre-test en post-test. Resultaten wijzen uit dat de maze-taak betrouwbaar is voor het meten van leesvaardigheid. Er is steun gevonden voor validiteit en sensitiviteit van de SIT maze-taak voor groei.

## **Inleiding**

Om het onderwijs in Nederland te verbeteren heeft de Nederlandse overheid in 2007 als speerpunt op de agenda het opbrengstgericht werken gezet (Rijksoverheid, 2010). Het doel van het opbrengstgericht werken is om zo veel mogelijk scholen te laten voldoen aan de standaarden voor kwaliteitszorg. Ook is het streven van de overheid dat in over één á twee jaar ruim de helft van de middelbare scholen opbrengstgericht gaan werken (Rijksoverheid, 2011a). Dit houdt in: ‘als een school systematisch en doelgericht werkt aan het maximaliseren van de prestaties van haar leerlingen, er sprake is van opbrengstgericht werken’ (Rijksoverheid, 2010, p.4). Het doel van de overheid is om basisvaardigheden van de Nederlandse taal en rekenen van leerlingen te verbeteren door middel van het stellen van doelen, het planmatig en resultaatgericht werken en zicht hebben op de leerresultaten van leerlingen (Rijksoverheid, 2010). Om de opbrengsten van een school te kunnen evalueren moeten scholen doelen opstellen, zodat sterke en zwakke plekken in leerresultaten kunnen worden opgespoord. Uit het onderzoek van de Rijksoverheid (2010) is gebleken dat veel scholen de ambitie hebben om doelen te stellen en streven naar goede resultaten, maar deze doelen nog niet hebben vastgelegd in hun beleidsplannen. Daarnaast wordt over het algemeen binnen de klas aandacht gericht op de evaluatie van het resultaat van een leerstrategie, in plaats van op de evaluatie van de vooruitgang in het leerproces.

Om zicht te hebben op leerresultaten heeft een school een goede toetsinstrumentarium nodig. Uit het onderzoek van de Rijksoverheid (2010) is gebleken dat scholen over het algemeen beschikken over goede instrumenten om leerresultaten te volgen, maar ook is gebleken dat een groot deel van de scholen geen vaste toetsingscriteria hanteren om leerlingen met leerproblemen op te sporen. Ook worden er nauwelijks probleemanalyses gemaakt door docenten naar aanleiding van de toetsen en daardoor krijgen veel leerlingen met leer- of gedragsproblemen vaak onvoldoende aanvullende hulp.

Om opbrengst gericht te kunnen werken moet een school op verschillende niveaus te investeren (Rijksoverheid, 2011b), waaronder groepsniveau en leerlingniveau. Deze niveaus zijn onderdeel van een dynamisch systeem dat kan zorgen voor een optimalisering van de schoolkwaliteit. Docenten spelen een belangrijke rol binnen deze niveaus. In de ideale situatie stellen docenten bij het groepsniveau doelen op en geven feedback over het proces waarop leerlingen tot hun resultaat zijn

gekomen. Voor zwakke leerlingen passen zij waar nodig hun instructie en de leerstof aan. Op leerlingniveau is het van belang dat een docent een leerling kan identificeren die een leerachterstand heeft en vervolgens hulp te bieden die daarbij past.

### **Leerlingvolgsystemen**

De Rijksoverheid heeft als doel gesteld dat middelbare scholen op korte termijn verplicht gaan worden om te werken met een leerlingvolgsysteem (Rijksoverheid, 2011a; Onderwijsraad, 2012). In Nederland kunnen scholen op dit moment gebruik maken van het Cito Volgstelsel in het voortgezet onderwijs om zicht te krijgen op het niveau en de ontwikkeling van leerlingen in het voortgezet onderwijs (CITO, 2013). Dit is een volgsysteem voor de niveaus vmbo, havo en vwo. Het doel van de Cito is om de voortgang van leerlingen te meten, te kunnen adviseren en kwaliteitszorg van het onderwijs te waarborgen dat wordt aangeboden. Het volgsysteem van de Cito bestaat uit een aantal toetsen waarmee de kernvaardigheden Nederlands leesvaardigheid en taalschat, taalverzorging, Engels leesvaardigheid en taalschat en Wiskunde van leerlingen van de 1<sup>e</sup> tot en met de 3<sup>e</sup> klas worden gemeten. Docenten kunnen de resultaten van deze toetsen online invoeren om zo de individuele voortgang van de leerling bij te houden en te vergelijken met landelijke normen. Met een percentielscore wordt nagegaan hoeveel procent de leerling van hetzelfde niveau hetzelfde of lager scoort dan andere leerlingen in de vergelijkingsgroep. Een nadeel van het Cito Volgstelsel is dat er sprake is van een klein aantal meetmomenten (bijvoorbeeld 1 tot 3) in een schooljaar om te kunnen zien op welk niveau een bepaalde leerling functioneert. Hierdoor krijgt de docent nauwelijks de gelegenheid om zwakke leerlingen te volgen en daar waar nodig zijn instructie aan te passen. Daarom zou een leerlingvolgsysteem met meerdere meetmomenten efficiënter zijn voor het bijhouden van de voortgang van een leerling.

Een alternatief systeem om leerlingen te kunnen volgen met betrekking tot hun schoolprestaties, is aan de hand van een volgsysteem voor jonge leerlingen dat is ontwikkeld in de Verenigde Staten (Deno, 1985). Dit volgsysteem wordt Curriculum Based Measurement (CBM) genoemd en heeft als doel de groei van een individuele leerling te meten. Aan de hand van CBM kunnen docenten op een overzichtelijke manier de voortgang en resultaten van de leerlingen bijhouden en waar nodig hun instructie aanpassen. Vakken zoals lezen, schrijven, spelling en wiskunde kunnen worden bijgehouden. CBM heeft verschillende eigenschappen die van belang zijn (Stecker, Fuchs & Fuchs, 2005). Ten eerste kan CBM worden gebruikt als formatieve assessment om lange termijn doelen te stellen. Ten tweede worden de scores van de leerlingen regelmatig geordend in een grafiek om zo te kunnen bijhouden of zij hun lange termijn doel kunnen halen. Daarnaast is CBM een volgsysteem dat betrouwbaar en valide is bevonden. CBM kan ook gebruikt worden om leerlingen met leerproblemen op te sporen (Deno, 1992).

Voordelen van het CBM zijn de tijdsduur en het gemak waarmee het gebruikt kan worden (Deno, 2003). De CBM taken zijn doorgaans niet langer dan 3 minuten en de procedures zijn

makkelijk te leren aan docenten en andere professionals. Zo kan bijvoorbeeld een maze-taak worden afgenomen voor een duur van twee minuten, en een week later weer één van twee minuten. Om de resultaten te kunnen generaliseren worden parallelle versies aangeboden, zodat er geen sprake kan zijn van een oefeneffect. Door het gemak waarmee CBM taken kunnen worden afgenomen is het mogelijk leerlingen regelmatig te testen. Hierdoor is er veel data beschikbaar voor docenten. De resultaten van de CBM taken worden in een grafiek gezet zodat de vooruitgang van de leerling nauwkeurig kan worden gevolgd. Aan de hand van de grafiek kunnen korte en lange termijn doelen worden opgesteld en kan worden gekeken of de leerling vooruitgang boekt (Deno, 2003; Fuchs & Fuchs, 1992). Wanneer er geen sprake is van vooruitgang kan de leraar zijn eigen instructieniveau evalueren en hier eventueel aanpassingen in maken. Deno (2003) geeft aan dat bij het CBM in acht genomen moet worden dat elke school het materiaal moet aanpassen aan het curriculum. Daarom bevat de handleiding van het CBM specificaties betreffende de duur van de taken, de instructies voor leerlingen, regels voor de scoring en de manier waarop de resultaten bijgehouden worden. Daarnaast is het CBM een betrouwbaar en valide meetinstrument voor het meten van voortgang van leerlingen (Espin, Wallace, Lembke, Campbell & Long, 2010; Tichá, Espin & Wayman, 2009; Fuchs & Fuchs, 1999).

### **Curriculum Based Measurement voor leesvaardigheid**

Om het leesniveau van leerlingen te kunnen meten maakt CBM gebruik van twee soorten taken, namelijk hardop lezen en maze-taken (Wayman, Wallace, Wiley, Tichá & Espin, 2007). Hardop lezen is een taak waarbij leerlingen gedurende 1 minuut een stuk tekst hardop voorlezen (Deno, 1985). Bij deze taak worden de woorden die correct worden gelezen gescoord en wanneer een leerling iets toevoegt, vervangt, weglaat, aarzelt of een woord verkeerd uitspreekt, wordt dit als een fout gerekend. De scores die de leerlingen hebben gehaald worden in een grafiek gezet en geanalyseerd. De docent bestudeert de grafiek om te kijken of het niveau van de leesvaardigheid van leerling toeneemt, gelijk blijft of afneemt. De docent past dan waar nodig de instructies aan. De hardop lezen taak wordt het meest gebruikt bij het CBM als indicator voor het algemene leesniveau op basisschoolniveau (Wayman, et al., 2007). Of de taak een goede indicator is voor het leesniveau op middelbare scholieren is niet bekend, omdat weinig onderzoek hiernaar is gedaan.

De hardop lezen taak heeft nadelen, namelijk dat het tijdsintensief is en gebrek heeft aan face validiteit als maatstaf voor leesbegrip (Faykus & McCurdy, 1998, zoals beschreven in Madelaine & Wheldall, 2004). Als alternatief voor het hardop lezen kunnen maze-taken worden gebruikt. Maze-taken worden sinds de jaren '70 al onderzocht als instrument dat leesvaardigheid meet (Wayman, et al., 2007). Maze-taken zijn taken waarbij leerlingen teksten moeten lezen binnen een bepaalde tijd en bij elk zevende woord een selectie moeten maken uit een drietal woorden en uiteindelijk het juiste woord moeten kiezen (Yeo, 2010). Twee van de drie woorden zijn afleiders en zijn duidelijk fout (zie Figuur 1).

Je kunt bewegen, en je kunt sporten. Het klinkt heel raar, maar deze **(ziel / kuur / twee)** dingen verschillen van elkaar. Onder bewegen **(verstaan / bezoeker / hovenier)** we rustig fietsen of wandelen. Dit **(lef / rib / kun)** je het beste dertig minuten per **(dag / hit / gum)** doen. Je lichaam blijft gezond en **(je / af / ei)** spieren blijven soepel. Omdat je spieren **(bestek / soepel / opgave)** blijven, krijg je minder last van **(basketbal / koffiepot / blessures)**. Maar bewegen is niet alleen goed **(snap / riep / voor)** je spieren.

*Figuur 1.* Voorbeeldtekst Mazes-online ontwikkeld door Espin & Chung, Universiteit Leiden.

Het lezen van de maze-taak gebeurt in stilte en duurt ongeveer 1 tot 3 minuten (Fuchs & Fuchs, 1992). Het aantal correcte antwoorden wordt daarbij gescoord. Ook bij deze taak worden de scores die de leerlingen hebben gehaald in een grafiek gezet en geanalyseerd. Ook hier bestudeert de docent de grafiek om te kijken of het niveau van de leesvaardigheid van leerling toeneemt, gelijk blijft of afneemt. De docent past de instructie aan wanneer dit nodig is. Onderzoek naar maze-taken laat zien dat deze goede psychometrische eigenschappen heeft (Wayman et al., 2007). Er is een goede betrouwbaarheid en validiteit van de maze-taken gevonden als voorspellers voor de prestaties en voortgang in leesvaardigheden (Tichá et al., 2009). Een voordeel is dat de taken geschikt zijn voor oudere leerlingen en kunnen worden afgenomen in een groep of via de computer (Wayman et al., 2007). Bovendien laten maze-taken sneller groei zien in leesvaardigheid in verhouding tot andere meetinstrumenten die leesvaardigheid meten (Fuchs & Fuchs, 1992). Ook geven zowel docenten als leerlingen aan maze-taken prettig te vinden, omdat het een leuke methode is en omdat meerdere aspecten van leesvaardigheid worden gemeten zoals decoderen, begrip en snelheid.

## **Huidige studie**

Omdat in de Verenigde Staten scholen aan hoge eisen moeten voldoen (Shapiro, Keller, Lutz, Santoro & Hintze, 2006), is veel onderzoek gedaan naar CBM (Hintze & Silbergliitt, 2005). Onderzoekers hebben onderzocht of CBM betrouwbaar, valide en objectief is (Fuchs & Fuchs, 1999).

Er zijn echter maar drie studies uitgevoerd die de psychometrische eigenschappen van de maze-taak hebben onderzocht op middelbare scholen. In de eerste studie door Tichá et al. (2009) werd de betrouwbaarheid, validiteit en de sensitiviteit voor groei van leesvaardigheid van de scores op de maze-taak en hardop lezen onderzocht. Vijfendertig leerlingen kregen tien weken lang een maze-selectie naast een hardop lezentoets. De scores van de maze-taak en de hardop lezentoets bleken een goede betrouwbaarheid en validiteit te hebben. De score op de maze-taak was significant voor groei, met een gemiddelde groei van 1.29 correcte antwoorden per week. De hardop lezen taak liet nagenoeg geen groei zien. Echter is deze studie uitgevoerd met een kleine steekproef waarbij afgevraagd moet

worden of de resultaten generaliseerbaar zijn naar de populatie. Ook wordt aangeraden andere materialen te gebruiken ter controle van de betrouwbaarheid en validiteit van de scores van de maze-taken en hardop leestaak.

De tweede studie gedaan door Espin et al. (2010) onderzocht 236 leerlingen uit het middelbaar onderwijs in de Verenigde Staten. Het doel van het onderzoek was om te kijken of de scores op de maze-taak en de hardop lezen taak een betrouwbare en valide voorspeller waren voor de scores op de State screeningstest in Minnesota. Uit dit onderzoek bleek dat de scores van de maze-taak en de hardop leestaak een betrouwbaar en valide voorspeller waren voor de scores op de nationale screeningstest. Alleen de maze-taak liet groei zien gedurende de meetmomenten. Hoewel deze steekproef groter was dan de eerste, was de steekproef niet heel groot en de leerlingen waren allen van dezelfde middelbare school. Hierdoor is het moeilijk om de uitkomsten van dit onderzoek te generaliseren naar de populatie. Daarnaast hebben de leerlingen die zijn onderzocht de leeftijd van ongeveer elf tot dertien jaar. Daarom wordt door de onderzoekers aangeraden dit onderzoek ook uit te voeren met oudere leerlingen.

In het derde onderzoek gedaan door Tolar et al. (2012) worden niet alleen de betrouwbaarheid en de validiteit van de scores op de maze-taken onderzocht, maar ook de oefeneffecten. Er zijn leerlingen van het middelbare onderwijs in de Verenigde Staten onderzocht die voor de studie in drie groepen waren verdeeld: gemiddelde lezers die een interventie kregen, zwakke lezers die een interventie kregen en zwakke lezers die geen interventie kregen. Zij kregen allen een aantal dezelfde maze-taken aangeboden en een aantal paralleltesten. In deze studie is gevonden dat zowel voor dezelfde als de nieuwe maze-taken een hoge betrouwbaarheid en validiteit is gevonden. Daarnaast was gevonden dat het oefeneffect sterker is bij gemiddelde lezers dan bij de zwakke lezers. In dit onderzoek hebben niet alle leerlingen een nieuwe en een parallelle maze-taak gemaakt. Daarom kunnen er bij deze studie moeilijk uitspraken worden gedaan over het oefeneffect. Daarnaast is de steekproef van deze studie gestratificeerd, waardoor de resultaten met bepaalde voorzichtigheid moeten worden geïnterpreteerd.

Uit bovengenoemde studies blijkt dat er steun is voor betrouwbaarheid, validiteit en sensitiviteit voor groei in leesvaardigheid bij de scores van maze-taak op middelbare scholen. Echter zijn dit maar drie studies die allen in de Verenigde Staten zijn uitgevoerd. In het huidige onderzoek zal de focus liggen op het onderzoeken van de betrouwbaarheid en validiteit van de CBM maze-taken voor middelbare scholieren in Nederland.

In deze studie zijn maze-taken worden gebruikt die beschikbaar zijn gesteld door het Centrum voor Educatieve Dienstverlening (CED-groep). De CED-groep is een onderwijsinstantie die zich inzet om professionals in het onderwijs te adviseren en te trainen (CED-groep, 2014). Ook ontwikkelt de CED-groep materialen die gebruikt kunnen worden in het onderwijs. Door middel van het ontwikkelen van nieuwe materialen wil de CED-groep bijdragen om de kwaliteit in het onderwijs te verbeteren. Op dit moment is de CED-groep bezig met het ontwikkelen van een leerlingvolgsysteem voor in het

voortgezet onderwijs. Op basis van eerder werk bij de Universiteit Leiden in samenwerking met Espin en Chung, hebben zij Sneltoets Informatieve Teksten (SIT) ontwikkeld om eens in de zes weken de voortgang van leerlingen te kunnen meten op leesvaardigheid (CED-groep, 2012). Deze SIT-toetsen zijn maze-taken met elk een ander onderwerp. Omdat de SIT-toetsen deel uit maken van een methode die is ontworpen om leerlingen te ondersteunen in het aanleren van leesvaardigheden, vindt de CED-groep het belangrijk om te weten wat de psychometrische eigenschappen zijn van deze toetsen. Daarnaast zal voor dit onderzoek een begrijpend lezentoets worden gebruikt, die ook is ontworpen door de CED-groep. Deze studie zal aan de hand van de SIT-toetsen en de begrijpend lezentoets als criteriumvariabele onderzoeken of de scores van de maze-taken betrouwbaar, valide en sensitief voor groei in leesvaardigheid zijn. Met het oog op CBM zouden de materialen van de CED-groep eventueel kunnen worden gebruikt als leerlingvolgsysteem.

Uit bovenstaande vloeit de volgende onderzoeksvraag voort, waarbij de bijbehorende deelvragen worden beantwoord: *Is de score van de SIT maze-taak een betrouwbaar en valide meetinstrument voor leesvaardigheid van middelbare scholieren van de 1<sup>e</sup> tot en met de 3<sup>e</sup> klas in Nederland?*

- a): Zijn de scores van de maze-taken betrouwbaar?
- b): Zijn de scores van de maze-taken valide indicatoren voor leesvaardigheid?
- c): Zijn de scores voor de maze-taken sensitief voor verandering in leesvaardigheid?

Voor het beantwoorden van de onderzoeksvraag en de deelvragen zijn drietal hypothesen opgesteld.

*Betrouwbaarheid scores maze-taken.* Er wordt een sterke paralleltest betrouwbaarheid verwacht van de maze-teksten van  $r = .80$ . Dit houdt in dat de twee teksten waar elke maze-taak uit bestaat een hoge correlatie zullen hebben. En er wordt een klein verschil verwacht per tijdstip van afname. In het onderzoek van Tichá et al. (2009) worden coëfficiënten van boven  $r = .80$  gevonden voor maze-taken. In hun onderzoek zijn ook kleine verschillen in tijdstip van afname gevonden van  $r = .84$  tot  $.88$ , maar dit gold niet voor elke tekst. In een ander onderzoek van Espin et al. (2010) is een groter verschil gevonden in afname van tijdstip, namelijk  $r = .80$  tot  $.88$ . Daarom wordt in dit onderzoek ervan uit gegaan dat er een sterke paralleltest betrouwbaarheid wordt gevonden.

*Scores maze-taken valide indicatoren voor leesvaardigheid.* Er wordt een gemiddelde tot hoge correlatie tussen de scores van de maze-taken en de scores van de criteriumvariabele verwacht. In het onderzoek van Tichá et al. (2009) is onderzoek gedaan naar de correlatie tussen de maze-taken met één criteriumvariabele. Hieruit bleek dat de correlaties van de maze-taak met de criteriumvariabele sterk waren gecorreleerd met correlaties van  $r = .80$  tot  $.85$ . Ook in het onderzoek van Espin et al. (2010) is gevonden dat de scores van de maze-taken valide indicatoren zijn voor leesvaardigheid, waarbij de correlaties varieerden van  $r = .75$  tot  $.80$ . Daarom zal in deze studie een gemiddelde tot hoge correlatie worden verwacht tussen de scores van de maze-taak en de scores van de criteriumvariabele, in verhouding tot voorgaande studies.

*Scores maze-taken sensitief voor verandering in leesvaardigheid.* Er wordt een verwacht dat de scores van de maze-taken sensitief zijn voor groei, waarbij het verschil in scores van de 1<sup>e</sup> t/m de 4<sup>e</sup> afname van de maze-taken per leerling van gemiddeld 1 correct woord per afname zal zijn. Uit de studie van Shin, Deno & Espin (2000) is gebleken dat maze-taken een goed alternatief waren voor betrouwbare meetinstrumenten zoals bijvoorbeeld de hardop lezen taak. Daarnaast zijn maze-taken sensitief in het meten van groei in de leesvaardigheid. Shin et al. (2000) hebben gevonden dat de vooruitgang niet groot is met een groei van gemiddeld 1.07 correcte woorden per maand. Zowel in de studie van Tichá et al. (2009) als in het onderzoek van Espin et al. (2010) is gevonden dat de scores van de maze-taken groei laten zien van ongeveer 1.29 tot 2.88 correcte woorden per week. Daarom zal in deze studie worden verwacht dat er sprake is voor sensitiviteit voor groei.

## **Methoden**

### **Participanten**

Participanten waren 856 leerlingen, waarvan 383 jongens (44,7%) en 369 meisjes (43,1%) van het middelbare onderwijs in Nederland. De gemiddelde leeftijd van de leerlingen was 13,06 jaar (SD = 1.06). Drieënvijftig procent van de leerlingen zat in de eerste klas, 16,7% in de tweede en 4,7% van de leerlingen in de derde klas.

Voor dit onderzoek hebben studenten van de Universiteit Leiden vier middelbare scholen in Nederland bezocht, waar zij data hebben verzameld. Deze data kwamen van leerlingen van het voortgezet onderwijs van verschillende onderwijsniveaus, namelijk het vmbo (praktijk onderwijs, vmbo-bbl, vmbo-kb en vmbo-kgt), havo en het vwo van de 1<sup>e</sup> tot en met de 3<sup>e</sup> klas. De scholen lagen verspreid over Nederland.

De steekproef was een gelegenheidssteekproef, waarbij scholen werden benaderd die al gebruik maakten van Nieuwsbegrip (Nieuwsbegrip, 2014a). Medewerkers van de CED-groep hebben deze scholen benaderd en een licentie voor een uitgebreide versie van Nieuwsbegrip als beloning in het vooruitzicht gesteld.

### **Meetinstrumenten**

Voor dit onderzoek zijn twee meetinstrumenten gebruikt, namelijk maze-taken op papier, ook wel Sneltoets Informatieve Teksten (SIT), en als criteriumvariabele de Nieuwsbegrip leestoets. Beide meetinstrumenten zijn ontwikkeld door het Centrum voor Educatieve Dienstverlening (CED-groep).

*SIT maze-taken.* In dit onderzoek werd gebruik gemaakt van SIT maze-teksten van de CED-groep, die zij hebben ontwikkeld in samenwerking met de Universiteit Leiden. Deze maze-taken werden ontwikkeld om te meten of groei zichtbaar was in de leesvaardigheid van een leerling. De teksten die zijn afgenomen bestonden uit ongeveer zestig tot tachtig woorden per tekst. Elke tekst had een eigen actueel onderwerp met een dieper liggende boodschap zoals bijvoorbeeld ruimteafval,



gamen en verstandig snoepen. De teksten werden geschreven door de CED-groep en zij hebben in samenwerking met Espin en Chung van de Universiteit Leiden de teksten omgevormd tot maze-taken aan de hand van de criteria van Fuchs en Fuchs (1992). Deze criteria beschrijven dat het van belang is dat er passages worden gebruikt die buiten het curriculum vallen. Deze passages moeten passen bij het niveau van de leerlingen. Vervolgens moet de eerste zin intact gelaten worden, terwijl voor de andere zinnen elk zevende woord wordt vervangen door drie alternatieven. Eén woord is correct en de andere twee woorden zijn afleiders. Vervolgens moet de tekst lang genoeg worden zodat een leerling deze niet kan afmaken binnen de vooraf aangegeven tijd.

De SIT maze-taken zijn in vijf verschillende niveaus gemaakt, namelijk niveau AA (groep 4); niveau A (groep 5/6); niveau B (groep 7/8 en onderbouw vmbo en mbo niveau 1); niveau C (leerjaar 2 vmbo, onderbouw havo/vwo en mbo niveau 2/3) en niveau D (leerjaar 4 vmbo-t, leerjaar 3/4 havo/vwo en mbo niveau 3/4). Deze SIT maze-taken kunnen zeven keer per schooljaar worden afgenomen.

In dit onderzoek werd gebruik gemaakt van vier van de vijf beschikbare niveaus van de maze-taken, namelijk niveau A, B, C en D. De maze-taken werden één keer in de zes á zeven weken afgenomen. Bij elke afname werden twee maze-teksten achter elkaar afgenomen. Het gemiddelde en de standaarddeviatie van de twee teksten zijn gebruikt voor de analyse. De maze-taken zijn vanaf november beschikbaar gesteld voor de scholen, maar pas vanaf medio februari en maart gebruikt voor dit onderzoek, afhankelijk van de school. De SIT-toetsen werden voor het eerst gebruikt tijdens het onderzoek op de scholen. De maze-taken werden klassikaal afgenomen op papier en waren niet tijdsintensief. Docenten namen de taken af nadat hij/ zij instructies had gekregen over de afname. De maze-taken werden op papier gemaakt en de leerlingen moesten na twee minuten stoppen met het maken van de taak. De leerlingen hadden consequent hetzelfde niveau moeten maken, maar dit is niet op elke school gedaan. Hierdoor verschilden de samples van leerlingen bij de analyse.

*Nieuwsbegrip leestoets.* De criteriumvariabele in dit onderzoek was de Nieuwsbegrip Leestoets. De CED-groep biedt scholen een begrijpend lezen methode aan die zij Nieuwsbegrip noemen. Met deze methode kunnen docenten leerlingen wekelijks laten oefenen met leesteksten en opdrachten (Nieuwsbegrip, 2014b). Bij het aanleren en oefenen van het begrijpend lezen maakt Nieuwsbegrip gebruik van evidence-based strategieën voor begrijpend lezen. Elke week staat een andere leesstrategie centraal (voorspellen, ophelderend van onduidelijkheden, vragen stellen, relaties en verwijzwoorden en samenvatten). De Nieuwsbegrip leesteksten bevatten actuele onderwerpen en er zijn vijf niveaus beschikbaar (niveau AA, A, B, C en D). Wanneer een school kiest voor uitbreiding van Nieuwsbegrip (Nieuwsbegrip XL) kunnen docenten ook gebruik maken van de Nieuwsbegrip leestoetsen. De leestoetsen bestaan uit teksten met actuele onderwerpen, waarbij leerlingen worden getoetst op hun leesvaardigheid. Docenten kunnen aan de hand van een begin- en eindtoets de voortgang van hun leerlingen bijhouden.

In huidig onderzoek is gebruik gemaakt van leestoetsen die zijn samengevoegd tot twee versies die als pre- en posttest zijn gebruikt. Er werd gebruik gemaakt van A-niveau teksten en een B-niveau teksten. Hierbij was de tekst van het A-niveau van een lager niveau dan het B-niveau, zodat zowel de slechte als de goede lezers deze toetsen konden maken. De leerlingen kregen 30 minuten de tijd om de leestoets te maken op de computer. Om verbetering te kunnen meten in de tijd, werden twee versies van de leestoets gemaakt: één die gebruikt werd als pre-test en een ander die gebruikt werd als post-test, maar sommige scholen hebben de verkeerde versie als pre-of post-test afgenomen. Er was een grote sample met leerlingen die versie 1 hadden gemaakt als pre-test en daardoor was er een veel kleinere sample leerlingen die versie 2 hadden gemaakt als post-test. Daarom zijn alleen de data van de pre-test gebruikt.

## **Procedure**

Ter voorbereiding op de afnames hebben de studenten aan de Universiteit Leiden een uitgebreide introductie gehad over het gebruik en de scoring van de SIT maze-taken. Ook hebben zij een korte instructie gehad over het observeren van de docenten en de leerlingen wanneer de SIT maze-taken worden afgenomen. Tevens is per leerling toestemming voor deelname aan het onderzoek gevraagd aan hun ouders.

Gedurende zes maanden is de leesvaardigheid van leerlingen van het voortgezet onderwijs gevolgd. Medio februari en maart werden de SIT maze-taken afgenomen in de klassen 1 tot en met 3 van middelbare scholen in Nederland. De pre-test van de Nieuwsbegrip Leestoets werd vanaf april afgenomen en de posttest aan het einde van het schooljaar. Tijdens afnames hebben studenten van de Universiteit Leiden de docenten geobserveerd of de instructie juist was gegeven en of taken juist waren afgenomen. De SIT maze-taak is om de zes á zeven weken afgenomen. De SIT maze-taken bestonden per afname uit twee teksten en werden klassikaal, onder begeleiding van vaste docenten worden afgenomen. De docenten hebben vooraf een instructie ontvangen die zij moeten voorlezen aan de klas. De Nieuwsbegrip Leestoets nam 30 minuten in beslag, terwijl de SIT maze-taken twee minuten duurden per tekst. De tijd werd door de docent bijgehouden. De studenten van de Universiteit Leiden zijn steekproefsgewijs langsgaan bij de scholen om te kijken of de taken op de juiste manier worden afgenomen.

De Nieuwsbegrip Leestoets en de SIT maze-taken werden door de studenten van de Universiteit Leiden gescoord en ingevoerd in het statistische computerprogramma SPSS 21. Hierna zijn de gegevens, inclusief demografische gegevens geanalyseerd.

## **Data-analyse**

Om de onderzoeksvraag te kunnen beantwoorden werden verschillende analyses uitgevoerd aan de hand van het statistische computerprogramma SPSS 21. De data werd ingevoerd en geanalyseerd. Per deelvraag is een selectie gemaakt van de data, omdat voor elke deelvraag een andere

dataset nodig was om de vraag te kunnen beantwoorden. Eerst werd gekeken hoe de data verdeeld was, voldeed aan de aannames van de toetsen die gebruikt zouden worden en of er geen missende of extreme waarden waren.

Om de hoofdvraag te kunnen beantwoorden of de score van de maze-taak een betrouwbaar en valide meetinstrument voor leesvaardigheid van middelbare scholieren van de 1<sup>e</sup> tot en met de 3<sup>e</sup> klas in Nederland, worden verschillende analyses uitgevoerd per deelvraag. Om de eerste deelvraag te kunnen beantwoorden werd met de Pearson's correlatiecoëfficiënt gekeken of er sprake was van betrouwbaarheid van de scores van de maze-taken. Bij de tweede deelvraag of de scores van de maze-taken valide indicatoren waren voor leesvaardigheid werd aan de hand van de Pearson's correlatiecoëfficiënt de correlaties berekend tussen de maze-taken en de criteriumvariabele Nieuwsbegrip Leestoets. Om te onderzoeken of de scores van de maze-taken sensitief waren voor verandering in leesvaardigheid werd een Repeated Measure ANOVA uitgevoerd.

## **Resultaten**

### **Data-inspectie**

In onderstaand stuk zijn de statistische analyses omschreven per hypothese. Voorafgaand aan de analyse werd gekeken naar een algemene data-inspectie per hypothese, omdat elke hypothese een andere dataset had. Per analyse zal worden besproken welke stappen werden gedaan.

### **Betrouwbaarheid**

Om de betrouwbaarheid van de scores op de maze-taken te berekenen is gebruik gemaakt van de Pearson's correlatiecoëfficiënt. Op deze wijze wordt gekeken of er een verband was tussen de twee teksten waaruit elk meetmoment bestaat.

Voorafgaand aan de analyse is naar de algemene gegevens gekeken, per meetmoment en per tekst, om te controleren of er geen sprake was van afwijkende gemiddelden en afwijkende standaard deviaties. In Tabel 1 is te zien dat het aantal meegewogen cases verschilde per afnamemoment. Dit komt omdat niet alle leerlingen hadden deelgenomen aan alle meetmomenten. Wanneer de gegevens in Tabel 1 nader werden bestudeerd was te zien dat de gemiddelde scores van de twee teksten op meetmoment 2 en 3 nagenoeg hetzelfde waren. Bijvoorbeeld was te zien dat op meetmoment 2 de gemiddelde scores voor niveau C 23.9 en 24 correcte antwoorden waren in twee minuten. Op meetmoment 3, waren de gemiddelde scores 21.2 en 22.7 correcte antwoorden. Echter was te zien dat op het eerste en vierde meetmoment de gemiddelde scores substantieel meer verschilden. Dit kan te maken hebben met het feit dat bij het eerste meetmoment de leerlingen nog niet wisten wat er van ze verwacht werd, terwijl bij meetmoment 4 zij dit wel wisten. Op meetmoment 4 voor niveau C was te zien dat de gemiddelde scores met verschil van 5.6 van elkaar afwijken. Deze resultaten betekenen dat de teksten mogelijk in moeilijkheidsgraad verschilden.

Tabel 1  
*Beschrijvende data tekst I en II per SIT meetmoment en niveau*

SIT		SIT Meetmoment								
		1		2		3		4		
		Tekst I	Tekst II	Tekst I	Tekst II	Tekst I	Tekst II	Tekst I	Tekst II	
SIT niveau	A	Mean	14.68	15.15	19.08	24.22	19.20	18.81	17.22	20.37
		Stand.dev.	6.33	5.65	6.66	8.77	7.47	7.63	6.88	7.46
		N	53	53	50	50	59	59	46	46
	B	Mean	14.06	18.42	20.06	18.13	19.20	19.59	19.05	22.58
		Stand.dev.	3.76	4.87	6.05	5.76	5.57	6.93	5.98	5.54
		N	139	139	89	89	138	138	132	132
	C	Mean	15.29	19.13	23.86	24.03	21.22	22.74	22.93	28.59
		Stand.dev.	5.02	6.69	6.51	6.64	7.89	7.08	6.44	7.21
		N	217	216	325	311	171	170	160	157
	D	Mean	21.01	23.87	25.44	25.35	--	--	--	--
		Stand.dev.	5.41	6.26	6.07	6.71	--	--	--	--
		N	70	70	66	66	--	--	--	--

*Note:* ontbrekende data wordt aangegeven met --.

Na de algemene analyse is ook gekeken of er sprake was van missende data en extreme waarden die invloed zouden kunnen hebben op de correlatie. Er waren in totaal  $N = 11$  cases weggelaten uit de analyse vanwege extreme waarden. Deze cases hadden een extreem hoge of lage score, met veel fouten in de tekst. (zie Appendix, Tabel 2B). In Tabel 2B is af te lezen dat de correlaties voor niveau B op meetmoment 1 en 4 en voor niveau C op meetmoment 3, lager waren wanneer de extreme waarden meegenomen werden in de analyse.

Nadat de extreme waarden waren verwijderd uit de dataset werden correlaties berekend tussen de twee teksten voor elk meetmoment en de correlaties werden ook bekeken per niveau. De correlatie van de scores, gerapporteerd in Tabel 2, bevinden zich in een range van  $r = .61 - .88$ . Hiervan bevonden twaalf van de veertien scores zich boven de correlatiecoëfficiënt van  $r = .70$  en zeven van de veertien bevonden zich boven de correlatiecoëfficiënt van  $r = .80$ . Alle scores zijn significant en boven de  $r = .60$ .

Tabel 2  
*Correlatie tekst I en II per SIT meetmoment en niveau*

SIT		SIT Meetmoment				
		1	2	3	4	
SIT niveau	A	Correlatie	.889**	.857**	.756**	.811**
		N	53	50	59	46
	B	Correlatie	.615**	.823**	.727**	.707**
		N	136	89	138	130
	C	Correlatie	.801**	.803**	.692**	.776**
		N	216	311	163	157
	D	Correlatie	.743**	.803**	--	--
		N	70	65	--	--

*Note:* \*\*  $p < .01$ .

Ontbrekende data wordt aangegeven met --.

### Valide indicator voor leesvaardigheid

Om de validiteit van de scores op de maze-taken te berekenen is gebruik gemaakt van de Pearson's correlatiecoëfficiënt om het verband tussen het tweede meetmoment van de maze-taken en de pre-test van de Nieuwsbegrip leestoets te berekenen. Voor deze analyse werd data gebruikt van het tweede meetmoment van de maze-taak en de pre-test van de Nieuwsbegrip leestoets. Hiervoor is gekozen, omdat het tijdstip van afname van het tweede meetmoment en de pre-test het dichtste bij elkaar lagen. Op deze wijze kon een representatieve weergave worden gegeven van de correlatie tussen de maze-taak en de leestoets. Niveau D werd niet meegenomen in de analyse omdat te weinig leerlingen de maze-taak van dat niveau hadden gemaakt in combinatie met de Nieuwsbegrip leestoets.

Voorafgaand aan de analyse werd naar de algemene gegevens gekeken van het meetmoment van de maze-taak en de pre-test van de Nieuwsbegrip leestoets, om te controleren of er geen sprake was van afwijkende gemiddelden en afwijkende standaard deviaties. In Tabel 3 is te zien dat de leerlingen met een hoger leesniveau hogere scores hadden voor de Nieuwsbegrip leestoets dan leerlingen met een lager leesniveau. Zo scoorden leerlingen die de maze-taak van niveau C hadden gemaakt gemiddeld 22.7 correcte woorden van in totaal 83 woorden binnen twee minuten. Op de Nieuwsbegrip leestoets scoorden zij gemiddeld 26.8 correcte antwoorden van in totaal 38 opgaven in dertig minuten. Leerlingen die niveau B hadden gemaakt scoorden op de maze-taak gemiddeld 18.7 correcte antwoorden van in totaal 65 woorden binnen twee minuten. Voor de Nieuwsbegrip leestoets haalden zij een gemiddelde score van 20.1 correcte antwoorden van in totaal 38 opgaven in dertig minuten. Het verschil in scores op de Nieuwsbegrip leestoets van de leerlingen die van de maze-taak niveau B en niveau C hebben gemaakt was 6.6 correcte antwoorden. Een verklaring voor deze resultaten zou zijn dat leerlingen die maze-taken van niveau C hadden gemaakt de Nieuwsbegrip

leestoets eenvoudiger vonden dan leerlingen die maze-taken van niveau B hadden gemaakt, wanneer alleen wordt gekeken naar de gemiddelde en standaarddeviaties.

Tabel 3  
*Beschrijvende data Nieuwsbegrip leestoets met de SIT maze-taak*

		SIT maze	Nieuwsbegrip leestoets	
SIT niveau	A	Mean	21.48	15.92
		Stand.dev.	7.35	5.15
		N	24	24
	B	Mean	18.68	20.17
		Stand.dev.	6.07	5.82
		N	35	35
	C	Mean	22.67	26.78
		Stand.dev.	5.34	6.54
		N	90	90

Daarnaast is onderzocht of er sprake was van missing data en extreme waarden. Voor deze analyse werd alleen de data van de leerlingen die zowel de Nieuwsbegrip leestoets hadden gemaakt als de maze-taken meegenomen. Het aantal meegewogen cases is klein per niveau (zie Tabel 3).

In Tabel 4 zijn de correlaties gerapporteerd van de correlaties tussen de maze-taak en de Nieuwsbegrip leestoets. De correlaties voor niveaus A en B waren  $r = .41$  en  $.56$ . Allebei de correlaties waren significant. De correlatie van Niveau C was  $r = .19$  en niet significant. Het zou kunnen zijn dat er sprake was van een plafondefect voor niveau C. Om te onderzoeken of er een plafondefect was, is gekeken naar de maximale score van de maze-taak en de leestoets. Bij de maze-taak van niveau C is de maximaal haalbare score 83 goede woorden en bij de Nieuwsbegrip leestoets een maximaal haalbare score van 38 goede antwoorden. Wanneer wordt gekeken naar de gemiddelde scores en de standaarddeviaties van de leerlingen die beide toetsen hebben gemaakt, is te zien dat het gemiddelde en de standaarddeviatie van de maze-taak ( $M = 22.67$ ,  $SD = 5.34$ ) en de leestoets ( $M = 26.78$ ,  $SD = 6.54$ ) niet groot genoeg was om te spreken van een plafondefect.

Tabel 4  
*Correlatie tekst Nieuwsbegrip leestoets met de SIT maze-taak*

SIT		Nieuwsbegrip leestoets
SIT niveau	A	Correlatie N
		.412* 24
	B	Correlatie N
		.567** 35
	C	Correlatie N
		.190 90

Note: \*  $p < .05$ ; \*\*  $p < .01$

### Sensitief voor verandering in leesvaardigheid

De sensitiviteit van de maze-taken voor groei in lezen was onderzocht door te bepalen of de scores van de maze-taak zouden toenemen gedurende de meetmomenten. Om deze vraag te kunnen beantwoorden werd gebruik gemaakt van Repeated Measure ANOVA.

Voor deze analyse was het van belang dat de leerlingen datapunten hadden voor alle vier de meetmomenten. Echter hadden maar 3 leerlingen voor niveau C aan alle vier de meetmomenten deelgenomen. Daarom is gekozen om drie meetmomenten te gebruiken voor de analyse, in plaats van vier. De grootste sample per niveau was gevonden voor meetmomenten 2, 3 en 4. Uiteindelijk werd de data voor meetmoment 2, 3 en 4 voor niveau A tot en met C gebruikt voor de analyse (zie Tabel 5). In Tabel 5 zijn de gemiddelde scores en standaarddeviaties gerapporteerd van de maze-taken per niveau voor elk meetmoment.

Tabel 5  
*Beschrijvende data SIT meetmoment 2, 3 en 4*

		SIT Meetmoment			
SIT		2	3	4	
SIT niveau	A	Mean	21.37	18.72	17.78
		Stand.dev.	7.30	7.23	6.49
		N	32	32	32
	B	Mean	19.48	19.51	21.05
		Stand.dev.	6.30	5.06	5.10
		N	60	60	60
	C	Mean	22.61	22.12	26.12
		Stand.dev.	5.16	6.95	6.44
		N	136	136	136

Voor niveau B en C namen de scores toe gedurende de meetmomenten, terwijl voor niveau A de scores afnamen. Om de significanties van deze verschillen te kunnen berekenen werd een Repeated

Measure uitgevoerd. Om een Repeated Measure te kunnen uitvoeren moest worden voldaan aan de aanname van sphericiteit. Uit Mauchly's test is gebleken dat de varianties niet significant waren ( $\chi^2(2) = 1.14, p = .567$ ), wat inhoudt dat de aanname van sphericiteit niet was geschonden en de Repeated Measure kon worden uitgevoerd. Uit de analyse bleek dat er sprake was van een hoofdeffect voor het meetmoment  $F(2, 194) = 30.93, p < .001$  en een interactie effect voor het meetmoment en het niveau  $F(4, 194) = 4.61, p = .001$ . In Figuur 2 (zie Appendix) zijn de gemiddelde scores visueel weergegeven in een grafiek. Men kan zien dat bij niveau B en C de scores op de maze-taak groeide naarmate de leerlingen verder kwamen in het leerjaar. Bij niveau A was er geen sprake van groei. Dit zou kunnen worden verklaard door het kleine aantal cases.

Om bovenstaande uitkomsten nader te onderzoeken zijn drie single Repeated Measures als follow-up test uitgevoerd voor elk niveau apart. Hieruit kwam naar voren dat bij niveau A wel een effect in tijd te zien is ( $F(2,30) = 8.52, p = .001$ ) en sprake was van een negatief lineair verband ( $F(1,31) = 17.58, p < .001$ ). Aan de hand van de LSD post-hoc toets waren voor niveau A significante verschillen gevonden tussen de tijdstippen 2 en 3 van  $p = .007$  en tussen tijdstippen 2 en 4 van  $p < .001$ . Bij niveau B was een effect in tijd te zien ( $F(2,59) = 5.37, p = .007$ ) en hier was sprake van een lineair verband ( $F(1,60) = 7.37, p = .009$ ). Tussen tijdstip 2 en 4 was een significant verschil gevonden aan de hand van de LSD post-hoc toets van  $p = .009$ . Bij niveau C was er een effect voor tijd gevonden ( $F(2,135) = 47.70, p < .001$ ) en hier was sprake van zowel een lineair ( $F(1,136) = 75.74, p < .001$ ) als een kwadratisch verband ( $F(1,136) = 26.59, p < .001$ ). Wanneer wordt gekeken naar de significanties tussen de tijdstippen was te zien dat tussen tijdstip 2 en 4 ( $p < .001$ ) en tijdstip 3 en 4 ( $p < .001$ ) een significant verschil was gevonden aan de hand van de LSD post-hoc toets.

## Discussie

Het doel van dit onderzoek was om te onderzoeken of de SIT maze-taak een betrouwbaar, valide en sensitief meetinstrument is voor het meten van leesvaardigheid. Wanneer wordt gekeken naar de analyses kan worden geconcludeerd dat voor elke hypothese steun is gevonden. Er is gebleken dat de SIT maze-taak een betrouwbaar meetinstrument is voor het meten van leesvaardigheid. De andere twee hypothesen kunnen voor dit onderzoek deels worden aangenomen. De maze-taak is valide bevonden voor niveaus A en B en bij niveaus B en C is steun gevonden voor sensitiviteit van groei voor het meten van leesvaardigheid.

Bij de analyse van de eerste hypothese over betrouwbaarheid van de scores van de maze-teksten, is gebleken dat er sprake is van een redelijk hoge correlatie met de meeste correlaties boven de  $r = .70$ , waarvan zeven van de veertien correlaties een correlatiecoëfficiënt had van  $r = .80$ . Wanneer dit wordt vergeleken met de correlaties ( $r = .80$ ) die gevonden zijn in het onderzoek van Tichá et al. (2009), kan worden gesteld dat de gevonden correlaties binnen de range van voorgaande onderzoeken liggen en daarmee de correlaties die zijn gevonden in dit onderzoek vrij hoog genoemd kunnen worden.



Bij de hypothese over de validiteit van de scores van de SIT maze-taken als indicator van de leesvaardigheid, is gekeken of er een relatie bestaat tussen de maze-taken en de criteriumvariabele. In dit onderzoek is alleen de pre-test van de Nieuwsbegrip leestoets gebruikt als criteriumvariabele. Uit de analyse is gebleken dat de correlatie positief is, maar niet sterk zoals gevonden in voorgaand onderzoek van Espin et al. (2010) waar de correlaties varieerden van  $r = .75$  tot  $.80$ . Daarnaast is gebleken dat het moeilijkste niveau (niveau C) niet significant is. Een reden voor deze verschillen kan de grootte van de sample zijn, die vrij klein was in dit onderzoek. Daarentegen is een meer waarschijnlijke reden de beperkte omvang van de scores binnen elk niveau. Wanneer men deze reden overweegt, zijn de correlaties die zijn gevonden voor niveaus A en B redelijk hoog en kan er gesproken worden van steun voor validiteit voor de maze-taak. Echter is de correlatie van  $r = .19$  voor niveau C laag, zelfs voor een sample met een beperkt aantal scores. Een mogelijke reden voor deze lage correlatie was het bereiken van een plafondeffect op of de maze-taak of de Nieuwsbegrip leestoets, maar dit plafondeffect was niet gevonden. Voor deze studie moet men concluderen dat er geen steun is gevonden voor validiteit op de maze-taak voor leerlingen die een hoger leesniveau hebben.

Tenslotte is de derde hypothese onderzocht, waarbij is gekeken of de scores van de maze-taken sensitief zijn voor verandering in leesvaardigheid. Deze analyse verschaft steun voor de maze-taak als meetinstrument voor groei van leerlingen voor niveaus B en C, maar niet voor niveau A. Voor niveau A namen de scores af gedurende de meetmomenten. Het is vreemd dat de scores voor de leerlingen die niveau C hadden gemaakt veranderden tussen meetmoment 3 en 4, vooral omdat de correlaties tussen de scores op meetmoment 2 en de Nieuwsbegrip leestoets laag waren en niet significant. Als de Nieuwsbegrip leestoets tegen het einde van het onderzoek was afgenomen (dichter bij meetmoment 4), dan was de relatie tussen de maze-taak en de Nieuwsbegrip leestoets waarschijnlijk sterker geweest. De dalende van de scores van de leerlingen die niveau A hadden gemaakt is lastig te verklaren. Het is mogelijk dat de leesvaardigheid van de leerlingen afnam gedurende de tijd, maar de aanname waarmee deze analyse is gedaan was dat het gebruik van het Nieuwsbegrip programma de leesvaardigheid van de leerlingen juist zou verbeteren. Een zwak punt van deze studie was het gebrek aan een onafhankelijke pre- en post leestoets. Zoals eerder genoemd was het gebruik van de pre- en posttest een onderdeel van dit onderzoek, maar de juiste versie van de posttest werd niet op elke school gebruikt. Voor deze studie moet men concluderen dat de SIT maze-taak niet sensitief was voor groei van leesvaardigheid van de leerlingen op niveau A. Geadviseerd wordt om in toekomstig onderzoek een onafhankelijke pre- en post-leestest mee te nemen in het onderzoek. Samenvattend kan worden gesteld dat de resultaten van dit onderzoek enige steun bieden voor de SIT-maze taak als een indicator voor leesvaardigheid en groei in lezen, maar de resultaten waren niet consistent genoeg in meetmoment en niveau.

Er kan worden geconcludeerd dat de data van dit onderzoek betrouwbaar waren voor paralleltest betrouwbaarheid, maar dat de gemiddelde scores niet altijd gelijk waren. Dit kan de vraag

oproepen of er geen sprake was van een oefeneffect. De correlaties voor validiteit waren redelijk goed voor de niveaus A en B, maar niet voor niveau C. Daarbij werd een positieve verandering gevonden in de scores van niveaus B en C, maar niet voor niveau A. Het is van belang dat dit onderzoek wordt gerepliceerd met grotere samples en data die zijn samengenomen wat betreft de niveaus. Eén van deze oplossingen zal bijdragen aan een grotere verdeling van de scores en zal ook bijdragen een beter beeld van de karakteristieken van het meetinstrument. Als aanvulling zal het gebruik van een onafhankelijke pre- en post-leestest een betere wijze zijn om de sensitiviteit van de groei van de scores gedurende de meetmomenten te kunnen meten.

### **Implicaties voor de maatschappij**

De uitkomsten van dit onderzoek kunnen positief zijn met het oog op het opbrengstgericht werken dat door de Rijksoverheid als doel is gesteld (Rijksoverheid, 2010). Zeker met het oog op de toekomst waarbij het gebruik van leerlingvolgsystemen verplicht gaat worden voor scholen in Nederland (Rijksoverheid, 2011a; Onderwijsraad, 2012). CBM geeft de mogelijkheid aan leraren om aan de hand van dit systeem op een overzichtelijke manier de voortgang en resultaten van de leerlingen bij te houden en, waar nodig, hun instructie aan te passen (Deno 1985). Deze mogelijkheid van het CBM is wat docenten nodig hebben om doelen te kunnen stellen, planmatig en resultaatgericht te werken. Doordat CBM resultaten van de leerlingen bijhoudt, evenals de aangepaste instructie, wordt het voor scholen makkelijker om sterke en zwakke plekken in de leerresultaten van de leerlingen op te sporen. Uit het onderzoek van de Rijksoverheid (2010) bleek dat scholen over het algemeen over een goed toetsinstrumentarium beschikken, maar nog geen vaste criteria hanteert om leerlingen met leerproblemen op te sporen. Vanuit het CBM kan voor leesvaardigheid de maze-taak worden gebruikt om op deze wijze criteria te kunnen stellen.

### **Sterke punten onderzoek**

Sterke punten van dit onderzoek zijn dat de vraagstelling vernieuwend is op het gebied van CBM in Nederland. Er is weinig onderzoek gedaan naar CBM voor leesvaardigheid op middelbare scholen in Nederland aan de hand van maze-taken. Ook is uit dit onderzoek gebleken dat, ondanks de limitaties, er bewijs is voor betrouwbaarheid en steun voor validiteit en sensitiviteit voor groei van leesvaardigheid. Doordat het gebruik van CBM weinig tijd vraagt van de docent en de leerling (Deno, 2003), zou dit eventueel ook ingezet kunnen worden als leerlingvolgsysteem in Nederland om op deze wijze leerproblemen eerder op te sporen en aan te pakken. Een ander sterk punt van dit onderzoek is dat gebruik is gemaakt van al bestaande methoden die reeds in de praktijk worden gebruikt. Dit onderzoek heeft bijgedragen aan het geven van inzicht in de reeds bestaande materialen. De CED-groep heeft in samenwerking met de Universiteit Leiden de resultaten van het onderzoek naar de psychometrische kenmerken van de meetinstrumenten geëvalueerd. De CED-groep gaat aan de hand

van de gevonden resultaten verder met het ontwikkelen en verbeteren van hun methoden en materialen.

### **Beperkingen**

Dit onderzoek heeft een aantal beperkingen. Het onderzoek is uitgevoerd aan de hand van een gelegenheidssteekproef, waarbij alleen scholen die al gebruik maakten van de Nieuwsbegrip methode hebben deelgenomen aan het onderzoek. Daarnaast was het optioneel voor de scholen om gebruik te maken van de Nieuwsbegrip leestoets. Hierdoor was het aantal leerlingen die zowel de maze-taak hadden gemaakt als de Nieuwsbegrip leestoets klein. Per leerling werd een niveau voor de maze-taak aangewezen en dit bleek lastig te zijn voor de docent om bij te houden. Hierdoor hebben niet alle leerlingen consequent hetzelfde niveau gemaakt of aan alle meetmomenten deelgenomen.

### **Suggesties voor toekomstig onderzoek**

Uit de beperkingen vloeien een aantal suggesties voort voor toekomstig onderzoek om de hypothesen van dit onderzoek een betere wetenschappelijke onderbouwing te geven. Aangeraden wordt om de onderzoeksgroep te vergroten, evenals de mogelijkheid tot het aanpassen van de meetinstrumenten. Het aantal niveaus zal kunnen worden geminimaliseerd en er zou een efficiëntere manier kunnen worden ontworpen om de maze-taken te maken en bij te houden. Men zou bijvoorbeeld gebruik kunnen maken van maze-taken die via de computer worden aangeboden, zodat elke leerling dezelfde taak maakt en daarbij ook consequent hetzelfde niveau. Daarnaast zou het aantal meetmomenten moeten worden vergroot, zodat er meer datapunten zijn en ook gekeken kan worden naar het leereffect dat over de tijd zou kunnen ontstaan.

Ook zou het constructief zijn wanneer de Universiteit Leiden nauw blijft samenwerken met onderwijsinstanties. Op deze wijze kan onderling informatie worden uitgewisseld over de ontwikkeling van materialen rondom een nieuw leerlingvolgsysteem in het voortgezet onderwijs in Nederland.

## Literatuur

- Brown-Chidsey, R., Johnson, P.jr. & Fernstrom, R. (2005). Comparison of Grade-level Controlled and Literature based Maze CBM Reading Passages. *School Psychology Review*, 34 (3), 387-394.
- CED-groep (datum onbekend). *Nieuwsbrief: Bloklessen niet (meer) als toetsmoment*. Verkregen op 30 juni 2014, van [http://www.nieuwsbegrip.nl/uploads/media\\_items/informatie-blokles.original.pdf](http://www.nieuwsbegrip.nl/uploads/media_items/informatie-blokles.original.pdf).
- CED-groep (2012). *Handleiding SIT: Niveau A, Versie 1*. Vertrouwelijk document verkregen van de CED-groep omstreeks oktober 2012.
- CED-groep (2014). Verkregen op 23 mei 2014, van <http://www.cedgroep.nl>.
- CITO (2013). *Resultaten PISA-2012*. Verkregen op 27 mei, 2014, van [http://www.cito.nl/onderzoek%20en%20wetenschap/deelname\\_int\\_onderzoek/pisa/resultaten](http://www.cito.nl/onderzoek%20en%20wetenschap/deelname_int_onderzoek/pisa/resultaten).
- CITO (2013). *Cito Volgstelsysteem voortgezet onderwijs: voortgang, advisering en kwaliteitszorg*. Verkregen op 25 februari, 2013, van [http://www.cito.nl/Onderwijs/Voortgezet%20onderwijs/cito\\_volgstelsysteem\\_vo.aspx](http://www.cito.nl/Onderwijs/Voortgezet%20onderwijs/cito_volgstelsysteem_vo.aspx).
- Deno, S.L. (1985). Curriculum-Based Measurement: The Emerging Alternative. *Exceptional Children*, 52, 219-232.
- Deno, S.L. (1992). The Nature and Development of Curriculum-Based Measurement. *Preventing School Failure*, 36, 5-10.
- Deno, S.L. (2003). Developments in Curriculum-Based Measurement. *The Journal of Special Education*, 37 (3), 184-192.
- Espin, C., Wallace, T., Lembke, E., Campbell, H. & Long, J.D. (2010). Creating a Progress-Monitoring System in Reading for Middle-School Students: Tracking Progress Toward Meeting High-Stakes Standards. *Learning Disabilities Research & Practice*, 25 (2), 60-75.
- Fuchs, L.S., & Fuchs, D. (1992). Identifying a Measure for Monitoring Student Reading Progress. *School Psychology Review*, 21, 45-58.
- Fuchs, L.S., & Fuchs, D. (1999). Monitoring Student Progress towards the Development of Reading Competence: A review of three forms of classroom-based assessment. *School Psychology Review*, 28, 659-671.
- Hintze, J. M., & Silbergliitt, B. (2005). A Longitudinal Examination of the Diagnostic Accuracy and Predictive Validity of R-CBM and High-Stakes Testing. *School Psychology Review*, 34 (3), 372-386.
- Madelaine, A. & Wheldall, K. (2004). Curriculum-Based Measurement of Reading: recent advances. *International Journal of Disability, Development and Education*, 51 (1), 57-82.

- Nieuwsbegrip (2014a), Verkregen op 18 juli 2014, van [www.nieuwsbegrip.nl](http://www.nieuwsbegrip.nl).
- Nieuwsbegrip (2014b). Verkregen op 25 februari 2014, van <http://www.nieuwsbegrip.nl/over-nieuwsbegrip/inhoud-didactiek>.
- Onderwijsraad (2012). *Advies: Leerlingvolgsysteem en tussentoets*. Verkregen op 28 juni van <http://www.onderwijsraad.nl/publicaties/2012/leerlingvolgsysteem-en-tussentoets/volledig/item217>.
- Rijksoverheid (2010). *Opbrengstgericht werken in het basisonderwijs*. Verkregen op 16 januari 2014, van <http://www.rijksoverheid.nl/documenten-en-publicaties/brochures/2011/01/20/opbrengstgericht-werken-in-het-basisonderwijs.html>.
- Rijksoverheid (2011b). *Nieuwsbericht: Naar een ambitieuze leercultuur*. Verkregen op 28 juni 2014, van <http://www.rijksoverheid.nl/nieuws/2011/05/23/naar-een-ambitieuze-leercultuur.html>.
- Rijksoverheid (2011b). *Opbrengsten maak er werk van!* Verkregen op 16 januari 2014, van <http://www.rijksoverheid.nl/documenten-en-publicaties/brochures/2011/02/10/opbrengsten-maak-er-werk-van.html>.
- Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-Based Measures and Performance on State Assessment and Standardized Tests: Reading and Math Performance in Pennsylvania. *Journal of Psychoeducational Assessment*, 24(1), 19–35.
- Shin, J., Deno, S.L. & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based-measurement of reading growth. *The Journal of Special Education*, 34, 164-172.
- Stecker, P.M., Fuchs, L.S. & Fuchs, D. (2004). Using Curriculum-Based Measurement to improve student achievement: review of research. *Psychology in the Schools*, 42 (8), 795-819.
- Tichá, R., Espin, C.A. & Wayman, M.M. (2009). Reading Progress Monitoring for Secondary-School Students: Reliability, Validity and Sensitivity to Growth of Reading-Aloud and Maze-Selection Measures. *Learning Disabilities Research & Practice*, 24 (3), 132-142.
- Tolar, T.D., Barth, A.E., Francis, D.J., Fletcher, J.M., Stuebing, K.K. & Vaughn, S. (2012). Psychometric Properties of Maze Tasks in Middle School Students. *Assessment for Effective Intervention*, 37 (3), 131-146.
- Wayman, M.M., Wallace, T., Wiley, H.I., Tichá, R., & Espin, C.A. (2007). Literature Synthesis on Curriculum-Based Measurement in Reading. *The Journal of Special Education*, 41, 85-120.
- Yeo, S. (2010). Predicting Performance on State Achievement Tests Using Curriculum-Based Measurement in Reading: A Multilevel Meta-Analysis. *Remedial and Special Education* 31(6), 412-422.

## Appendix

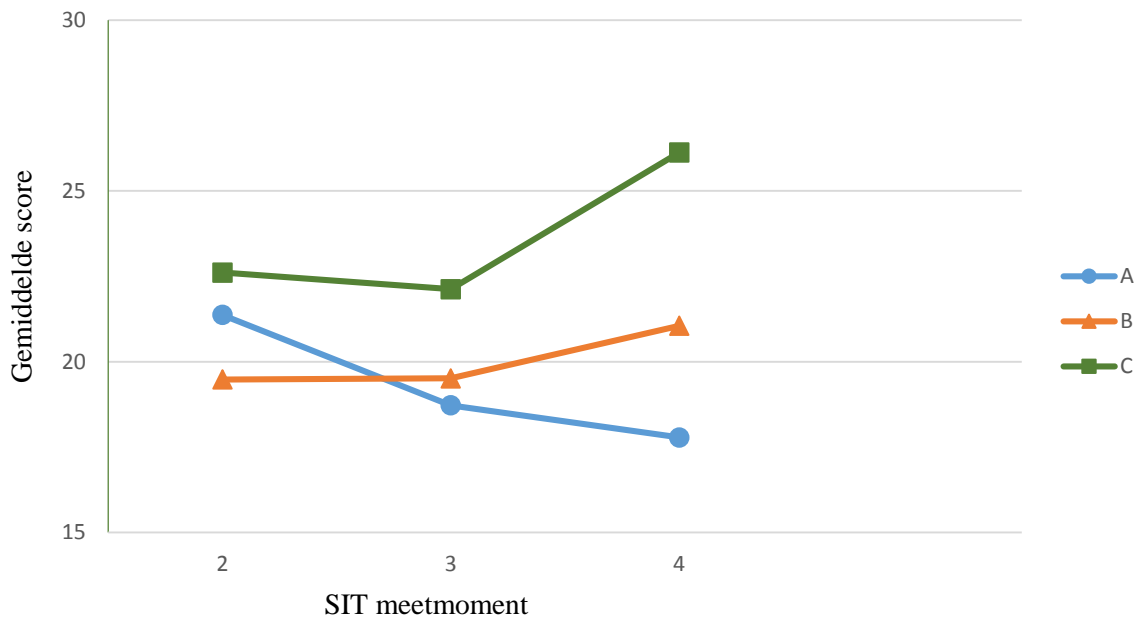
Tabel 2B

*Correlatie tekst 1 en 2 per SIT meetmoment en niveau met de extreme waarden*

		SIT meetmoment				
		1	2	3	4	
SIT niveau	A	Correlatie	.889**	.857**	.756**	.811**
		N	53	50	59	46
	B	Correlatie	<b>.595**</b>	.823**	.727**	<b>.665**</b>
		N	<b>139</b>	89	138	<b>132</b>
	C	Correlatie	.801**	.803**	<b>.628**</b>	.776**
		N	216	311	<b>169</b>	157
	D	Correlatie	.743**	.803**	--	--
		N	70	65	--	--

Note: \*\* p < .01

Ontbrekende data wordt aangegeven met --.



Figuur 2. Gemiddelde groei van de SIT meetmomenten per niveau