



**Universiteit Utrecht**

**Masterscriptie**

## **Leesvaardigheid: uni- of multidimensioneel?**

Een analyse van leesvaardigheid aan de hand van toetsen van  
Cito Volgsysteem Voortgezet Onderwijs met het oog op  
mogelijke diagnostische informatie.

Naam: BA N.D. Dijkstra

Studentnummer: 5501237

MA Taal, Mens en Maatschappij

Universiteit Utrecht

Begeleider: prof. dr. H. H. van den Bergh

Tweede lezer: Nicole Alberts

December 2015

## Voorwoord

Voor u ligt de scriptie ‘Leesvaardigheid: uni- of multidimensioneel?’. Een onderzoek gehouden op het gebied van leesvaardigheid onder leerlingen in de onderbouw van het voortgezet onderwijs. De scriptie is geschreven in het kader van mijn afstuderen aan de opleiding Taal, Mens en Maatschappij aan de Universiteit Utrecht (UU). Van juli tot en met december ben ik bezig geweest met het onderzoek en het schrijven van deze scriptie.

De scriptie is geschreven in opdracht van mijn stagebedrijf, Cito B.V. Samen met mijn stagebegeleidster, Nicole Alberts, is de onderzoeksopdracht tot stand gekomen. Leesvaardigheid was voor mij een enigszins nieuw onderwerp en bracht daardoor extra uitdaging met zich mee. Gelukkig vond ik het onderwerp erg interessant en was Nicole altijd bereid om te helpen waar nodig. Een andere uitdaging was het analyseprogramma Mplus, dat voor dit onderzoek gebruikt is en waar ik nog nooit mee gewerkt had. Met behulp van Anke Weekers en Marije den Ouden zijn de resultaten geworden tot wat ze nu zijn. Daarnaast kon ik gelukkig voor vragen en feedback, zowel voor de analyses als de inhoud, ook altijd terecht bij mijn begeleider vanuit de UU, de heer H. van den Bergh.

Bij dezen wil ik graag mijn begeleiders bedanken voor de fijne begeleiding en ondersteuning tijdens dit traject. Ook wil ik natuurlijk Cito bedanken voor alle data die ik heb kunnen gebruiken voor het onderzoek en alle hulpmiddelen die ik bij Cito tot mijn beschikking had. Zonder deze medewerking had ik dit onderzoek nooit uit kunnen voeren.

Tenslotte wil ik graag mijn ouders, mijn drie broers en mijn vriendinnen bedanken voor hun steun en de nodige ontspanning tijdens het schrijven van mijn scriptie. Het was niet altijd gemakkelijk, maar ook mede dankzij hen is de scriptieperiode tot een goed einde gekomen. Heel erg bedankt hiervoor.

Ik wens u veel leesplezier toe.

Nynke Dijkstra

Amsterdam, december 2015.

## Samenvatting

In dit onderzoek is gekeken naar de mogelijkheden tot het verkrijgen van diagnostische informatie betreffende leesvaardigheid Nederlands in de onderbouw van het voortgezet onderwijs. Hiervoor zijn de volgende onderzoeksvragen opgesteld: In hoeverre kan er diagnostische informatie verkregen worden over leesvaardigheid Nederlands in de onderbouw van het voortgezet onderwijs? In hoeverre zijn de veronderstelde dimensies te onderscheiden in de toetsen leesvaardigheid Nederlands van het Volgsysteem Voortgezet Onderwijs (VVO)? De veronderstelde dimensies behorend bij leesvaardigheid zijn ingedeeld in de categorieën Begrijpen en Interpreteren, waarbij bij zowel Begrijpen als bij Interpreteren nog een indeling in vier subcategorieën is gemaakt (oftewel acht subcategorieën in totaal).

Naar aanleiding van de literatuur is de verwachting ontstaan dat de veronderstelde dimensies in de toetsen leesvaardigheid Nederlands van het Cito VVO niet te onderscheiden zijn als gevolg van de complexiteit van leesvaardigheid, waarbij meerdere processen met elkaar interacteren (o.a. Alderson, 2000). Voor het beantwoorden van de onderzoeksvragen zijn confirmatieve factoranalyses uitgevoerd met behulp van Mplus op de toetsresultaten van leerlingen van verschillende opleidingsniveaus. Deze leerlingen hebben deelgenomen aan het onderdeel leesvaardigheid Nederlands van Toets 1 en Toets 3 van het Cito VVO. Voorafgaand aan de factoranalyses zijn de items uit de toetsen gecategoriseerd volgens de bovengenoemde veronderstelde dimensies, oftewel deelvaardigheden. Om te achterhalen of de veronderstelde deelvaardigheden vastgesteld zouden kunnen worden zijn drie modellen met elkaar vergeleken: een ééndimensionaal factormodel, een tweedimensionaal factormodel (met gecorreleerde factoren) en een achtdimensionaal factormodel (met gecorreleerde factoren).

De resultaten hebben laten zien dat de drie veronderstelde modellen aannemelijke modellen zijn, maar dat er nauwelijks verschil is tussen de modellen. De correlaties tussen de verschillende factoren hebben aangetoond dat er veel samenhang is tussen de deelvaardigheden: de factoren lijken hetzelfde te meten. Om die reden is het dus weinig zinvol om onderscheid te maken in de verschillende deelvaardigheden en is het moeilijk om diagnostische informatie te verkrijgen. Het is daarom raadzaam voor toetsontwikkelaars en onderzoekers om voorzichtig te zijn in hun beweringen dat leestoetsen gebruikt kunnen worden om diagnostische informatie te geven.

## Inhoudsopgave

1.	Inleiding .....	1
2.	Leesvaardigheid en diagnostische informatie.....	3
2.1	Theorieën leesvaardigheid: vaardigheden en processen.....	3
2.2	Onderzoek naar verschillende dimensies in leesvaardigheid .....	5
2.2.1	<i>Deelvaardigheden, processen en dimensies</i> .....	6
2.3	Het toetsen van leesvaardigheid.....	7
2.4	Nederlands leesvaardigheid in Cito Volgstelsysteem VO .....	8
3.	Methode.....	11
3.1	Participanten .....	11
3.2	Procedure .....	11
3.2.1	<i>Itemcategorisatie</i> .....	12
3.3	Analyses.....	13
4.	Resultaten.....	14
4.1	Descriptieve statistieken .....	14
4.2	Vaststellen van dimensies .....	14
5.	Conclusie .....	18
6.	Literatuurlijst .....	22
Bijlage 1	Overzicht van de besproken processen in sectie 1.1. ....	28
Bijlage 2	Subcategorieën bij ‘Begrijpen’ met voorbeelden en voorbeelditems. ....	29
Bijlage 3	Subcategorieën bij ‘Interpreteren’ met voorbeelden en voorbeelditems.....	30
Bijlage 4	De drie confirmatieve factor modellen in beeld. ....	31
Bijlage 5	Modelpassing statistieken behorend bij de drie confirmatieve factor modellen.....	32
Bijlage 6	Correlaties Model 2 en Model 3 behorende bij Toets 1 en Toets 3.....	33
Bijlage 7	Ranges van gestandaardiseerde regressiecoëfficiënten van Toets 1 en Toets 3, behorend bij de drie confirmatieve factor modellen.....	34

## 1. Inleiding

Vanuit het onderwijsveld is er een toenemende vraag naar diagnostische leestoetsen (Onderwijsraad, 2014; Onderwijsraad, 2011). Dit is onder andere het gevolg van enkele onderwijsveranderingen in Nederland, waaronder het invoeren van de referentieniveaus en de ambities uit het Actieplan Beter Presteren. Zowel in het referentiekader, als in het Actieplan Beter Presteren staat opbrengstgericht werken centraal, om de prestaties van leerlingen in het primair en het voortgezet onderwijs te verbeteren (Van Bijsterveldt-Vliegenthart, 2011). Ook de prestaties voor Nederlands leesvaardigheid, behorende bij het kernvak Nederlands, moeten verbeterd worden. Verscheidene onderzoeken van onder andere de Nederlandse Inspectie van het Onderwijs (2008), Van den Broek (2010) en het PPO-rapport van Cito (Cito, 2014) hebben namelijk aangetoond dat veel leerlingen in het primair onderwijs de vaardigheid in lezen, en specifiek in het begrijpen van teksten, onvoldoende beheersen. Hierdoor zullen zij in het voortgezet onderwijs ook onvoldoende in staat zijn teksten zelfstandig en met begrip te lezen.

Naar aanleiding van het Actieplan is er in 2013 een wetsvoorstel ingediend bij de Tweede Kamer voor de invoering van een verplicht leerlingvolgsysteem en een diagnostische tussentijdse toets aan het eind van de onderbouw van het voortgezet onderwijs. Door middel van een leerlingvolgsysteem zullen scholen, docenten, leerlingen en ouders meer inzicht krijgen in het niveau van de leerlingen. Met een diagnostische tussentijdse toets kunnen docenten vervolgens achterhalen welke aspecten leerlingen beheersen en welke aspecten nog niet beheerst worden. Op de aspecten die nog niet beheerst worden kan verder gedifferentieerde ondersteuning geboden worden.

Volgens onderzoek van het Kohnstamm Instituut zijn er nog weinig diagnostische toetsingsinstrumenten voorhanden die praktisch inzetbaar zijn op scholen (Ledoux, Meijer, van der Veen & Breetvelt, 2013), zo ook voor leesvaardigheid. Het ontwikkelen van dergelijke toetsingsinstrumenten blijkt echter nog niet zo gemakkelijk te zijn. Om te kunnen achterhalen welke aspecten van leesvaardigheid leerlingen beheersen en welke niet, is er namelijk informatie nodig over de verschillende vaardigheden die er mogelijk in leesvaardigheid te onderscheiden zijn. In een diagnostische toets worden deze verschillende vaardigheden, zogenoemde deelvaardigheden, vervolgens door middel van verschillende items bevraagd. De prestaties op deze verschillende vaardigheden kunnen de sterke en zwakke aspecten van een leerling binnen leesvaardigheid aantonen, waarop een leerling dan onderwijs op maat kan krijgen en extra kan oefenen op de zwakke deelvaardigheden.

In verscheidene diagnostische leestoetsen die in Nederland ontwikkeld zijn is het onderscheid in deelvaardigheden in leesvaardigheid terug te zien. In het kader van het Actieplan

Beter Presteren is in opdracht van het ministerie van OCW door het College voor Toetsen en Examens (CvTE) bijvoorbeeld een diagnostische tussentijdse toets (DTT) ontwikkeld. Deze DTT toetst leesvaardigheid aan de hand van opgaven die betrekking hebben op de volgende deelvaardigheden van het lezen: retorische (op doel en publiek gerichte), tekststructurele, linguïstische (met betrekking tot woord- en zinsniveau) en strategische vaardigheden. Voor deze deelvaardigheden geldt dat via een aantal opgaven wordt vastgesteld in hoeverre de leerling deze onder, op of boven niveau beheerst (CvTE, 2014). De toets DiaTekst, een gestandaardiseerd instrument om het tekstbegrip van leerlingen te meten en behorend bij een leerlingvolgsysteem voor de onderbouw van het voortgezet onderwijs, maakt een onderscheid in drie niveaus van tekstverwerking: begrip op micro- (woord- en zinsniveau), meso- (combineren van informatie uit verschillende zinnen en/of tekstgedeelten voornamelijk binnen de alinea) en macroniveau (tekst als geheel) (Meyer, 1984; Kintsch & Van Dijk, 1978). Op basis van de scores van leerlingen op deze niveaus wordt diagnostische informatie gegeven: er wordt een lezersprofiel opgesteld, waarmee de lezer gekarakteriseerd wordt als een bepaald type lezer (Hacquebord, 1997).

Toetsen zoals Diataal en de DTT hebben als uitgangspunt dat leesvaardigheid verschillende aspecten of deelvaardigheden kent, waarbij ook verondersteld wordt dat deze deelvaardigheden bij lezers in verschillende mate zijn ontwikkeld. Er is echter veel discussie of de deelvaardigheden van leesvaardigheid wel empirisch te onderscheiden zijn. En zo ja, welke deelvaardigheden wel en welke niet onderscheidbaar zijn (vgl. Alderson, 2000; Kalifa & Weir, 2009; Weir & Porter, 1994). Sommige onderzoekers menen aangetoond te hebben dat leesvaardigheid een multidimensioneel construct is (Davis, 1968; Spearritt, 1972; Thorndike, 1974, e.a.), terwijl anderen zich afvragen of een onderscheid tussen verschillende deelvaardigheden wel zinvol is (Lunzer, Waite & Dolan, 1979; Schedl, Gordon, Carey & Tang, 1996; Meijer & Van Gelderen, 2002; Van Steensel, Oostdam & Van Gelderen, 2014). Als leesvaardigheid opgevat moet worden als multidimensioneel construct, waarbij leesvaardigheid onderverdeeld kan worden in verschillende deelvaardigheden, kan er gemakkelijker diagnostische informatie gegeven worden over de onderscheiden deelvaardigheden. Indien leesvaardigheid daarentegen opgevat zou moeten worden als een unidimensioneel construct is het moeilijk om diagnostische informatie aan leerlingen te geven die specifieke leesmoeilijkheden benoemt. Immers, er wordt dan maar één vaardigheid gemeten: leesvaardigheid als een ondeelbaar geheel.

Naar aanleiding van deze discussie en de toenemende vraag naar diagnostische informatie, zal in dit onderzoek gekeken worden naar de mogelijkheden tot het verkrijgen van diagnostische informatie betreffende leesvaardigheid Nederlands in de onderbouw van het voortgezet onderwijs. De volgende onderzoeksvraag staat hierbij centraal:

*In hoeverre kan er diagnostische informatie verkregen worden over leesvaardigheid Nederlands in de onderbouw van het voortgezet onderwijs?*

## 2. Leesvaardigheid en diagnostische informatie

### 2.1 Theorieën leesvaardigheid: vaardigheden en processen

De mogelijkheid om diagnostische informatie te geven door middel van toetsing, zoals bij DiaTaal en de DTT, veronderstelt dat er afzonderlijke deelvaardigheden binnen leesvaardigheid te onderscheiden zijn. Deze deelvaardigheden zijn voortgevloeid uit vele verschillende onderzoeken naar en theorieën over vaardigheden en processen in leesvaardigheid. Hieronder zullen enkele theorieën besproken worden om meer inzicht te verkrijgen in de mogelijke deelvaardigheden. In de literatuur wordt er vaak geen duidelijk onderscheid gemaakt in vaardigheden, capaciteiten en strategieën (Alderson, 2000). Een belangrijk onderscheid dat wel veelal gemaakt wordt bij het benoemen van deelvaardigheden is tussen het *product* (wat de lezer weet en begrijpt na het lezen van een tekst) en de *processen* (de cognitieve acties van een lezer om het product te bereiken). Een focus op de onderliggende cognitieve processen behorend bij leesvaardigheid, zou nuttig kunnen zijn voor het beschrijven, verklaren en het bieden van gedifferentieerde ondersteuning aan zwakke lezers (Rapp, Van den Broek, McMaster, Kendeou & Espin, 2007). In dit onderzoek zal dan ook meer de nadruk op de processen gelegd worden, alleen zal veelal de term deelvaardigheden gebruikt.

Leesvaardigheid is een complexe vaardigheid, die afhangt van de uitvoering en het integreren van verschillende linguïstische en cognitieve processen (Kendeou, Van den Broek, Helder & Karlsson, 2014). Lezers proberen een geschreven tekst op verschillende levels te begrijpen: lexicaal, syntactisch, semantisch en pragmatisch. Tijdens het lezen is er daarnaast ook sprake van een complexe interactie tussen onder andere de inhoud van de tekst zelf, de doelen van de lezer en de kennis van de wereld (Pressley & Afflerbach, 1995; De Jong, 2011).

De algemene opvatting is dat lezen verdeeld kan worden in twee vaardigheden, namelijk de vaardigheid 'decoderen' (woordherkenning) en de vaardigheid 'begrip' (Gough, Ehri & Treiman, 1992). In lijn met deze opvatting wordt in het Nederlandse basisonderwijs ook een onderscheid gemaakt tussen technisch lezen en begrijpend lezen. In het voortgezet onderwijs in Nederland wordt hierin geen onderscheid meer gemaakt, maar wordt er 'leesvaardigheid Nederlands' gemeten. Nederlandse en Amerikaanse onderzoeken hebben aangetoond dat de relatie tussen technisch lezen en begrijpend lezen vanaf de leeftijd van zeven jaar tot in het voortgezet onderwijs sterk afneemt (Aarnoutse & Van Leeuwe, 1988; Kintsch & Kintsch, 2004; Thorndike, 1973-1974; Van Gelderen, 2012). Verondersteld wordt dan ook dat de vaardigheid in het decoderen in het voortgezet onderwijs al grotendeels verworven is. In dit onderzoek wordt daarom met leesvaardigheid voornamelijk de vaardigheid 'begrip' bedoeld.

Voor het verkrijgen van diagnostische informatie in leesvaardigheid wordt vaak een onderscheid gemaakt in de volgende basisvaardigheden, die allen belangrijk zijn voor succesvol

lezen: fonologisch bewustzijn, decoderen, vlotheid en woordenschatkennis (National Reading Panel [NRP], 2000). Echter is gebleken dat kinderen problemen kunnen hebben met leesvaardigheid, terwijl ze goed lijken te zijn in deze deelvaardigheden (Alvermann, Fitzgerald & Simpson, 2006; Englert & Thomas, 1987; Taylor & Williams, 1983). Er zou dus sprake moeten zijn van andere onderliggende vaardigheden om een succesvolle lezer te worden (Landi, 2010). Deze uitkomst heeft geleid tot een cognitieve benadering van leesvaardigheid waarbij er sprake is van een indeling in lagere-orde-processen en hogere-orde-processen. Met lagere-orde-processen wordt verwezen naar het 'vertalen' van het geschrevene in betekenisvolle eenheden, en daarmee ook naar de bovengenoemde basisvaardigheden. De hogere-orde-processen, de overige onderliggende vaardigheden, combineren deze eenheden tot een betekenisvolle en coherente mentale representatie (Kendeou, Rapp & Van den Broek, 2003) en omvatten het begrijpen en interpreteren van een tekst (NRP, 2000; Rapp et al., 2007). De lagere-orde-processen (basisvaardigheden) en hogere-orde-processen ontwikkelen zich mogelijk eerder simultaan dan sequentieel en een ontwikkeling in het ene proces kan zorgen voor ontwikkeling in het andere proces (o.a. Tannenbaum, Torgesen & Wagner, 2006).

In het interactieve model voor leesvaardigheid wordt leesvaardigheid omschreven als een continue wisselwerking tussen tekstgestuurde (bottom-up) en kennisgestuurde (top-down) verwerking om tot begrip van een tekst te komen. In dit interactieve model is tekstgestuurde verwerking gericht op de inhoud van de tekst. Er wordt hierbij gebruik gemaakt van woorden, passages en hun onderlinge betekenisrelaties om tot begrip te komen. Kennisgestuurde verwerking duidt op de inzet van de eigen kennis van de wereld en kennis van teksten, van taal en van contexten (bijv. Aarnoutse, 1988; Kamil, Mosenthal, Pearson & Barr, 2000; Reitsma & Walraven, 1991; Goldman & Rakestraw, 2000; Alderson, 2000).

Kintsch en Kintsch (2005) maken een onderscheid in termen van representatie. In deze theorie wordt een onderscheid gemaakt tussen drie niveaus van begripsprocessen die met elkaar interacteren: decodering, micro- en macrostructuur en het maken van een situatiemodel. Bij het decoderen van een tekst zijn alleen de perceptuele en conceptuele processen betrokken die leiden tot woord- en zinsbetekenissen in het hoofd van een lezer. De lezer vormt proposities, ook wel betekenseenheden genoemd, op basis van het decoderen van de tekst.

Het tweede niveau, de micro- en macrostructuur van een tekst, vormen samen de 'textbase'. De microstructuur bestaat uit het gehele netwerk van de individuele proposities die bij het decoderen ontstaan zijn, en hun onderlinge syntactische relaties. Hierbij gaat het om begrip op het lokale niveau van de tekst. De macrostructuur omvat de globale structuur van een tekst: de microstructuur met alle kleine betekenseenheden wordt hierbij teruggebracht tot een aantal overkoepelende betekenseenheden. De micro- en macrostructuur worden voornamelijk bepaald door wat de auteur heeft geschreven en worden dus bepaald door de tekst zelf.



Het derde niveau bestaat uit het maken van een situatiemodel, een mentaal model van de situatie als omschreven in de tekst. Hierbij spelen inferenties, gevolgtrekkingen over wat bedoeld wordt die verder gaan dan de informatie in de tekst, een belangrijke rol (De Jong, 2011). Om een tekst daadwerkelijk goed te kunnen begrijpen moet een lezer de informatie in een tekst integreren met eerdere informatie en ervaringen. Het gebruik van informatie buiten de tekst betekent ook dat het situatiemodel niet alleen uit proposities bestaat, maar ook andere componenten bevat, zoals bijvoorbeeld kennis, emoties en herinneringen. Het model omvat dus zowel tekstuele informatie, als achtergrondkennis (Kintsch, 1988; Van den Broek, Kendeou, Kremer, Lynch, Butler & White, 2005). De constructie van het situatiemodel is dan ook een dynamisch proces en ontwikkelt zich tijdens het lezen van een tekst. Het model wordt steeds aangepast als er nieuwe informatie wordt gelezen (De Jong, 2011), met als gevolg dat het situatiemodel veel minder voorspelbaar is dan de textbase.

Bovenstaande theorieën laten zien dat leesvaardigheid een complex proces is en dat er verschillen zijn tussen de diverse theorieën, maar dat er ook overeenkomsten zijn. Deze overeenkomsten zijn in Bijlage 1 ook te zien, waar de besproken benaderingen naast elkaar zijn gezet. In de tabel is er overlap te zien tussen de verschillende processen en de processen binnen de theorieën vertonen een wisselwerking. Leesvaardigheid kan dan ook omschreven worden als een complexe, dynamische vaardigheid, waarbij door middel van verschillende processen op verschillende niveaus in wisselwerking met elkaar informatie wordt verwerkt om tot begrip van een tekst te komen.

## **2.2 Onderzoek naar verschillende dimensies in leesvaardigheid**

Uit de vorige sectie is gebleken dat leesvaardigheid opgebouwd is uit een veelheid aan processen, die mogelijk corresponderen met zogenoemde deelvaardigheden. Het onderscheiden van verschillende deelvaardigheden in leesvaardigheid, zowel producten als processen, is echter al lange tijd reden tot onderzoek. Er bestaat veel discussie over de mogelijkheid om verschillende deelvaardigheden binnen leesvaardigheid empirisch te onderscheiden en te classificeren. Het is onduidelijk (a) welke deelvaardigheden bestaan en (b) hoe ze gecategoriseerd moeten worden (Alderson, 2000). De vele onderzoeken naar mogelijke deelvaardigheden geven ook verschillende uitkomsten, waarbij er over het algemeen twee uitgangsposities zijn:

- *leesvaardigheid bestaat uit meerdere (deel)vaardigheden en is dus multidimensioneel;*
- *leesvaardigheid bestaat uit één vaardigheid en is daarmee unidimensioneel.*

Een voorbeeld van een onderzoek waarbij meerdere deelvaardigheden gevonden zijn, is het onderzoek van Davis (1968). In dit onderzoek wordt aangetoond dat aan leesvaardigheid ontleed kan worden in vijf deelvaardigheden: woordbegrip, het maken van inferenties, tekststructuur, herkennen

van doel, toon en houding van de schrijver en het beantwoorden van expliciet of impliciet gestelde vragen. In onderzoek van Spearritt (1972) worden de eerste vier vaardigheden van Davis ondersteund, maar is er wel maar één overduidelijk unieke vaardigheid terug gevonden: woordbegrip. Bij de overige drie vaardigheden zijn hoge correlaties gevonden, oftewel met alle drie vaardigheden is waarschijnlijk hetzelfde gemeten. Thorndike (1974) heeft vervolgens ook onderzoek naar de verschillende vaardigheden gedaan en heeft twee factoren gevonden: algemeen begrijpend lezen en woordbetekenis.

Onderzoeken waarbij geen evidentie is gevonden voor het bestaan van verschillende deelvaardigheden binnen leesvaardigheid, zijn onder andere het onderzoek van Lunzer, Waite & Dolan (1979) en het onderzoek van Schedl, Gordon, Carey en Tang (1996). Schedl et al. (1996) veronderstelden twee deelvaardigheden te meten – ‘reasoning’ en ‘overig’ – maar hebben dit niet terug gevonden in hun analyses. Volgens beide onderzoeken bestaat lezen dan ook uit één globale geïntegreerde vaardigheid. Meijer en Van Gelderen (2002) voerden analyses uit op data van het Nederlandse PISA onderzoek en vonden daarbij hoge intercorrelaties (rond .95) tussen de drie veronderstelde deelvaardigheden: Zoeken & Vinden, Integreren & Interpreteren en Reflecteren & Evalueren. De deelvaardigheden lijken dus hetzelfde te meten, oftewel leesvaardigheid. In een onderzoek van Van Steensel, Oostdam en Van Gelderen (2014) werden soortgelijke deelvaardigheden verondersteld (Vinden, Interpreteren en Reflecteren), maar ook zij vonden geen bewijs voor multidimensionaliteit.

### *2.2.1 Deelvaardigheden, processen en dimensies*

Bovenstaande secties tonen aan dat er in de bestaande theorieën verschillende processen beschreven worden (2.1), en dat deze processen in de genoemde onderzoeken op verschillende manieren vertaald zijn naar deelvaardigheden (2.2). De deelvaardigheden die in verschillende onderzoeken gevonden zijn, zowel producten als processen, zijn steeds net anders gedefinieerd en geclusterd, maar zijn grofweg volgens dezelfde denklijnen gekozen. De verschillen in processen tonen aan dat het proces om lezend tot begrip te komen erg complex is, waar de verschillen in deelvaardigheden doen vermoeden dat het moeilijk is om duidelijk omschreven deelvaardigheden te kunnen onderscheiden en empirisch aan te tonen. Deze verschillen hebben tot gevolg dat er tot op heden nog steeds geen consensus is over het toetsen van uni- of multidimensionaliteit binnen leesvaardigheid. De literatuur en de onderzoeken lijken echter vaker unidimensionaliteit te ondersteunen.

### **2.3 Het toetsen van leesvaardigheid**

Om te achterhalen of een lezer een tekst begrijpt, kunnen leesvaardigheidstoetsen ingezet worden. Het toetsen van leesvaardigheid is echter, net als leesvaardigheid zelf, complex. Er zijn vele factoren waar rekening mee gehouden moet worden bij het maken van een toets, zoals bijvoorbeeld met de lezer zelf, de tekst, het item en de interactie met de tekst. Daarnaast moet eerst ook duidelijk zijn wat er nu verstaan wordt onder het begrijpen van teksten, zodat met de toets ook daadwerkelijk het begrip van de tekst gemeten wordt. Voor het ontwikkelen van een toets is daarom altijd een construct nodig: het moet duidelijk zijn wat er wordt gemeten met een toets. Constructen zijn niet direct waarneembare (latente) kwaliteiten (Cronbach & Meehl, 1955; Cohen, Manion & Morrison, 2007), zoals in dit geval bijvoorbeeld de vaardigheid tekstbegrip. Een dergelijk construct wordt aan de hand van theoretisch gegronde kennis geoperationaliseerd. Gestandaardiseerde leesvaardigheidstoetsen kunnen het construct op verschillende manieren definiëren. Zo kan een definitie van het construct leesvaardigheid bijvoorbeeld gebaseerd zijn op veronderstelde cognitieve processen (Grabe & Jiang, 2014) of op de basisvaardigheden zoals genoemd in sectie 2.1 (NRP, 2000).

Aan de hand van de beschrijvingen van enkele gestandaardiseerde leesvaardigheidstoetsen worden de veronderstelde dimensies duidelijk zichtbaar. In de inleiding werden de toetsen Diatekst en de DTT beide al even genoemd, maar ook in de beschrijvingen van de leesvaardigheidstoets in het PISA-onderzoek, de PIRLS en in die van de DARC-test worden meerdere dimensies in leesvaardigheid onderscheiden. In het PISA-onderzoek worden er drie mentale strategieën onderscheiden voor het begrijpen van een tekst: Zoeken & Vinden, Integreren & Interpreteren en Reflecteren & Evalueren (OECD, 2013; Cito, 2010). De PIRLS maakt een onderscheid in de volgende begripsprocessen: focus en terug vinden van expliciete informatie in een tekst, het maken van eenvoudige inferenties, het interpreteren en integreren van ideeën en informatie, het onderzoeken en beoordelen van de inhoud, het taalgebruik en de tekstuele elementen in een tekst (Mullis, Martin, Kennedy, Trong & Sainsbury, 2009). De DARC-test is ontwikkeld op basis van onderzoeken van Potts en Peterson (1985) en Hannon en Daneman (2001). Hierbij worden vier deelprocessen verondersteld voor succesvol tekstbegrip: tekstgeheugen, het maken van inferenties, toegang tot kennis en kennisintegratie (Francis, Snow, August, Carlson, Miller & Iglesias, 2006).

Voor het maken van een toets moet een toetsmatrijs gemaakt worden, waarmee aangegeven wordt hoe de opgaven in een toets verdeeld zijn over de te onderscheiden dimensies. Afhankelijk van de leerdoelen, wordt bepaald op welk niveau (welk denkproces) de leerdoelen getoetst moeten worden, welke toetsvorm zich daarvoor het beste leent en hoeveel vragen er over de verschillende leerdoelen gesteld gaan worden. De items van leesvaardigheidstoetsen zijn vaak ingedeeld op basis van de locatie van de te verkrijgen informatie (lokaal/globaal tekstniveau) en de mate waarin een leerling informatie expliciet of impliciet uit een tekst kan halen, resulterend in

itemcategorieën zoals het achterhalen van de hoofdgedachte, het vinden van de juiste informatie en het maken van gevolgtrekkingen (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, 2009; Davis, 1968; Goldman & Durán, 1988; OECD, 2013; Rosenshine, 1980; Rouet, Vidal-Abarca, Erboul & Millogo, 2001; Song, 2008; Vidal-Abarca, Gilabert & Rouet, 1998).

De hierboven genoemde toetsen laten zien dat een construct op verschillende manieren bepaald kan worden, die allen gebaseerd zijn op verschillende theorieën. Een construct van een toets kan dan wel bestaan uit meerdere dimensies, maar is het ook mogelijk om deze dimensies door middel van toetsing te onderscheiden? Of wordt er bij leesvaardigheidstoetsen maar één construct getoetst, namelijk leesvaardigheid?

## **2.4 Nederlands leesvaardigheid in Cito Volgsysteem VO**

In dit onderzoek zal de dimensionaliteit van het construct behorende bij de toets Nederlands leesvaardigheid in het Cito Volgsysteem Voortgezet Onderwijs (VVO) onderzocht worden. De toetsen zijn beschikbaar voor de volgende niveaus: VMBO-BB, VMBO-GT en HAVO/VWO. De leesvaardigheidstoetsen Nederlands bestaan allen uit twee taken en in totaal 44 tot 54 meerkeuze items, passend bij de verschillende teksten. De toetsen brengen op vier verschillende momenten, gedurende de onderbouw van het voortgezet onderwijs, het niveau en de voortgang van leerlingen in beeld. De toetsen van Toets 0 tot en met Toets 3 hebben vier functies: determinatie, progressiebepaling, diagnostiek en kwaliteitsbewaking. Met de diagnostiek wordt hier bedoeld het in kaart brengen van de eventuele zwakke en sterke kernvaardigheden (Rekenen/Wiskunde, Nederlands en Engels) van een leerling en de onderlinge invloeden (Cito, 2015).

In Nederland zijn de kerndoelen en het 'Referentiekader doorlopende leerlijnen taal en rekenen' (Meijerink, Letschert, Rijlaarsdam, Van den Bergh & Van Streun, 2009) in principe het uitgangspunt voor toetsontwikkeling, ook voor het Cito Volgsysteem VO. De prestatiecriteria behorende bij het referentiekader zijn ook gebaseerd op constructen op basis van geaccepteerde theorieën. Hierbij wordt er voor leesvaardigheid voor enkele deelvaardigheden gespecificeerd wat een leerling moet weten en kunnen: Begrijpen, Interpreteren, Evalueren/Reflecteren, Samenvatten en Opzoeken. Hier wordt dus verondersteld dat leesvaardigheid uit tenminste vijf aspecten bestaat.

Voor de toetsconstructie hanteert het VVO voor leesvaardigheid een categorie-indeling waarbij twee (deel)vaardigheden, vergelijkbaar met die van het referentiekader, worden onderscheiden: het begrijpen van geschreven teksten en het interpreteren van geschreven teksten. De vaardigheid Begrijpen heeft betrekking op de verwerking van informatie die expliciet in de tekst vermeld staat. Bij de vaardigheid Interpreteren gaat het erom dat de lezer de informatie uit de tekst verbindt aan zijn eigen kennis, waaronder zijn kennis van de wereld en kennis van taaltekstkenmerken. Evalueren/Reflecteren en Samenvatten en Opzoeken worden in de categorie-

indeling niet meegenomen. De vaardigheid Evalueren/Reflecteren is minder objectief en het kerndoel Reflecteren wordt alleen opgevat om het taalleerproces te bevorderen ((Expertgroep Doorlopende leerlijnen Taal en Rekenen, 2008). De vaardigheden Samenvatten en Opzoeken behoren niet tot de toetsdoelen in het VVO.

Om ervoor te zorgen dat alle vaardigheden die getoetst moeten worden ook daadwerkelijk in de toets voorkomen, is er een categorie-indeling gemaakt op basis van de twee deelvaardigheden. Voor beide vaardigheden is een indeling gemaakt in vier subcategorieën, in totaal acht. Deze subcategorieën staan weergegeven in Tabel 1. De verschillen tussen de categorieën zijn vetgedrukt weergegeven. In Bijlage 2 en Bijlage 3 zijn voorbeelden en voorbeeldvragen te zien van de subcategorieën behorend bij respectievelijk Begrijpen en Interpreteren.

**Tabel 1: Categorie-indeling deelvaardigheden Leesvaardigheid bij constructieproces VVO.**

Cat.	Begrijpen	Cat.	Interpreteren
<b>BEGR1</b>	Opgaven die vragen naar de betekenis van een woord, woordgroep of zin(nen) die <b>expliciet</b> in de tekst vermeld wordt.	<b>INTER1</b>	Opgaven die vragen naar het <b>afleiden</b> van de betekenis van een woord, woordgroep of zin(nen).
<b>BEGR2</b>	Opgaven die vragen naar specifieke inhoudselementen die <b>expliciet</b> in de tekst aan de orde gesteld worden.	<b>INTER2</b>	Opgaven die vragen naar het <b>afleiden</b> van informatie uit de tekst op <b>lokaal niveau</b> . De leerling moet zijn voorkennis inzetten naast de informatie die de tekst geeft.
<b>BEGR3</b>	Opgaven die vragen naar <b>eenvoudige expliciete verbanden op lokaal niveau</b> . Verbanden kunnen worden gelegd op basis van inhoudelijke en/of structurele elementen zoals signaalwoorden.	<b>INTER3</b>	Opgaven die vragen naar het <b>afleiden</b> van informatie op het <b>globale niveau</b> van de tekst. Verbanden kunnen worden gelegd op basis van inhoudelijke en/of structurele elementen.
<b>BEGR4</b>	Opgaven die vragen naar <b>complexe expliciete verbanden over grotere tekstdelen</b> heen. Verbanden kunnen worden gelegd op basis van inhoudelijke en structurele elementen.	<b>INTER4</b>	Opgaven die vragen naar het <b>taalgebruik en schrijfstijl van de schrijver</b> en waar <b>verbanden moeten worden gelegd tussen tekstuele informatie en kennis van het taalsysteem</b> .

## 2.5 Onderzoeksvragen

In de voorafgaande secties is geprobeerd een overzicht te geven van de verschillende leesvaardigheidstheorieën, de mogelijke constructen bij leesvaardigheidstoetsen en de verschillende onderzoeken die gedaan zijn naar de uni- en multidimensionaliteit van leesvaardigheid. Uit dit overzicht is ook min of meer te herleiden waar het construct van Cito VVO vandaan komt. Geconcludeerd kan worden dat leesvaardigheid meerder processen omvat, die terug te vinden zijn in de verschillende toetsconstructen. Hiermee wordt gesuggereerd dat het mogelijk is om diagnostische informatie te geven in leesvaardigheid door middel van toetsing. Er is echter nog steeds geen consensus over het toetsen van leesvaardigheid als uni- of multidimensioneel construct, oftewel ook nog geen eenduidig antwoord over de mogelijkheid tot het geven van diagnostische informatie in leesvaardigheid. Om bij te kunnen dragen aan een eenduidiger antwoord, zijn aan de hand van dit overzicht en de omschrijving van het construct dat gebruikt wordt bij de toetsen van het VVO van Cito (sectie 2.4) de volgende onderzoeksvragen opgesteld:

### ***In hoeverre is het mogelijk om diagnostische informatie te verkrijgen betreffende leesvaardigheid Nederlands in de onderbouw van het voortgezet onderwijs?***

*In hoeverre zijn de veronderstelde dimensies te onderscheiden in de toetsen leesvaardigheid Nederlands van het VVO?*

Om bovenstaande vragen te kunnen beantwoorden, zullen factoranalyses uitgevoerd worden op de testresultaten van leerlingen, waarbij het onderdeel leesvaardigheid Nederlands van Cito VVO (tweede generatie) is afgenomen. Op basis van de voorgaande literatuur en bevindingen is de verwachting voor dit onderzoek dat de veronderstelde dimensies in de toetsen leesvaardigheid Nederlands van het VVO niet te onderscheiden zijn. Dit zal een gevolg zijn van de complexiteit van leesvaardigheid, waarbij meerdere processen met elkaar interacteren. Alderson (2000) geeft hiervoor een mogelijke verklaring: vanaf het moment dat kinderen het decoderen beheersen en de tekst begrijpen, wordt het gehele leesproces zodanig geïntegreerd, dat de mogelijk onderliggende deelvaardigheden waaruit leesvaardigheid bestaat, niet meer van elkaar te onderscheiden zijn. De verschillende veronderstelde dimensies zullen daardoor moeilijk te onderscheiden zijn, met als gevolg dat diagnostische informatie van leesvaardigheid ook moeilijk te verkrijgen is. Er kan immers niets gezegd worden over de afzonderlijke dimensies.

### 3. Methode

#### 3.1 Participanten

In Tabel 2 is te zien van hoeveel leerlingen per niveau de data zijn meegenomen in de analyses voor dit onderzoek. Deze leerlingen hebben in leerjaar 2011-2012 deelgenomen aan Toets 1 en Toets 3 van Cito VVO. Ze bevinden zich verspreid over heel Nederland en zaten op het moment van afname respectievelijk aan het eind van de eerste of de derde klas van het voortgezet onderwijs, verschillende niveaus.

In Tabel 3 staan gegevens over de betrouwbaarheid van zowel Toets 1 als Toets 3 (Cito, 2015). De coëfficiënt alpha van de toetsen ligt tussen 0.71 en 0.80, waarmee kan worden aangenomen dat de toets voldoende betrouwbaar is (COTAN, 2010).

Tabel 2: Aantal leerlingen per niveau en per toets.

Niveau	Toets 1	Toets 3
VMBO BB	8614	5936
VMBO KB/GT	24193	19878
HAVO/VWO	20815	16381

Tabel 3: Gegevens betrouwbaarheid Toets 1 en Toets 3 (Cito, 2015).

Toets + niveau	Coëfficiënt alpha	Standaard meetfout	Gemiddelde score
<b>Toets 1</b>			
VMBO BB	0,80	2,94	28,5
VMBO GT	0,78	3,00	34,4
HAVO/VWO	0,71	3,05	34,8
<b>Toets 3</b>			
VMBO BB	0,77	3,11	25,1
VMBO GT	0,74	3,27	30,3
HAVO/VWO	0,73	3,16	36,5

#### 3.2 Procedure

De papieren toetsen zijn klassikaal afgenomen en elke leerling heeft een uniek antwoordblad. De leerlingen hebben voorafgaand aan elke toets een korte klassikale instructie gekregen, waarna ze individueel de toets hebben gemaakt. Het afnemen van de leesvaardigheid toets, bestaande uit twee taken, kostte ongeveer twee keer één lesuur. De toetsen mochten op verschillende momenten worden afgenomen: tijdens de lessen van de betreffende vakken, tijdens blokuren in bijvoorbeeld een toetsweek of tijdens mentoruren. Wenselijk was wel om de taken twee keer onder dezelfde conditie af te nemen. Daarnaast mochten de toetsen in willekeurige volgorde worden afgenomen. Na afloop van de afname zijn de antwoordbladen teruggestuurd naar Cito en automatisch gescoord.

Vervolgens zijn de antwoordbladen klaargemaakt voor onderzoek en zijn ze gebruikt voor de analyses.

### 3.2.1 Itemcategorisatie

Voorafgaand aan het uitvoeren van de analyses, zijn alle 294 items van Toets 1 en Toets 3, alle niveaus, gecategoriseerd op basis van de veronderstelde deelvaardigheden in Tabel 1: per item is beoordeeld welke deelvaardigheid er met het item getoetst wordt. In Bijlage 2 en 3 zijn per categorie voorbeelditems gegeven.

Het indelen van items naar veronderstelde deelvaardigheden is niet altijd zonder problemen. Verschillende beoordelaars kunnen makkelijk tot verschillende oordelen komen. Aan de hand van Bijlage 3 is bijvoorbeeld te zien dat items waarbij er naar de hoofdgedachte gevraagd wordt, in te delen zijn in zowel categorie INTER2 als INTER3. Het verschil tussen deze twee categorieën is dat er bij items met de categorie INTER2 gevraagd wordt naar het afleiden van de hoofdgedachte op lokaal niveau en bij INTER3 naar het afleiden van de hoofdgedachte op globaal niveau. Maar in het geval van een korte tekst, is er dan sprake van een lokaal of een globaal niveau? Om consequent te zijn bij het categoriseren, is de categorisatie van de items vergeleken met de categorisatie van een tweede beoordelaar. De items waarover geen overeenstemming was zijn na overleg ingedeeld in een categorie. Hierdoor zijn dezelfde denklijnen ontstaan bij de categorisatie.

De categorisatie heeft in de volgende itemsets geresulteerd, te zien in Tabel 4 en Tabel 5. Per toets en per niveau is te zien hoe vaak de veronderstelde deelvaardigheden voorkomen, zowel voor de categorisatie in Begrijpen en Interpreteren, als voor de categorisatie in de acht subcategorieën. Daarnaast staat ook het totale aantal items per toets en per niveau weergegeven. Categorieën waarbij minder dan 2 items gevonden zijn, zijn weggelaten uit de analyses. In onderstaande tabellen zijn deze cellen lichtgrijs gemarkeerd en vetgedrukt weergegeven.

**Tabel 4: Aantal items per factor voor Toets 1, alle opleidingsniveaus.**

Toets 1	Begrijpen	B1	B2	B3	B4	Interpreteren	I1	I2	I3	I4	Totaal items
<b>BB</b>	16	3	4	8	1	28	8	7	13	0	44
<b>KB/GT</b>	24	4	4	15	1	26	7	8	11	0	50
<b>HAVO/VWO</b>	24	2	5	13	4	26	6	7	12	1	50

**Tabel 5: Aantal items per factor voor Toets 3, alle opleidingsniveaus.**

Toets 3	Begrijpen	B1	B2	B3	B4	Interpreteren	I1	I2	I3	I4	Totaal items
<b>BB</b>	20	0	3	11	6	25	0	9	16	0	44
<b>KB/GT</b>	23	1	7	13	2	28	1	15	11	1	50
<b>HAVO/VWO</b>	26	1	5	13	7	28	3	8	15	2	50



### 3.3 Analyses

Om te achterhalen of de veronderstelde deelvaardigheden betreffende leesvaardigheid vastgesteld kunnen worden, zijn er confirmatieve factoranalyses uitgevoerd met behulp van Mplus. Hiervoor zijn drie modellen met elkaar vergeleken. Een weergave van de modellen is te zien in Bijlage 4. Model 1 is een eendimensionaal factormodel waarbij verondersteld wordt dat met de toetsen van VVO één vaardigheid wordt gemeten, namelijk leesvaardigheid. In Model 2 wordt een tweedimensionaal factormodel (met gecorreleerde factoren) gespecificeerd waarbij verondersteld wordt dat de test twee dimensies meet, namelijk Begrijpen en Interpretieren. In Model 3 wordt een achtdimensionaal factormodel (met gecorreleerde factoren) gespecificeerd waarbij verondersteld wordt dat de test acht dimensies meet, namelijk BEGR1, BEGR2, BEGR3, BEGR4 en INTER1, INTER2, INTER3 en INTER4. De inhoud van deze veronderstelde deelvaardigheden is eerder al genoemd en te zien in Tabel 1.

Voor het beoordelen van de passing van de drie modellen zijn verschillende indicatoren gebruikt (Geiser, 2013; Bentler, 1990; Hu & Bentler, 1999): de Chi-kwadraat test, die niet significant behoort te zijn; de comparative fit index (CFI) en de Tucker-Lewis Index (TLI), die beide boven 0.90 behoren te zijn voor een redelijk goede passing en boven 0.95 voor een goede passing; de root mean square error of approximation (RMSEA), die lager dan 0.05 behoort te zijn.

Naast het beoordelen van de passingsmaten, is ook gekeken naar de gestandaardiseerde regressiecoëfficiënten en de correlaties, om de mate van samenhang tussen de factoren te bepalen.

## 4. Resultaten

### 4.1 Descriptieve statistieken

In Tabel 6 zijn de descriptieve statistieken van de scores op Toets 1 en Toets 3 weergegeven. De gegevens betreffen alleen de scores van de leerlingen die betrokken zijn geweest bij dit onderzoek, en omvat dus niet de gehele populatie die landelijk deel heeft genomen aan Toets 1 en Toets 3.

Tabel 6: Scorereange, gemiddelde, standaarddeviatie en kurtosis van de scores op Toets 1 en Toets 3.

Toets	Niveau	Range	Gemiddelde	SD	Kurtosis (SE)
1	BB	6 – 44	28.49	6.49	-.20 (.05)
	KB/GT	4 – 50	34.44	6.38	.52 (.03)
	HAVO/VWO	7 – 50	34.84	5.57	.58 (.03)
3	BB	5 – 43	25.14	6.42	-.35 (.06)
	KB/GT	5 – 49	30.45	6.45	.00 (.04)
	HAVO/VWO	10 – 53	36.54	5.97	.26 (.04)

### 4.2 Vaststellen van dimensies

De modelpassingen die bij de confirmatieve factor modellen (Bijlage 4) horen, zijn te vinden in Bijlage 5. Hier staan in Tabel 7, 8 en 9 de modelpassingen per Toets en per niveau weergegeven. De strikte Chi-kwadraat test behorende bij Model 1, 2 en 3, is significant voor zowel Toets 1 als Toets 3, als voor alle opleidingsniveaus. Dit zou betekenen dat de modellen niet goed bij de geobserveerde data passen. De significantie van de Chi-kwadraat zou ook een gevolg kunnen zijn van het grote aantal participanten in dit onderzoek. Bij een groot aantal is de Chi-kwadraat snel significant (Tabachnick & Fidell, 2013). Hoewel de modellen strikt genomen niet heel goed bij de geobserveerde data passen, kunnen de passingsverschillen tussen de modellen wel geïnterpreteerd worden. Deze passingsverschillen zijn weergegeven in Tabel 7.

Tabel 7: Vergelijking modelpassingen Toets 1 en Toets 3.

Niveau		BB		KB/GT		HAVO/VWO	
Model		1 vs. 2	2 vs. 3	1 vs. 2	2 vs. 3	1 vs. 2	2 vs. 3
Toets 1	x2	35.96	281.59	3.14	482.15	12.70	232.09
	df	1	48	1	62	1	68
	p(X2)	<.001	<.001	<.001	<.001	<.001	<.001
Toets 3	x2	86.32	10.11	28.55	1638.42	9.55	405.50
	df	1	9	1	153	1	72
	p(X2)	<.001	<.001	<.001	<.001	<.001	<.001

Op grond van de passingsverschillen moet geconcludeerd worden dat het unidimensionale factor model significant minder goed bij de data past dan het twee-, of achtdimensionale factor model. Dit geldt voor zowel Toets 1 als Toets 3 en alle opleidingsniveaus. Naar aanleiding van deze gegevens zou een keuze gemaakt moeten worden voor Model 3, waarbij in leesvaardigheid een onderscheid wordt gemaakt in acht deelvaardigheden.

Er zijn echter twee bezwaren waarom deze conclusie wellicht overhaast is. Ten eerste zijn er naast de Chi-kwadraat andere passingsmaten waarmee modellen beoordeeld kunnen worden, zoals (ook in sectie 3.3 genoemd): CFI, TLI, RMSEA en p(RMSEA). De passingsmaten in Bijlage 5 laten zien dat de waardes van CFI en TLI allen boven 0.90 zijn, bij HAVO/VWO Toets 1 zelfs boven 0.95. De RMSEA waardes zijn lager dan 0.05 en de bijbehorende p-waarde is  $>0.05$ , oftewel de RMSEA waardes zijn significant. Op basis van deze gegevens kan daarom gesproken worden van drie aannemelijke modellen (Schermelleh-Engel, Moosbrugger & Müller, 2003).

Ook voor deze passingsmaten kan gekeken worden naar de passingsverschillen. Deze zijn weergegeven in Tabel 8. De CFI geeft weer of het beoogde model beter bij de geobserveerde data past dan het eenvoudigere model (Geiser, 2013). Hoe hoger deze maat is, hoe beter het model bij de geobserveerde data past. In Tabel 8 is te zien dat er nauwelijks verschil is tussen de CFI van Model 1 en Model 2. Ook de vergelijking tussen Model 2 en Model 3 laat weinig vooruitgang zien.

De TLI vergelijkt de passing van het beoogde model met de passing van het eenvoudigere model (Geiser, 2013). Ook bij deze maat geldt: hoe hoger de maat, hoe beter het model bij de geobserveerde data past. Tabel 9 laat ook voor TLI weinig vooruitgang zien tussen de modellen.

De RMSEA waarde geeft een benadering van het model. Een goed model zou een RMSEA waarde lager dan 0.05 moeten hebben, oftewel: hoe lager de maat, hoe beter het model bij de geobserveerde data past. Ook voor de RMSEA is in Tabel 9 te zien dat er nauwelijks verschillen zijn tussen de modellen. Ook de p(RMSEA), die gebruikt kan worden om de hypothese te toetsen dat de RMSEA waarde  $\leq 0.05$  is, geeft geen verschillen weer tussen de modellen.

**Tabel 8: Vergelijking modelpassing Toets 1 en Toets 3.**

Niveau	Toets	BB		KB/GT		HAVO/VWO	
Model		1 vs. 2	2 vs. 3	1 vs. 2	2 vs. 3	1 vs. 2	2 vs. 3
CFI	1	0	0.003	0	-0.002	0	0.001
	3	0.003	0	0	0.011	0	0.006
TLI	1	0.001	0.002	0	-0.003	0	0.001
	3	0.003	-0.001	0.001	0.011	0	0.005
RMSEA	1	0	0	0	0	0.01	0
	3	0	0	0	-0.001	0	-0.001
p(RMSEA)	1	$>0.999$	$>0.999$	$>0.999$	$>0.999$	$>0.999$	$>0.999$
	3	$>0.999$	$>0.999$	$>0.999$	$>0.999$	$>0.999$	$>0.999$

Ten tweede is het verstandig om de parameterschattingen in de overwegingen te betrekken. Voor de geanalyseerde modellen en voor de onderzoeksvraag, zijn de correlaties tussen de acht factoren in het best passende model, Model 3, het meest relevant. In Bijlage 6 wordt een compleet overzicht van de correlaties tussen de factoren in beide modellen per opleiding weergegeven. Voor Model 2 is hierin te zien dat de correlaties tussen de factoren ‘Begrijpen’ en ‘Interpreteren’ gemiddeld voor alle niveaus hoog zijn,  $r = 0.97$  voor Toets 1 en  $r = 0.95$  voor Toets 3. De bijbehorende p-waardes zijn  $p < .001$ , oftewel er is sprake van een significante correlatie.

In onderstaande tabel, Tabel 10, worden de correlaties samengevat en staan de gemiddelde correlaties (over de opleidingen) tussen de acht factoren voor Toets 1 (onder de diagonaal) en voor Toets 3 (boven de diagonaal) weergegeven. De cellen met daarin een vraagteken (in Bijlage 6 de lege cellen) zijn combinaties van items die niet in de toets voorkomen. Deze combinaties komen ook overeen met de gegevens in Tabel 4 en 5, die het aantal items per factor voor beide toetsen weergeven. Voor deze combinaties zijn dus geen correlaties tussen de factoren gevonden.

**Tabel 9: Gemiddelde correlaties (over opleidingen) tussen de acht factoren voor Toets 1 en Toets 3.**

	BEGR1	BEGR2	BEGR3	BEGR4	INTER1	INTER2	INTER3	INTER4
BEGR1		?	?	?	?	?	?	?
BEGR2	0.97		0.89	0.88	0.94	0.86	0.9	0.55
BEGR3	0.9	0.97		0.95	1.08	0.97	0.9	0.49
BEGR4	0.84	0.83	0.94		1.13	0.9	0.87	0.49
INTER1	1.12	1	?	?		0.99	1.18	0.56
INTER2	0.97	0.91	0.94	0.88	0.97		0.96	0.42
INTER3	0.98	0.96	0.94	?	0.96	0.96		0.55
INTER4	?	?	?	?	?	?	?	

Uit Tabel 10 blijkt dat de correlatie bij Toets 1 tussen BEGR1 en BEGR2 0.97 is, oftewel erg hoog. Ook de overige correlaties zijn hoog, zowel bij Toets 1 als bij Toets 3. Enkele correlaties laten zelfs een opvallende waarde  $r = >0.999$  zien. Onderzoek van Deegan (1978) toont aan dat zowel gestandaardiseerde regressiecoëfficiënten als correlaties met een waarde groter dan 1 toegestaan zijn. Dit zou kunnen wijzen op een hoge mate van multicollineariteit in de data, oftewel een hoge mate van samenhang tussen de factoren. Opvallend zijn de correlaties tussen INTER4 en de overige factoren voor Toets 3, deze correlaties zijn vergeleken met de overige correlaties vrij laag. Dit is mogelijk een gevolg van het kleine aantal items dat ingedeeld is bij de factor INTER4, namelijk 2, het minimale aantal items dat nodig is voor een analyse.

In Bijlage 7, Tabel 8 en 9, staan de ranges van de gestandaardiseerde regressiecoëfficiënten weergegeven behorend bij Model 2 en Model 3. Deze waarden voorspellen in hoeverre de variabelen

BEGR1, BEGR2, BEGR3, BEGR4 en INTER1, INTER2, INTER3, INTER4 de scores op de items voorspellen. In Tabel 5 zijn enkele lege cellen zichtbaar, een gevolg van te weinig items per factor. De lege cellen komen overeen met de gegevens in Tabel 4 en 5, die het aantal items per factor voor beide toetsen weergeven: bijvoorbeeld Toets 1, niveau BB, omvat maar één item BEGR4, waardoor er geen informatie over te verkrijgen is. De overige waardes in de tabel zijn overwegend positief, waarmee aangetoond wordt dat de items meten wat ze behoren te meten, oftewel de veronderstelde subcategorieën. De bijbehorende p-waardes zijn ook overwegend  $<.001$ , oftewel ook hier is sprake van een significant verband. Bij HAVO/VWO Toets 1 zijn echter negatieve waarden in de drie modellen te vinden. De significantie is hier ook afwijkend:  $p>.05$ . Mogelijk heeft dit te maken met een sleutfout in de toets.

## 5. Conclusie

Dit onderzoek is verricht met als doel de mogelijkheden tot het verkrijgen van diagnostische informatie in leesvaardigheid te onderzoeken, in de onderbouw van het voortgezet onderwijs. Hiervoor zijn de toetsresultaten van Toets 1 en Toets 3 gebruikt van het Cito Volgsysteem Voortgezet Onderwijs. Om deze mogelijkheden tot diagnostische informatie te kunnen onderzoeken is de volgende deelvraag opgesteld: In hoeverre zijn de veronderstelde dimensies te onderscheiden in de toetsen leesvaardigheid Nederlands van het VVO?

De resultaten van dit onderzoek laten zien dat de drie veronderstelde modellen aannemelijke modellen zijn. Een eerste vergelijking van de drie modellen laat zien dat Model 1, waarbij leesvaardigheid als unidimensionale factor wordt weergegeven, slechter bij de data van het VVO past dan Model 2, dat significant weer slechter past dan Model 3. De veronderstelde acht dimensies in de toetsen leesvaardigheid Nederlands van het VVO kunnen op basis van dit resultaat onderscheiden worden. Aan de hand van dit resultaat zou dan ook diagnostische informatie verkregen kunnen worden. De gestandaardiseerde regressiecoëfficiënten zijn overwegend positief en geven daarmee weer dat de items meten wat ze behoren te meten, namelijk één van de dimensies.

Een tweede vergelijking, door middel van de passingsmaten CFI, TLI, RMSEA en  $p(RMSEA)$ , toont aan dat alle modellen aannemelijk zijn, maar dat er nauwelijks verschil is tussen de modellen. Model 2 en Model 3 passen dus niet beter bij de data dan Model 1 en in dat geval zou het beter zijn om het minst problematische model te handhaven (Geiser, 2013): Model 1. Indien dit model als best passende model wordt aangenomen, worden er in leesvaardigheid geen verschillende dimensies onderscheiden.

De resultaten die doorslaggevend zijn voor dit onderzoek zijn de correlaties tussen de verschillende factoren, oftewel tussen de veronderstelde dimensies. De hoge correlaties tussen de factoren tonen aan dat er veel samenhang is tussen de deeltaaligheden: de factoren lijken hetzelfde te meten. Om die reden is het weinig zinvol om onderscheid te maken in de verschillende deeltaaligheden. Op basis van deze resultaten is het dan ook moeilijk om diagnostische informatie te verkrijgen.

### 5.1 Discussie

De resultaten van dit onderzoek komen enigszins overeen met de resultaten in de besproken studies (sectie 2.2.). Het is enerzijds wel mogelijk om voor leesvaardigheid een onderscheid te maken in verschillende dimensies, alleen is met de hoge correlaties aangetoond dat er blijkbaar erg veel samenhang is tussen deze dimensies. Hiermee wordt het erg lastig om diagnostische informatie te verkrijgen voor leesvaardigheid. Deze conclusie komt overeen met de bevindingen in het theoretisch

kader waarin aangetoond wordt dat leesvaardigheid een complex interactief proces is, bestaande uit meerdere deelvaardigheden en processen. Ondanks dat deze complexiteit theoretisch erkend wordt, is het moeilijk om aan de hand van leesvaardigheidstoetsen iets te zeggen over de afzonderlijke aspecten. Een mogelijke verklaring, naast de verklaring van Alderson (2000) in sectie 2.5, voor het niet kunnen onderscheiden van deelvaardigheden wordt gegeven in het Interactive Compensatory Model van Stanovich (1980). In dit model is sprake van een interactieproces tussen de verschillende dimensies, waarbij lezers zwakke dimensies compenseren met betere dimensies. Een zwakke lezer die bijvoorbeeld moeite heeft met woordbetekenissen zal dit mogelijk compenseren door meer beroep te doen op de contextuele factoren. Deze compensatie maakt het lastiger om diagnostische informatie te geven, doordat er sprake is van een wisselwerking tussen de verschillende processen. Ook McNamara en Kendeou (2011) zeggen in hun onderzoek dat er bij het beantwoorden van toetsopgaven leesvaardigheid zo veel verschillende processen aan het werk zijn, dat het niet mogelijk is om op basis van de toetsuitslag te bepalen welke specifieke deelvaardigheden een leerling heeft ingezet om een opgave te beantwoorden. De resultaten op toetsen geven wel inzicht in het leesbegrip van leerlingen, maar kunnen geen informatie geven over hun sterke en zwakke aspecten. Dit maakt het geven van diagnostische informatie dan ook lastig.

Belangrijk om te vermelden is dat de uitkomst van dit onderzoek niet hoeft te gelden voor andere toetsen: andere toetsen kunnen een andere uitkomst vertonen. Cutting en Scarborough (2006) hebben bijvoorbeeld aangetoond dat de keuze van de toets invloed heeft op de gemeten relatie tussen begrip en andere vaardigheden. De leesvaardigheidstoetsen vertonen niet allemaal dezelfde onderliggende vaardigheden en cognitieve processen, waardoor de toetsen verschillend beïnvloed kunnen worden (Landi, 2010). De veronderstelde deelvaardigheden die in deze studie onderzocht zijn zouden daarom misschien ook door middel van een andere toets onderzocht moeten worden, voor het kunnen geven van een eenduidigere uitkomst.

De veronderstelde deelvaardigheden leiden tot een tweede aspect om rekening mee te houden: de mate waarin items daadwerkelijk ingedeeld konden worden in de verschillende categorieën. Vooral bij Model 3 gaf dit problemen. Niet elke categorie kwam namelijk even vaak in de toetsen voor, waardoor er bij enkele factoren te weinig items voorkwamen en het nauwelijks mogelijk was om hier informatie uit te halen. Dit heeft mogelijk ook gezorgd voor de hoge correlaties bij Model 3 en de correlaties groter dan 1. Het zou dus waarschijnlijk voor de analyses beter zijn om per categorie even veel items te hebben. Daarnaast zijn de items na het maken van de toetsen gecategoriseerd, waardoor ze misschien niet overduidelijk één bepaalde categorie meten en mogelijk ook in een andere categorie ingedeeld hadden kunnen worden. Dit is onder andere een gevolg van het invoeren van de referentieniveaus pas tijdens het constructieproces van Toets 1 en Toets 3. Er is in deze toetsen zoveel mogelijk rekening gehouden met de referentieniveaus, alleen garandeert dit

niet dat elk item één bepaalde categorie representeert. Dit is in de resultaten mogelijk terug te vinden bij de gestandaardiseerde regressiecoëfficiënten. De waarden hiervan zijn overwegend positief, maar niet heel hoog, waarmee aangegeven wordt dat de items niet overduidelijk de deelvaardigheden meten die ze behoren te meten. Opvallend zijn de negatieve regressiecoëfficiënten, een gevolg van de categorie-indeling of van de inhoud van het item zelf. Laatstgenoemde is een aspect waar in dit onderzoek geen invloed op uitgeoefend had kunnen worden, dit is namelijk al tijdens het constructieproces van de toetsen bepaald.

Een derde punt van discussie is het gegeven dat de itemsoort een rol kan spelen. De toetsen van het VVO omvatten alleen meerkeuze items, echter is aangetoond dat het gebruik van meerkeuze items het leesproces kan beïnvloeden. Een studie van Rupp, Ferne en Choi (2006) heeft aangetoond dat lezers bij het maken van een toets de tekst in stukken opdelen aan de hand van de vragen die hen gesteld worden. Hierdoor zal een lezer eerder gefocust zijn op de microstructuur dan de macrostructuur en minder gebruik maken van hogere-orde inferenties. Een ander soort items zou dus andere resultaten kunnen geven.

Voor een vervolgstudie wordt daarom aangeraden om analyses uit te voeren op een evenwichtiger categorie-indeling, waarbij alle categorieën even veel items bevatten. Ook zou bijvoorbeeld onderzoek gedaan kunnen worden naar zowel meerkeuze items als open vragen, zodat voor verschillende soorten items de veronderstelde deelvaardigheden onderzocht kunnen worden. Daarnaast omvatten de VVO toetsen maar een beperkt aantal items. Om meer informatie over de items te kunnen krijgen, is het aan te raden om meer items in de analyses te betrekken.

Tot slot, wat betekenen de conclusies van dit onderzoek voor het verkrijgen van diagnostische informatie in leesvaardigheid Nederlands? Het is in ieder geval raadzaam voor toetsontwikkelaars en onderzoekers om voorzichtig te zijn in hun beweringen dat leestoetsen gebruikt kunnen worden om diagnostische informatie te geven (Kalifa & Weir, 2009). Een reden om voorzichtig te zijn is de aanhoudende discussie over leesvaardigheid als uni- of multidimensioneel construct en het toetsen van de mogelijk onderliggende vaardigheden. Daarnaast is interventiestudies onder andere ook naar voren gekomen dat instructie in afzonderlijke deelvaardigheden weliswaar een positief effect heeft op de beheersing van de betreffende deelvaardigheden, maar dat er geen generaliserend effect op de overkoepelende vaardigheid kan worden vastgesteld (Walraven, 1995). Zo heeft bijvoorbeeld vooruitgang in prestaties op de deelvaardigheid 'vinden van de hoofdgedachte' geen positieve invloed op de prestaties op de algemene leesvaardigheid toetsen. Is het dan wel relevant om diagnostische informatie te geven op basis van deelvaardigheden, door middel van toetsing?

Enkele alternatieven voor het verkrijgen van diagnostische informatie zijn een diagnostisch gesprek en 'modeling' ('hardop-denkend-voordoën'). Door middel van een diagnostisch gesprek



proberen leerkrachten inzicht te krijgen in het denkproces van leerlingen tijdens leesvaardigheid. Door zicht te krijgen op de denkstappen en het begripsniveau van leerlingen, kan de leerkracht inspelen op de behoeftes van de leerlingen (Mortel & Förner, 2013). Modeling kan in de klas gebruikt worden door leerkrachten en leerlingen, maar het zou ook gebruikt kunnen worden bij het ontwikkelingsproces van een toets, om na te gaan of de items daadwerkelijk de onderliggende processen aan het licht brengen (Kalifa & Weir, 2009). Voor vervolgstudies zou de nadruk dus mogelijk ook meer op andere manieren van diagnosticeren kunnen liggen.

## 6. Literatuurlijst

- Aarnoutse, C.A.J. & Leeuwe, J.F.J. van (1988). Het belang van technisch lezen, woordenschat en ruimtelijke intelligentie voor begrijpend lezen. *Pedagogische Studiën* 65, 49-59.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.
- Alvermann, D.E., Fitzgerald, J., & Simpson, M. (2006). Teaching and learning in reading. In P.A. Alexander & P. H. Winne (Red.), *Handbook of educational psychology* (2nd ed., pp. 427-455). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246.
- Bijsterveldt-Vliegthart, M. van (2011). Naar passend onderwijs. [6 paragrafen] *Ministerie van OCW* [Brief/Kamerstuk]. Beschikbaar via: <http://www.rijksoverheid.nl/onderwerpen/passend-onderwijs/documenten-en-publicaties/kamerstukken/2011/01/31/naar-passend-onderwijs.html>
- Broek, P. van den (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, 328 (5977), 453-456.
- Broek, P. van den, Kendeou, P., Kremer, K., Lynch, J.S., Butler, J., White, M. J., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. Paris & S. Stahl (Red.), *Children's reading comprehension and assessment*, 107-130, Mahwah, NJ: Erlbaum.
- Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, 19, 13-27.
- Chen, H., & Chen, J. (2015). Exploring reading comprehension skill relationships through the G-DINA model. *Educational Psychology*, 1-20.
- Cito (2010). *Resultaten PISA-2009, praktische kennis en vaardigheden van 15 jarigen, Nederlandse uitkomsten van het Programme for International Student Assessment (PISA) op het gebied van leesvaardigheid, wiskunde en natuurwetenschappen in het jaar 2009*. Arnhem: Cito.
- Cito (2015). *Wetenschappelijke verantwoording Toets 0 t/m 3, tweede generatie*. Arnhem: Cito.
- Cohen, L., Manion, L. & Morrison, K. (2007). *Research Methods in Education*. Londen/New York: Routledge.

- College voor Toetsen en Examens (2014). *Publieksversie Toetswijzer diagnostische tussentijdse toets voor Nederlands, Engels en wiskunde*. Geraadpleegd op 11 november 2015, van [https://www.hetcvte.nl/item/diagnostische\\_tussentijdse\\_toets](https://www.hetcvte.nl/item/diagnostische_tussentijdse_toets).
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cutting, L. E., & Scarborough, H. s. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10(3), 277-299.
- Davis, F. B. (1968). Research in comprehension in reading. *Reading Research Quarterly*, 3, 499-545.
- Deegan, J. (1978). On the occurrence of standardized regression coefficients greater than one. *Educational and Psychological Measurement*, 38(4), 873-888.
- Dijk, T. A. van, & Kintsch, W. (1983). *Strategies of discourse competence*. New York: Academic Press.
- Dijk, T. A. van, & Kintsch, W. (1983). *The notion of macrostructure*. In: Dijk van, T.A. and Kintsch, W. (red.). *Strategies of discourse comprehension*. New York: Academic Press, pp. 189–223.
- Englert, C.S., & Thomas, C.C. (1987). Sensitivity to text structure in reading and writing: A comparison between learning disabled and non-learning disabled students. *Learning Disability Quarterly*, 10, 93-105.
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN. Online via: <http://www.psynip.nl/website-openbaar-documenten-nip-algemeen/beoordelingssysteem.pdf>.
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008). *Over de drempels met taal: De niveaus voor de taalvaardigheid*. Enschede: OCW/SLO.
- Farr, R., & Carey, R. F. (1986). *Reading: What can be measured?* (2<sup>nd</sup> ed.). Newark: DE International Reading Association.
- Förrer, M. & Mortel, K. van de (2010). *Lezen... Denken... Begrijpen*. Handboek begrijpend lezen in het basisonderwijs. Amersfoort: CPS.
- Francis, D. J, Snow, C.E., August, D., Carlson, C.D., Miller, J., & Iglesias, A. (2006). Measures of Reading Comprehension: A Latent Variable Analysis of the Diagnostic Assessment of Reading Comprehension. *Scientific Studies of Reading*, 10(3), 301-322.
- Geiser, C. (2013). *Data Analysis with Mplus*. New York: The Guilford Press.

- Gelderen, A. van (2012). 'Basisvaardigheden' en het onderwijs in lezen en schrijven. *Levende Talen Tijdschrift*, 13, 1, 3-15.
- Goldman, S. R., & Durán, R. P. (1988). Answering questions from oceanography texts: Learner, task, and text characteristics. *Discourse Processes*, 11(4), 373-412.
- Goldman, S.R., & Rakestraw, J.A. jr. (2000). Structural aspects of constructing meaning from text. *Handbook of Reading Research*, 3, 311-335.
- Gough, P., Ehri, L., & Treiman, R. (Red.) (1992). *Reading Acquisition*. Hillsdale, NJ: Erlbaum.
- Grabe, W., & Jiang, X. (2014). Assessing Reading. In: Kunnan, A.J. (Ed.) *The Companion to Language Assessment*, 1, 185-200. London: John Wiley and Sons.
- Hacquebord, H. I. (1999). A Dutch comprehension test for identifying reading problems in L1 and L2 students. *Journal of Research in Reading*, 22(3), 299-303.
- Hacquebord, H.I. (1997). Constructie en evaluatie van een tekstbegriptoets voor de brugklas. *Tijdschrift voor onderwijsresearch*, 2, 37-54.
- Hannon, B., & Daneman, M. (2001). A new tool for understanding individual differences in the component of processes of reading comprehension. *Journal of Educational Psychology*, 93, 103-128.
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Inspectie van het onderwijs (2008). *Basisvaardigheden taal in het voortgezet onderwijs. Resultaten van een inspectieonderzoek naar taalvaardigheid in de onderbouw van het vmbo en praktijkonderwijs*, Utrecht: Inspectie van het Onderwijs.
- Jong, P. de (2011). Begrijpend lezen. In P. de Jong & H. Koomen (Red.), *Interventie bij onderwijsleerproblemen* (pp. 25-39). Antwerpen: Garant.
- Kamil, P.B., Mosenthal, P.B., Pearson, P.D., & Barr, R. (Red.) (2000). *Handbook of Reading Research*. Mahwah, NJ: Erlbaum.
- Kendeou, O., Broek, P. van den, Helder, A., & Karlsson, A.K. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, 29(1), 10-16.
- Kendeou, P., Rapp, D.N., & Broek, P. van den. (2003). The influence of readers' prior knowledge on text comprehension and learning from text. *Progress in education*, 13, 189-209.

- Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge: Cambridge University Press.
- Kintsch W. & Dijk, T. van (1978). Towards a model of discourse comprehension and production. *Psychological Review*, 85(5), 363-394.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Kintsch, W., & Kintsch, E. (2005). Comprehension. In: S. G. Paris and S. A. Stahl (Red.), *Children's reading comprehension and assessment*, 71-92, Mahwah, NJ: Lawrence Erlbaum Associates.
- Kuhlemeier, H., Jolink, A., Jongen, I., Krämer, I, Hemker, B., & Bechger, T. (2014). *Balans van de leesvaardigheid in het basis- en speciaal basisonderwijs 2. Uitkomsten van de peiling in het schooljaar 2011-2012 in groep 8, groep 5 en de eindgroep van het SBO* (PPON-reeks nummer 54). Arnhem: Cito.
- Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading and Writing*, 23(6), 701-717.
- Ledoux, G., Meijer, J., Veen, I. van der, & Breetvelt, I. (2013). *Meetinstrumenten voor sociale competenties, metacognitie en advanced skills. Een inventarisatie*. Amsterdam: Kohnstamm Instituut.
- Lunzer, E., Waite, M., & Dolan, T. (1979). Comprehension and comprehension tests. In: E. Lunzer and K. Garner (red.), *The effective use of reading* (37-71). London: Heinemann Educational Books.
- McNamara, D. S., & Kendeou, P. (2011). Translating advances in reading comprehension research to educational practice. *International Electronic Journal of Elementary Education*, 4 (1), 33-46.
- Meijer, J., & Gelderen, A. J. S. van (2002). *Lezen voor het leven: Een empirische vergelijking van een nationale en een internationale leesvaardigheidspeiling*. Amsterdam: SCO-Kohnstamm Instituut.
- Meijerink, H. P., Letschert, J.F., Rijlaarsdam, G.C.W., Bergh, H.H. van den, & Streun, A. van (2009). *Referentiekader taal en rekenen. De referentieniveaus*. Enschede: OCW/SLO.
- Meyer, B. J. (1984). Organizational aspects of text: Effects on reading comprehension and applications for the classroom. *Promoting reading comprehension*, 113-138.

- Mullis, I.V.S., Martin, M.O., Kennedy, A., Trong, K., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.
- OECD (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, PISA, OECD Publishing, Paris.
- Onderwijsraad (2011). *Naar hogere leerprestaties in het voortgezet onderwijs*. Den Haag: Onderwijsraad.
- Onderwijsraad (2014). *Een eigentijds curriculum*. Den Haag: Onderwijsraad.
- Paris, S.G. & Stahl, S. A. (2005). *Children's reading comprehension and assessment*. Mahwah, NJ: Erlbaum.
- Potts, G.R., & Peterson, S.B. (1985). Incorporation versus compartmentalization in memory for discourse. *Journal of Memory and Language*, 24, 107-118.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum.
- Rapp, D. N., Broek, P. van den, McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, 11(4), 289-312.
- Reitsma, P., & Walraven, M. (1991). Inleiding: ontwikkelingen in theorie en onderzoek. In: P. Reitsma, & M. Walraven (Red.), *Instructie in begrijpend lezen*. Delft, Eburon.
- Rosenshine, B. V. (1980). Skills hierarchies in reading comprehension. In: R.J. Spiro, B.C. Bruce, and W.F. Brewer (Red.), *Theoretical issues in reading comprehension*, 535-554, Hillsdale, N.J.: Erlbaum.
- Rouet, J. F., Vidal-Abarca, E., Erboul, A. B., & Millogo, V. (2001). Effects of information Search Tasks on the Comprehension of Instructional Text. *Discourse Processes*, 31(2), 163-186.
- Schedl, M., Gordon, A., Carey, P. A., & Tang, K. L. (1996). An analysis of the dimensionality of TOEFL reading comprehension items. *ETS Research Report Series*, 1995(2), i-26.

- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures, *Methods of Psychological Research Online*, 8(2), 23-74.
- Song, M. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing*, 25(4), 435-464.
- Spearritt, D. (1972). Identification of subskills of reading comprehension by maximum likelihood factor analysis. *Reading Research Quarterly*, 92-111.
- Steensel, R. van, Oostdam, R., & Gelderen, A. van (2012). Assessing reading comprehension in adolescent low achievers: Subskills identification and task specificity. *Language Testing*, 30(1), 3-21.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics, 6th ed.* Boston: Pearson.
- Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, 10, 381–398.
- Taylor, M. B., & Williams, J. P. (1983). Comprehension of learning-disabled readers: Task and text variations. *Journal of Educational Psychology*, 75, 743-751.
- Thorndike, R. L. (1973–1974). Reading as reasoning. *Reading Research Quarterly*, 9, 135–147.
- Vidal-Abarca, E., Gilabert, R., & Rouet, J. F. (1998). The role of question type on learning from scientific text. *Meeting on Comprehension and Production of Scientific Texts*, Aveiro, Portugal.
- Weir, C. J., & Porter, D. (1994) The Multi-Divisible or Unitary Nature of Reading: The language tester between Scylla and Charybdis. *Reading in a Foreign Language*, 10(2), 1–19.
- Weir, C., Hawkey, R., Green, A., & Devi, S. (2006). The cognitive processes underlying the academic reading construct as measured by IELTS. *IELTS Collected Papers 2: Research in Reading and Listening Assessment*, 2, 212.

**Bijlage 1      Overzicht van de besproken processen in sectie 1.1.**

Gough et al. (1992)		Linguïstische processen		Cognitieve processen		Kintsch & Kintsch (2005)	
Decoderen	Woord-herkenning	Tekstgestuurde verwerking (bottom-up-processing)	Is gericht op de inhoud van de tekst. Er wordt gebruik gemaakt van woorden, passages en hun onderlinge betekenisrelaties om tot begrip te komen.	Lagere-orde-processen	Vertalen van geschreven in betekenisvolle eenheden.	Decoderen	Perceptuele en conceptuele processen die leiden tot woord- en zinsbetekenissen.
Begrip	Begrijpen van zinnen in een tekst. Opbouw tekststructuur. Integreren voorgaande in tekst met kennis die lezer al heeft.			Kennisgestuurde verwerking (top-down-processing)	Duidt op de inzet van de eigen kennis van de wereld en kennis van teksten, van taal en van contexten.	Hogere-orde-processen.	Combineren van taaleenheden tot een betekenisvolle en coherente mentale representatie. Omvatten begrijpen en interpreteren van een tekst.
						Situatiemodel	Mentaal model van de situatie als omschreven in de tekst en eigen eerdere informatie en ervaringen.



Categorie	Begrijpen		
<b>BEGR1</b>	Opgaven die vragen naar de betekenis van een woord, woordgroep of zin(nen) die expliciet in de tekst vermeld wordt.		
	<table border="1"> <tr> <td><i>Betekenis woordgroep</i></td> <td><i>Wat betekent volgens de tekst '..... zijn'?</i></td> </tr> </table>	<i>Betekenis woordgroep</i>	<i>Wat betekent volgens de tekst '..... zijn'?</i>
<i>Betekenis woordgroep</i>	<i>Wat betekent volgens de tekst '..... zijn'?</i>		
<b>BEGR2</b>	Opgaven die vragen naar specifieke inhoudselementen die expliciet in de tekst aan de orde gesteld worden. Dit zijn bijvoorbeeld feiten en meningen, voorwerpen, aantallen, een plaats van handeling of tijdsperioden, (hoofd)personen.		
	<table border="1"> <tr> <td><i>Feiten en meningen</i></td> <td> <i>Een schrijver kan gebruik maken van feiten en meningen in zijn tekst.</i>  <i>1. De tekst bevat feiten.</i>  <i>2. De tekst bevat meningen.</i> </td> </tr> </table>	<i>Feiten en meningen</i>	<i>Een schrijver kan gebruik maken van feiten en meningen in zijn tekst.</i> <i>1. De tekst bevat feiten.</i> <i>2. De tekst bevat meningen.</i>
<i>Feiten en meningen</i>	<i>Een schrijver kan gebruik maken van feiten en meningen in zijn tekst.</i> <i>1. De tekst bevat feiten.</i> <i>2. De tekst bevat meningen.</i>		
<b>BEGR3</b>	Opgaven die vragen naar eenvoudige expliciete verbanden op lokaal niveau. Verbanden kunnen worden gelegd op basis van inhoudelijke en/of structurele elementen zoals signaalwoorden. Voorbeelden van verbanden zijn verwijzingen, vergelijkingen, tegenstellingen, generalisaties en voorbeelden, oorzaak en gevolg, vraag en antwoord, reden en verklaring, middel en doel, deel-/geheelrelaties, conclusie en argumenten.		
	<table border="1"> <tr> <td><i>Verwijzing</i></td> <td><i>Wat wordt er bedoeld met: Ze (regel 2)?</i></td> </tr> </table>	<i>Verwijzing</i>	<i>Wat wordt er bedoeld met: Ze (regel 2)?</i>
	<i>Verwijzing</i>	<i>Wat wordt er bedoeld met: Ze (regel 2)?</i>	
	<table border="1"> <tr> <td><i>Reden en verklaring</i></td> <td><i>Waarom voeren de studenten actie voor ....?</i></td> </tr> </table>	<i>Reden en verklaring</i>	<i>Waarom voeren de studenten actie voor ....?</i>
<i>Reden en verklaring</i>	<i>Waarom voeren de studenten actie voor ....?</i>		
<table border="1"> <tr> <td><i>Doel</i></td> <td><i>Wat is het doel van de eerste alinea?</i></td> </tr> </table>	<i>Doel</i>	<i>Wat is het doel van de eerste alinea?</i>	
<i>Doel</i>	<i>Wat is het doel van de eerste alinea?</i>		
<b>BEGR4</b>	Opgaven die vragen naar complexe expliciete verbanden over grotere tekstdelen heen. Verbanden kunnen worden gelegd op basis van inhoudelijke en structurele elementen. Voorbeelden van verbanden zijn vergelijkingen, tegenstellingen, volgordes, generalisaties en voorbeelden, oorzaak en gevolg, vraag en antwoord, reden en verklaring, middel en doel, deel/geheelrelaties, conclusie en argumenten.		
	<table border="1"> <tr> <td><i>Volgordes</i></td> <td><i>Het artikel beschrijft verschillende gebeurtenissen. In welke volgorde vinden de hieronder genoemde gebeurtenissen plaats?</i></td> </tr> </table>	<i>Volgordes</i>	<i>Het artikel beschrijft verschillende gebeurtenissen. In welke volgorde vinden de hieronder genoemde gebeurtenissen plaats?</i>
<i>Volgordes</i>	<i>Het artikel beschrijft verschillende gebeurtenissen. In welke volgorde vinden de hieronder genoemde gebeurtenissen plaats?</i>		

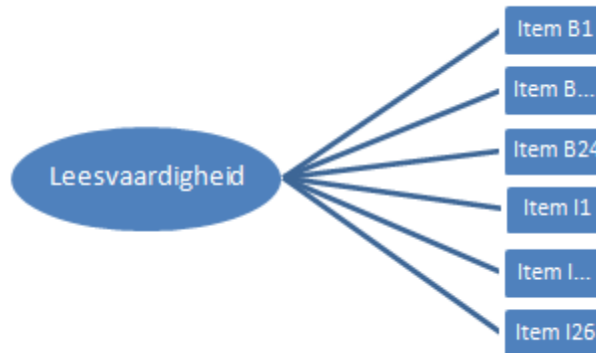
## Bijlage 3

## Subcategorieën bij 'Interpreteren' met voorbeelden en voorbeelditems.

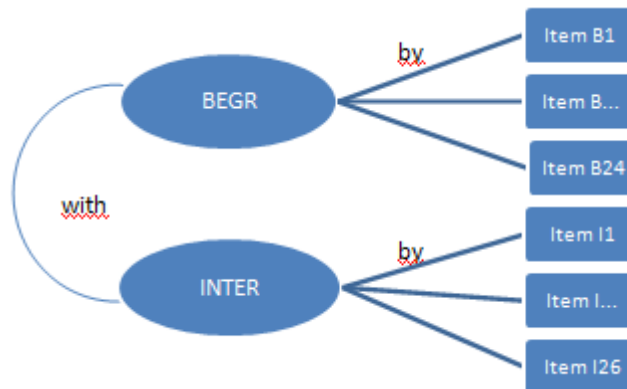
Categorie	Interpreteren	
<b>INTER1</b>	Opgaven die vragen naar het afleiden van de betekenis van een woord, woordgroep of zin(nen).	
	<i>Betekenis woord</i>	<i>In de derde alinea wordt het woord '.....' gebruikt. Wat betekent dit woord?</i>
<b>INTER2</b>	Opgaven die vragen naar het afleiden van informatie uit de tekst op lokaal niveau. De leerling moet zijn voorkennis inzetten naast de informatie die de tekst geeft. Voorbeelden hiervan zijn opgaven over onderwerp, thema, hoofdlijnen, hoofdgedachte, hoofdpersoon, setting en doel en doelgroep van de tekst, het vertelperspectief en de tekststructuur.	
	<i>Hoofdgedachte</i>	<i>Wat is de hoofdgedachte van alinea 3?</i>
	<i>Onderwerp</i>	<i>Welk kopje past het best bij alinea 2?</i>
<b>INTER3</b>	Opgaven die vragen naar het afleiden van informatie op het globale niveau van de tekst. Verbanden kunnen worden gelegd op basis van inhoudelijke en/of structurele elementen. Voorbeelden hiervan zijn opgaven over onderwerp, thema, hoofdlijnen, hoofdgedachte, setting en doel en doelgroep van de tekst, hoofdpersoon, het vertelperspectief en de tekststructuur.	
	<i>Hoofdgedachte</i>	<i>Welke zin past het <b>best</b> bij deze tekst als geheel? Wat is de hoofdgedachte van deze tekst?</i>
	<i>Onderwerp</i>	<i>Wat is het belangrijkste onderwerp van deze tekst?</i>
	<i>Doel</i>	<i>Wat is het doel van deze tekst?</i>
<b>INTER4</b>	Opgaven die vragen naar het taalgebruik en schrijfstijl van de schrijver en waar verbanden moeten worden gelegd tussen tekstuele informatie en kennis van het taalsysteem. Voorbeelden hiervan zijn vragen tekstsoort, naar het gebruik van aanhalingstekens of vragen naar specifieke woordkeuze waarbij, register, sociale en culturele conventies een rol spelen.	
	<i>Gebruik van aanhalingstekens</i>	<i>Waarom staat '.....' hier tussen aanhalingstekens?</i>

## Bijlage 4 De drie confirmatieve factor modellen in beeld.

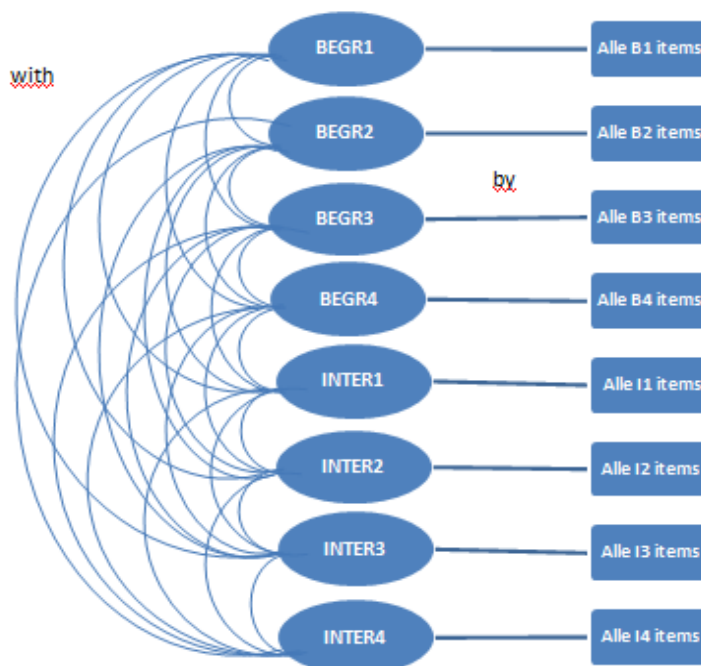
Model 1: Leesvaardigheid als unidimensionale vaardigheid.



Model 2: Leesvaardigheid als tweedimensionaal factormodel met gecorreleerde factoren.



Model 3: Leesvaardigheid als achtdimensionaal factormodel met gecorreleerde factoren.



**Bijlage 5 Modelpassing statistieken behorend bij de drie confirmatieve factor modellen.**

**Tabel 1: Modelpassing Model 1: 'Unidimensionaal' van Toets 1 en Toets 3.**

	Toets 1			Toets 3		
	BB	KB/GT	HAVO/VWO	BB	KB/GT	HAVO/VWO
$\chi^2$	4049.327	10070.042	3985.578	3028.331	8224.518	4488.686
Df	902	1175	1175	945	1224	1377
$p(\chi^2)$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
CFI	0.938	0.929	0.959	0.931	0.918	0.936
TLI	0.934	0.926	0.957	0.928	0.914	0.934
RMSEA	0.020	0.018	0.011	0.019	0.017	0.012
$p(\text{RMSEA})$	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999

**Tabel 2: Modelpassing Model 2: 'Tweedimensionaal met gecorreleerde factoren' van Toets 1 en Toets 3.**

	Toets 1			Toets 3		
	BB	KB/GT	HAVO/VWO	BB	KB/GT	HAVO/VWO
$\chi^2$	4013.366	10066.904	3972.875	2942.015	8195.966	4479.138
Df	901	1174	1174	944	1223	1376
$p(\chi^2)$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
CFI	0.938	0.929	0.959	0.934	0.918	0.936
TLI	0.935	0.926	0.957	0.931	0.915	0.934
RMSEA	0.020	0.018	0.011	0.019	0.017	0.012
$p(\text{RMSEA})$	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999

**Tabel 3: Modelpassing Model 3: 'Acht dimensionaal met gecorreleerde factoren' van Toets 1 en Toets 3.**

	Toets 1			Toets 3		
	BB	KB/GT	HAVO/VWO	BB	KB/GT	HAVO/VWO
$\chi^2$	3731.773	9584.750	3740.781	2931.909	6557.543	4073.634
df	853	1112	1106	935	1070	1304
$p(\chi^2)$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
CFI	0.941	0.927	0.960	0.934	0.929	0.942
TLI	0.937	0.923	0.958	0.930	0.926	0.939
RMSEA	0.020	0.018	0.011	0.019	0.016	0.011
$p(\text{RMSEA})$	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999

**Bijlage 6 Correlaties Model 2 en Model 3 behorende bij Toets 1 en Toets 3.**

**Tabel 4: Model 2 - Correlaties Toets 1 en Toets 3 per niveau.**

		BB		KB/GT		HAVO/VWO	
<b>BEGR with INTER</b>	<b>Toets 1</b>	0.955	p<0.001	0.992	p<0.001	0.976	p<0.001
	<b>Toets 3</b>	0.911	p<0.001	0.976	p<0.001	0.974	p<0.001

**Tabel 5: Model 3 - Correlaties Toets 1.**

Factoren		BB	KB/GT	HAVO/ VWO
<b>BEGR2 with</b>	<b>BEGR1</b>	1.02	0.99	0.91
<b>BEGR3 with</b>	<b>BEGR1</b>	0.88	0.93	0.88
	<b>BEGR2</b>	0.85	1.07	0.98
<b>BEGR4 with</b>	<b>BEGR1</b>	-	-	0.84
	<b>BEGR2</b>	-	-	0.83
	<b>BEGR3</b>	-	-	0.94
<b>INTER1 with</b>	<b>BEGR1</b>	1.13	1.14	1.097
	<b>BEGR2</b>	0.99	0.94	1.08
	<b>BEGR3</b>	0.90	0.95	1.096
	<b>BEGR4</b>	-	-	0.98
<b>INTER2 with</b>	<b>BEGR1</b>	0.99	1.05	0.88
	<b>BEGR2</b>	0.84	0.96	0.92
	<b>BEGR3</b>	0.85	0.97	0.99
	<b>BEGR4</b>	-	-	0.88
	<b>INTER1</b>	0.93	0.9	1.08
<b>INTER3 with</b>	<b>BEGR1</b>	1.13	0.97	0.84
	<b>BEGR2</b>	0.93	0.99	0.94
	<b>BEGR3</b>	0.86	0.99	0.97
	<b>INTER1</b>	0.97	0.93	0.98
	<b>INTER2</b>	0.94	0.96	0.98

**Tabel 6: Model 3 - Correlaties Toets 3.**

Factoren		BB	KB/GT	HAVO/ VWO
<b>BEGR3 with</b>	<b>BEGR2</b>	1.03	0.82	0.81
<b>BEGR4 with</b>	<b>BEGR2</b>	1.07	0.81	0.77
	<b>BEGR3</b>	0.99	0.90	0.95
<b>INTER1 with</b>	<b>BEGR2</b>	-	-	0.94
	<b>BEGR3</b>	-	-	1.08
	<b>BEGR4</b>	-	-	1.13
<b>INTER2 with</b>	<b>BEGR2</b>	0.95	0.85	0.77
	<b>BEGR3</b>	0.98	0.998	0.94
	<b>BEGR4</b>	0.93	0.83	0.94
	<b>INTER1</b>	-	-	0.99
<b>INTER3 with</b>	<b>BEGR2</b>	1.04	0.82	0.83
	<b>BEGR3</b>	0.89	0.89	0.92
	<b>BEGR4</b>	0.91	0.68	1.03
	<b>INTER1</b>	-	-	1.18
	<b>INTER2</b>	1.01	0.95	0.93
<b>INTER4 with</b>	<b>BEGR2</b>	-	-	0.55
	<b>BEGR3</b>	-	-	0.49
	<b>BEGR4</b>	-	-	0.49
	<b>INTER1</b>	-	-	0.56
	<b>INTER2</b>	-	-	0.42
	<b>INTER3</b>	-	-	0.55

**Bijlage 7 Ranges van gestandaardiseerde regressiecoëfficiënten van Toets 1 en Toets 3, behorend bij de drie confirmatieve factor modellen.**

**Tabel 7: Model 1 - Ranges van gestandaardiseerde regressiecoëfficiënten Toets 1 en Toets 3 per niveau.**

		BB	KB/GT	HAVO/VWO
<b>Toets 1</b>	<b>BLE by</b>	0.1 - 0.65	0.1 - 0.60	-0.06 - 0.60
<b>Toets 3</b>	<b>BLE by</b>	0.04 - 0.59	-0.13 - 0.58	0.06 - 0.46

**Tabel 8: Model 2 - Ranges van gestandaardiseerde regressiecoëfficiënten Toets 1 en Toets 3 per niveau.**

		BB	KB/GT	HAVO/VWO
<b>Toets 1</b>	<b>INTER by</b>	0.10 - 0.55	0.13 - 0.52	-0.06 - 0.50
	<b>BEGR by</b>	0.16 - 0.67	0.10 - 0.61	-0.001 - 0.60
<b>Toets 3</b>	<b>INTER by</b>	0.09 - 0.60	-0.13 - 0.58	0.17 - 0.46
	<b>BEGR by</b>	0.04 - 0.60	0.07 - 0.48	0.06 - 0.45

**Tabel 9: Model 3 - Ranges van gestandaardiseerde regressiecoëfficiënten Toets 1 en Toets 3 per niveau.**

		BB	KB/GT	HAVO/VWO
<b>Toets 1</b>				
BEGR	BEGR1 by	0.22 - 0.40	0.1 - 0.39	0.31 - 0.67
	BEGR2 by	0.36 - 0.43	0.3 - 0.39	0.2 - 0.48
	BEGR3 by	0.29 - 0.55	0.21 - 0.6	-0.002 - 0.52
	BEGR4 by	-	-	0.18 - 0.43
INTER	INTER1 by	0.29 - 0.55	0.29 - 0.51	0.17 - 0.33
	INTER2 by	0.11 - 0.48	0.13 - 0.53	-0.062 - 0.38
	INTER3 by	0.29 - 0.47	0.25 - 0.45	0.100 - 0.50
<b>Toets 3</b>				
BEGR	BEGR1 by	-	-	-
	BEGR2 by	0.14 - 0.34	0.08 - 0.49	0.27 - 0.42
	BEGR3 by	0.09 - 0.59	0.19 - 0.48	0.15 - 0.46
	BEGR4 by	0.04 - 0.60	0.22 - 0.39	0.06 - 0.36
INTER	INTER1 by	-	-	0.22 - 0.41
	INTER2 by	0.19 - 0.53	-0.57 - 0.13	0.27 - 0.38
	INTER3 by	0.09 - 0.61	0.18 - 0.40	0.17 - 0.45
	INTER4 by	-	-	0.48 - 0.52