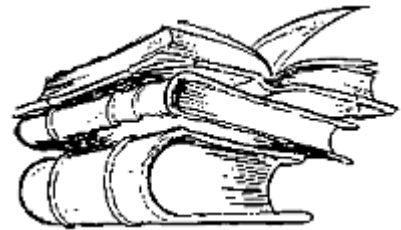


De leesbaarheid van schoolteksten

*Een experimenteel onderzoek naar de invloed van woordmoeilijkheid
op het tekstbegrip van zwakke en sterke lezers*



Bacheloreindwerkstuk Nederlandse taal en cultuur

Naam	Wendy van Dooren
Studentnummer	3342026
Begeleider	R. Kraf
Datum	1 juli 2011

Inhoudsopgave

1	Inleiding	4
2	Theoretisch kader.....	6
2.1	Tekstbegrip	6
2.2	Woordmoeilijkheid.....	8
2.3	Cloze-toets	9
2.3.1	Wat voor soort cloze?	10
2.3.2	Factoren die de prestaties op cloze-items beïnvloeden	11
2.3.3	Welke scoringsmethode?	13
2.4	Belang van het onderzoek	13
3	Experimentele opzet	14
3.1	Onderzoeksopzet	14
3.1.1	Proefpersonen	14
3.1.2	Materiaal.....	15
3.1.3	Procedure.....	16
3.1.4	Scoringsmethode	17
3.2	Hypotheses	17
4	Resultaten	20
4.1	Verloop van het onderzoek	20
4.2	Beschrijving van de resultaten	21
4.2.1	Scores op de verschillende toetsen	21
4.2.2	Betrouwbaarheid	23
4.2.3	Correlaties tussen de verschillende toetsen	26
4.2.4	Woordenschat en leesvaardigheid	27
4.2.5	Cloze-toets	32
4.2.6	Woordenschat en leesvaardigheid als covariaat	34
4.2.7	Moeilijkheid van de cloze-items	35
4.3	Verschillende scoringsmethodes	43
4.4	Terugkoppeling hypothesen.....	45
5	Conclusie	48
6	Discussie.....	49

Literatuurlijst.....	52
----------------------	----

Bijlagen

Bijlage 1	Oorspronkelijke tekst
Bijlage 2	Makkelijke tekst
Bijlage 3	Moeilijke tekst.....
Bijlage 4	Woordmanipulaties en cloze-gaten
Bijlage 5	Tabel met woordfrequenties van de woordmanipulaties per versie
Bijlage 6	Regels voor cloze-ontwerp
Bijlage 7	Makkelijke cloze-toets
Bijlage 8	Moeilijke cloze-toets
Bijlage 9	Toetsverdeling over de vier klassen
Bijlage 10	Beoordeling cloze-toets.....

1 Inleiding

Een conversatie in de klas tussen een docent en een leerling verloopt als volgt:

Docent: “Zou je me eens in een paar zinnen kunnen zeggen waar deze tekst over gaat?”

Leerling: “Euhhh... ik snapte het eigenlijk niet zo goed.”

Docent: “En waardoor komt dat, denk je?”

Leerling: “Er staan veel moeilijke woorden in de tekst die ik niet ken... ”

Dit voorbeeld laat een situatie zien die zich in het hedendaagse onderwijs nog al eens voordoet: leerlingen begrijpen een tekst niet omdat er te veel moeilijke woorden in staan (Bonset & Braaksma, 2008). Dit kan grote gevolgen hebben, zoals ook de resultaten op begripstoetsen laten zien: door het te grote aantal moeilijke woorden zijn leerlingen niet in staat om tot de kern van een tekst door te dringen. De woordmoeilijkheid verspert als het ware hun weg naar de werkelijke inhoud van een tekst. Het gevolg is dat de tekst door deze leerlingen niet begrepen wordt, laat staan dat ze er iets van leren.

Gezien de centrale plaats die teksten in het onderwijs innemen, is onderzoek naar de leesbaarheid van schoolteksten van belang. Dit soort onderzoek is niet alleen waardevol voor leerlingen en docenten, maar ook voor de samenstellers van lesmethoden. Het niveau van schoolteksten moet immers aansluiten op dat van de leerlingen. Om passende lesmethoden te kunnen ontwikkelen, is het belangrijk dat men weet wat nu precies het effect is van woordmoeilijkheid op tekstbegrip. Om een tekst te kunnen begrijpen is het wel of niet hebben van een grote woordenschat, naast o.a. wereldkennis en grammaticale kennis, namelijk een cruciale factor (Stahl, 2003). Bovendien moet de vraag gesteld worden of woordmoeilijkheid eenzelfde invloed heeft op zwakke als op sterke lezers. Naar gelang de verschillen tussen deze lezerstypen in de omgang met moeilijke woorden, kunnen vervolgens teksten worden samengesteld die qua woordmoeilijkheid aansluiten op het niveau van de leerlingen, zodat scholieren meer vat krijgen op schoolteksten en deze ook beter gaan begrijpen.

In dit onderzoek staat de volgende vraag centraal:

Wat is de invloed van woordmoeilijkheid op het tekstbegrip van zwakke en sterke lezers?

Dit onderwerp wordt in het hierna volgende theoretisch kader (hoofdstuk 2) verder uitgewerkt en toegelicht. Ook zal het daar bekeken worden in het licht van ander gerelateerd onderzoek. Vervolgens zal in hoofdstuk 3 de experimentele opzet van dit onderzoek besproken worden. Deze sectie bevat o.a. informatie over de proefpersonen, de materialen en de gehanteerde scoringsmethode. Ook zal daar een overzicht van de hypotheses worden gegeven. Hoofdstuk 4 wordt vervolgens gewijd aan de resultaten van het onderzoek. En in de hoofdstukken 5 en 6 volgen ten slotte de conclusie en de discussie.

2 Theoretisch kader

In deze sectie worden aan de hand van relevante literatuur tekstbegrip en de daarmee samenhangende factor woordmoeilijkheid nader toegelicht. Centraal hierin staan de termen *tekstbegrip*, *woordmoeilijkheid* en *cloze-toets*.

2.1 Tekstbegrip

Laten we om te beginnen eens duidelijk voor ogen krijgen wat nu eigenlijk verstaan wordt onder de term *tekstbegrip*. Wanneer we een tekst lezen construeren we in ons hoofd een bepaalde voorstelling van die tekst. Deze weergave wordt ook wel een mentale representatie van zo'n tekst genoemd (Sanders & Van Wijk, 2002). Je kunt je natuurlijk wel voorstellen dat zo'n tekstrepresentatie enorm kan variëren van persoon tot persoon: iedereen interpreteert een tekst immers naar eigen mogelijkheden en op een eigen manier. Dé tekstrepresentatie bestaat dan ook niet. Het is bovendien zo dat er in beginsel eigenlijk al sprake is van 3 verschillende, maar toch aan elkaar gerelateerde representatieniveaus. Volgens Fletcher (1994) is er bij de oppervlakterepresentatie sprake van een constructie van de letterlijke vorm en woorden van een tekst, gaat het bij de representatie op tekstbasisniveau om de betekenis van de tekst en staat in het situatiemodel voornamelijk het integreren van tekstuele informatie met voorkennis centraal. Vooral dit laatste representatieniveau, het situatiemodel, is een belangrijke factor voor tekstbegrip: pas wanneer men in staat is belangrijke tekstuele informatie te koppelen aan eigen kennis, wereldkennis en context en dit samen kan brengen in een samenhangend netwerk, is er werkelijk sprake van tekstbegrip (Land, 2009). Het is dus vooral dit niveau dat van belang is als het gaat om het meten van tekstbegrip.

Nu we wat meer zicht hebben op wat tekstbegrip inhoudt, moeten we ons afvragen welke factoren tekstbegrip kunnen beïnvloeden. Daarbij kan een onderscheid gemaakt worden tussen tekstfactoren en factoren die afkomstig zijn van de lezer (Stahl e.a., 1989). Tot de tekstfactoren behoren o.a. aspecten als woordmoeilijkheid en zinsbouw, maar ook structurele kenmerken als de mate van coherentie in een tekst en de lay-out behoren hiertoe. Voorkennis en woordenschat zijn daarentegen typische lezerskenmerken. Nu kunnen deze factoren afzonderlijk of in combinatie met elkaar van grote invloed zijn op het begrip van een tekst. Zo laat Stahl (2003) zien dat kennis van woorden gepaard gaat met

wereldkennis en kennis over de verschillende contexten waarin woorden kunnen worden gebruikt. Veel kennis van woorden en de wereld zorgen dientengevolge dus voor een positief effect op tekstbegrip. Echter tekstbegrip kan ook negatief worden beïnvloed. Zo kan de context in een tekst incompleet of misleidend zijn en op die manier ongunstige effecten hebben op tekstbegrip (Konopak, 1988).

Om nu het tekstbegrip van een tekst te voorspellen, wat in het onderwijs nogal belangrijk is in de keuze voor een bepaalde tekst, maakt men gebruik van zogenaamde leesbaarheidsformules (Stahl, 2003). Deze formules geven aan de hand van enkele factoren een schatting van de moeilijkheid van een tekst. Vaak wordt er dan gekeken naar zinslengte en woordfactoren als woordlengte en woordfrequentie (Stahl, 2003); hoe langere woorden en zinnen en hoe minder frequent een woord voorkomt, des te moeilijker wordt over het algemeen een tekst ingeschat. Naast het feit dat langere woorden en zinnen niet altijd moeilijker hoeven te zijn (bijvoorbeeld *uitnodiging* tegenover *smart*), richten dergelijke leesbaarheidsformules zich enkel op tekstfactoren (Stahl, 2003). En dan zelfs op slechts een klein aantal tekstkenmerken, want terwijl de focus ligt op woord- en zinsniveau wordt er in deze leesbaarheidsformules aan andere tekstfactoren (bijvoorbeeld coherentie) geen aandacht besteed. Aangezien tekstbegrip daarnaast ook beïnvloed wordt door lezerskenmerken (Stahl e.a., 1989), zou een methode die de leesbaarheid van een tekst voorspelt idealiter ook met dit type factoren rekening moeten houden: naast aspecten als woordmoeilijkheid en zinsbouw moet ook aandacht geschonken worden aan de voorkennis, de woordenschat en het leesvaardigheidsniveau van de lezer. Deze lezerskenmerken variëren van persoon tot persoon, waardoor verschillende lezerstypen ontstaan; door bijvoorbeeld een verschil in voorkennis over de ruimtevaart zal een tekst over dit onderwerp voor iemand met veel voorkennis hiervan gemakkelijker zijn dan voor een lezer die van dit onderwerp weinig af weet. Een leesbaarheidsformule lijkt hier geen rekening mee te houden en gaat bij het bepalen van de leesbaarheid van een tekst uit van een gemiddelde lezer (Stahl, 2003). Het feit dat lezerskenmerken bij het inschatten van de tekstmoeilijkheid vrijwel geen aandacht krijgen, heeft vooral te maken met de werkwijze van leesbaarheidsformules: met een leesbaarheidsformule moet snel kunnen worden bepaald of een lezer een tekst zou kunnen begrijpen en het uitgebreid toetsen van iemands leesvaardigheid en woordenschat zal in zo'n geval ten koste gaan van deze snelheid.

Tot slot maakt de leesbaarheidsformule duidelijk dat het moeilijker maken van de woorden in een tekst leidt tot een slechter tekstbegrip (Stahl e.a., 1989). Woordmoeilijkheid heeft dus invloed op de leesbaarheid van een tekst. Echter in de meeste experimentele onderzoeken wordt de invloed van woordmoeilijkheid op tekstbegrip vaak bewust in combinatie met andere factoren onderzocht. Zo onderzochten Stahl e.a. (1989) de interactie tussen voorkennis en woordmoeilijkheid als effect op tekstbegrip. Echter gezien het belang van de moeilijkheidsgraad van woorden in een tekst voor het tekstbegrip en de plaats die teksten in het onderwijs innemen, is onderzoek naar hetgeen woordmoeilijkheid als onafhankelijke factor doet met tekstbegrip van belang. Nu ontcom je er hierbij niet aan dat naast woordmoeilijkheid ook lezerskenmerken als leesvaardigheid en woordenschat van invloed kunnen zijn op tekstbegrip; verschillende lezerstypen begrijpen een tekst immers ieder op hun eigen manier. Echter in het hier gepresenteerde onderzoek zal de factor woordmoeilijkheid wel zoveel mogelijk centraal staan.

2.2 Woordmoeilijkheid

Om iets te kunnen zeggen over de invloed van woordmoeilijkheid op tekstbegrip is het allereerst van belang om vast te stellen wat voor soort kenmerk woordmoeilijkheid eigenlijk is: een tekstfactor of een lezerskenmerk? In de vorige paragraaf werd woordmoeilijkheid kortweg beschouwd als een tekstkenmerk. En in zekere zin klopt dat ook: in een tekst kunnen heel concreet de moeilijke woorden worden aangewezen. Echter wanneer het gaat om tekstbegrip speelt, zoals gezegd, ook het type lezer een rol (Stahl e.a., 1989); de woorden die voor de ene lezer moeilijk zijn, zijn dat voor een andere juist weer niet. Dit vanwege verschillen in voorkennis, woordenschat en leesvaardigheid. Er is dus met andere woorden een link tussen de betekenis van een woord en de kennis van de lezer over dat woord; weinig kennis bij de lezer, bijvoorbeeld bij een woord in een specifieke context, zorgt voor woordmoeilijkheid. Woordmoeilijkheid zou dus eigenlijk gezien moeten worden als een interactie tussen tekst en lezer (Stahl e.a., 1989). Echter over het algemeen beschouwt men het meestal als een tekstkenmerk. En als zodanig zal woordmoeilijkheid in dit onderzoek ook behandeld worden. Belangrijk nog om te weten is dat Stahl e.a. (1989) geen interactie vonden tussen voorkennis en woordmoeilijkheid. De voorkennis van een lezer bij een bepaalde tekst heeft dus geen effect op het effect van woordmoeilijkheid. Vandaar dat in dit

onderzoek het aspect voorkennis achterwege zal blijven en de aandacht zal liggen op woordenschat en leesvaardigheid als lezerskenmerken.

Een ander belangrijk punt is de vraag hoe de woordmoeilijkheid van een tekst het beste gemeten kan worden. Vaak gebeurt dit door te kijken naar de woordlengte en de woordfrequentie in een tekst (Stahl, 2003). Echter dit levert nogal wat problemen op. Vooral woordlengte is een zwakke maat voor de woordmoeilijkheid van een tekst (Stahl, 2003): zoals het voorbeeld met de woorden *uitnodiging* en *smart* al liet zien, hoeven langere woorden lang niet altijd moeilijker te zijn dan kortere woorden. Om te voorkomen dat een tekst op deze manier onterecht als moeilijk wordt bestempeld, zal woordlengte in dit onderzoek niet gebruikt worden als maat voor woordmoeilijkheid. Het kijken naar de woordfrequentie, dus hoe frequent een woord over het algemeen voorkomt, is daarentegen ook niet geheel probleemloos. Zo houdt woordfrequentie geen rekening met meerdere betekenissen van een woord (Stahl, 2003); een woord dat vaak voorkomt heeft wellicht in een bepaalde context een hele andere, vaak minder bekende, betekenis. Omdat het woord dus vaak voorkomt, wordt het ook in die ene context niet beschouwd als een moeilijk woord. Daarnaast worden ook eigennamen, vanwege hun vaak lage frequentie, gezien als moeilijke woorden, terwijl ze voor de lezer nauwelijks tot problemen leiden (Stahl, 2003). Bovendien kun je discussiëren over wanneer een woord nu wel of niet frequent genoemd kan worden; het wel of niet in een woordfrequentielijst staan is daarvoor een ruwe methode (Stahl, 2003). Hoewel er aan woordfrequentie als maat voor de woordmoeilijkheid van een tekst dus enkele nadelen kleven, is het in het hedendaagse onderzoek een nog altijd veelgebruikte methode. En aangezien woordmoeilijkheid in dit onderzoek slechts op hele grove wijze zal worden bepaald (er wordt enkel onderscheid gemaakt tussen makkelijk en moeilijk), zal er hier van de woordfrequentiemethode weinig nadeel worden ondervonden.

2.3 Cloze-toets

Nu we wat meer zicht hebben gekregen op de termen *tekstbegrip* en *woordmoeilijkheid* kan de vraag gesteld worden hoe de invloed van woordmoeilijkheid op tekstbegrip het beste gemeten kan worden. Zoals ook het onderzoek van Kamalski (2007) laat zien, zijn er verschillende manieren om tekstbegrip te toetsen: via begripsvragen, een cloze-toets, een sorteertaak of aan de hand van een mental model taak. Belangrijk bij het meten van tekstbegrip is het gegeven dat de begripstoets zo objectief mogelijk moet zijn en dat men

niet te veel productieve vaardigheden van een lezer vraagt, het gaat immers om tekstbegrip en niet om tekstproductie. Begripsvragen vallen in dat opzicht al af. Hoewel het in het onderwijs een veelgebruikte methode is, verliezen begripsvragen snel hun objectiviteit doordat niet alle vragen even moeilijk zijn. Het valt dan ook sterk te betwijfelen of iedere vraag wel echt tekstbegrip meet (Kamalski, 2007). Een cloze-toets, waarbij er op de open plekken in een tekst een woord moet worden ingevuld, is als begripstoets daarentegen een stuk objectiever (Kamalski, 2007). De vragen zijn in dit geval de gaten, dus de moeilijkheid van de vragen speelt hier geen rol. Wel kun je je bij deze methode afvragen of het lezen en begrijpen van een tekst een beroep doet op dezelfde cognitieve processen als die gebruikt worden bij het invullen van de gaten in een tekst. Het antwoord hierop is: nee, niet helemaal. Het invullen van gaten vergt natuurlijk een bepaalde mate van taalproductie, terwijl het lezen en begrijpen van een tekst receptieve vaardigheden zijn. Echter geen enkele toetsmethode meet voor de volle 100 % tekstbegrip. In die zin is er dus op iedere methode wel iets aan te merken. Het belangrijkste is, en dat laat Kamalski (2007) in haar onderzoek ook zien, dat er met de cloze-toets op een betrouwbare en valide manier tekstbegrip op situatiemodelniveau kan worden gemeten. Een valide meetmethode, die meet wat je ook echt wilt meten, is volgens Kamalski (2007) de voorwaarde om betrouwbare conclusies te kunnen trekken over het effect van een onafhankelijke variabele op de afhankelijke variabele. Een bepaalde score zou bij het gebruik van een valide meetmethode dus een reflectie moeten zijn van het onderliggende theoretische construct dat je als onderzoeker wilt meten (Kamalski, 2007). Gezien daarnaast het feit dat de sorteertaak en de mental model taak nog vrij nieuw zijn in dit onderzoeksveld en men met de cloze-toets bekend is in het onderwijs, wordt in dit onderzoek de cloze-toets gebruikt om tekstbegrip te meten.

2.3.1 Wat voor soort cloze?

Met de keuze om tekstbegrip te gaan meten met een cloze-toets zijn we er nog niet helemaal. Er zijn namelijk 3 typen clozetoetsen, die variëren wat betreft de presentatie en keuze van de gaten: de fixed-ratio cloze, de rational fill-in cloze en de rational multiple-choice cloze (Abraham & Chapelle, 1992). Bij de fixed-ratio cloze worden gaten gevormd door enkel steeds het zoveelste woord in een tekst weg te laten. De rational fill-in cloze let erop of er voor de weggehaalde woorden nog voldoende aanwijzingen in de tekst staan om ze in te vullen en of er voor het merendeel van de gaten ook echt naar informatie buiten de

zin moet worden gezocht (Abraham & Chapelle, 1992). Ook bij de rational multiple-choice cloze worden de gaten in een tekst op deze manier geselecteerd. Daarbij heeft de lezer bij het invullen van de gaten steeds de keuze uit 4 antwoordmogelijkheden (Abraham & Chapelle, 1992). Aangezien het gaat om het meten van tekstbegrip ligt de voorkeur bij de rational fill-in cloze. Doordat bij deze cloze-toets niet steeds het zoveelste woord wordt weggelaten, maar juist heel zorgvuldig naar geschikte gaten wordt gezocht, ben je beter in staat om de tekst zelf te sturen. Op die manier kunnen gaten die voorkennis of grammaticale kennis van de lezer vragen, en dus losstaan van tekstbegrip, vermeden worden. Door nu gaten te kiezen waardoor de lezer min of meer gedwongen wordt om de context in de tekst te gebruiken, zal deze zelf verbanden moeten zien te leggen tussen tekstdelen en wordt er op die manier veel sterker tekstbegrip gemeten (Abraham & Chapelle, 1992). Dit is nog meer het geval wanneer je de lezer zelf een antwoord laat invullen dan wanneer deze de keuze heeft uit een aantal antwoordmogelijkheden. Op dat punt heeft de rational fill-in cloze dus de voorkeur boven de rational multiple-choice cloze. De vaak geuite kritiek op cloze-toetsen dat ze tekstbegrip te lokaal zouden meten (Stahl e.a., 1989), wordt hiermee toch enigszins afgezwakt. Het meten van tekstbegrip op een globaler niveau wordt in ieder geval niet bereikt wanneer steeds het zoveelste woord in een tekst wordt weggelaten, omdat de kans dan immers groot is dat er ook gaten bij zitten die niet hoofdzakelijk tekstbegrip meten. Een cloze-gat zou bijvoorbeeld terecht kunnen komen op een lidwoord, waarbij de lezer in dat geval aan het erop volgende zelfstandige naamwoord al genoeg heeft om een correct antwoord in te vullen. Op die manier meet je geen tekstbegrip, maar wordt er een beroep gedaan op grammaticale kennis; buiten die ene zin is de tekst immers verder niet nodig om het cloze-gat in te vullen. Ook wordt er geen tekstbegrip gemeten bij gaten die de lezer op basis van voorkennis zonder de context in de tekst kan invullen. Met andere woorden, het is bij het meten van tekstbegrip van belang om gaten te kiezen waarbij bij het invullen ervan gebruik moet worden gemaakt van de context en de grotere verbanden in de tekst. Van de hier besproken cloze-toetsen meet de rational fill-in cloze dan ook het beste tekstbegrip.

2.3.2 Factoren die de prestaties op cloze-items beïnvloeden

Ieder gat in een cloze-toets heeft zijn eigen kenmerken, waardoor er onder de gaten van een tekst verschillende typen ontstaan. De kenmerken van deze items kunnen mede met de keuze van de gaten invloed hebben op de moeilijkheid van een in te vullen cloze-item

(Abraham & Chapelle, 1992). Naast tekst- en lezerskenmerken spelen dus ook itemkenmerken een rol. Volgens Abraham en Chapelle (1992) gaat het hierbij o.a. om de volgende factoren: de plaats van contextuele aanwijzingen ten opzichte van het in te vullen gat, de zinslengte van de zin waarin het gat staat, de frequentie van het in te vullen woord in de tekst, het feit dat functiewoorden makkelijker in te vullen zijn dan inhoudswoorden, de woordlengte van het in te vullen woord, het aantal mogelijke antwoorden binnen een gat, het feit dat woorden ook nog verbogen kunnen worden en tot slot de vraag of er gekozen kan worden uit een aantal opties (multiple-choice) of dat er zelf een woord moet worden ingevuld. Dit zijn allemaal kenmerken die losstaan van de (voor)kennis van de lezer en enkel uit de tekst zelf te halen zijn. Abraham en Chapelle (1992) noemen ze ook wel intrinsieke factoren. In hun onderzoek vonden ze bij de rational fill-in cloze een vijftal itemkenmerken die invloed hadden op de toetsresultaten: de beschikbaarheid van de context, woordlengte, functie- en inhoudswoorden, het aantal mogelijke antwoorden en het aantal mogelijke vormen. Aangezien het in dit onderzoek gaat om het meten van tekstbegrip zou je vooral van itemkenmerken als context, woordfrequentie in de tekst en inhoudswoorden willen zien dat ze toetsresultaten beïnvloeden, omdat er vooral bij deze kenmerken verder moet worden gekeken dan die ene zin waarin het cloze-item staat; voor een goed begrip van de tekst heeft men hier echt de tekst als geheel nodig. En dat is waar het om gaat bij tekstbegrip: hoe goed is een lezer in staat om aanwijzingen uit de context te halen.

De manier waarop al dit soort itemkenmerken worden vormgegeven, bepaalt in grote mate de moeilijkheid van een in te vullen cloze-item en daarmee dus ook de prestaties op de cloze-toets (Abraham & Chapelle, 1992). Met andere woorden, itemkenmerken kunnen al dan niet sturend werken op tekstbegrip; hoe vaker een woord in de tekst voorkomt, hoe makkelijker het als cloze-item is in te vullen. Op die manier kunnen cloze-items in moeilijkheid verschillen, wat vervolgens van invloed kan zijn op tekstbegrip. Echter er zijn ook enkele itemkenmerken die tekstbegrip niet beïnvloeden; het aantal mogelijke vormen van een woord en de lengte van het in te vullen woord zeggen niets over het feit of een tekst nu wel of niet begrepen is. Itemkenmerken kunnen dus de moeilijkheid van een cloze-item beïnvloeden, maar zijn lang niet altijd van invloed op tekstbegrip. Naast de intrinsieke factoren zijn tot slot bij het invullen van de cloze-items ook de (voor)kennis van de lezer en zijn strategieën om achter het juiste woord te komen als extrinsieke factoren van belang (Abraham & Chapelle, 1992). Ook hier werken tekstfactoren en lezerskenmerken dus samen.

2.3.3 Welke scoringsmethode?

De manier waarop cloze-toetsen worden gecorrigeerd is cruciaal voor het trekken van conclusies over het tekstbegrip van een lezer. Over het algemeen worden er 3 verschillende scoringsmethoden onderscheiden: de exact-word methode, de methode waarbij antwoorden zowel semantisch als syntactisch acceptabel moeten zijn en de methode waarbij antwoorden enkel semantisch goed moeten zijn (Kobayashi, 2002). Als het gaat om het meten van tekstbegrip is de exact-word methode in dit geval niet zo geschikt als scoringsmethode. Een lezer die in een cloze-toets het woord *woning* invult daar waar *huis* bedoeld was, krijgt ondanks het feit dat hij heeft laten zien de tekst begrepen te hebben onterecht toch een punt aftrek. Vooral voor lezers met een hoge taalvaardigheid is dit een nadelige methode (Kobayashi, 2002): een grote woordenschat en veel taalkennis brengt ze geen voordelen, sterker nog ze worden er zelfs op afgestraft. Het is dus sowieso verstandiger om meerdere juiste antwoorden goed te rekenen. Er moet dan alleen nog beslist worden of men de antwoorden al dan niet syntactisch correct wil zien. In het licht van het meten van tekstbegrip geeft Kobayashi (2002) aan dat het verstandiger is om antwoorden die syntactisch gezien fout zijn maar qua betekenis wel aansluiten op de context, als correct te beschouwen; grammaticale en syntactische fouten zeggen immers niks over iemands tekstbegrip. De methode waarbij antwoorden enkel semantisch correct moeten zijn, is dus het meest geschikt om tekstbegrip te meten bij verschillende lezerstypen.

2.4 Belang van het onderzoek

Wat tekstbegrip betreft is de invloed van woordmoeilijkheid hierop in onderzoeken uit het verleden voornamelijk bekeken in combinatie met een andere factor, bijvoorbeeld voorkennis van de lezer. Gezien het belang van woordmoeilijkheid in een tekst voor het tekstbegrip en de plaats die teksten in het onderwijs innemen, is het van belang om de aandacht eens wat specifiek te richten op wat woordmoeilijkheid doet met tekstbegrip. Bovendien moet hierbij een onderscheid gemaakt worden tussen zwakke en sterke lezers en moet er bekeken worden of woordmoeilijkheid op deze lezerstypen al dan niet eenzelfde invloed uitoefent, om zo ook lezerskenmerken in het onderzoek te betrekken. De resultaten ervan kunnen uiteindelijk niet alleen belangrijk zijn voor docent en leerling, maar ook voor de samenstellers van lesmethoden. Dit onderzoek heeft dus zeker zijn waarde.

3 Experimentele opzet

In dit hoofdstuk zal de opzet van het onderzoek nader worden toegelicht. Achtereenvolgens worden het type onderzoek, de proefpersonen, de materialen en de gehanteerde werkwijzen rondom de afname en de analyse van het experiment besproken. Tot besluit zal nog een overzicht van de hypothesen worden gegeven.

3.1 Onderzoeksopzet

Het hier gepresenteerde onderzoek naar de invloed van woordmoeilijkheid op het tekstbegrip van zwakke en sterke lezers is een experimenteel onderzoek. In een 2x2-design is de variatie in woordmoeilijkheid en lezerstype vormgegeven: naast een cloze-toets met makkelijke woorden ook een cloze-toets met moeilijke woorden en dit verspreid over leerlingen uit de 2^e klas van het vmbo en het vwo. De onafhankelijke variabele in dit onderzoek is dus woordmoeilijkheid, waarvan er een effect verwacht wordt op de afhankelijke variabele tekstbegrip. Om nu te kunnen bepalen of een effect op tekstbegrip ook echt ontstaat vanwege de manipulatie van de woordmoeilijkheid en niet vanwege onderlinge verschillen tussen bijvoorbeeld de vwo-klassen wat betreft leesvaardigheidsniveau, zijn er in dit onderzoek naast de cloze-toets nog twee toetsen afgenomen, waarvan de afgeleide variabelen dienen als covariaat: een leesvaardigheidstoets (ontwikkeld door het CITO) en een woordenschattoets (ontwikkeld door de Rijksuniversiteit Groningen). Bovendien zal hieruit blijken of vwo-leerlingen ook echt sterkere lezers zijn dan vmbo-leerlingen.

3.1.1 Proefpersonen

Voor dit onderzoek is er een beroep gedaan op 84 leerlingen uit vier 2^e klassen (2 vmbo- en 2 vwo-klassen) van het voortgezet onderwijs (in totaal 42 vmbo-leerlingen en 42 vwo-leerlingen). Het onderzoek heeft plaatsgevonden op de Philips van Horne Scholengemeenschap, een middelbare school in het Limburgse Weert. In totaal ging het om 43 jongens en 41 meisjes, allen met een leeftijd van 13 of 14 jaar. Alle leerlingen maakten dezelfde leesvaardigheidstoets en woordenschattoets. De makkelijke en de moeilijke versie van de cloze-toets zijn aselekt over de klassen verdeeld, zo dat de ene vmbo-klas de makkelijke versie kreeg en de andere vmbo-klas de moeilijke. Hetzelfde gold voor de vwo-

klassen. Tussen de twee vmbo-klassen zijn er verder geen significante verschillen wat betreft woordenschat ($t(29) = -1.14$, $p = .263$) en leesvaardigheid ($t(29) = -1.75$, $p = .090$). Ook de twee vwo-klassen verschillen niet significant in woordenschat ($t(37) = -2.01$, $p = .052$) en leesvaardigheid ($t(37) = -1.21$, $p = .235$). Tijdens het afnemen van het experiment waren 14 leerlingen gedeeltelijk of geheel niet aanwezig in de klas (6 jongens en 5 meisjes uit de vmbo-klassen en 1 jongen en 2 meisjes uit de vwo-klassen). Deze leerlingen zijn in het onderzoek verder buiten beschouwing gelaten, wat resulteert in een overgebleven groep van in totaal 31 vmbo-leerlingen en 39 vwo-leerlingen (36 jongens en 34 meisjes).

3.1.2 Materiaal

Als basis voor de cloze-toets is gebruik gemaakt van een geschiedenis tekst uit de Nederlandse Historische Canon, op het niveau van de onderbouw van het voortgezet onderwijs. De gebruikte tekst in dit onderzoek heeft als titel 'Koning Willem I' (bijlage 1). Om van deze tekst een geschikte cloze-toets te maken, is allereerst het aantal woorden behoorlijk verminderd; van 393 woorden naar maximaal 270 woorden. Vervolgens zijn er op deze tekst woordmanipulaties toegepast, waardoor er een makkelijke en een moeilijke versie van de tekst is ontstaan (bijlage 2 en 3). Wat betreft deze manipulatie van de woordmoeilijkheid (bijlage 4) zijn er enkel inhoudswoorden gemanipuleerd; zelfstandige naamwoorden, werkwoorden en adjectieven zijn door hun semantische inhoud moeilijker direct uit het geheugen te halen dan de op syntactische informatie gebaseerde functiewoorden en hebben dus voor hun correcte interpretatie de hulp van de tekst meer nodig (Abraham & Chapelle, 1992). Door nu juist dit type woorden te manipuleren, wordt er een sterker beroep gedaan op tekstbegrip.

Bij de manipulatie is verder uitgegaan van minimaal 1 woordmanipulatie per 10 woorden, waarbij er hoofdzakelijk naar synoniemen voor de te vervangen woorden is gezocht. Op deze manier is het aantal woorden en de syntactische structuur van de makkelijke en de moeilijke versie zoveel mogelijk gelijk gehouden. Daarnaast is bij de manipulaties ook gekeken naar de woordfrequentie (bijlage 5): in de moeilijke tekst is grotendeels gekozen voor woordsynoniemen die in ons taalgebruik over het algemeen minder frequent voorkomen, maar wel nog natuurlijk taalgebruik opleveren. Echter aangezien er met woordfrequentie qua definiëring en operationalisering nogal wat problemen kunnen optreden (Stahl, 2003), is deze methode bij de woordmanipulaties slechts als extra check gebruikt om makkelijke en

moeilijke woorden te onderscheiden. Het lastige aan de woordfrequentiemethode is namelijk het feit dat woorden die relatief veel voorkomen niet altijd ook makkelijker zijn; *buffer* komt veel frequenter voor dan *beschermingsgebied* (bijlage 5), terwijl van dit laatste woord de betekenis veel gemakkelijker is af te leiden. Het probleem zit hem hier in het feit dat *buffer* in verschillende contexten weer andere betekenissen heeft. Tussen deze verschillende betekenissen maakt de woordfrequentiemethode echter geen onderscheid, met als gevolg dat het woord *buffer* hier eigenlijk een veel te hoge frequentie heeft en zo een vertekend beeld geeft van de woordmoeilijkheid. Daarnaast is het bij de woordfrequentiemethode niet duidelijk onder welke woordfrequentiegrens een woord voor een 2^e-klasser op het vmbo of vwo moeilijk wordt. Het is dus raadzaam om woordfrequentie bij de woordmanipulaties met de nodige voorzichtigheid te benaderen.

Na het manipuleren van de woordmoeilijkheid zijn de makkelijke en de moeilijke tekstversie omgegoten tot een rational fill-in cloze-toets. Om de keuze van de gaten en ook het aantal te bepalen, is gebruik gemaakt van de regels voor het ontwerpen van cloze-toetsen zoals Pander Maat en Kraf die hebben opgesteld (bijlage 6). Toepassing van deze regels heeft geleid tot een cloze-toets waarin verspreid over de tekst 20 woorden zijn weggelaten (bijlage 4). Dit zijn voor zowel de makkelijke als de moeilijke versie dezelfde cloze-items (bijlage 7 en 8). De beide tekstversies verschillen dus alleen wat betreft woordmoeilijkheid. Naast een makkelijke of moeilijke versie van de cloze-toets kreeg elke leerling zoals gezegd ook nog een leesvaardigheidstoets (ontwikkeld door het CITO) en een woordenschattoets (ontwikkeld door de Rijksuniversiteit Groningen).

3.1.3 Procedure

Het onderzoek heb ik samen met mijn medestudente Jos Guldemonnd op drie opeenvolgende dagen afgenomen. Alle klassen kregen twee volle lessen van 50 minuten de tijd om de toetsen te maken. Hierbij kreeg een van de vwo-klassen alle toetsen op één dag, weliswaar met een tussenuur tussen de twee toetsmomenten. De andere drie klassen maakten de toetsen verspreid over twee dagen. Roostertechisch gezien was dit de enige mogelijkheid. Aan het begin van ieder uur kreeg de klas steeds een korte instructie waarin verteld werd wat er van hen verwacht werd en hoeveel tijd er voor de toetsen gepland stond. Aangezien de toetsen in de verschillende klassen in duo's zijn afgenomen, maakte elke leerling naast de gezamenlijke leesvaardigheidstoets en woordenschattoets twee cloze-toetsen: de

makkelijke versie van de ene persoon en de moeilijke versie van de andere, of dit juist net omgekeerd (bijlage 9). Om een mogelijk volgorde-effect te voorkomen, zijn de klassen in tweeën verdeeld: de ene helft kreeg eerst de tekst over *Koning Willem I* en de andere helft startte met de tekst van Jos. In dit onderzoek zullen haar cloze-toets en de bijbehorende resultaten verder buiten beschouwing worden gelaten.

3.1.4 Scoringsmethode

Voor de beoordeling van de cloze-toetsen is uitgegaan van de methode waarbij antwoorden enkel semantisch correct moeten zijn, aangezien deze scoringsmethode het meest geschikt is om tekstbegrip te meten bij verschillende lezerstypen (Kobayashi, 2002). Bijlage 10 geeft, naast de exacte antwoorden van de cloze-toets, een overzicht van de antwoorden die als semantisch correct zijn goedgekeurd. Voor de verdere analyse van de resultaten op de cloze-toetsen is, naast de codering 0 = niet gemaakt/fout antwoord en 1 = semantisch goed/letterlijk goed, ook nog de volgende codering toegepast:

0 = niet gemaakt/fout antwoord

1 = semantisch goed

2 = letterlijk goed

Op deze manier kan later in de resultatenanalyse altijd nog een onderscheid worden gemaakt tussen antwoorden die letterlijk goed zijn en antwoorden die semantisch acceptabel zijn, wat bijvoorbeeld iets kan zeggen over welke scoringsmethode de betrouwbaarste cloze-toetsen oplevert.

3.2 Hypotheses

In dit onderzoek wordt dus gekeken naar de invloed van woordmoeilijkheid op het tekstbegrip van zwakke en sterke lezers, waarbij deze verschillende lezerstypen zijn vertaald naar respectievelijk vmbo- en vwo-leerlingen. Met behulp van eerdere onderzoeksresultaten en kennis op dit gebied kunnen hier de volgende hypothesen worden geformuleerd:

H1: *Teksten met makkelijke woorden worden beter begrepen dan teksten met moeilijke woorden, zowel door vmbo- als door vwo-leerlingen.*

Deze veronderstelling is gebaseerd op het feit dat ook eerdere onderzoeken laten zien dat het moeilijker maken van de woorden in een tekst leidt tot een slechter tekstbegrip (o.a. Stahl e.a., 1989). Ook in de veelgebruikte leesbaarheidsformules wordt de moeilijkheid van een tekst in grote mate bepaald door de factor woordmoeilijkheid (Stahl, 2003). De verwachting is hier dan dus ook dat er ongeacht het schoolniveau op de makkelijke cloze-toets een hogere score zal worden behaald dan op de moeilijke cloze-toets en dat woordmoeilijkheid dus invloed heeft op de leesbaarheid van een tekst.

H2: *Gemiddeld behalen vwo-leerlingen zowel op de makkelijke als op de moeilijke cloze-toets betere resultaten dan vmbo-leerlingen.*

Deze hypothese vindt haar ondersteuning in het gegeven dat vmbo-leerlingen meer moeite hebben met het lezen van schoolteksten dan vwo-leerlingen, omdat zij over het algemeen minder goede lezers zijn (Jansma, Van Kleunen & Leenders, 2011). Naar verwachting zullen de vwo-leerlingen als sterkere lezers op beide tekstversies dan ook hoger scores.

H3: *De resultaten op de makkelijke en de moeilijke cloze-toets zullen bij vmbo-leerlingen meer van elkaar verschillen dan bij vwo-leerlingen.*

Gezien het feit dat vmbo-leerlingen in het algemeen getypeerd kunnen worden als zwakke lezers (Jansma, Van Kleunen & Leenders, 2011), is de verwachting dat een manipulatie van de woordmoeilijkheid op dit lezerstype meer invloed heeft dan op vwo-leerlingen. Vmbo'ers hebben vaak te kampen met woordenschatproblemen (Jansma, Van Kleunen & Leenders, 2011), waardoor er een gebrek is aan kennis van woorden en kennis over de verschillende contexten waarin woorden kunnen worden gebruikt. En laat veel kennis van woorden nu net zorgen voor een positief effect op tekstbegrip (Stahl, 2003). Hieruit volgt de verwachting dat vmbo-leerlingen op de moeilijke cloze-toets veel lager zullen scoren dan op de makkelijke cloze-toets. Aangezien vwo-leerlingen als sterkere lezers beschikken over een betere woordenschat zullen ze niet alleen op de makkelijke cloze-toets goede resultaten behalen, maar ook op de moeilijke. Vandaar de verwachting dat bij vmbo-leerlingen de resultaten op de beide cloze-toetsen meer van elkaar zullen verschillen dan bij vwo-leerlingen.

Het zou mooi zijn als deze hypotheses door het hier gepresenteerde onderzoek zouden kunnen worden bevestigd en de hierop aansluitende nulhypothese kan worden verworpen:

H0: *Woordmoeilijkheid heeft geen invloed op het tekstbegrip van zwakke en sterke lezers.*

4 Resultaten

In deze sectie zullen de resultaten van het onderzoek worden besproken, waarbij er ook een terugkoppeling naar de hypothesen zal worden gemaakt. Alvorens dat, zal hier eerst nog kort worden ingegaan op het verloop van het onderzoek.

4.1 Verloop van het onderzoek

Buiten het feit dat er tijdens de afname van het experiment een aantal leerlingen gedeeltelijk of geheel afwezig was (in totaal 14 leerlingen), is het onderzoek verder goed verlopen. De docenten van de betreffende klassen waren goed op de hoogte van wat er stond te gebeuren. Ze hadden hun leerlingen verteld dat er toetsen afgenomen zouden worden waarmee het leesvaardigheidsniveau kon worden bepaald. Hieraan zouden verder geen cijfers worden verbonden, maar het zou de docent wel een goed beeld geven van hoe ver een leerling al is op het gebied van leesvaardigheid. Op deze manier gingen de leerlingen serieus met de toetsen aan de slag, waardoor er in de klas gedurende het experiment een prettige werksfeer was. Dit gold zowel voor de vmbo- als voor de vwo-klassen.

Elke klas had twee lesuren van 50 minuten de tijd om in totaal vier toetsen te maken; twee toetsen per lesuur. In elke les werd steeds begonnen met de cloze-toets, waarvoor de leerlingen 15 minuten de tijd hadden. Pas na dit kwartier of als iedereen klaar was, werd begonnen met de andere toets. Dit was steeds eerst de woordenschattoets en pas in het tweede lesuur werd de leesvaardigheidstoets gemaakt (bijlage 9). Voor deze toetsen werd dan de resterende lestijd gebruikt. Iedereen had alle toetsen binnen de tijd af, op drie leerlingen na wat betreft de leesvaardigheidstoets.

Vanwege het feit dat het voor de leerlingen duidelijk was wat er van hen verwacht werd, hebben alle klassen serieus aan dit onderzoek meegewerkt. Elk toetsmoment gingen de leerlingen vrijwel direct aan het werk, wat de indruk geeft dat ze het toch wel leuk vonden om aan dit experiment mee te doen. Ook het feit dat het qua tijd voor het merendeel van de leerlingen haalbaar was om de toetsen te maken, draagt hiertoe bij. Door de goede medewerking van zowel de docenten als de leerlingen is de afname van het onderzoek op de Philips van Horne Scholengemeenschap in Weert dus vlot verlopen.

4.2 Beschrijving van de resultaten

De resultaten van dit onderzoek zijn uiteindelijk gebaseerd op de toetsprestaties van 70 leerlingen uit de 2^e klas van het voortgezet onderwijs. Deze leerlingen waren gedurende het gehele experiment aanwezig en hebben dus ook alle toetsen gemaakt. Hierbij gaat het in totaal om 31 vmbo-leerlingen en 39 vwo-leerlingen (36 jongens en 34 meisjes). Al deze leerlingen maakten dus dezelfde leesvaardigheidstoets en woordenschattoets. Wat de cloze-toets betreft, maakten 37 leerlingen de makkelijke versie (14 vmbo'ers en 23 vwo'ers) en 33 leerlingen de moeilijke versie (17 vmbo'ers en 16 vwo'ers). Tabel 4.1 geeft een totaaloverzicht van het aantal leerlingen per cloze-toetsversie en schoolniveau.

Tabel 4.1 Aantal leerlingen per cloze-toetsversie en schoolniveau

		Schoolniveau		Totaal
		Vmbo	Vwo	
Cloze-toetsversie	Makkelijk	14	23	37
	Moeilijk	17	16	33
	Totaal	31	39	70

4.2.1 Scores op de verschillende toetsen

De leesvaardigheidstoets en woordenschattoets bestaan respectievelijk uit 25 en 50 multiple-choice vragen. Elk goed antwoord is gecodeerd met een 1 en aan een fout antwoord is een 0 gekoppeld. Bij de drie leerlingen die de leesvaardigheidstoets niet helemaal af hadden, zijn de niet gemaakte vragen gecodeerd als 99. Bij de cloze-toets, die uit 20 cloze-items bestaat, zijn de niet ingevulde gaten samen met de foute antwoorden gecodeerd als 0. Daar waar het namelijk bij de leesvaardigheidstoets heel duidelijk steeds de laatste drie tot vier vragen waren die bij drie leerlingen oningevuld bleven, ging het bij de cloze-toets ook om items in het begin of in het midden van de tekst. Een niet ingevuld cloze-gat kan in dit geval dan ook beschouwd worden als een fout antwoord; de ene leerling die het antwoord op een cloze-item niet weet, probeert toch een gok te wagen, terwijl een andere leerling besluit om niets in te vullen. Dit is een verschil met de leesvaardigheidstoets. Aangezien het daar bij drie leerlingen ging om het ontbreken van een antwoord op de laatste paar vragen, terwijl de overige leerlingen deze toets wel volledig gemaakt hebben, kunnen we er hier vanuit gaan dat deze drie leerlingen aan deze vragen niet meer zijn

toegekomen en ze daarom niet gemaakt hebben. Met andere woorden, bij de leesvaardigheidstoets kunnen de niet gemaakte vragen beschouwd worden als niet af vanwege tijdgebrek, terwijl aan niet ingevulde cloze-items wel tijd is besteed, maar deze door een leerling uiteindelijk niet gemaakt zijn. Vandaar dat een niet gemaakte vraag bij de leesvaardigheidstoets dus gecodeerd wordt met 99 en een niet ingevuld cloze-item met een 0. Wat betreft de goede antwoorden op de cloze-toets is er zoals gezegd een nog verder onderscheid gemaakt tussen antwoorden die semantisch goed zijn (gecodeerd als 1) en antwoorden die letterlijk goed zijn (gecodeerd als 2). Echter als basis is in eerste instantie steeds uitgegaan van de volgende codering:

0 = niet gemaakt/fout antwoord

1 = semantisch goed/letterlijk goed

Op deze manier wordt de nadruk vooral gelegd op het wel of niet goed hebben van een antwoord. De verdere onderverdeling naar de aard van een goed antwoord kan in een later stadium nog gebruikt worden als informatiebron over welke scoringsmethode bijvoorbeeld de betrouwbaarste cloze-toetsen oplevert.

De scores op de verschillende toetsen zijn bepaald door te kijken naar het aantal goed beantwoorde vragen. Tabel 4.2 laat deze scores per toets zien.

Tabel 4.2 Gemiddelde scores, standaarddeviaties en laagste en hoogste scores van de vier toetsen

Toets	Aantal leerlingen	Gemiddelde score	Standaarddeviatie	Laagste score	Hoogste score
Woordenschattoets	70	42,53	4,67	30	50
Leesvaardigheidstoets	70	18,56	3,96	8	25
Cloze-toets makkelijk	37	15,27	2,63	9	20
Cloze-toets moeilijk	33	12,88	3,51	6	19

Voor de woordenschattoets konden maximaal 50 punten worden behaald, wat één leerling ook daadwerkelijk gelukt is. De laagste score op de woordenschattoets was 30 punten. Ook dit resultaat kwam bij één leerling voor. Gemiddeld is er op deze toets een score van 42,53

behaald (SD = 4,67). Voor de leesvaardigheidstoets konden in totaal 25 punten worden gehaald, wat ook één leerling gelukt is. Ook de laagste score, 8 punten, kwam bij één leerling voor. Op de leesvaardigheidstoets werd verder een gemiddelde score van 18,56 behaald (SD = 3,96). Wat de cloze-toets betreft, is alleen de makkelijke versie door een leerling foutloos gemaakt. De laagste score op de cloze-toets, 6 punten, werd behaald op de moeilijke versie. Gemiddeld scoorde men voor de makkelijke cloze-toets 15,27 punten (SD = 2,63) en voor de moeilijke versie 12,88 punten (SD = 3,51).

4.2.2 Betrouwbaarheid

Om de stabiliteit van de meting naar het tekstbegrip van zwakke en sterke lezers te garanderen, is het belangrijk om te kijken naar hoe betrouwbaar de verschillende toetsen in dit onderzoek zijn. De betrouwbaarheid van een toets is dan ook een soort maat voor hoe constant een bepaalde toets is in zijn metingen; hoe meer dezelfde uitkomsten bij herhaalde metingen, des te betrouwbaarder een toets is. Tabel 4.3 geeft een overzicht van de betrouwbaarheid van de in dit onderzoek gebruikte toetsen.

Tabel 4.3 **Betrouwbaarheid van de verschillende toetsen**

Toets	Betrouwbaarheid (Alpha)	Aantal items	Buiten beschouwing gelaten items
Leesvaardigheidstoets	0,746	25	0
Cloze-toets makkelijk	0,602	20	0
Cloze-toets moeilijk	0,719	20	0
Cloze-toets (zonder versieonderscheid)	0,704	20	0

Met een betrouwbaarheid van ($\alpha = .75$) is de leesvaardigheidstoets voldoende betrouwbaar en wat betreft het meten van leesvaardigheid dus behoorlijk stabiel. Alle 25 vragen kunnen dan ook meegenomen worden in de analyse.

Bij de cloze-toets zien we iets opvallends: met een betrouwbaarheid van ($\alpha = .72$) voor de moeilijke cloze-toets is niets mis, echter de makkelijke cloze-toets heeft met ($\alpha = .60$) een veel lagere betrouwbaarheid. De betrouwbaarheid van de makkelijke versie zit, zoals tabel 4.3 laat zien, beneden de veilige ondergrens van ($\alpha = .70$). Wanneer item 7, dat slechts door

de helft van het aantal leerlingen goed beantwoord is, in de betrouwbaarheidsanalyse buiten beschouwing wordt gelaten, stijgt de betrouwbaarheid van de makkelijke versie naar ($\alpha = .63$) en die van de moeilijke versie naar ($\alpha = .73$). Ook de betrouwbaarheid van de cloze-toets als geheel neemt toe en stijgt van ($\alpha = .70$) naar een betrouwbaarheid van ($\alpha = .73$). Hoewel het weglaten van dit item de classificatie van de betrouwbaarheid niet verandert (deze blijft bij de makkelijke cloze-toets liggen tussen 0.6 en 0.7 en bij de moeilijke cloze-toets tussen 0.7 en 0.8), zal verderop in deze paragraaf worden toegelicht waarom toch besloten is om item 7 buiten de analyse te laten. Verder zijn alle vragen meegenomen in de analyse (tabel 4.4); enkel het verwijderen van de items 4, 5 en 9 zou de betrouwbaarheidscoëfficiënt van de beide cloze-toetsen nog kunnen verhogen, echter dan gaat het om slechts kleine verschillen in getal en niet om een echte verandering van de betrouwbaarheid.

Tabel 4.4 Betrouwbaarheid van de verschillende toetsen na weglating van item 7

Toets	Betrouwbaarheid (Alpha)	Aantal items	Buiten beschouwing gelaten items
Leesvaardigheidstoets	0,746	25	0
Cloze-toets makkelijk	0,631	20	1 (item 7)
Cloze-toets moeilijk	0,729	20	1 (item 7)
Cloze-toets (zonder versieonderscheid)	0,732	20	1 (item 7)

De betrouwbaarheid van de woordenschattoets wordt hier verder buiten beschouwing gelaten, omdat er bij deze toets in de analyse enkel met de totaalscore is gewerkt en niet met de 50 multiple-choice vragen apart.

De betrouwbaarheid van de makkelijke cloze-toets is opvallend laag, terwijl de betrouwbaarheid van zowel de leesvaardigheidstoets als de moeilijke cloze-toets wel voldoende is (de betrouwbaarheid ligt hoger dan ($\alpha = .70$)). Wanneer de makkelijke cloze-toets meerdere malen zou worden afgenomen, zouden we bij deze toets dus niet zo vaak dezelfde uitkomsten zien op bepaalde items als bij de moeilijke cloze-toets. Dat blijkt bijvoorbeeld ook uit het feit dat item 7, naast het gegeven dat het slechts door de helft van de leerlingen (35) goed beantwoord is, door maar liefst 23 leerlingen in de makkelijke versie

fout is gedaan en maar door 12 leerlingen in de moeilijke versie. Als je puur naar de versie kijkt, zou je dit contrast eerder andersom verwachten. Waarom dit item nu juist in de makkelijke cloze-toets zoveel is fout gedaan, is moeilijk te verklaren. De kans is hier in ieder geval groot dat item 7 van de cloze-toets niet homogeen is in datgene wat het meet: het construct tekstbegrip wordt niet echt gemeten. Vandaar ook dat bij weglating van dit item de betrouwbaarheidscoëfficiënt van de cloze-toetsen een stuk stijgt. Doordat het antwoord op item 7 niet zo gemakkelijk uit de tekst valt te halen (bijlage 7), zouden vwo-leerlingen vanwege hun grotere woordenschat hiervan nog weleens nadeel kunnen ondervinden; zij hebben immers buiten de tekst de keuze uit een groot aantal woorden. Als we de verdeling van de makkelijke cloze-toets over de leerlingen bekijken, zou dit vermoeden wel eens kunnen kloppen; 14 vmbo-leerlingen en 23 vwo-leerlingen maakten de makkelijke versie van de cloze-toets (tabel 4.1). Hiervan hadden 8 vmbo'ers item 7 fout en maar liefst 15 vwo'ers. Bij de moeilijke cloze-toets was de verdeling vanwege minder afwezigheid tijdens het experiment een stuk eerlijker: 17 vmbo- leerlingen en 16 vwo- leerlingen (tabel 4.1). Hier zien we dan ook slechts 12 fouten bij item 7, die gemaakt zijn door 6 vmbo'ers en 6 vwo'ers. Het lijkt er dus op dat item 7, vanwege het nauwelijks verbonden zijn met de tekst, vooral voor vwo-leerlingen problemen heeft opgeleverd. Dit beeld zien we versterkt terug in de betrouwbaarheid op de makkelijke cloze-toets, omdat deze versie door relatief veel vwo'ers is gemaakt. Het is in dit geval dan ook beter om item 7 uit de analyse te verwijderen.

Tot slot laat tabel 4.5 zien wat de betrouwbaarheid van de cloze-toetsen is wanneer alleen exacte antwoorden als correct worden beschouwd.

Tabel 4.5 Betrouwbaarheid van de cloze-toetsen bij alleen exact antwoord goed

Toets	Betrouwbaarheid (Alpha)	Aantal items	Buiten beschouwing gelaten items
Cloze-toets makkelijk	0,435	20	0
Cloze-toets moeilijk	0,715	20	0
Cloze-toets (zonder versieonderscheid)	0,625	20	0

Voor zowel de makkelijke en de moeilijke cloze-toets als de cloze-toets als geheel neemt de betrouwbaarheid af wanneer alleen exacte antwoorden op de cloze-items worden

goedgekeurd. Opvallend hierbij is dat de betrouwbaarheid van de moeilijke cloze-toets met ($\alpha = .715$) nauwelijks daalt. Dit betekent dat er bij deze versie naast semantisch correcte antwoorden ook veel exacte antwoorden zijn gegeven, waardoor het bij deze cloze-toets voor de betrouwbaarheid niet uitmaakt welke scoringsmethode gehanteerd is. Echter bij de makkelijke cloze-toets is duidelijk te zien dat de betrouwbaarheid een stuk lager wordt, ($\alpha = .44$), wanneer enkel exacte antwoorden goed zijn (tabel 4.5). Alle cloze-toetsen zijn in dit geval dus het betrouwbaarst wanneer ook synoniemen goed worden gerekend. Op deze manier wordt tekstbegrip ook zo eerlijk mogelijk gemeten.

4.2.3 Correlaties tussen de verschillende toetsen

Om te bepalen of de scores op de verschillende toetsen onderling samenhangen, is gekeken naar de correlaties tussen deze toetsen. Op die manier wordt duidelijk of een leerling die op de ene toets hoog heeft gescoord, dat op de andere toetsen ook heeft gedaan. Tabel 4.6 geeft een overzicht van deze correlaties.

Tabel 4.6 Correlaties tussen de verschillende toetsen

	Woordenschattoets	Leesvaardigheidstoets	Cloze-toets
Woordenschattoets	-	0,686**	0,631**
Leesvaardigheidstoets	0,686**	-	0,703**
Cloze-toets	0,631**	0,703**	-

** $p < 0,01$ (Significante correlatie)

In één oogopslag is te zien dat alle correlaties tussen de verschillende toetsen significant zijn. Bovendien is er sprake van een sterke samenhang tussen de toetsen. Zo is de correlatie tussen de scores op de woordenschattoets en op de leesvaardigheidstoets sterk positief ($r = .69$; $p < .001$). Dit wil zeggen dat leerlingen die hoog hebben gescoord op de woordenschattoets, ook een goed resultaat op de leesvaardigheidstoets hebben behaald en andersom. Ook de correlatie tussen de resultaten op de woordenschattoets en op de cloze-toets is sterk positief ($r = .63$; $p < .001$). Leerlingen met een goede score op de woordenschattoets hebben dus ook de cloze-toets goed gemaakt, en andersom hebben leerlingen met een goed resultaat op de cloze-toets dit ook gehaald op de woordenschattoets. De laatste correlatie, die tussen de scores op de leesvaardigheidstoets

en op de cloze-toets, is ook sterk positief ($r = .70$; $p < .001$). Dit betekent dat leerlingen die hoog hebben gescoord op de leesvaardigheidstoets, ook een hoge score op de cloze-toets hebben behaald en andersom. Vanwege het feit dat alle correlaties tussen de verschillende toetsen sterk positief zijn, kan gezegd worden dat een leerling die op de ene toets hoog scoort dit over het algemeen op de andere toets ook zal doen. Dit is fijn om te weten. Niet alleen omdat er vanwege deze stabiliteit voorspellingen gedaan kunnen worden over iemands toetsresultaten, maar vooral ook omdat er door de samenhang in scores op de verschillende toetsen reden is om aan te nemen dat deze toetsen stabiel zijn in datgene wat ze meten (in dit geval tekstbegrip). Alle toetsen claimen ditzelfde te meten en dat wordt hier door de sterke correlaties tussen de resultaten op de verschillende toetsen bevestigd. Met andere woorden, er is hier sprake van convergente validiteit. De door leerlingen behaalde resultaten kunnen dan ook gekoppeld worden aan hun vermogen om een tekst te begrijpen; immers beïnvloeding van de score door factoren als concentratie en de hoeveelheid tijd zullen bij een leerling niet gauw zo constant zijn op drie verschillende toetsen.

4.2.4 Woordenschat en leesvaardigheid

Om iets te kunnen zeggen over de resultaten op de woordenschat- en leesvaardigheidstoets en over de eventuele verschillen in de prestaties hierop tussen vmbo- en vwo-leerlingen, is het allereerst nuttig om naar de gemiddelde scores op deze toetsen te kijken. Tabel 4.2 gaf al een overzicht van de totaal gemiddelden. In tabel 4.7 wordt er bij de gemiddelde scores een onderscheid gemaakt tussen schoolniveau.

Tabel 4.7 Gemiddelde scores, standaarddeviaties en laagste en hoogste scores van de woordenschat- en leesvaardigheidstoets per schoolniveau

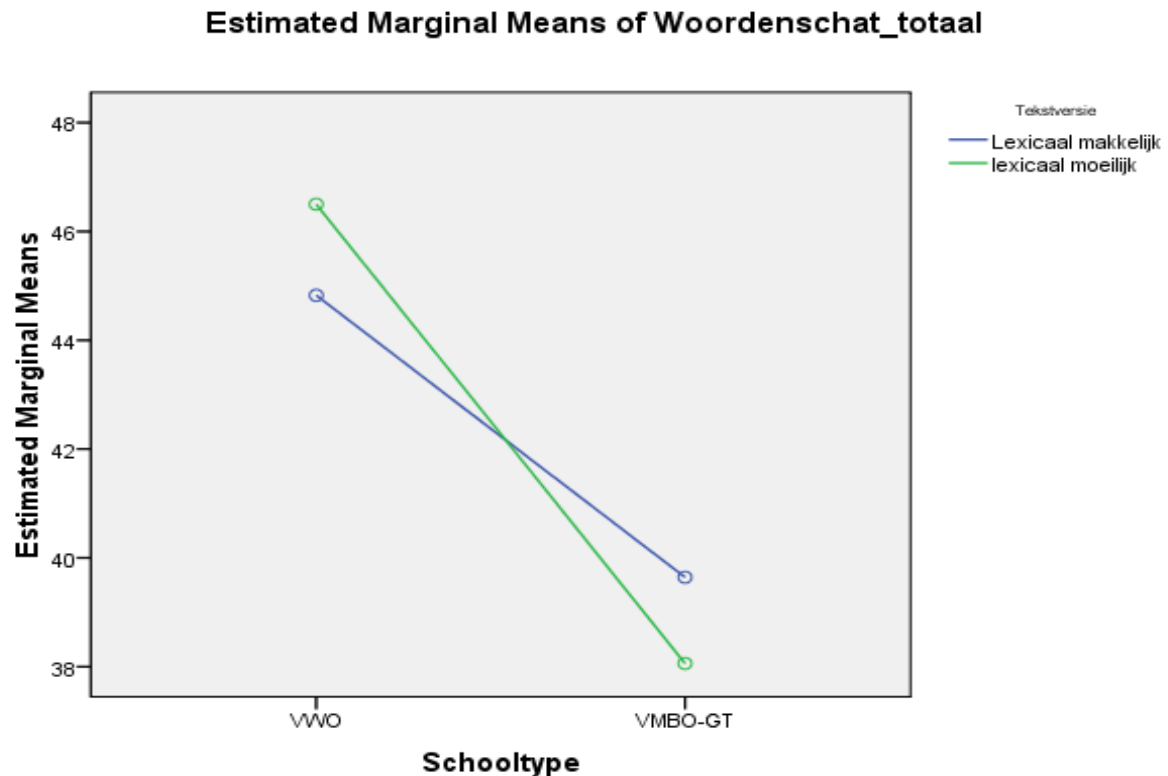
Toets	Schoolniveau (aantal leerlingen)	Gemiddelde score	Standaarddeviatie	Laagste score	Hoogste score
Woordenschattoets	Vmbo (31)	38,77	3,86	30	47
	Vwo (39)	45,51	2,66	35	50
Leesvaardigheidstoets	Vmbo (31)	15,55	3,47	8	22
	Vwo (39)	20,95	2,39	14	25

Een ANOVA-analyse (Univariate) laat zien dat de resultaten op de woordenschattoets naar gelang het schoolniveau van elkaar verschillen ($F(1, 66) = 77.24, p < .001$). De gemiddelde score op de woordenschattoets behaald door vmbo-leerlingen ($M = 38,77$ en $SD = 3,86$) verschilt dus significant van de gemiddelde score die vwo-leerlingen op deze toets hebben behaald ($M = 45,51$ en $SD = 2,66$). De onafhankelijke variabele schoolniveau verklaart in dit geval 53,9 % van de variantie in de score, zodat gesproken kan worden van een groot effect. Vwo-leerlingen scoren dus significant hoger op de woordenschattoets dan vmbo-leerlingen. Gezien het feit dat 95 % van de vwo'ers een score had tussen de 38 en 50 punten is deze woordenschattoets voor hen wellicht wat te eenvoudig. In ieder geval, aangezien woordenschat vaak een indicator is voor wat voor type lezer iemand is (Jansma, Van Kleunen & Leenders, 2011), kan hier vastgesteld worden dat op basis van deze resultaten op de woordenschattoets vwo-leerlingen de sterke lezers zijn en vmbo-leerlingen de zwakke.

Daarnaast is gekeken naar hoe het zit met een eventueel tekstversie-effect (makkelijke of moeilijke cloze-toets) op de score op de woordenschattoets. De verwachting hierbij is dat er geen effect van tekstversie te zien zal zijn op de scores op de woordenschattoets. De leerlingen zijn immers los van hun woordenschatniveau per klas over de tekstversies verdeeld, waarbij er tussen de vmbo-classes en tussen de vwo-classes geen significante verschillen waren wat betreft woordenschat. Bovendien is er geen enkele relatie tussen de woorden uit de beide cloze-toetsversies en die uit de woordenschattoets. Met andere woorden, de cloze-toets staat los van de woordenschattoets. Het zou daarom bijvoorbeeld ook niet zo mogen zijn dat de leerlingen die de makkelijke cloze-toets hebben gemaakt significant hoger scoren op de woordenschattoets dan de leerlingen die de moeilijke cloze-toets hebben gemaakt. En dus verwacht je ook geen effect van cloze-toetsversie op de scores op de woordenschattoets. Uit de ANOVA-analyse blijkt deze verwachting te kloppen: de scores op de woordenschattoets verschillen niet significant per versie ($F(1, 66) = .003, p = .954$). De verschillen tussen de gemiddelde scores op de woordenschattoets van de leerlingen met verschillende cloze-toetsversies zijn dus gebaseerd op toeval.

Wel is er sprake van een interactie-effect tussen schooltype en tekstversie ($F(1, 66) = 4.42, p = .039$). Dit houdt in dat de onafhankelijke variabelen schooltype en tekstversie samen van invloed zijn op de afhankelijke variabele, in dit geval de score op de woordenschattoets. Figuur 4.1 laat deze interactie duidelijk zien.

Figuur 4.1 Interactie-effect tussen schooltype en tekstversie op de woordenschattoets



Uit deze figuur blijkt dat vwo-leerlingen die de moeilijke versie hebben gemaakt beter scoren op de woordenschattoets dan vwo'ers met de makkelijke versie. Voor vmbo-leerlingen geldt juist dat wanneer zij de makkelijke versie hebben gemaakt zij een hoger resultaat behalen op de woordenschattoets dan wanneer zij de moeilijke versie hebben gekregen. Bij de vwo-leerlingen zorgt de moeilijke cloze-toets dus voor hogere resultaten op de woordenschattoets en bij de vmbo-leerlingen doet de makkelijke cloze-toets dit. Dit is opvallend, aangezien tekstversie hier in combinatie met schooltype wel invloed heeft op de resultaten van de woordenschattoets. Naar een verklaring hiervoor blijft het gissen. Vmbo-leerlingen zouden door hun kleinere woordenschat meer baat kunnen hebben bij een tekst met makkelijke woorden, waarbij deze gevoeligheid voor tekstversie-effecten in dit geval ook zijn uitwerking heeft op de scores op de woordenschattoets. De vwo-leerlingen doen het op de woordenschattoets juist beter wanneer ze de moeilijke cloze-toets hebben gemaakt. Hoewel er tussen de vwo-klassen geen significante verschillen waren wat betreft woordenschat en leesvaardigheid, zou het toch zo kunnen zijn dat er in de moeilijke versie net wat meer vaardige lezers zaten, wat vervolgens zijn afspiegeling heeft op de resultaten

van de woordenschattoets. Hoewel dit alles aannemelijk klinkt, blijft het gissen naar een verklaring voor dit interactie-effect tussen schooltype en tekstversie.

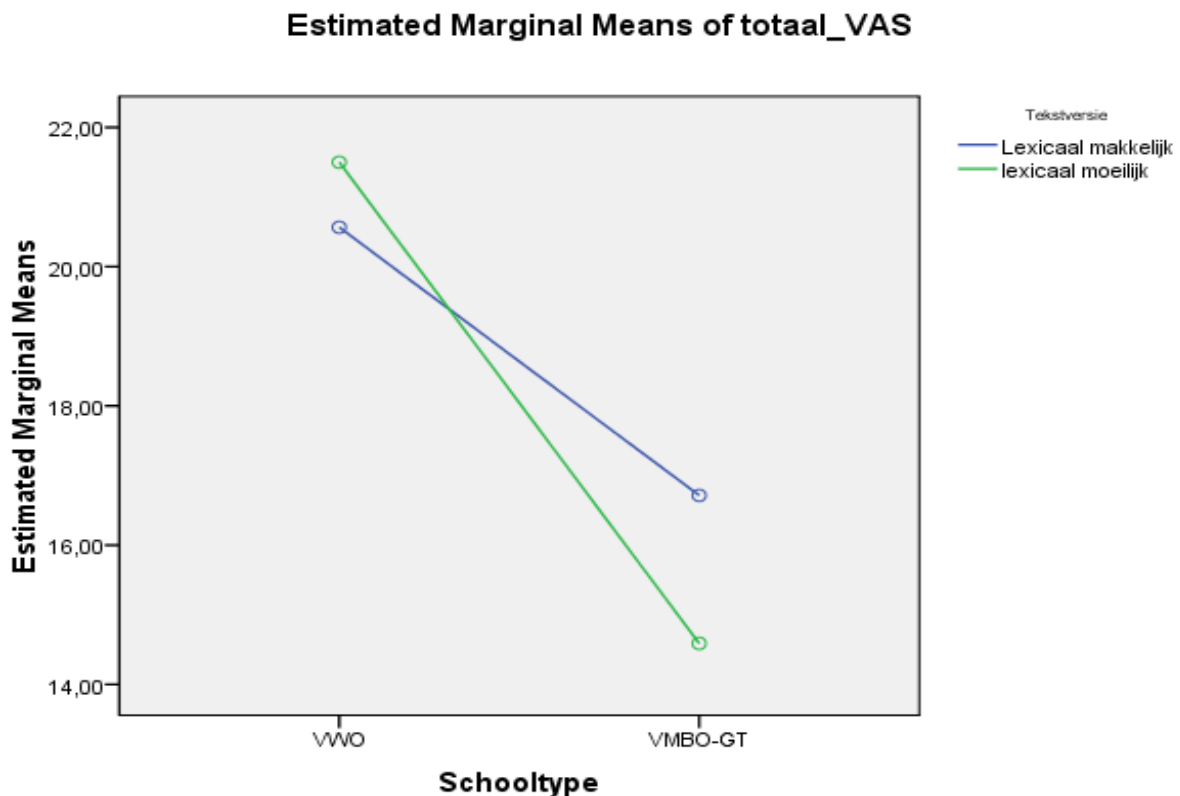
In figuur 4.1 is verder het hoofdeffect van schooltype duidelijk te zien; de vwo-leerlingen scoren hoger op de woordenschattoets dan de vmbo-leerlingen, ongeacht de cloze-toetsversie. Het feit dat vmbo- en vwo-leerlingen bij de woordenschattoets ieder profiteren van een andere tekstversie, laat nogmaals zien dat er van een hoofdeffect van tekstversie geen sprake is. Om er nu voor te zorgen dat bij de verdere analyse van de resultaten op de cloze-toets geen hinder ondervonden wordt van verschillen tussen leerlingen binnen een groep wat betreft woordenschat, worden deze verschillen door het opnemen van woordenschat als covariaat in een variantie-analyse gecorrigeerd.

Wanneer er voor de leesvaardigheidstoets eenzelfde ANOVA-analyse (Univariate) wordt uitgevoerd, dan zien we dat ook de resultaten op deze toets wat betreft het schoolniveau van elkaar verschillen ($F(1, 66) = 60.31, p < .001$). De gemiddelde score op de leesvaardigheidstoets behaald door vmbo-leerlingen ($M = 15,55$ en $SD = 3,47$) verschilt dus significant van de gemiddelde score die vwo-leerlingen op deze toets hebben behaald ($M = 20,95$ en $SD = 2,39$). De onafhankelijke variabele schoolniveau verklaart hier 47,7 % van de variantie in de score, waardoor ook in dit geval gesproken kan worden van een groot effect. Opnieuw scoren vwo-leerlingen dus significant hoger dan vmbo-leerlingen en kunnen zij ook hier beschouwd worden als de sterke lezers.

Ook op de score op de leesvaardigheidstoets blijft een effect van tekstversie uit ($F(1, 66) = .74, p = .393$). De resultaten van de leesvaardigheidstoets kunnen dus niet verklaard worden door de cloze-toetsversie. Dit is prettig, aangezien er ook hier verwacht werd dat de scores op de leesvaardigheidstoets niet significant zouden verschillen per versie. De leerlingen zijn namelijk los van hun leesvaardigheidsniveau per klas verdeeld over de twee versies, waarbij er tussen de vmbo-klassen en tussen de vwo-klassen geen significante verschillen waren wat betreft leesvaardigheid. Bovendien staat ook hier de cloze-toets in feite los van de leesvaardigheidstoets, aangezien deze laatste toets bestaat uit enkele korte teksten met hele andere onderwerpen, structuren en woorden dan die bij de cloze-toets. Het zou hier dan dus ook tegen de verwachting in zijn wanneer er bij een van de twee tekstversies significant hogere scores op de leesvaardigheidstoets zouden worden gehaald.

Wel is er opnieuw sprake van een interactie-effect tussen schooltype en tekstversie ($F(1, 66) = 4.88, p = .031$), zoals ook figuur 4.2 laat zien.

Figuur 4.2 Interactie-effect tussen schooltype en tekstversie op de leesvaardigheidstoets



Ook voor de leesvaardigheidstoets geldt dat vwo-leerlingen die de moeilijke versie hebben gemaakt beter scoren dan vwo'ers met de makkelijke versie. En voor vmbo-leerlingen geldt juist weer dat wanneer zij de makkelijke versie hebben gemaakt zij een hoger resultaat behalen op in dit geval de leesvaardigheidstoets dan wanneer zij de moeilijke versie hebben gekregen. Opnieuw is het dus opvallend dat tekstversie in combinatie met schooltype invloed heeft op de resultaten van in dit geval de leesvaardigheidstoets. Net zoals bij de woordenschattoets blijft het ook hier gissen naar een verklaring voor dit feit en kan min of meer uitgegaan worden van eenzelfde uitleg. Vmbo-leerlingen zouden als zwakkere lezers meer baat kunnen hebben bij een tekst met makkelijke woorden, waarbij deze gevoeligheid voor tekstversie-effecten in dit geval ook zijn uitwerking heeft op de scores op de leesvaardigheidstoets. De vwo-leerlingen doen het op de leesvaardigheidstoets juist beter

wanneer ze de moeilijke cloze-toets hebben gemaakt. Hoewel er tussen de vwo-klassen geen significante verschillen waren wat betreft leesvaardigheid en woordenschat, zou het toch zo kunnen zijn dat er in de moeilijke versie net wat meer vaardige lezers zaten, wat vervolgens zijn afspiegeling heeft op de resultaten van de leesvaardigheidstoets. Echter ook hier blijft het dus gissen naar de precieze verklaring voor dit interactie-effect tussen schooltype en tekstversie.

Figuur 4.2 laat verder het hoofdeffect van schooltype duidelijk zien; de vwo-leerlingen scoren hoger op de leesvaardigheidstoets dan de vmbo-leerlingen, ongeacht de cloze-toetsversie. Het feit dat vmbo- en vwo-leerlingen bij de leesvaardigheidstoets ieder profiteren van een andere tekstversie, laat nogmaals zien dat er van een hoofdeffect van tekstversie geen sprake is. Ook hier zal leesvaardigheid als covariaat in een variantie-analyse worden opgenomen, waardoor eventuele verschillen tussen leerlingen binnen een groep wat betreft leesvaardigheid worden gecorrigeerd.

4.2.5 Cloze-toets

Ook bij de cloze-toets kan gekeken worden naar verschillen in de prestaties tussen vmbo- en vwo-leerlingen en naar de invloed van de makkelijke en de moeilijke cloze-toets. Tabel 4.8 geeft per schoolniveau een overzicht van de gemiddelde scores op de cloze-toetsen.

Tabel 4.8 Gemiddelde scores, standaarddeviaties en laagste en hoogste scores van de makkelijke en moeilijke cloze-toets per schoolniveau

Toets	Schoolniveau (aantal leerlingen)	Gemiddelde score	Standaarddeviatie	Laagste score	Hoogste score
Cloze-toets makkelijk	Vmbo (14)	13,29	2,84	8	17
	Vwo (23)	15,87	1,91	10	19*
Cloze-toets moeilijk	Vmbo (17)	9,94	2,68	5	17
	Vwo (16)	14,69	2,27	11	18
Cloze-toets	Vmbo (31)	11,45	3,19	5	17
	Vwo (39)	15,38	2,12	10	19*

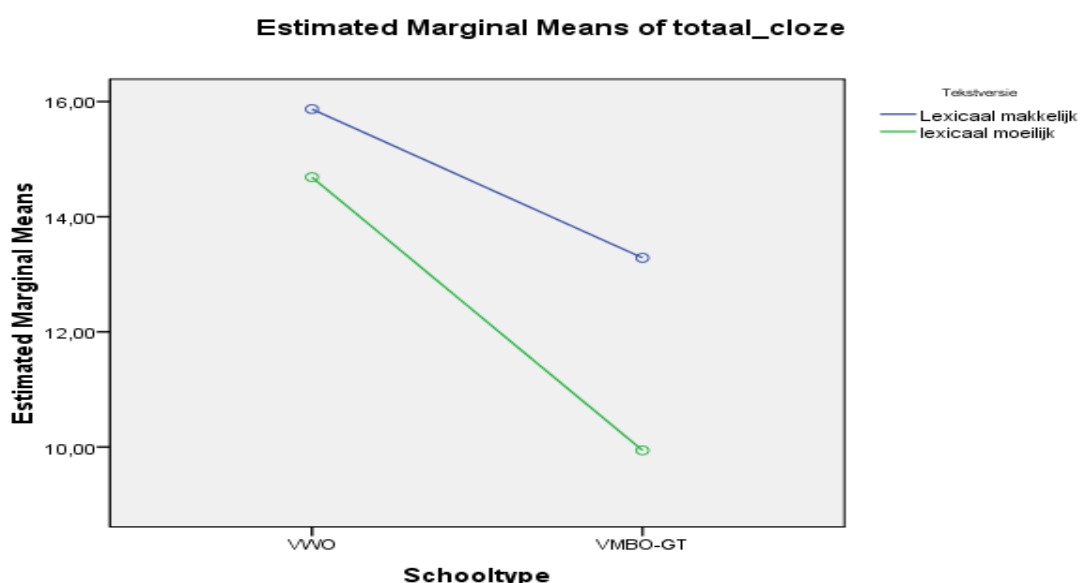
* Vanwege het buiten de analyse houden van item 7 is de maximale score op de cloze-toets nu 19 punten

Een ANOVA-analyse (Univariate) laat zien dat de resultaten op de cloze-toets naar gelang het schoolniveau van elkaar verschillen ($F(1, 66) = 39.73, p < .001$). De gemiddelde score op de cloze-toets behaald door vmbo-leerlingen ($M = 11,45$ en $SD = 3,19$) verschilt dus significant van de gemiddelde score die vwo-leerlingen op deze toets hebben behaald ($M = 15,38$ en $SD = 2,12$). De onafhankelijke variabele schoolniveau verklaart in dit geval 37,6 % van de variantie in de score, wat een groot effect inhoudt. Vwo-leerlingen scoren dus significant hoger op de cloze-toets dan vmbo-leerlingen, wat ze wederom typeert als sterkere lezers.

Ook is er een hoofdeffect van tekstversie op de score op de cloze-toets ($F(1, 66) = 15.15, p < .001$). Dit betekent dat de leerlingen die de makkelijke cloze-toets hebben gemaakt significant hoger scoren dan de leerlingen die de moeilijke cloze-toets hebben gemaakt. De gemiddelde scores bevestigen dit gegeven ($M = 14,89$ voor de makkelijke versie en $M = 12,24$ voor de moeilijke versie).

Bij de cloze-toets laat de ANOVA-analyse geen significant interactie-effect zien tussen schooltype en tekstversie ($F(1, 66) = 3.46, p = .067$). Dit houdt in dat de onafhankelijke variabelen schooltype en tekstversie samen de score op de cloze-toets niet echt beïnvloeden. Echter er is hier sprake van een net niet significant interactie-effect ($p = .067$). Figuur 4.3 laat dan ook zien dat de lijnen elkaar dreigen te kruisen, wat betekent dat er wel degelijk gesproken kan worden over enige vorm van interactie.

Figuur 4.3 Een net niet significant interactie-effect tussen schooltype en tekstversie op de cloze-toets



Goed te zien is allereerst dat er op de moeilijke cloze-toets slechter wordt gescoord (ongeacht schoolniveau) dan op de makkelijke cloze-toets, en dat daarnaast vwo-leerlingen de cloze-toets in het algemeen beter maken dan vmbo-leerlingen. Echter het feit dat figuur 4.3 deze twee hoofdeffecten niet weergeeft als parallelle lijnen, geeft aan dat schooltype en tekstversie samen toch wel enige invloed hebben op de prestaties op de cloze-toets. Ook al is deze interactie dan niet zo sterk, uit figuur 4.3 blijkt uit de naar elkaar toegroeiende lijnen dat de beide schooltypes verschillend reageren op de tekstversies: de vmbo-leerlingen met de makkelijke cloze-toets scoren heel duidelijk hoger dan de vmbo'ers met de moeilijke cloze-toets, terwijl het verschil in scores op de cloze-toetsversies bij vwo-leerlingen veel kleiner is. Ook vwo'ers scoren hoger op de makkelijke cloze-toets, echter de dreiging van de lijnen om elkaar te kruisen geeft aan dat de vwo-leerlingen ook met de moeilijke cloze-toets goed uit de voeten kunnen. Ze scoren op deze versie dan ook niet voor niets een gemiddelde score van 14,69, tegenover 15,87 op de makkelijke cloze-toets (tabel 4.8). Uit het feit dat de scores op de beide cloze-toetsversies bij vwo-leerlingen zo dicht bij elkaar liggen en dat vmbo'ers vooral op de makkelijke versie hoog scoren, lijken vmbo-leerlingen de neiging te hebben gevoeliger te zijn voor tekstversie-effecten dan vwo-leerlingen. Echter gezien het ontbreken van een daadwerkelijk significant interactie-effect kan hier het feit dat een vmbo'er meer dan een vwo'er zou profiteren van de makkelijke versie niet worden hard gemaakt.

4.2.6 Woordenschat en leesvaardigheid als covariaat

Om verschillen in woordenschat en leesvaardigheid tussen leerlingen binnen een groep te reduceren, worden hier de resultaten op de woordenschat- en leesvaardigheidstoets als covariaat in de variantie-analyse opgenomen. Er is dan geen sprake van een interactie-effect tussen schooltype en tekstversie ($F(1, 64) = .27, p = .604$), vanwege een correctie van de verschillen tussen leerlingen binnen een groep wat betreft woordenschat en leesvaardigheid. Hierna blijkt dat de moeilijke cloze-toets (ongeacht het schoolniveau) slechter gemaakt wordt dan de makkelijke cloze-toets. Daar waar een interactie-effect uitblijft, wordt een tekstversie-effect na de covariantie-analyse echter wel gerealiseerd ($F(1, 64) = 16.93, p < .001$). Hierbij kunnen we met een effectgrootte van 0,21 spreken van een middelmatig effect. Uit het feit dat de resultaten op de woordenschat- en leesvaardigheidstoets als covariaat allebei significant zijn, respectievelijk ($F(1, 64) = 4.24, p =$

.043) en ($F(1, 64) = 13.21, p = .001$), blijkt dat deze toetsen een deel van het verschil in de scores verklaren. Leesvaardigheid is in dit geval dan de sterkste variabele in de covariaat: hoe groter dus je leesvaardigheid hoe beter je bestand bent tegen de moeilijkheid van een tekst. Het feit dat er na de covariantie-analyse een echt hoofdeffect van tekstversie overblijft, geeft aan dat er in eerste instantie meer vaardige lezers bij de moeilijke versie zaten. Na de correctie hierop blijkt het hoofdeffect voor tekstversie dan ook nog groter te zijn geworden: van een klein effect (effectgrootte 0,19) naar een middelmatig effect (effectgrootte 0,21). Dit geldt echter niet voor het schooltype ($F(1, 64) = .63, p = .429$), waarvan we nu zelfs helemaal geen hoofdeffect meer zien. Dit betekent dat de verschillen in scores bij de twee schooltypen worden verklaard door de resultaten op de als covariaat opgenomen woordenschat- en leesvaardigheidstoets. Met andere woorden, het verschil tussen vmbo en vwo is een verschil in woordenschat en leesvaardigheid. Dit bevestigt wederom het gegeven dat vmbo-leerlingen getypeerd kunnen worden als zwakke lezers en vwo-leerlingen als sterke lezers.

4.2.7 Moeilijkheid van de cloze-items

In deze paragraaf wordt de vraag gesteld welke items uit de cloze-toets beschouwd kunnen worden als moeilijk en welke als relatief makkelijk. Oftewel, welke items zorgen voor het tekstversie-effect? Ten eerste is het zo dat er een aantal cloze-items significant is wat betreft verschil in tekstversie (tabel 4.9). Dit betekent dat er bij deze items sprake is van een sterke manipulatie tussen de makkelijke en de moeilijke versie. Op deze items is in de makkelijke versie dan ook significant hoger gescoord dan op dezelfde items in de moeilijke versie.

Tabel 4.9 Cloze-items die significant verschillen wat betreft cloze-toetsversie

Cloze-item	Significantie (Pearson Chi-square)
10	$p < .001$
14	$P = .004$
15	$P = .035$
16	$P = .036$
18	$p < .001$

Laten we deze cloze-items eens wat concreter bekijken door hun context in de verschillende tekstversies met elkaar te vergelijken. Dit ziet er voor item 10 als volgt uit:

Makkelijke cloze-toets:

De 9)..... van de koloniën leverden ten slotte de dure tropische goederen. Om het 10)..... van deze goederen makkelijker te maken, liet de koning tussen Noord en Zuid kanalen en wegen bouwen.

Moeilijke cloze-toets:

De 9)..... van de koloniën leverden ten slotte de kostbare tropische goederen. Om het 10)..... van deze goederen te simplificeren, liet de koning tussen Noord en Zuid kanalen en wegen construeren.

Wanneer we de scores op dit item in beide tekstversies bekijken, dan valt meteen het significante verschil op ($p < .001$): in de makkelijke cloze-toets is dit item door alle 37 leerlingen (14 vmbo'ers en 23 vwo'ers) goed beantwoord (semantisch of letterlijk), terwijl 12 van de 33 leerlingen (17 vmbo'ers en 16 vwo'ers) dit item bij de moeilijke cloze-toets fout hadden. Bekijken we daartoe de context rond item 10, dan blijkt dat de woorden *makkelijker maken/simplificeren* en *bouwen/construeren* hier cruciaal zijn als het gaat om het correct kunnen invullen van cloze-gat 10. Nu valt ook het scoreverschil op dit item tussen de beide versies te begrijpen: woorden als *makkelijker maken* en *bouwen* zullen geen problemen opleveren en maken het cloze-item in dit geval misschien zelfs iets te gemakkelijk, echter de betekenis van *simplificeren* en *construeren* is lang niet bij iedere 2^e-klasser bekend. Bij de moeilijke cloze-toets verschenen, naast goede antwoorden als *vervoeren* en *leveren* (bijlage 10), dan ook antwoorden als *geld*, *water* en *kostbare*. In totaal hadden 11 vmbo-leerlingen en slechts 1 vwo-leerling dit item fout, wat duidelijk maakt dat het moeilijker maken van de woorden rondom het cloze-gat hier vooral de vmbo-leerlingen treft.

Eenzelfde analyse kan gegeven worden voor de items 14, 15, 16 en 18, die allen behoren tot dezelfde alinea:

Makkelijke cloze-toets:

Ondanks zijn economische activiteiten, viel de 12)..... bij de Belgen niet in de smaak. 13)..... zagen in hem een heerser die absolute 14)..... wilde hebben en meer 15)..... van de elite niet toestond. In 1830 lieten de Brusselaars zich door een opera in de schouwburg 16)..... tot een opstand. Willem I stuurde er een 17)..... op af, maar dat hielp niet. 18)..... verkreeg zijn onafhankelijkheid. Pas in 1839 accepteerde de koning de Belgische 19)..... . Het jaar erop nam 20)..... teleurgesteld afscheid als koning.

Moeilijke cloze-toets:

Ondanks zijn economische bedrijvigheid, stond de 12)..... de Belgen niet aan. 13)..... zagen in hem een monarch die absolute 14)..... wilde nastreven en meer 15)..... van de elite niet duldde. In 1830 lieten de Brusselaars zich door een opera in de schouwburg 16)..... tot een mouterij. Willem I stuurde er een 17)..... op af, maar dat mocht niet baten. 18)..... verwierf zijn autonomie. Pas in 1839 erkende de koning de Belgische 19)..... . Het jaar erop nam 20)..... gedesillusioneerd afscheid als koning.

In deze alinea draait het om de verhouding tussen de Belgen en Willem I, waarbij deze laatste een Belgische opstand tevergeefs de kop probeert in te drukken en vervolgens aftreedt als koning. Item 18 is op dit punt cruciaal en laat zien of iemand de tekst daadwerkelijk heeft begrepen. *België* of *Brussel* zijn hier de enige goede antwoorden (bijlage 10). Aanwijzingen voor dit antwoord volgen niet alleen stapsgewijs uit deze alinea, maar staan daarnaast nog eens heel duidelijk in de zinnen rondom item 18; uit de zinnen met item 16 en 17 valt te lezen dat Willem I er niet in slaagde de Belgische opstand te onderdrukken en uit het woord *onafhankelijkheid/autonomie* en de zinnen na de zin met item 18 valt af te leiden dat België erkend wordt als staat en Willem I als verliezer afscheid neemt als koning. De leerlingen die bij item 18 antwoorden invulden als *Willem I* en *Hij* hebben dit gedeelte van de tekst dus heel duidelijk niet begrepen. In de makkelijke versie werd dit item door 10 leerlingen fout gedaan en in de moeilijke versie door maar liefst 25 van de 33 leerlingen. In deze moeilijke versie zal men dus waarschijnlijk moeite hebben gehad met woorden als *monarch*, *mouterij*, *autonomie* en *gedesillusioneerd*. In totaal maakten 7 vmbo'ers en 3 vwo'ers in de makkelijke versie item 18 fout, tegenover 15 vmbo'ers en 10 vwo'ers in de

moeilijke versie. Item 18 blijkt dus vooral in de moeilijke cloze-toets een moeilijk item te zijn. Wat dat betreft heeft de woordmanipulatie haar werk goed gedaan.

Ook de items 14, 15 en 16 laten een significant hogere score zien bij de makkelijke cloze-toets (tabel 4.9). Item 14 zal voor leerlingen die het woord *monarch* niet kennen moeilijk te beantwoorden zijn, terwijl met het woord *heerser* nauwelijks problemen mogen worden verwacht. In de makkelijke versie maakten dan ook slechts 3 leerlingen item 14 fout (1 vmbo'er en 2 vwo'ers) en in de moeilijke versie 12 leerlingen (10 vmbo'ers en 2 vwo'ers). Item 15 daarentegen wordt moeilijk door het woord *duldde*, terwijl er op die plek in de makkelijke versie het eenvoudigere woord *toestond* staat. 12 Leerlingen in de makkelijke versie beantwoordden dit item fout (8 vmbo'ers en 4 vwo'ers) en 19 leerlingen in de moeilijke versie (16 vmbo'ers en 3 vwo'ers). De leerlingen die bij dit item *geduld* invulden, bleken dus heel duidelijk het woord *duldde* niet te kennen en hebben om die reden deze zin niet goed begrepen. Om vervolgens cloze-item 16 goed te kunnen beantwoorden moet de betekenis van het woord *opstand/muiterij* duidelijk zijn. Voor 6 vmbo-leerlingen en 4 vwo-leerlingen leverde dit in de makkelijke versie problemen op en in de moeilijke versie voor 13 vmbo'ers en 4 vwo'ers. Uit de resultaten op deze items 14, 15 en 16 blijkt opnieuw dat vmbo-leerlingen de meeste hinder ondervinden van de moeilijker gemaakte woorden rondom de gaten in de moeilijke cloze-toets. En toevallig waren dit net items met cruciale woorden in de bijbehorende zinnen.

De items 10, 14, 15, 16 en 18 zijn dus significant slechter gemaakt in de moeilijke cloze-toets (tabel 4.9). Moeilijke woorden in de context van deze cloze-items hebben het invullen ervan moeilijker gemaakt en soms zelfs te moeilijk wanneer het correct invullen van een item afhing van het kennen van een specifiek woord (zoals bij item 14). Steeds komt hierbij het gegeven naar voren dat vooral de vmbo-leerlingen moeite hebben met de moeilijke cloze-toets en dus gebaat zijn bij een tekst met makkelijke woorden. In ieder geval kan hier gezegd worden dat deze laatste alinea van de tekst door de woordmanipulaties significant moeilijker is gemaakt, wat ook gebleken is uit het scoreverschil op de beide tekstversies. De items 14, 15, 16 en 18 in deze passage van de tekst hebben dan ook zeker een bijdrage geleverd aan het tekstversie-effect.

Daar waar er sprake is van lage percentages bij de score op de cloze-items in beide versies van de cloze-toets, betekent dit dat het een moeilijk item moet zijn geweest. Immers de score op zo'n item is zowel in de makkelijke als in de moeilijke cloze-toets relatief laag. Dit is heel duidelijk het geval voor de items 1 en 3, hier gepresenteerd in hun context:

Makkelijke cloze-toets:

*Na de Franse overheersing keerde de zoon van de stadhouder Willem V in 1813 terug naar Nederland om er koning te worden. Dat was een duidelijke **1)**..... met het verleden. Willem I werd niet, zoals zijn vader, **2)**..... in alle provincies, maar koning van één groot rijk. **(nieuwe alinea)** In 1815 werden de voormalige Oostenrijkse Nederlanden (het tegenwoordige België) met het **3)**..... van de Oude Republiek samengevoegd om te dienen als beschermingsgebied tegen het verslagen Frankrijk.*

Moeilijke cloze-toets:

*Na de Franse overheersing keerde de zoon van de landvoogd Willem V in 1813 terug naar Nederland om er het koningschap te aanvaarden. Dat was een onmiskenbare **1)**..... met het verleden. Willem I werd niet, zoals zijn vader, **2)**..... in alle gewesten, maar koning van een eenheidsstaat. **(nieuwe alinea)** In 1815 werden de voormalige Oostenrijkse Nederlanden (het huidige België) met het **3)**..... van de Oude Republiek verenigd om te dienen als buffer tegen het verslagen Frankrijk.*

Wat betreft item 1 was er in de makkelijke versie slechts 45,9 % van de antwoorden correct en in de moeilijke versie slechts 24,2 %. Hoewel er in de moeilijke cloze-toets op dit item ten opzichte van de makkelijke cloze-toets erg laag is gescoord, is dit verschil niet significant ($p = .059$). Het is echter opvallend dat dit item in beide versies zoveel is fout gedaan (door 12 vmbo'ers en 8 vwo'ers in de makkelijke versie en door 14 vmbo'ers en 11 vwo'ers in de moeilijke versie), immers de zin erna bevat met de woorden *Willem I werd niet,, maar* de aanwijzing voor de hier bedoelde tegenstelling (met als correcte antwoorden *breuk, confrontatie, verschil, tegenstelling* en *verandering*). Leerlingen die antwoordden met *overeenkomst*, wat in zowel de makkelijke als de moeilijke cloze-toets voorkwam, hebben dit begin van de tekst niet goed begrepen. Bij de moeilijke versie, ook vanwege de lagere score hierop, heeft dit waarschijnlijk te maken met de moeilijke woorden in de context rondom

item 1, zoals het woord *onmiskenbare* waarna direct het cloze-item gevraagd wordt. Het niet kennen van de betekenis van dit woord zou dus al voor problemen kunnen zorgen. Echter in de makkelijke cloze-toets kan hiervan met het woord *duidelijke* geen sprake zijn. De lage score op dit item moet dus op een andere manier verklaard worden. Wellicht is item 1 als eerste cloze-gat te moeilijk, aangezien het onderwerp van de tekst dan nog geïntroduceerd moet worden, de leerlingen er nog in moeten komen en het toch een specifiek woord is dat er van hen gevraagd wordt. Gezien het feit dat er op dit item geantwoord werd met veel verschillende woorden, die soms qua betekenis in de buurt kwamen van het correcte antwoord (zoals *stap*, *beslissing* en *gebeurtenis*), was dit een lastig item om na te kijken.

Bij item 3 werd in de makkelijke en moeilijke versie achtereenvolgens 43,2 % en 33,3 % goed beantwoord. Hierbij maakten 21 leerlingen dit item fout in de makkelijke cloze-toets (9 vmbo'ers en 12 vwo'ers) en 22 leerlingen in de moeilijke cloze-toets (13 vmbo'ers en 9 vwo'ers). De verklaring voor de vele foute antwoorden kan bij de moeilijke cloze-toets gezocht worden in het moeilijke woord *buffer*. Echter in de makkelijke cloze-toets staat er op deze plek het woord *beschermingsgebied*, wat de leerlingen al deels zou kunnen leiden naar het goede antwoord (*grondgebied*, *gebied*, *rijk* of *land*). Gezien ook het grote aantal fouten in de makkelijke versie is dit op de een of andere manier niet gelukt. Waarschijnlijk is het zo dat dit cloze-item de leerlingen qua antwoordmogelijkheden in beide tekstversies op een te breed spoor heeft gezet; immers bij item 3 lijkt een woord als *leger*, in een context met het woord *beschermingsgebied/buffer*, ook prima te passen. Echter de op item 3 volgende zin, uit in dit geval de makkelijke cloze-toets, laat zien dat dit niet kan:

Zo **4)**..... het Verenigd Koninkrijk: een middelgroot **5)**..... met veel overzeese gebieden.

Er blijkt hieruit dat het gaat om het vormen van een land (en niet een leger), dat dus moet dienen als *beschermingsgebied/buffer* tegen het verslagen Frankrijk. Waarschijnlijk hebben de leerlingen bij de zin waarin item 3 staat en die tevens een nieuwe alinea opent, zich te veel gericht op de zin zelf en niet zozeer op de erop volgende informatie. Gezien het feit dat het goede antwoord uit de context zelf te halen valt, zou je bij een focus hierop immers meer correcte antwoorden verwachten.

In ieder geval blijken item 1 en 3 dus moeilijke items te zijn geweest. Dit gold zowel voor de makkelijke en de moeilijke cloze-toets als voor de vmbo- en vwo-leerlingen.

Wat verder nog opvalt is het feit dat bij de cloze-items 4, 12 en 19 het percentage correcte antwoorden (semantisch of letterlijk) in de moeilijke cloze-toets hoger ligt dan in de makkelijke. Bij item 4 gaat het hierbij om 73 % goed in de makkelijke versie tegenover 75,8 % in de moeilijke versie. Bij de items 12 en 19 zijn de verschillen tussen de percentages correcte antwoorden op de makkelijke en moeilijke cloze-toets nog iets groter, respectievelijk 59,5 % en 69,7 % bij item 12 en 62,2 % en 69,7 % bij item 19. Dit zijn verder echter lang geen significante verschillen ($p = .790$ bij item 4, $p = .372$ bij item 12 en $p = .507$ bij item 19) en kunnen daardoor ook als toevallig worden beschouwd.

Verder kan hier nog vermeld worden dat in de makkelijke cloze-toets de items 10 en 13 door alle leerlingen goed zijn beantwoord (semantisch of letterlijk). In de moeilijke versie is respectievelijk 63,6 % en 97 % van deze items correct beantwoord. Dit geeft duidelijk aan dat item 13 in beide versies een te makkelijk item is geweest. Het feit dat er hier slechts geantwoord hoefde te worden met het woordje *Zij* lijkt dit dan ook te bevestigen:

Makkelijke cloze-toets:

*Ondanks zijn economische activiteiten, viel de **12)**..... bij de Belgen niet in de smaak. **13)**..... zagen in hem een heerser die (enz.).*

Moeilijke cloze-toets:

*Ondanks zijn economische bedrijvigheid, stond de **12)**..... de Belgen niet aan. **13)**..... zagen in hem een monarch die (enz.).*

Het is duidelijk dat het bij item 13 over *de Belgen* gaat en dat hiernaar moet worden terugverwezen met een woordje als *Zij*, *Ze*, of *Velen*. Tekstbegrip wordt hier dus vooral lokaal gemeten en aangezien het antwoord daardoor in beide tekstversies als het ware voor het grijpen ligt, is item 13 eigenlijk een te makkelijk item.

Tot slot van deze paragraaf wordt nog even teruggekomen op item 7: het item dat niet zozeer tekstbegrip zou meten, het item waarmee vooral de vwo-leerlingen vanwege hun

grotere woordenschat moeite hadden en het item dat significant beter gemaakt werd in de moeilijke cloze-toets ($p = .031$) en om deze redenen in de analyse verder buiten beschouwing is gelaten. Laten we de context van item 7 in beide tekstversies eens nader bekijken:

Makkelijke cloze-toets:

6)..... probeerde de oude economische bloei terug te krijgen door in zijn land de sterke kanten van de economie te ondersteunen. Het **7)**....., waar al technologische vernieuwingen hadden plaatsgevonden, moest zich richten op het maken van dagelijkse producten. De handelaren uit het noorden moesten die producten vervolgens de **8)**..... over brengen.

Moeilijke cloze-toets:

6)..... probeerde de oude economische ervaring te herstellen door in zijn land de sterke kanten van de economie te stimuleren. Het **7)**....., waar al een Industriële Revolutie had plaatsgevonden, moest zich toeleggen op het maken van consumptiegoederen. De koopmannen uit het noorden moesten die producten vervolgens de **8)**..... over brengen.

Het woord *zuiden* is het woord dat exact correct zou zijn. In totaal kwamen 6 van de 70 leerlingen met dit antwoord (3 vwo'ers in de makkelijke versie en 1 vmbo'er en 2 vwo'ers in de moeilijke versie). De enige aanwijzing voor het woord *zuiden* in de tekst valt te halen uit het woord *noorden* in de zin met item 8, waarin de tegenstelling *noord/zuid* herkend kan worden. Echter het woord *noorden* hoeft helemaal niet te wijzen op deze tegenstelling en daarom zijn antwoorden als *land*, *rijk* en *gebied* ook goed gerekend (bijlage 10). Echter op deze manier wordt het de leerlingen niet makkelijk gemaakt om item 7 met behulp van de tekst in te vullen. Ook de semantisch correcte antwoorden vallen namelijk niet zo goed uit de tekst te halen; er wordt nergens informatie gegeven over waar de Industriële Revolutie al heeft plaatsgevonden. Hierdoor is het begrijpelijk dat leerlingen ook komen met antwoorden als *centrum* en *noorden*.

In paragraaf 4.2.2 werd reeds de veronderstelling geopperd dat vwo-leerlingen vanwege hun grotere woordenschat nadeel zouden kunnen ondervinden van het feit dat het antwoord op item 7 niet gemakkelijk uit de tekst te halen valt. Zij hebben immers buiten de

tekst de keuze uit een groot aantal woorden. Gezien het feit dat in de makkelijke cloze-toets vooral vwo-leerlingen dit item fout beantwoordden (15 van de 23 tegenover 8 van de 14 vmbo'ers), zou dit vermoeden wel eens kunnen kloppen. Het lijkt er dus op dat item 7, vanwege het nauwelijks verbonden zijn met de tekst, vooral voor vwo-leerlingen problemen heeft opgeleverd. Dit beeld is met name sterk terug te zien bij de makkelijke cloze-toets, omdat deze versie door relatief veel vwo'ers is gemaakt. Nu we ook de context van item 7 hebben bekeken en met nog meer zekerheid kunnen zeggen dat dit item nauwelijks tekstbegrip meet, is het een verstandig besluit geweest om item 7 uit de analyse te verwijderen.

4.3 Verschillende scoringsmethodes

Voordat er overgegaan wordt tot het evalueren van de hypothesen zal er in deze paragraaf eerst nog gekeken worden naar het verschil tussen de exact-word scoringsmethode en de methode waarbij antwoorden enkel semantisch goed moeten zijn (Kobayashi, 2002); worden er andere effecten gevonden wanneer de exact-word scoringsmethode wordt toegepast? Hoewel de data in dit onderzoek voor beide scoringsmethodes zijn gecodeerd (paragraaf 3.1.4), zijn tot nu toe alle analyses uitgevoerd aan de hand van de scoringsmethode waarbij antwoorden semantisch goed moeten zijn. Interessant is het nu om te kijken of de hierbij gevonden effecten standhouden wanneer met de exact-word methode wordt gewerkt; gezien het feit dat synoniemen bij deze scoringsmethode fout worden gerekend, is deze methode in het corrigeren een stuk strenger. In paragraaf 4.2.2 is al gekeken naar het effect van de exacte scoringsmethode op de betrouwbaarheid van de cloze-toets. Hier bleek dat voor zowel de makkelijke en de moeilijke cloze-toets als de cloze-toets als geheel de betrouwbaarheid afneemt wanneer alleen exacte antwoorden op de cloze-items worden goedgekeurd (tabel 4.5). De cloze-toets meet in dit geval dus stabiel tekstbegrip wanneer ook synoniemen goed worden gerekend.

Laten we nu eens kijken of de exact-word methode gevolgen heeft voor de in dit onderzoek gevonden effecten op de cloze-toets. Allereerst is er ook bij het goed rekenen van enkel exacte antwoorden de verwachting dat vwo-leerlingen hoger zullen scoren op de cloze-toets dan vmbo-leerlingen, immers verschillen tussen deze leerlingen wat betreft woordenschat en leesvaardigheid zijn verschillen die los van de scoringsmethode in de werkelijkheid gewoonweg bestaan. We zien dan inderdaad ook bij deze exacte

scoringmethode een hoofdeffect van schooltype ($F(1, 66) = 33.50, p < .001$). Wel valt hierbij op dat de gemiddelde scores een stuk lager liggen dan wanneer ook synoniemen goed worden gerekend ($M = 4,32$ en $SD = 2,65$ bij de vmbo-leerlingen en $M = 7,28$ en $SD = 1,57$ bij de vwo-leerlingen). Gezien de strengere manier van corrigeren bij de exact-word methode was dit te verwachten. Het in verhouding toch enigszins lage gemiddelde bij de vwo-leerlingen (gemiddeld 7,28 punten van de 19) zou verklaard kunnen worden vanuit hun grotere woordenschat; vwo-leerlingen zullen vanwege hun brede woordkennis vaak meerdere synoniemen weten voor een bepaald woord, waardoor het exacte antwoord lang niet altijd gegeven wordt. De exact-word methode zal dit soort synoniemen echter fout rekenen, ondanks dat de vwo-leerlingen hebben laten zien de tekst begrepen te hebben. Vandaar ook dat het eerlijker is om tekstbegrip te beoordelen aan de hand van semantisch correcte antwoorden, vooral wanneer je te maken hebt met taalvaardige lezers (Kobayashi, 2002). Ook bij de exacte scoringmethode is het schooltype-effect een groot effect. De onafhankelijke variabele schoolniveau verklaart in dit geval 33,7 % van de variantie in de score. Vwo-leerlingen scoren dus nog steeds significant hoger op de cloze-toets dan vmbo-leerlingen wanneer alleen exacte antwoorden op de cloze-toets goed zijn. Vwo-leerlingen zijn dus echt de sterkere lezers. Het feit dat ook nu het schooltype-effect verdwijnt wanneer woordenschat en leesvaardigheid als covariaat aan de variantie-analyse worden toegevoegd ($F(1, 64) = .92, p = .340$), laat nogmaals zien dat de verschillen tussen vmbo en vwo verschillen in woordenschat en leesvaardigheid zijn. Uit het feit dat de resultaten op de woordenschat- en leesvaardigheidstoets als covariaat allebei significant zijn, respectievelijk ($F(1, 64) = 4.34, p = .041$) en ($F(1, 64) = 5.55, p = .022$), blijkt dat deze toetsen een deel van het verschil in de scores verklaren. Leesvaardigheid is in dit geval weer de sterkste variabele in de covariaat: hoe groter dus je leesvaardigheid hoe beter je bestand bent tegen de moeilijkheid van een tekst.

Ook het tekstversie-effect blijft bij de exact-word methode overeind ($F(1, 66) = 8.08, p = .006$). De effectgrootte is daarbij iets gedaald (van 0,19 naar 0,11), alhoewel we nog steeds te maken hebben met een klein effect. Op de makkelijke cloze-toets werd een gemiddelde score van 6,76 gehaald ($SD = 2,13$) en op de moeilijke cloze-toets een gemiddelde score van 5,09 ($SD = 2,77$). Ook na de correctie op woordenschat en leesvaardigheid is er een tekstversie-effect ($F(1, 64) = 8.07, p = .006$), met wederom een effectgrootte van 0,11. Het feit dat er ook bij de exacte scoringmethode een hoofdeffect van tekstversie te zien is, geeft

aan dat de leerlingen in de makkelijke versie relatief veel exact correcte antwoorden hebben gegeven, waardoor hier nog steeds geldt dat de makkelijke cloze-toets (ongeacht schoolniveau) significant beter gemaakt wordt dan de moeilijke cloze-toets. Wanneer de leerlingen in de makkelijke versie daarentegen vooral met synoniemen zouden hebben geantwoord, zou er immers van het tekstversie-effect waarschijnlijk nog maar weinig zijn overgebleven.

Ook bij de exacte scoringsmethode zien we een net niet significant interactie-effect tussen schooltype en tekstversie ($F(1, 66) = 3.44, p = .068$). Opnieuw blijken de scoreverschillen tussen de tekstversies bij vmbo-leerlingen (verschil van 2,28 punten) groter te zijn dan die bij vwo-leerlingen (verschil van 0,48 punten). Vmbo-leerlingen blijken dan ook in de moeilijke versie minder exact correcte antwoorden te geven ($M = 3,29$ en $SD = 2,37$ tegenover $M = 7,00$ en $SD = 1,67$ bij de vwo-leerlingen). Echter gezien het ontbreken van een daadwerkelijk significant interactie-effect kunnen hier net als bij de semantisch correcte scoringsmethode verder geen waardevolle uitspraken worden gedaan. In ieder geval is er ook hier na de covariantie-analyse geen interactie-effect tussen schooltype en tekstversie ($F(1, 64) = .54, p = .464$), vanwege een correctie van de verschillen tussen leerlingen binnen een groep wat betreft woordenschat en leesvaardigheid.

Na deze toepassing van de exact-word methode op de resultaten op de cloze-toetsen, kan gezegd worden dat de behaalde scores op de beide tekstversies zowel bij vmbo- als bij vwo-leerlingen gemiddeld een stuk lager liggen dan de gemiddelde scores verkregen bij de semantisch correcte scoringsmethode. Uiteraard is het gegeven dat synoniemen bij de exacte methode fout gerekend worden hiervoor een verklaring. Ondanks dit verschil in gemiddelden zien we bij deze scoringsmethode verder wel dezelfde effecten terug: een hoofdeffect voor zowel schooltype als tekstversie, maar geen significant interactie-effect. Toch is er hier goed aan gedaan om in dit onderzoek ook synoniemen goed te rekenen; het levert niet alleen betrouwbaardere cloze-toetsen op, maar het is daarnaast ook een eerlijke manier om tekstbegrip te meten.

4.4 Terugkoppeling hypothesen

Uit de verkregen resultaten, die het onderzoek naar de invloed van woordmoeilijkheid op het tekstbegrip van zwakke en sterke lezers naar voren heeft gebracht, kunnen we de

nulhypothese verwerpen en daarop aansluitend de H1-hypothese aannemen:

H0: *Woordmoeilijkheid heeft geen invloed op het tekstbegrip van zwakke en sterke lezers.*

H1: *Teksten met makkelijke woorden worden beter begrepen dan teksten met moeilijke woorden, zowel door vmbo- als door vwo-leerlingen.*

De scores op de makkelijke cloze-toets zijn significant hoger dan die op de moeilijke cloze-toets ($F(1, 66) = 15.15, p < .001$). Na het opnemen van woordenschat en leesvaardigheid als covariaat in een variantie-analyse, om zo verschillen tussen leerlingen binnen een groep te reduceren, wordt het tekstversie-effect zelfs nog versterkt ($F(1, 64) = 16.93, p < .001$): van een klein effect (effectgrootte 0,19) naar een middelmatig effect (effectgrootte 0,21). Hierbij is er tussen de makkelijke en moeilijke tekstversie geen verschil in woordenschat ($F(1, 66) = .003, p = .954$) en leesvaardigheid ($F(1, 66) = .74, p = .393$). Aangezien woordmoeilijkheid dus het enige punt is waarop de beide cloze-toetsversies van elkaar verschillen, kan er hier geconcludeerd worden dat woordmoeilijkheid invloed heeft op tekstbegrip en teksten met makkelijke woorden dus beter begrepen worden dan teksten met moeilijke woorden.

Ook de H2-hypothese kan aangenomen worden:

H2: *Gemiddeld behalen vwo-leerlingen zowel op de makkelijke als op de moeilijke cloze-toets betere resultaten dan vmbo-leerlingen.*

Vwo-leerlingen behaalden zowel op de woordenschattoets ($F(1, 66) = 77.24, p < .001$), de leesvaardigheidstoets ($F(1, 66) = 60.31, p < .001$) als op de cloze-toets ($F(1, 66) = 39.73, p < .001$) ten opzichte van vmbo-leerlingen significant hogere resultaten. Het feit dat na het corrigeren van verschillen tussen leerlingen wat betreft woordenschat en leesvaardigheid het schooltype-effect verdwenen is ($F(1, 64) = .63, p = .429$), geeft aan dat de verschillen tussen vmbo- en vwo-leerlingen juist verschillen zijn in woordenschat en leesvaardigheid. Hierdoor is nog maar eens bewezen dat vwo-leerlingen over het algemeen betere lezers zijn

dan vmbo-leerlingen, wat Jansma, Van Kleunen en Leenders (2011) in hun onderzoek ook al constateerden.

Tot slot wordt de H3-hypothese besproken:

H3: *De resultaten op de makkelijke en de moeilijke cloze-toets zullen bij vmbo-leerlingen meer van elkaar verschillen dan bij vwo-leerlingen.*

De vmbo-leerlingen behaalden op de makkelijke cloze-toets gemiddeld 13,29 punten (SD = 2,84) en op de moeilijke cloze-toets gemiddeld 9,94 punten (SD = 2,68), wat gemiddeld een verschil van 3,35 punten tussen de scores laat zien. De vwo-leerlingen daarentegen scoorden op de makkelijke versie gemiddeld 15,87 punten (SD = 1,91) en gemiddeld 14,69 punten (SD = 2,27) op de moeilijke versie, wat gemiddeld een verschil van 1,18 punten tussen de scores inhoudt. Hoewel het verschil in scores tussen de beide cloze-toetsen bij de vmbo-leerlingen een stuk groter is dan bij de vwo-leerlingen, is er toch net geen sprake van een significant interactie-effect tussen schooltype en tekstversie ($F(1, 66) = 3.46, p = .067$). Het is dus wel zo dat de gemiddelde scores op de makkelijke en de moeilijke cloze-toets bij vmbo-leerlingen meer van elkaar verschillen dan bij vwo-leerlingen (respectievelijk een verschil van 3,35 punten tegenover een verschil van 1,18 punten), en in die zin hebben de vmbo'ers door een beperkter woordenschat- en leesvaardigheidsniveau de neiging gevoeliger te zijn voor tekstversie-effecten dan vwo'ers. Echter dit kan hier niet hard gemaakt worden, aangezien er geen sprake is van een significant interactie-effect. Er mag hier dan ook niet geconcludeerd worden dat vmbo-leerlingen meer profiteren van de makkelijke versie, ondanks het feit dat het scoreverschil tussen de tekstversies bij deze groep leerlingen groter is. De derde hypothese kan hier dan dus ook niet worden aangenomen.

5 Conclusie

Het hier gepresenteerde onderzoek had tot doel de invloed van woordmoeilijkheid op het tekstbegrip van zwakke en sterke lezers te bepalen. Conform aan eerdere onderzoeken zijn vmbo-leerlingen hierbij getypeerd als de zwakke lezers en vwo-leerlingen als de sterke lezers. Deze indeling in lezerstypen blijkt terecht gekozen: ook dit onderzoek laat zien dat vwo'ers betere lezers zijn dan vmbo'ers doordat er door deze leerlingen zowel op de woordenschat- en leesvaardigheidstoets als op de beide cloze-toetsversies significant hoger werd gescoord. De hier gevonden verschillen in toetsresultaten tussen vmbo- en vwo-leerlingen zijn dan ook verschillen in woordenschat en leesvaardigheid. Daarnaast geldt voor deze beide schooltypes dat het tekstkenmerk woordmoeilijkheid hun tekstbegrip beïnvloedt. Dit bleek allereerst uit het feit dat de twee cloze-toetsversies, enkel gemanipuleerd op woordmoeilijkheid, wat score betreft van elkaar verschilden: de cloze-toets met makkelijke woorden werd significant beter gemaakt dan de cloze-toets met moeilijke woorden, zowel door vmbo- als door vwo-leerlingen. De aanname dat het de woordmoeilijkheid is die hier van invloed is op tekstbegrip werd ten tweede bevestigd door het feit dat er tussen de makkelijke en moeilijke cloze-toets geen verschillen waren in woordenschat en leesvaardigheid. Woordmoeilijkheid blijkt dus het enige punt te zijn waarop deze beide cloze-toetsversies van elkaar verschillen, waardoor de verkregen resultaten kunnen worden geïnterpreteerd in het licht van deze onafhankelijke variabele. Dit onderzoek bevestigt dan ook de verwachting dat woordmoeilijkheid van invloed is op tekstbegrip. Echter hierbij kunnen geen uitspraken gedaan worden over het precieze effect van woordmoeilijkheid op het type lezer. Want hoewel de resultaten op de makkelijke en de moeilijke cloze-toets bij vmbo-leerlingen meer van elkaar verschilden dan bij vwo-leerlingen, was hier net geen significant interactie-effect zichtbaar tussen schooltype en tekstversie. Vmbo-leerlingen neigen door een beperkter woordenschat- en leesvaardigheidsniveau net even wat gevoeliger te zijn voor tekstversie-effecten dan vwo-leerlingen, al zal dat hier door het niet hard kunnen maken van deze veronderstelling een vermoeden blijven. Echter voor het ontwikkelen van geschikte lesmethodes voor vmbo- en vwo-leerlingen is informatie over het effect van woordmoeilijkheid op verschillende lezerstypen noodzakelijk. Alleen dan kan het hier gevonden effect van woordmoeilijkheid op tekstbegrip concreet in schoolteksten en passend bij het niveau van een leerling worden toegepast.

6 Discussie

Naar aanleiding van de resultaten uit dit onderzoek kan gezegd worden dat teksten met makkelijke woorden beter begrepen worden dan teksten met moeilijke woorden. Hierbij scoren vwo-leerlingen als sterkere lezers op beide tekstversies beter. Echter vervolgens laat dit onderzoek tegen de verwachting in geen significant interactie-effect zien tussen schooltype en tekstversie; de resultaten op de makkelijke en de moeilijke cloze-toets bij vmbo-leerlingen verschillen niet significant van die van vwo-leerlingen, ondanks dat die verschillen bij de vmbo'ers wel groter zijn. Een verklaring voor dit afwijkende resultaat moet dan ook gezocht worden in een klein hoekje, immers er waren geen significante verschillen tussen de vmbo-leerlingen en tussen de vwo-leerlingen in woordenschat en leesvaardigheid. Al gauw kom je dan bij het toetsmateriaal terecht; wat betreft de woordmanipulaties op de cloze-toets hoeven er maar net enkele items een minder sterke manipulatie te hebben ondergaan, waardoor de makkelijke en moeilijke versie ongewild dichter bij elkaar zijn komen te staan. Met andere woorden, hoe sterker de manipulatie van de woordmoeilijkheid, hoe meer de beide cloze-toetsen op dit punt van elkaar zullen verschillen en hoe groter de kans is op significante scoreverschillen op de cloze-toetsversies tussen vmbo- en vwo-leerlingen, aannemende dat vmbo-leerlingen als zwakke lezers meer moeite zullen hebben met moeilijke woorden in een tekst. Het net niet significante interactie-effect maakt het lastig te bepalen om welke items het hier dan precies gaat. Een pilot study van de cloze-toetsen vooraf aan dit onderzoek was dan ook niet verkeerd geweest. Door de cloze-toetsen door vergelijkbare proefpersonen (vmbo- en vwo-leerlingen uit de 2^e klas) te laten maken, zouden eventuele zwakke manipulaties kunnen worden opgespoord, waardoor de kwaliteit van de cloze-toetsen nog verbeterd had kunnen worden. In dit onderzoek zijn de cloze-toetsen wegens tijdgebrek slechts beoordeeld door enkele medestudenten. Echter zij kijken met een veel hoger woordenschat- en leesvaardigheidsniveau natuurlijk heel anders naar zo'n toets dan middelbare scholieren. Bovendien speelt hier ook nog het feit mee dat deze studenten op de hoogte waren van wat er met de cloze-toets bereikt zou moeten worden. Een pilot study van de cloze-toetsen vooraf aan dit onderzoek zou in ieder geval nuttig zijn geweest en had wellicht door een verbetering van het toetsmateriaal een significant interactie-effect tussen schooltype en

tekstversie laten zien. Echter dit zijn uiteraard slechts speculaties waarop dit onderzoek verder geen antwoord kan geven.

Wat de verdeling van het toetsmateriaal over de klassen betreft, zijn in dit onderzoek de makkelijke en de moeilijke versie van de cloze-toets aselekt over de klassen verdeeld. Hierbij kreeg de ene vmbo-klas de makkelijke versie en de andere vmbo-klas de moeilijke. Hetzelfde gold voor de vwo-klassen. Ondanks het feit dat er zowel tussen de vmbo-klassen als tussen de vwo-klassen geen verschillen waren in woordenschat en leesvaardigheid, was het nog beter geweest wanneer er binnen elke klas alle mogelijke versies verdeeld waren. Nu kan het immers nog zo zijn geweest dat het toetsmoment (vroeg of laat op de dag) invloed heeft gehad, immers een bepaalde versie werd door een bepaald schoolniveau in één keer gemaakt. Om invloed van dit soort factoren zoveel mogelijk uit te sluiten, is het beter om in het vervolg aan elke klas alle mogelijke toetsversies te geven.

Om de resultaten verkregen uit dit onderzoek echt generaliseerbaar te maken, zou dit onderzoek uitgevoerd moeten worden bij een groter aantal dan 84 leerlingen, waarna er door afwezigheid op een van de toetsdagen uiteindelijk zelfs nog maar 70 leerlingen van overbleven. Wanneer de toetsen door de leerlingen op één dag zouden worden gemaakt, zou er in ieder geval minder sprake zijn van uitval van proefpersonen. Echter dan zou weer gediscussieerd kunnen worden over de vraag of vier toetsen op één dag naast het reguliere schoolprogramma van de leerlingen niet een beetje te veel van het goede is. Hoe dan ook zouden meer proefpersonen de verkregen resultaten generaliseerbaarder maken. Bovendien gelden de hier gevonden resultaten enkel voor vmbo- en vwo-leerlingen uit de 2^e klas van het voortgezet onderwijs, waarbij vmbo in dit geval staat voor de gemengde theoretische leerweg (vmbo-gt). Wellicht zouden de resultaten heel anders zijn geweest wanneer een gymnasiumklas met een ander type vmbo-onderwijs zou zijn vergeleken. Op het gebied van de invloed van woordmoeilijkheid op tekstbegrip valt er dan ook nog genoeg te onderzoeken. Niet alleen door middel van meer proefpersonen en andere type leerlingen, echter ook door bijvoorbeeld het kiezen van een andere begriptoets. Hoewel dit onderzoek heeft laten zien dat cloze-toetsen wel degelijk tekstbegrip meten, bleek uit de item-analyse dat cloze-toetsen vaak vooral lokaal tekstbegrip meten. Het verbinden van grotere tekstdelen, waarvoor uiteraard ook tekstbegrip nodig is, krijgt bij een tekstbegriptoets als de

cloze-toets veel minder aandacht; het zijn vaak toch de woorden rondom een cloze-gat die het invullen ervan beïnvloeden. In vervolgonderzoek zou daarom tekstbegrip eens gemeten moeten worden aan de hand van methoden als de sorteertaak en de mental model taak, begripstoetsen die in dit onderzoeksveld nog tamelijk nieuw zijn en op dit gebied doorgaans ook nog niet vaak worden toegepast, maar wel uitgaan van het meten van tekstbegrip op een globaler niveau (Kamalski, 2007). Om ervoor te zorgen dat tekstbegrip volledig gemeten wordt, zou een combinatie van een locale en een globale begripstoets in dit geval ideaal zijn.

Tot slot kan er hier aan samenstellers van lesmethoden het advies meegegeven worden dat er bij het ontwikkelen van schoolteksten zeker gelet moet worden op de woordmoeilijkheid, aangezien dit onderzoek heeft aangetoond dat woordmoeilijkheid van invloed is op tekstbegrip. Daarbij is het van belang om te weten dat teksten met makkelijke woorden door zowel vmbo- als vwo-leerlingen beter worden begrepen. Vmbo-leerlingen hebben hierbij de neiging gevoeliger te zijn voor moeilijke woorden dan vwo-leerlingen. Het kan dan ook zeker geen kwaad wanneer schoolteksten voor vmbo'ers in verhouding meer makkelijke woorden bevatten dan schoolteksten voor vwo'ers. Willen we de schoolteksten ook in de toekomst leesbaar houden, dan is het belangrijk dat er op deze aspecten gelet wordt. Alleen zo kunnen teksten samengesteld worden die afgestemd zijn op het niveau van vmbo- en vwo-leerlingen.

Literatuurlijst

- **Abraham, R., & C. Chapelle** (1992). The meaning of cloze test scores: An item difficulty perspective. *The Modern Language Journal*, 76 (4), p. 468-479.
- **Bonset, H., & M. Braaksma** (2008). *Het schoolvak Nederlands opnieuw onderzocht: Een inventarisatie van onderzoek van 1997 tot en met 2007*. Enschede: Drukkerij Netzdruk.
- **Fletcher, C.R.** (1994). Levels of representation in memory for discourse. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. New York: Academic Press, p. 589 – 607.
- **Jansma, N., Kleunen van, E., Leenders, E.** (2011). *Lezen onder de loep. Voor docenten Nederlands en vakdocenten in het vmbo-mbo*. SLO: nationaal expertisecentrum leerplanontwikkeling. Enschede: SLO.
- **Kamalski, J.** (2007). How to measure situation model representations. In: *Coherence marking, comprehension and persuasion. On the processing and representation of discourse*. Dissertatie. Utrecht: Universiteit Utrecht, p. 83 – 106.
- **Kobayashi, M.** (2002). Cloze tests revisited: Exploring item characteristics with special attention to scoring methods. *The Modern Language Journal*, 86 (4), p. 571-586.
- **Konopak, B.C.** (1988). Effects of inconsiderate vs. considerate text on secondary students' vocabulary learning. In: *Journal of Literacy Research, 1554-8430*, Volume 20, Issue 1, p. 25 – 41.
- **Land, J.** (2009). *Zwakke lezers, sterke teksten? Effecten van tekst- en lezerskenmerken op het tekstbegrip en de tekstwaardering van vmbo-leerlingen*. Stichting Lezen reeks, deel 13. Delft: Eburon.
- **Sanders, T.J.M., & C. van Wijk** (2002). Taal en de cognitieve processen van productie en verwerking. In: T. Janssen (red.), *Taal in gebruik*. Den Haag: SDU Uitgevers, p. 45 – 60.
- **Stahl, Steven A.** (2003). Vocabulary and readability: How knowing word meanings affects comprehension. *Topics in Language Disorders*, 23 (3), p. 241 – 247.
- **Stahl, S.A., Jacobson, M.G., Davis, C.E., Davis, R.L.** (1989). Prior knowledge and difficult vocabulary in the comprehension of unfamiliar text. In: *Reading research Quarterly* 24 (1), p. 27 – 43.