

Taaltoetsen testen

De kwaliteit van toetsen binnen een methode
voor moderne vreemde talen.

Masterthesis onderwijskundig ontwerp en advisering.

Datum: 20 augustus 2008
Auteur: Femke van Laarhoven
Beoordelaars: Jeroen Jansen
Liesbeth Baartman
Studentennummer: 3102483

Inhoudsopgave

Voorwoord	pag. 3
Samenvatting	pag. 4
1. Inleiding	pag. 5
2. Theoretisch kader	pag. 7
2.1 Competentiegericht onderwijs	pag. 7
2.2 Het begrip competentie	pag. 8
2.3 Assessment van competenties	pag. 10
2.4 Twaalf kwaliteitscriteria	pag. 11
2.5 Het CEFR	pag. 20
2.6 Omschrijving methode Taalblokken	pag. 24
2.7 Doelstelling en vraagstelling	pag. 26
3. Methode	pag. 27
3.1 Onderzoeksonwerp	pag. 27
3.2 Onderzoeksgroep	pag. 27
3.3 Instrumenten	pag. 29
4. Resultaten	pag. 34
4.1 Wat is de mening van docenten over de kwaliteit van de eisen en voorwaarden van competentiegericht toetsen binnen de methode Taalblokken?	pag. 34
4.2 Wat is de mening van studenten over de kwaliteit van de eisen en voorwaarden van competentiegericht toetsen binnen de methode Taalblokken?	pag. 37
4.3 Op welke manier komen de kwaliteitscriteria van competentiegericht toetsen terug in de toetsen van de methode Taalblokken?	pag. 41
4.4 Welke keuzes heeft Malmberg gemaakt bij het ontwerpen van de methode Taalblokken gericht op de kwaliteitseisen van het competentiegericht toetsen en het common Europeaan framework of reference?	pag. 43
5. Conclusie en discussie	pag. 45
5.1 Korte samenvatting resultaten	pag. 45
5.2 Verklaring resultaten	pag. 47
5.3 Beperkingen onderzoek	pag. 48
5.4 Mogelijkheden voor verder onderzoek	pag. 50

7. Literatuurlijst		pag. 51
--------------------	--	---------

Bijlage

Bijlage 1	CEFR	pag. 56
Bijlage 2	Interview met docenten	pag. 58
Bijlage 3	Vragenlijst studenten	pag. 62
Bijlage 4	Beoordelingsinstrument voor toetsen	pag. 65
Bijlage 5	Overeenstemmingpercentages objectiviteitmeting beoordelingsinstrument	pag. 76
Bijlage 6	Resultaten vragenlijst studenten	pag. 77
Bijlage 7	Tabel uitkomst factoranalyse	pag. 78

Voorwoord

Deze thesis richt zich op de kwaliteit van de toetsen van de methode Taalblokken, voor moderne vreemde talen, van uitgeverij Malmberg. Dit onderzoek is uitgevoerd in opdracht van uitgeverij Malmberg als een verlengde van de stageopdracht voor de master opleiding onderwijskundig ontwerp en advisering. Binnen de stage is een advies uitgegeven om de toetsen van de methode Taalblokken te verbeteren. Deze thesis richt zich op de theoretische grondslag achter dit advies. Het onderzoek heeft plaatsgevonden in de periode januari 2008 tot augustus 2008.

Er zijn een aantal mensen betrokken geweest bij dit onderzoek. Allereerst wil ik Jeroen Janssen bedanken voor zijn goede en intensieve begeleiding. Je feedback en opbeurende gesprekken hebben me geholpen deze prestatie neer te kunnen zetten. Dan wil ik graag Lotte Govers, Marjon Dorjee en Alice van Laarhoven bedanken voor het voorzien van feedback op mijn conceptversies van mijn thesis. Dankzij jullie inzicht heb ik nog veel verbeteringen aan kunnen brengen. Ook wil ik graag de docenten en studenten van de ROC's bedanken voor hun tijd en inzet. Door jullie heeft het onderzoek vorm gekregen. Als laatste wil ik Claartje Bom en Hans Prins van uitgeverij Malmberg bedanken voor de begeleiding die ze mij hebben gegeven tijdens mijn stage. Tijdens het schrijven van mijn thesis heb ik daar veel profijt van gehad. Zonder jullie medewerking had deze thesis niet geschreven kunnen worden. Erg bedankt allemaal.

Femke van Laarhoven

Samenvatting

Binnen deze thesis wordt de kwaliteit van de toetsen van de methode Taalblokken van Malmberg besproken. Taalblokken is een methode voor moderne vreemde talen, opgezet aan de hand van het common European framework of reference for languages (CEFR). De kwaliteit van de toetsen is bepaald aan de hand van twaalf kwaliteitscriteria van Baartman, Kirschner, Prins en Van der Vleuten (2007) voor competentiegericht toetsen. De kwaliteit is vastgesteld door het afnemen van een interview bij twee docenten van ROC's. Hiernaast is een vragenlijst afgenomen bij 64 studenten van het beroepsonderwijs (MBO). Bij een medewerker van Malmberg is een interview afgenomen. Ten slotte zijn veertien toetsen beoordeeld aan de hand van een, voor dit onderzoek, ontworpen beoordelingsinstrument. Alle instrumenten zijn opgezet aan de hand van de operationalisering van de twaalf kwaliteitscriteria van Baartman en collega's (2007).

Uit het onderzoek is moeilijk te concluderen aan welke kwaliteitscriteria de methode Taalblokken *volledig* voldoet. Niet alle criteria zijn onderzocht door middel van interviews, een vragenlijst en het beoordelingsinstrument voor toetsen. Dit is te wijten aan praktische beperkingen zoals tijd en steekproefgrootte. Volgens de docenten voldoet de methode aan accepteerbaarheid, tijd en kosten, onderwijskundige gevolgen, eerlijkheid, geschiktheid voor onderwijsdoelen, betekenisvolheid en transparantie. Volgens de studenten voldoet de methode aan tijd en kosten, eerlijkheid en transparantie. Het beoordelingsinstrument voor toetsen geeft aan dat de methode voldoet aan vergelijkbaarheid en betekenisvolheid. Als laatste geeft de uitgever van Taalblokken aan dat de methode voldoet aan tijd en kosten. Dat niet alle criteria volledig naar voren komen kan verklaard worden door de grote veranderingen die plaats hebben gevonden binnen het MBO, gericht op competentiegericht onderwijs (CGO). Docenten en studenten moeten hier nog in groeien en het onderwijsconcept integreren in hun onderwijs. Daarnaast wordt er vanwege praktische redenen niet volledig gebruik gemaakt van alle handvaten die binnen de methode Taalblokken aanwezig zijn om CGO/ toetsen vorm te geven. Als laatste kan gezegd worden dat de methode op een aantal aspecten nog tekort schiet en daarop verbeterd kan worden. Voorbeelden van deze verbeteringen zijn: het duidelijker aanbieden van doelomschrijvingen bij de toetsen, uitbreiden van instructies bij toetsen, een betere docentenhandleiding en implementatie van de methode per ROC.

De instrumenten die gebruikt zijn binnen het onderzoek zijn niet optimaal betrouwbaar waardoor meerdere gevonden resultaten nietig verklaard kunnen worden. Voor vervolgonderzoek is het daarom een aanbeveling de instrumenten uit te breiden en concreter te maken. De dataverzameling zou plaats moeten vinden met een grotere steekproef. Door praktische beperkingen van dit onderzoek is de onderzoeksgroep gering gebleven wat de resultaten minder generaliseerbaar maakt.

1. Inleiding

De aanleiding voor dit onderzoek is een aantal maatschappelijke problemen. Deze zijn de aanleiding voor de ontwikkeling van nieuwe onderwijsmethoden en nieuwe programma's voor competentie assessment. In Nederland speelt het taalniveau van burgers een grote rol. Het niveau van de Nederlandse taallerende is niet voldoende. In een onderzoek naar het instroomniveau van MBO leerlingen op het Mondriaan college in Den Haag bleek dat 73% van de leerlingen lager scoren dan het verwachte instroomniveau (Taalbeheersing voorwaarde voor toetreding arbeidsmarkt, persbericht, 2007). Dit levert veel zorgen op binnen het onderwijs.

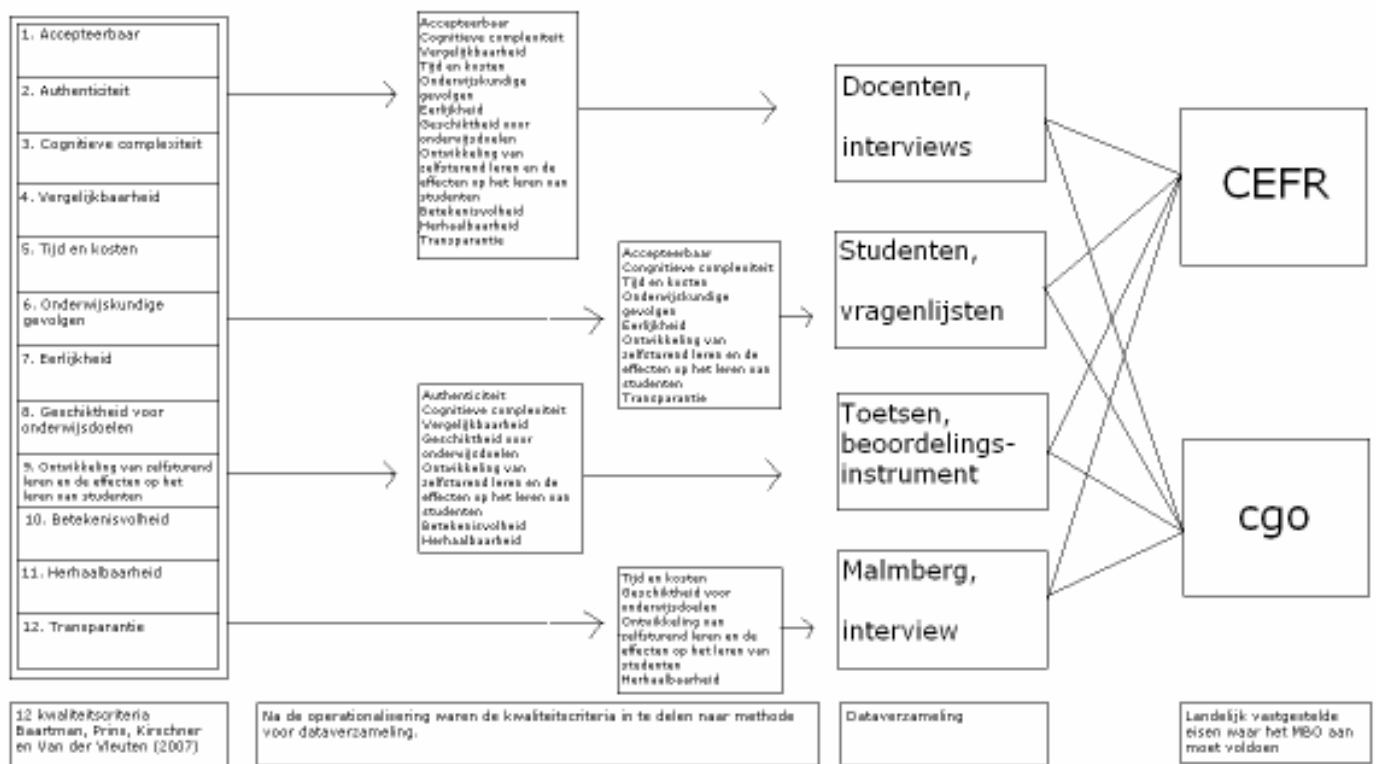
De ontwikkeling van taalcompetenties speelt ook binnen Europa een grote rol. De overheid wil dat de burger mobiel is en zichzelf kan redden met taal in andere landen. Om dit te realiseren is een document ontwikkeld dat taalcompetenties voor de Europese burger omschrijft. Dit bestand is ontwikkeld door de Council of Europe en omschrijft een zestal taalcompetentie niveaus, ook wel CEF (common European framework) niveaus genoemd. Binnen deze thesis zullen de termen CEF en CEFR door elkaar gebruikt worden. Het CEFR is het document dat alle aspecten omschrijft van de CEF- niveaus. Hiermee wordt ook bijvoorbeeld de toepasbaarheid en toetsing van de CEF- niveaus bedoeld. De CEF- niveaus zijn dus een onderdeel van het CEFR. Dit document fungeert als een basisdocument voor alle landen van Europa. Voor Nederland is dit basisdocument vertaald in een raamwerk 'Nederlands' en een raamwerk 'moderne vreemde talen'. Hierin staat specifiek omschreven wat Nederland verwacht van hun burgers binnen deze zes CEF- niveaus op het gebied van Nederlands en de moderne vreemde talen.

Naast dit raamwerk is het bestand 'Leren, loopbaan, burgerschap' (Stuurgroep Competentiegericht Beroepsonderwijs, 2007) een belangrijke leidraad voor het MBO in Nederland. Dit document beschrijft de kwalificatie-eisen op het gebied van burgerschap, leren en loopbaan van Nederlanders. Deze kwalificatie-eisen worden ook vormgegeven in de vorm van CEF- niveaus. Dit betekent dat een timmerman op het gebied van moderne vreemde talen een lager CEF- niveau behoort te behalen dan bijvoorbeeld een secretaresse. Deze eisen staan naast de beroepseisen die gesteld worden aan de Nederlandse burger om in een bepaald beroep aan de slag te kunnen. Die beroepseisen staan in het kwalificatiedossier. Een kwalificatiedossier beschrijft de inhoud van een beroep, de benodigde competenties en de benodigde kennis en vaardigheden voor een beginnende beroepsbeoefenaar. Het document 'Leren, loopbaan en burgerschap' en het kwalificatiedossier geven dus samen aan wat verwacht wordt van iemand die een MBO- diploma heeft (Stuurgroep Competentiegericht Beroepsonderwijs, 2007).

In dit onderzoek zal de kwaliteit van een competentie assessment programma (CAP) dat vorm gegeven is binnen de methode Taalblokken van uitgeverij Malmberg onderzocht worden. De kwaliteit van dit CAP wordt gemeten met behulp van de kwaliteitscriteria van Baartman en collega's (2007). Een overzicht van de onderdelen binnen dit onderzoek wordt weergegeven in Figuur 1. Hierbij staan de twaalf kwaliteitscriteria van Baartman en collega's (2007) voorop. De in dit onderzoek gebruikte instrumenten zijn ontworpen aan de hand van deze twaalf kwaliteitscriteria. Na een operationalisering per kwaliteitscriterium zijn passende vragen voor de interviews bij docenten en een medewerker van Malmberg geformuleerd, is een vragenlijst opgesteld voor studenten en is een beoordelingsinstrument ontwikkeld voor de analyse van de

toetsen. Verder zijn twee blokken met daarin het CEFR en CGO toegevoegd. Zoals eerder genoemd moet het MBO aan deze twee normen voldoen en dus ook het CAP van Taalblokken van Malmberg. Daarom is met deze normen rekening gehouden bij het bevragen van docenten, studenten, een medewerker van Malmberg en het analyseren van toetsen. Zoals al eerder genoemd zijn dit twee eisen waar het MBO aan moet voldoen en waaraan dus ook het CAP van Taalblokken van Malmberg aan moet voldoen.

In deze thesis zal er eerst een theoretisch kader gepresenteerd worden dat gericht is op het competentiegerichte onderwijs, competentiegerichte toetsen en het CEFR. Daarna zal omschreven worden welke methoden er worden gebruikt om de onderzoeksvraag te beantwoorden. De dataverzameling zal hierna besproken worden. Na de omschrijving van de dataverzameling zullen de resultaten die omschreven zijn gekoppeld worden aan de discussie en een conclusie.



Figuur 1 Conceptueel raamwerk onderdelen masterthesis.

2. Theoretisch kader

Binnen deze thesis staan een aantal aspecten centraal. Deze zijn respectievelijk, CGO, de evaluatie van competentiegerichte toetsen en het CEFR. Deze onderdelen zullen in het onderstaand theoretisch kader toegelicht worden.

2.1 Competentiegericht onderwijs

Het huidige onderwijs verschuift steeds meer van een leerkrachtgestuurd onderwijssysteem naar een leerlinggestuurd onderwijssysteem. Leerlingen moeten zich in de huidige maatschappij steeds beter zelf kunnen redden. Kennis en vaardigheden verouderen snel in de huidige maatschappij. Werkgevers hebben vooral belangstelling voor werknemers die vakinhoudelijk onderlegd zijn maar vooral ook over specifieke persoonlijke kwaliteiten beschikken. Die specifieke eigenschappen moeten ervoor zorgen dat de ontwikkeling en vernieuwing van een vakgebied soepel verloopt (Mulder, Wesselink, Biemans, Nieuwenhuis & Poell, 2003). Persoonlijke kwaliteiten blijken hierin een belangrijkere rol te spelen dan specifieke vakkennis. Daarnaast zijn studenten van tegenwoordig kritische consumenten geworden. De student van nu wil weten wat het nut is van wat hij leert. Hij wil best leren, maar wil ook weten waar dat voor nodig is in relatie tot zijn leerdoel. Het onderwijs is nu vooral gericht op het aanleren van kennisdelen en vaardigheden. Vervolgens stuurt het de student de (beroeps)praktijk in waar hij/zij zelf moet proberen al het geleerde te verbinden tot een samenhangend geheel.

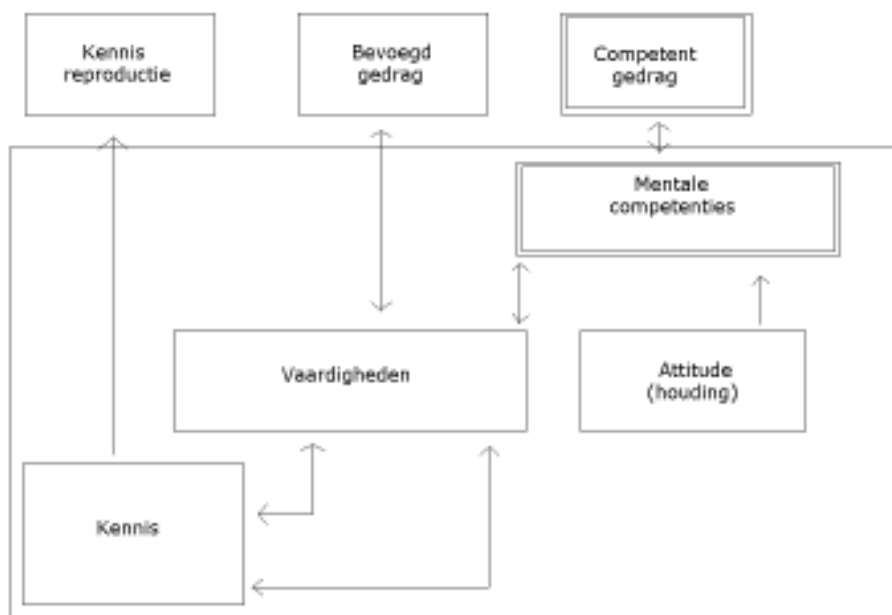
Het voorgaande probleem waarbij studenten zelf in de beroepspraktijk kennis en vaardigheden moet gaan koppelen wordt tegenwoordig steeds vaker binnen het beroepsonderwijs al aangeleerd door middel van competenties. Competenties bieden de mogelijkheid om kennis en vaardigheden aan elkaar te koppelen. Hiermee wordt bedoeld dat de onderdelen niet zozeer los van elkaar worden gezien maar meer als een geïntegreerd geheel. Tevens geven zij beter inzicht in de samenhang tussen wat geleerd moet worden (Competentie- en taalgericht toetsen, n.d.) en hoe dit geleerde in de praktijk tot uiting komt. De gedachte is dat met CGO een betere link gelegd kan worden tussen de nieuwe eisen van de samenleving, arbeid, arbeidscontracten en het onderwijs. Het idee hierbij is om met behulp van dit onderwijsconcept jongeren beter gereedschap mee te geven zodat ze meer zelfredzaam zijn binnen de maatschappij, en beter functioneren als medewerker in het bedrijfsleven en binnen hun loopbaan. Hierbinnen spelen competenties een grote rol (Mulder et al., 2003).

Deze hiervoor omschreven veranderingen hebben zich de afgelopen jaren binnen het MBO afgespeeld. De start van competentiegericht leren wordt omschreven in het voorstel "Koers BVE" (Sociaal-economische raad, 2001). Hierin wordt helder uiteengezet waarom er behoefte is aan competentiegericht leren binnen het MBO. Er is een vraag- en aanbodzijde in het beroepenveld. Aan de aanbodzijde is sprake van veroudering van de beroepsbevolking. Dit zorgt voor spanning op de arbeidsmarkt. Aan de vraagzijde zijn twee ontwikkelingen gaande die zijn begonnen aan het begin van 2001. De eerste ontwikkeling is de aanhoudende economische groei en werkgelegenheidsontwikkeling. De tweede ontwikkeling is die van een snelle 'upgrading' van het kwalificatieniveau van beroepen. Hiermee worden de snelle technologische ontwikkelingen, internationalisering van het werkveld en het toenemende belang van kennis in de

beroepsuitoefening bedoeld. Opleidingseisen nemen door deze veranderingen steeds meer toe en ook het niveau van de gevraagde competenties wordt steeds hoger. Werknemers zullen vaker hun kennis en vaardigheden moeten oprispen om bij te blijven met de ontwikkelingen die plaatsvinden in het beroepenveld. De kennis en vaardigheden die studenten nu opdoen in hun opleiding zijn niet voldoende om een leven lang te kunnen functioneren in het beroepenveld. Dit wijst erop dat veranderingen in de beroepspraktijk onvoldoende hun aansluiting vinden bij onderwijsprogramma's. Hierdoor zijn er nog teveel verschillen tussen de eindtermen van de kwalificatiestructuren en de inzetbaarheid van gediplomeerde arbeidskrachten. Om deze aansluiting beter te laten verlopen is een begin gemaakt met de invoer van competentiegericht leren in het MBO (Sociaal-economische raad, 2001).

2.2 Het begrip competentie

In de literatuur wordt geen duidelijke definitie van het begrip competentie gegeven. Het probleem bij het begrip competentie is dat er geen vaststaande definitie bestaat om het begrip helder te krijgen. Stoof, Martens en Merriënboer (2000) vragen zich in hun artikel af waardoor dit komt. De verklaring wordt gezocht in het semantische domein (de betekenis van woorden). Bij het verwoorden van het begrip competentie kan er niet gerefereerd worden aan een concreet voorbeeld zoals bij het begrip vogel. Een competentie kan niet aangewezen worden, het is een symbolisch begrip (Stoof et al., 2000; Eraut, 1994). Het begrip competentie kan dus in dezelfde lijn gezet worden als de begrippen liefde, intelligentie, schoonheid, et cetera. In deze thesis zal uitgegaan worden van het competentie model van Westera (2001), zie Figuur 2. Dit model bevat alle onderdelen van het begrip competentie zoals uiteengezet door verschillende onderzoekers. Naar aanleiding van het model van Westera (2001) zal besproken worden hoe alle componenten van competentie samenvallen.



Figuur 2 Competentie model volgens verschillende definities (Westera, 2001)

Omdat er geen vaststaande definitie bestaat van competentie, hebben veel onderzoekers een eigen definitie gecreëerd. Westera (2001) sluit zich aan bij de visie van Grondlund die competenties heel zwart wit ziet en onder het woord 'competentie' alleen de operationaliseerbare vaardigheden zoals lezen en schrijven verstaat. Dit zijn binnen het model van Westera (2001) de vaardigheden. Naast deze vaardigheden moet er bij competenties ook rekening gehouden worden met attitude (houding/ gedrag). Op het gebied van attitude bestaat echter een spanningsveld. Onderzoekers zijn het niet over eens of attitude bij competentie hoort. Westera (2001) vertaalt binnen dit raamwerk attitude als een 'gevoel/ houding'. Dit gevoel heeft invloed op de kennis en vaardigheden zoals ook in het model blijkt aan de dubbele omranding rondom 'mentale competenties'. Westera (2001) noemt hierbij de visie van Stephenson en Weil, die zeggen dat competenties ook componenten van attitude bevatten en dat ze dus veel breder zijn dan alleen operationaliseerbare vaardigheden. Kennis en vaardigheden zijn meetbaar en daarom relatief gemakkelijk aan te leren in tegenstelling tot attitude. Stoof en collega's (2000) noemen hierbij de visie van Spencer en Spencer, die zeggen dat attitude moeilijk te ontwikkelen is omdat het "onder de waterlijn" is, niet tastbaar. Van der Maesen de Sombreff en Schakel (1999) zeggen echter dat attitude geen onderdeel is van competentie. Competenties kunnen namelijk aanwezig zijn zonder dat je ze ziet. Een persoon kan actief zijn zonder dat die activiteit in gedrag observeerbaar is, bijvoorbeeld als de persoon nadenkt. Competenties zijn interne structuren die onafhankelijk zijn van concrete situaties waarop gereageerd wordt. Competenties zijn volgens Van der Maesen de Sombreff en Schakel (1999) een noodzakelijke voorwaarde voor gedrag, maar zijn niet gelijk te stellen met gedrag. Westera (2001) voegt hier het kenniscomponent aan toe en zegt dat een competentie een zeer gewaardeerde kwalificatie is die staat voor het effectief inzetten van kennis en vaardigheden in een specifieke, meestal complexe situatie waarin moeilijke beslissingen gemaakt moeten worden. Westera (2001) noemt ook de visie van Barnett die zich hierbij aansluit door competentie te definiëren als het maken van beslissingen in onvoorspelbare en onverwachte situaties. Bij deze situaties zijn in het model van Westera (2001) "mental competences" ook van belang. Dit zijn de processen die in de hersenen plaatsvinden bij het verwerken van kennis. Cognitieve vaardigheden zijn daarnaast ook verbonden met metacognitieve activiteiten zoals probleem oplossen, beredeneren, denken, toetsen, concluderen, et cetera. Bij deze activiteiten wordt gebruik gemaakt van mentale processen zoals analyseren, synthetiseren en evalueren (Bloom & Nemetz, 1957). White (1959) noemt bij mentale processen ook het gevoel van zelfverzekerdheid, het gevoel dat iemand heeft als hij een bepaalde taak of studie goed heeft afgerond. Deze mentale processen hebben weer invloed op het competent gedrag van een persoon, zoals ook blijkt uit de pijl van mentale processen naar competent gedrag. Kennis, vaardigheden en attitude zijn verbonden met elkaar in het model van Westera (2001). Dit blijkt uit het kader dat om deze begrippen staat. Dit houdt in dat bij het bekijken van CGO volgens Westera (2001) rekening gehouden moet worden met het geïntegreerd onderwijzen van kennis, vaardigheden en attitude. Hierbij moet rekening gehouden worden met de mentale processen die daardoor plaatsvinden. Als laatste noemt Westera (2001) de visie van Chomsky. Hij voegt hier aan toe dat een competentie een algemene en continue prestatie is die individueel en variabel plaatsvindt. Een competentie wordt niet afgerond, deze is altijd in ontwikkeling (Stoof et al., 2000). Het is dus een continu individueel en variabel proces. Ieder persoon heeft zijn eigen punten waar hij goed in is en interesse in heeft, deze zal hij

zelf beter ontwikkelen dan iemand anders. Je bent zelf verantwoordelijk voor de ontwikkeling van je competenties.

2.3 Assessment van competenties

Binnen deze thesis zal vooral gekeken worden naar het assessment van competenties binnen de methode Taalblokken van Malmberg, rekening houdend met het CEFR. Hier zal later een toelichting op volgen. Nu zal eerst worden ingegaan op competentiegericht toetsen.

Een belangrijk onderdeel van het onderwijs is het assessment. Assessment bestaat uit een aantal procedures die gebruikt worden om informatie te verzamelen over de prestaties van studenten (Johnsen, 1997). Met de verandering van het onderwijs naar CGO verandert ook het assessment. Competentiegerichte curricula vereisen verschillende benaderingen van assessment om competentieverwerving te bepalen. Competenties kunnen gezien worden als de capaciteit om een specifieke combinatie van kennis, vaardigheden en attitudes binnen een specifieke context te plaatsen. Een assessment zal gericht moeten zijn op de integratie van deze drie elementen (Baartman et al., 2007).

In de overgang van 'traditionele assessment' naar 'assessment van competenties' heeft de kwaliteit van het assessment een centrale rol gespeeld. Er is veel commentaar gekomen op het traditionele kennisassessment. Birenbaum, Breuer, Cascallar, Dochy, Dori, Ridgway, Wiesemes en Nickmans (2006) vinden dat de traditionele assessment teveel gefocust is op het assessment *van* het leren in plaats van het assessment *voor* het leren. Het assessment is beperkt in omvang en negeert individuele verschillen in het onderwijs. Het is nog teveel summatief gericht terwijl het assessment binnen CGO meer formatief gericht moet zijn. Het moet ook gericht zijn op de ontwikkeling van de studenten (formatief) en niet exclusief op het resultaat dat de studenten uiteindelijk moeten behalen (summatief).

Het assessment van competenties is erg complex omdat sprake is van een integratie van kennis, vaardigheden en attitudes. Vanwege deze complexiteit, is het onmogelijk om competentie te toetsen aan de hand van één assessmentmethode (Baartman, Bastiaens, Kirschner & Van der Vleuten, 2006). Van der Vleuten en Schuwirth (2005) zeggen bijvoorbeeld dat een assessment niet gezien moet worden als een psychometrisch probleem dat opgelost kan worden met één methode van assessment. Het moet meer gezien worden als een instructie- ontwerpprobleem dat omringd is door heel veel assessmentmethoden die gebruikt worden binnen één curriculum. Baartman en collega's (2006) pleiten ook voor de integratie van verschillende assessmentmethoden in één CAP. Hierbinnen wordt rekening gehouden met nieuwe methoden voor assessment in combinatie met meer klassieke methoden voor assessment. Het probleem is echter volgens Baartman en collega's (2006), dat de ontwikkeling van de assessmentmethoden wordt belemmerd omdat het niet duidelijk is wat de voorwaarden zijn voor dit soort assessment. Zijn de traditionele criteria zoals validiteit en betrouwbaarheid voor toetsen nog steeds toepasbaar en voldoende of zijn er aanvullende criteria nodig? Aan de hand van deze vraag hebben Baartman en collega's (2007) twaalf kwaliteitscriteria ontwikkeld voor een CAP waaraan competentiegerichte toetsen getoetst kunnen worden.

Bij het gebruik van deze kwaliteitscriteria is het belangrijk dat een CAP in zijn geheel wordt bekeken. Hiervoor wordt in het artikel van Baartman en collega's (2007) een aantal redenen genoemd. Ten eerste is het belangrijk omdat, zoals hierboven genoemd, een competentie een

geïntegreerd geheel is van kennis, vaardigheden en attitudes. Deze kunnen niet door één bepaald assessment tegelijkertijd getoetst worden (Van der Vleuten & Schuwirth, 2005) het is hierbij dus belangrijk om naar het geheel van assessmentmethoden te kijken.

Als tweede punt is het belangrijk om een brede benadering te hebben van het leerarrangement waarbij de aandacht gericht is op alles wat aanwezig is op het gebied van assessment in het complete onderwijsprogramma. Het voordeel van deze benadering is dat de druk op de betrouwbaarheid van de toetsen met lage consequenties voor de student (bijvoorbeeld formatieve toetsen) afneemt. Die middelen kunnen geïnvesteerd worden in de ontwikkeling van kostbare, duurzame en betrouwbare (en meer valide) assessments waar dat nodig is, met name bij toetsen met hoge consequenties voor de student (bijvoorbeeld summatieve toetsen).

Ten derde heeft een CAP zowel summatieve als formatieve doelen. Het hoofddoel van formatief toetsen is het leveren van feedback en het genereren van geschikte leeractiviteiten. Het doel van summatief toetsen is vooral het leveren van een eindbeoordeling. In het CAP wordt aan beiden aandacht besteed, wel wordt er rekening gehouden met de verschillen tussen beiden.

Aan de hand van de afwegingen die genoemd zijn, zijn twaalf kwaliteitscriteria door Baartman en collega's (2007) samengesteld en gevalideerd waaraan een CAP moet voldoen. In het volgende onderdeel van het theoretisch kader zal er ingegaan worden op deze twaalf kwaliteitscriteria. Naar aanleiding van de literatuur zullen deze onderbouwd en toegelicht worden. Daarnaast zal uitgelegd worden hoe deze later bij het onderdeel 'methode' concreet gemaakt worden zodat ze toepasbaar zijn binnen het beoordelingsinstrument dat in deze thesis is ontwikkeld.

2.4 Twaalf kwaliteitscriteria

Binnen deze thesis zal de methode Taalblokken van Malmberg onderzocht worden. Hierbij zal gekeken worden naar de twaalf kwaliteitscriteria zoals gesteld door Baartman en collega's (2007). Hieronder zullen deze twaalf kwaliteitscriteria verder toegelicht worden.

2.4.1 Accepteerbaarheid

Onder accepteerbaarheid (*acceptability*) wordt verstaan dat de doelen die gesteld worden binnen het CAP voor iedereen acceptabel zijn (Baartman et al., 2007). Het is hierbij belangrijk om rekening te houden met meningen, gevoelens en tradities. Deze bepalen tot welke mate een toetsmethode is geaccepteerd door zowel leerkrachten, leerlingen, als de beroepspraktijk. Het is van belang om hierop te letten bij het ontwerpen en implementeren van een assessmentmethode omdat het invloed kan hebben op de uitvoer hiervan (Van der Vleuten, Scherpbier, Dolmans, Schuwirth, Verwijnen en Wolfhagen, 2000). De invloed kan bijvoorbeeld bestaan uit gedemotiveerde studenten of docenten die het assessment programma niet op een goede wijze uitvoeren.

2.4.2 Authenticiteit

Authenticiteit (authenticity) is de mate waarin het CAP te vergelijken is met de toekomstige beroepspraktijk (Baartman et al., 2007). Gulikers, Bastiaens en Kirschner (2004, p. 69) geven de volgende definitie voor een authentieke assessment: "Een assessment dat ervoor zorgt dat

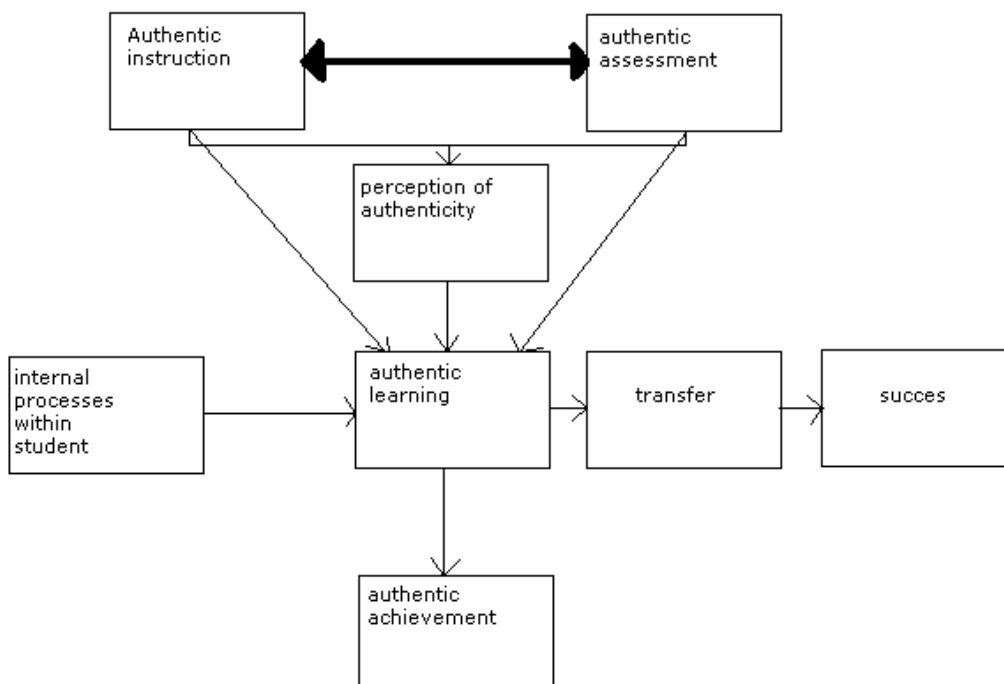
studenten dezelfde competenties, of combinatie van kennis, vaardigheden en attitudes, gebruiken die ze nodig hebben in een situatie in de beroepspraktijk”.

In het onderstaande model (zie Figuur 3) komen de drie elementen van authenticiteit naar voren:

1. Authentieke assessments moeten vooraf gegaan worden door authentieke instructie om de studenten positief te beïnvloeden;
2. Authentieke assessments moeten ervoor zorgen dat studenten hun competenties kunnen laten zien door een betekenisvolle en waardevolle prestatie;
3. Authenticiteit is subjectief, wat ervoor zorgt dat de mening van studenten erg belangrijk wordt met betrekking tot de invloed van het assessment op leren.

Naast deze drie elementen bestaan vijf dimensies van de authenticiteit van assessments, namelijk:

1. *De te toetsen taak*, deze moet betekenisvol en relevant zijn, zoveel mogelijk identiek zijn aan de opdracht die in de praktijk uitgevoerd wordt en moet rekening houden met de integratie van kennis, vaardigheden en attitude;
2. *De fysieke context* is de omgeving waarin de deelnemers de praktijkopdracht moeten uitvoeren, deze moet zoveel mogelijk reflecteren hoe kennis, vaardigheden en attitude gebruikt worden in de beroepspraktijk;
3. *De sociale context* zijn de sociale processen die tijdens de uitvoer van de opdracht een mogelijke rol zullen spelen, zoals samenwerken, overleggen, et cetera;
4. *Het resultaat of de vorm van de toets* moet een product of een prestatie zijn die later ook in de beroepspraktijk van de studenten gevraagd kan worden. Het moet een prestatie zijn die de mogelijkheid geeft om over de te toetsen competenties te oordelen, moet gepresenteerd kunnen worden aan andere studenten;
5. *De criteria voor toetsen* zijn de criteria waarop werknemers in de beroepspraktijk afgerekend worden (Gulikers et al., 2004).



Figuur 3 General framework (Gulikers et al., 2004)

Dekker en Sanders (2008) geven een definitie van authentiek toetsen: "Authentiek toetsen is een vorm van toetsen waarbij deelnemers moeten demonstreren over de competenties te beschikken die van beginnende beroepsbeoefenaars verlangd worden". Dekker en Sanders (2008) geven ook twee redenen voor het gebruik van authentieke toetsen. Ten eerste, het doel van ieder beoordelingsinstrument moet zijn om datgene te meten wat mensen met dat 'beoordelingsinstrument' willen meten. Wanneer iemand wil weten of een student een sollicitatiegesprek in het Engels kan voeren, zal een praktijksituatie een betere beoordeling van dit proces en bijbehorende competenties opleveren dan een schriftelijk examen met vragen over dit proces en de competenties. Ten tweede heeft onderzoek laten zien dat de manier waarop deelnemers beoordeeld worden invloed heeft op hoe en wat deelnemers leren. Als het beoordelen met authentieke examens of toetsen er inderdaad toe leidt dat deelnemers meer gemotiveerd zijn om de competenties te verwerven die ze in de beroepspraktijk nodig hebben, dan is het de taak van het onderwijs die authentieke toetsen te ontwikkelen en af te nemen.

Klarus (2001) geeft in zijn trap 8 dimensies (zie ook Figuur 4) van authenticiteit. Deze zullen hieronder kort toegelicht worden:

1. *Portfolio-inventarisatie*, deze toetsingsvorm is het minst authentiek. Hierbij wordt alleen gekeken naar de verzamelde opdrachten en verslagen in het portfolio, en deze worden vergeleken met de te bereiken doelen. Competenties worden hierdoor boven tafel gehaald maar niet erkend.
2. *Kennistoets*, bij CGO wordt de kennis nooit los van de context en het toepassen van kennis getoetst. Een kennistoets heeft dus een lage authenticiteit omdat het niet wordt geïntegreerd met vaardigheden en attitude.
3. *Beoordeling verslagen*, bij het beoordelen van verslagen moet rekening gehouden worden met de volgende kwaliteitseisen: authenticiteit, actualiteitswaarde, relevantie, kwantiteit en variatie. Bij deze punten is authenticiteit van doorslaggevende betekenis. Het is de vraag of een student voldoende bewijs heeft geleverd om te bewijzen dat hij voldoet aan de competentie die getoetst wordt. Als authenticiteit goed is maar de rest matig dan zijn de bewijzen maar in beperkte mate geschikt.
4. *Schriftelijke beoordeling volledige handelingen*, schriftelijke beoordelingen zijn niet ideaal gezien de generaliseerbaarheid en validiteit bij het toetsen van beroepsvaardigheden. Schriftelijke toetsen hebben echter toch een aantal praktische voordelen. Het is denkbaar dat een schriftelijke beoordeling de hoeveelheid beoordelingen (observaties) in de beroepspraktijk tot een acceptabele omvang terug brengt, zonder dat de transferwaarde en generaliseerbaarheid van de beoordelingsresultaten in gevaar komen. Daarnaast kunnen schriftelijke toetsen ook betrekking hebben op een volledige handeling en de arbeidssituatie kan tot op een zekere hoogte gesimuleerd worden. Een simulatie of een authentieke toetssituatie kan veel tijd en geld kosten. Een schriftelijke toets, mits op beroepscompetenties gericht, kan dan naast een simulatie een alternatief zijn.
5. *Criteriongericht interview*, hierbij komt een docent erachter wat voor competenties een student heeft. Het is een mondelinge toets van ervaringen. Dit gebeurt met een speciale interview techniek, het Situatie Taken Actie Resultaten model. Doordat er diep ingegaan wordt op de beroepssituatie is deze methode ook authentiek te noemen.
6. *Portfoliobeoordeling*, in een portfolio zijn er verschillende bewijsstukken aanwezig die variëren van heel concreet en betrouwbaar tot bewijsstukken waarvan de betrouwbaarheid moeilijk is vast

2.4.3 Cognitieve complexiteit

Cognitieve complexiteit (cognitive complexity) is belangrijk omdat binnen een CAP de cognitieve vaardigheden van de student naar boven moeten komen. Daarnaast moet het CAP in staat zijn om denkprocessen van leerlingen te beoordelen (Baartman et al., 2007). Naast de denkprocessen moet het CAP ook handvaten geven om kennis en vaardigheden van studenten te beoordelen. Het moet de docent in staat stellen een goede beoordeling te geven.

Cognitieve complexiteit is een begrip dat in 1950 in de psychologie ontstaan is. Eerder onderzoek laat zien dat individuen met een hoge cognitieve complexiteit beter zijn in het maken van beslissingen en beter zijn in de sociale omgang met mensen. Cognitieve complexiteit is gerelateerd aan het sociaal functioneren van mensen, sociale interactie en intelligentie (Winn & Bittner, 2005). Uit onderzoek van Kuiken en Vedder (2008) blijkt dat taken die cognitief complex zijn en veel vragen van de student, de aandacht afhouden van de inhoud van de toets. Dit zorgt dat de student niet meer genoeg aandacht heeft voor de inhoud van het assessment. Hoe hoger de complexiteit van de toets hoe moeilijker het wordt om te bepalen of je nog de goede competentie meet. Belangrijk is dus goed in de gaten te houden hoe complex de te toetsen taak gemaakt wordt. Daarnaast hangt de complexiteit af van de student. Wat gemeten wordt door een individueel item in een toets zal afhangen van de cognitieve processen die gestimuleerd worden door het beantwoorden van de vraag. Dit zal niet alleen afhangen van item en tijd maar ook van persoon tot persoon voor een enkel item. Een reactie op een item zal voor de ene persoon gericht zijn op herkenning en bij de ander op gewoon logisch denken. Dit heeft te maken met de mentale processen die Westera (2004) noemt in zijn artikel. Een opsomming van de toetsitems moet daarom zorgen voor een heterogeen geheel (Winn & Bittner, 2005).

2.4.4 Vergelijkbaarheid

Vergelijkbaarheid (comparability) is belangrijk omdat een CAP op een consistente en betrouwbare manier moet worden toegepast. De taak, criteria en werkcondities moeten vergelijkbaar zijn voor alle studenten (Baartman et al., 2007). De beoordeling die een student ontvangt zou alleen afhankelijk moeten zijn van de kwaliteit van de prestatie. Het zou niet moeten uitmaken wie er beoordeelt, er moet sprake zijn van een objectieve beoordeling. Dit is echter een ideaal dat alleen door een sterk gestandaardiseerde/ traditionele toets behaald zou kunnen worden.

Interbeoordelaarsbetrouwbaarheid is vaak een belangrijk issue bij assessmentmethoden. Dit betekent dat twee docenten hetzelfde assessment moeten kunnen afnemen en hieruit dezelfde conclusies en beoordelingen moeten kunnen trekken (Dekker & Sanders, 2008). Dit kan ondervangen worden door de condities waaronder het assessment moet worden uitgevoerd voor alle studenten hetzelfde te houden. Het scoren van dit assessment kan dan ook op een consistente manier gedaan worden. De assessmentmethoden zijn echter niet zo gestandaardiseerd als traditionele toetsen, dus de scores zullen vaak gebaseerd worden op persoonlijke meningen/ beoordelingen (Dierick & Dochy, 2001). Richtlijnen voor assessment zijn erg belangrijk, zoals antwoordmodellen en lijsten met criteria. Deze zorgen ervoor dat de mening van de beoordelaars geneutraliseerd worden. Daarnaast is het belangrijk om meerdere beoordelaars te gebruiken (Van der Vleuten et al., 2000). Studenten moeten onderling resultaten kunnen vergelijken, de uitvoering van het assessment moet dus zo consistent en verantwoordelijk mogelijk plaatsvinden (Uhlenbeck, Verloop en Beijaard, 2002).

2.4.5 Tijd en kosten

Tijd en kosten (cost and efficiency) zijn belangrijk omdat de tijd en middelen die nodig zijn voor de ontwikkeling en de uitvoer van het CAP in verhouding moeten zijn met de (leer)opbrengsten (Baartman et al., 2007). De tijd en bronnen die nodig zijn voor de toets moeten afgezet worden tegen de voordelen. Er moet dus wel verbetering plaatsvinden van leren, lesgeven en participeren in de beroepspraktijk (Baartman et al., 2006).

2.4.6 Onderwijskundige gevolgen

Onderwijskundige gevolgen (educational consequences) is de mate waarin een CAP een positief effect heeft op het leerproces van studenten en het onderwijzen van docenten (Baartman et al., 2007). Ieder assessment heeft een onderwijskundige consequentie. Deze zijn echter moeilijk te voorspellen. Een consequentie van een bepaald assessment zou kunnen zijn dat studenten gaan leren voor de toets, ze leren dan mogelijk heel beperkt (Broekkamp & van Hout-Wolters, 2007). Er wordt dan alleen nog maar geleerd naar aanleiding van de stof die bevraagd wordt in de toets. Dit is een uitdaging voor toetsontwikkelaars. Ze kunnen dit fenomeen gebruiken om gewenst leergedrag naar boven te halen (Van der Vleuten, 1996). Er is bewijs nodig over de positieve en negatieve effecten van het assessment en hoe studenten het doel van het onderwijs zien en hoe ze hun leren, en docenten hun lesgeven, hierop aanpassen. Zo kunnen we achter het 'backwash' oftewel 'washback' effect van het onderwijs komen. Dit zijn de bedoelde en niet bedoelde positieve en negatieve effecten van een competentiegericht programma op leren en instructie (Alderson & Wall, 1993). Naar aanleiding van deze bewijzen kan door ontwikkelaars van competentiegerichte programma's gekeken worden hoe de onderwijskundige consequenties zo positief mogelijk blijven.

2.4.7 Eerlijkheid

Eerlijkheid (fairness) is belangrijk omdat studenten een eerlijke en gelijkwaardige kans moeten krijgen om hun competenties te laten zien. Dit kan bijvoorbeeld door studenten zichzelf op verschillende manieren te laten uiten en ervoor te waken dat er geen vooroordelen zijn aan de kant van de beoordelaars (Baartman et al., 2007). De toets moet dus alle kennis, vaardigheden en attitude dekken die ervoor zorgen dat de student de taak goed kan uitvoeren. Het moet een goede weergave zijn van de kennis die de student op dat onderdeel heeft, de vaardigheden die daarbij passen en de attitude die de student ten aanzien van het onderwerp heeft (Frederiksen & Collins, 1989; Linn et al., 1991; Uhlenbeck et al., 2002). Daarnaast is het belangrijk dat de toets goed aansluit bij het niveau dat de student heeft en dat de culturele aspecten die binnen de oefening een rol spelen, bekend zijn bij de student (Dierick & Dochy, 2001). De beoordeling moet voor alle studenten zo veel mogelijk gelijk gehouden worden. Studenten moeten de toetsresultaten met elkaar kunnen vergelijken. Dit kan alleen als iedere student een eerlijke kans heeft gehad en rechtvaardig is beoordeeld.

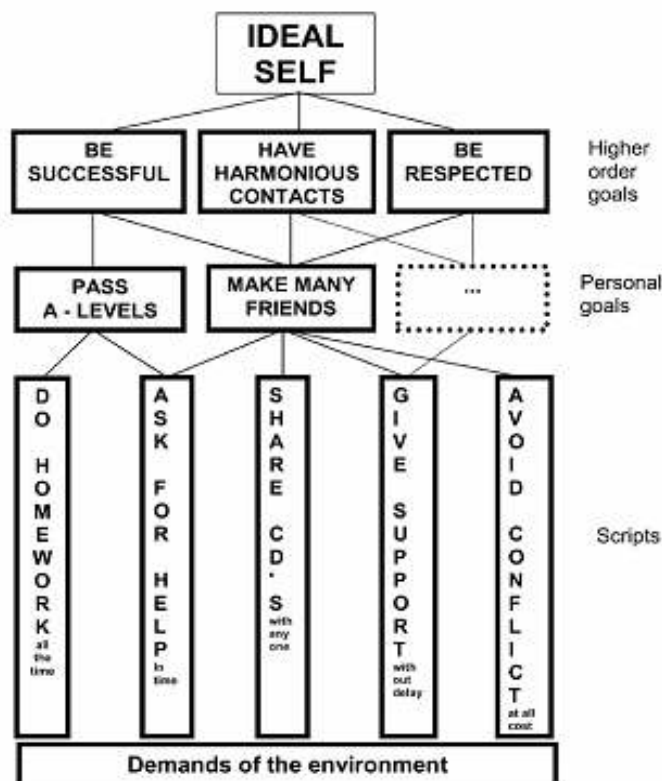
2.4.8 Geschiktheid voor onderwijsdoelen

Geschiktheid voor onderwijsdoelen (fitness for purpose) is belangrijk omdat de standaarden, curriculum, instructie en assessment met elkaar verbonden moeten zijn. De assessmentdoelen en -methoden moeten gelijk zijn aan de onderwijskundige doelen zoals gesteld in het curriculum (Baartman et al., 2007). Als methoden voor assessment veranderen, verandert ook het doel van

de toets. Het meest belangrijke element van de nieuwe assessmentmodellen is de reflectie over competenties die nodig zijn in de beroepspraktijk. Het doel is om ervoor te zorgen dat de succescriteria van het onderwijs hetzelfde blijven als die van de beroepspraktijk. Het is dus belangrijk om in de gaten te houden of het assessment nog steeds voldoet aan de onderwijsdoelen (Dierick & Dochy, 2001).

2.4.9 Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten

Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten (fitness for self assessment) is belangrijk omdat het CAP het zelfgereguleerd leren van studenten moet stimuleren. Daarnaast moet het studenten laten oefenen in zelfsturend leren en het geven en krijgen van feedback (Baartman et al., 2007). Zelfgereguleerd leren is geen nieuwe techniek voor assessment maar vooral een manier om studenten meer te betrekken bij hun eigen leerproces. Het wordt over het algemeen gebruikt als een formatieve assessmentmethode, als een vorm van reflectie op iemand zijn eigen leerproces en resultaten. Door middel van een portfolio kan zelfgereguleerd leren bijvoorbeeld gestimuleerd kunnen worden. Hierin houden studenten hun vorderingen bij en kijken ze hoe hun eigen leerproces vordert en ontwikkelt (Dierick & Dochy, 2001). Volgens Boekaerts (2002) is het van essentieel belang om persoonlijke leerdoelen van studenten te kennen om te begrijpen hoe en waarom ze hun leerproces vormgeven. Er zijn verschillende soorten doelen die een student kan hebben/ bereiken. Deze variëren van "higher-order goals" tot scripts die heel concreet aangeven hoe je iets gaat aanpakken. Carver en Scheier (2000) hebben hier een model van gemaakt (zie Figuur 5)



Figuur 5 Hiërarchie van leerdoelen (Carver & Scheier, 2000)

Binnen dit model wordt er onderscheid gemaakt tussen hoge doelen met een hoge abstractie, persoonlijke doelen en scripts die steeds concreter worden. Belangrijk hierin is dat, aan de hand van de doelen die leerlingen opstellen, te zien is hoe ze leren. Hiermee kan dus ook rekening gehouden worden bij het geven van een assessment en uiteindelijk voor de docent bij het geven van feedback op dit assessment.

2.4.10 *Betekenisvolheid*

Betekenisvolheid (meaningfulness) is belangrijk omdat het CAP voor alle betrokken partijen zinvol moet zijn (Baartman et al., 2007; Linn et al., 1991). Het assessment moet studenten betekenisvolle problemen geven die waardevol zijn voor hun onderwijservaring. Daarnaast is het belangrijk dat er een link gelegd wordt tussen het assessment en de persoonlijke interesse van de student. Zo krijgt het assessment meer betekenis voor de student. Ook docenten moeten de betekenis van het assessment duidelijk overzien. Zij moeten studenten aanzetten tot leren en daarbij is het belangrijk dat de docent het nut/ de betekenis inziet van het assessment.

2.4.11 *Herhaalbaarheid*

Herhaalbaarheid (reproducibility of decisions) is belangrijk omdat de beslissingen die gemaakt worden op basis van de resultaten van het CAP precies en constant moeten zijn voor situaties en toetsafnemers. Beslissingen mogen niet alleen afhangen van de beoordelaar en van een specifieke assessment situatie (Baartman et al., 2007; Van der Vleuten & Schuwirtz, 2005). Beslissingen die gemaakt worden over studenten moeten gebaseerd zijn op meerdere assessments, uitgevoerd door meerdere beoordelaars en op meerdere momenten. De term generaliseerbaarheid wordt ook wel gebruikt. Er moeten namelijk algemene conclusies getrokken kunnen worden over het leerproces van de studenten en niet over één bepaald aspect van dat leerproces (Linn et al., 1991).

2.4.12 *Transparantie*

Transparantie (transparency) is belangrijk omdat CAP's duidelijk en begrijpelijk moeten zijn voor alle betrokkenen. Externe controlebureaus moeten een helder beeld kunnen vormen over de manier waarop het CAP is ontwikkeld en uitgevoerd (Baartman et al., 2007; Dierick & Dochy, 2001; Frederiksen & Collins, 1989). De student moet de beoordelingscriteria kennen en ervan op de hoogte zijn. Daarnaast moet de student weten wie de beoordelaar is en wat het doel van de toets is. De student moet weten wat er van hem verwacht wordt zodat hij zich goed kan voorbereiden op het assessment en hierop zijn leerproces kan aanpassen (Frederiksen & Collins, 1989). Als een indicatie voor transparantie benoemen Baartman en collega's (2007) de theorie van Hambleton. Hij geeft aan dat leerlingen zichzelf net zo goed moeten kunnen beoordelen als hun docenten dat doen.

2.4.13 *Overlap tussen de kwaliteitscriteria*

Betrouwbaarheid en validiteit hebben een hele belangrijke betekenis binnen de psychometrie. Ze meten ethische belangen over wat gemeten wordt en de eerlijkheid daarvan. Ze moeten wel anders geoperationaliseerd worden om geschikt gevonden te worden voor CAP's en CGO (Baartman, 2008). Betrouwbaarheid is dan ook binnen het CAP door Baartman (2008) geoperationaliseerd als de criteria vergelijkbaarheid en herhaalbaarheid. Validiteit zit verweven in

ieder criterium omdat het een moeilijk aspect is om te operationaliseren, het speelt binnen ieder criterium een rol.

De 12 kwaliteitscriteria van Baartman en collega's (2007) overlappen elkaar op bepaalde punten. Er zijn bepaalde criteria die elkaar opvolgen of een sterk verband hebben met elkaar. Bijvoorbeeld de criteria accepteerbaarheid, vergelijkbaarheid, geschiktheid voor onderwijsdoelen, eerlijkheid, herhaalbaarheid en transparantie. Deze zes criteria vullen elkaar goed aan. Baartman (2008) operationaliseert zoals eerder genoemd betrouwbaarheid in vergelijkbaarheid en herhaalbaarheid. Deze worden apart genoemd in Tabel 1 omdat deze volgens Baartman (2008) de kern vormen van betrouwbaarheid. Alle criteria zorgen er echter allen voor dat het CAP duidelijk en begrijpbaar is voor docenten en studenten. Daarnaast toetsen ze of dit ook werkelijk zo is en of de betrouwbaarheid van het CAP hoog genoeg is. Omdat het criterium geschiktheid voor onderwijsdoelen toetst of alle doelen binnen een CAP aanwezig zijn, verhoogt dit de accepteerbaarheid. Als niet alle doelen aanwezig zouden zijn zouden docenten het CAP minder snel accepteren. Daarnaast zorgen de criteria vergelijkbaarheid, eerlijkheid, herhaalbaarheid en transparantie ervoor dat alle studenten gelijk beoordeeld worden. Dit staat dan weer in verband met de geschiktheid voor onderwijsdoelen en de accepteerbaarheid. Als niet alle doelen aanwezig zijn in een CAP worden studenten bijvoorbeeld landelijk niet gelijk getoetst omdat niet alle scholen met hetzelfde assessment programma werken. Zo toetsen deze zes criteria los van elkaar een ander aspect van een CAP maar houden ze ook verband met elkaar. Het ene criterium staat niet los van het andere criterium.

Dit geldt ook voor de criteria authenticiteit en betekenisvolheid. Een authentieke toetsomgeving geeft de toets meer betekenis voor de student. Een toets binnen de beroepspraktijk of een simulatie van de beroepspraktijk zorgt ervoor dat de toets gaat leven voor de student en hierdoor ook meer betekenis krijgt. Ook tussen cognitieve complexiteit, onderwijskundige gevolgen en ontwikkeling van zelfstandig leren en de effecten op het leren van studenten zit een verband. Dit verband is niet zo sterk maar wel aanwezig. Doordat rekening gehouden wordt met de cognitieve complexiteit van een toets, dus met de denkstappen die een student doorloopt heeft dit invloed op de onderwijskundige gevolgen. Als een leerkracht studenten aanstuurt op de denkstappen die hij doorloopt heeft dit gevolgen voor de manier waarop hij leert. Een student zal zijn studiestrategie misschien wel moeten aanpassen. En dit kan weer verband houden met de ontwikkeling van zelfstandig leren en de effecten op het leren van studenten. Studenten moeten binnen CGO leren om meer zelfstandig te werken. Doordat ze worden aangestuurd op het gebied van cognitieve complexiteit en onderwijskundige gevolgen leren studenten zich hierop aanpassen. Het bevordert de zelfstandigheid van studenten.

Tijd en kosten is een criterium dat eigenlijk verband houdt met alle criteria. Het heeft te maken met de praktische toepasbaarheid en totstandkoming van het CAP. Als dit om bepaalde redenen niet optimaal vorm wordt gegeven, kan dit de criteria beperken in hun volledigheid.

De punten binnen het raamwerk van Baartman en collega's (2007) zijn allemaal losstaande punten met hun eigen omschrijving en operationalisering. Veel punten vallen echter samen en overlappen elkaar. Het is om deze reden een sterk samenhangend geheel. Uit dit spanningsveld blijkt ook dat alle criteria naar voren moeten komen binnen de verschillende methoden die getoetst worden. Dit is belangrijk omdat criteria elkaar aanvullen en het ene criterium soms niet zonder het andere kan. Wel kunnen criteria compenseren door beter te scoren op het ene criterium dan op het andere. Ze kunnen elkaar aanvullen omdat ze veel met elkaar samenhangen.

Tabel 1 Samenhang tussen de verschillende kwaliteitscriteria

Twee belangrijke punten bij het meten van een CAP, zorgt voor ethische belangen over wat er gemeten wordt en de eerlijkheid daarvan.		Clusters binnen de kwaliteitscriteria met daarbinnen criteria die samenhangen en elkaar aanvullen.		
Tijd en kosten (basis van een goed CAP)				
Betrouwbaarheid	Validiteit	Geschiktheid voor onderwijsdoelen	Authenticiteit	Cognitieve complexiteit
↓	↓	↓	↓	↓
Vergelijkbaarheid Herhaalbaarheid	Zit verweven in ieder kwaliteitscriterium	Accepteerbaarheid	Betekenisvolheid	Onderwijskundige gevolgen
↓		↓		↓
Accepteerbaarheid Geschiktheid voor onderwijsdoelen Eerlijkheid Transparantie		Vergelijkbaar Eerlijkheid Herhaalbaarheid Transparantie		Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten

De twaalf kwaliteitscriteria zoals hierboven genoemd zullen later voor het beoordelingsinstrument verder geoperationaliseerd worden zodat ze toegepast kunnen worden op de toetsen Engels van Taalblokken, Malmberg.

Naast de vraag of de toetsen van Taalblokken voldoen aan de criteria van competentiegericht toetsen zal er ook gekeken worden of de toetsen aansluiten bij het CEFR. Het CEFR zal hieronder toegelicht worden.

2.5 Het CEFR

Het CEFR is een raamwerk voor taal dat als doel heeft de Europese burger mobieler te maken in zijn taalvaardigheden en om barrières te overbruggen op het gebied van communicatie. Het CEFR beschrijft op een veelomvattende manier wat de lerende moet leren om een taal te gebruiken voor communicatie. Daarnaast beschrijft het welke kennis en vaardigheden de lerende moet ontwikkelen om deze effectief te kunnen gebruiken. Professionals die werken in het veld van de moderne vreemde talen en voortkomen uit verschillende onderwijssettings kunnen door het CEFR beter met elkaar communiceren. Het zorgt voor Europese mobiliteit voor zowel taallerenden als werkenden (Council of Europe, 2001). De taalniveaus die in het CEFR worden gebruikt zijn voor iedereen gelijk. Dit zorgt ervoor dat alle lerenden van talen elkaar begrijpen op het niveau dat ze beheersen. Dit zorgt voor mobiliteit en zelfredzaamheid binnen de landen van Europa.

Het CEFR bestaat uit een basisdocument dat algemeen is en een basis verschaft voor alle landen in Europa. Het basisdocument is echter niet voor ieder land direct toepasbaar omdat het niet taalspecifiek is. Hierdoor bestaan er een aantal vertalingen/ varianten op het CEFR die gericht zijn op de verschillende talen binnen Europa. Binnen deze thesis zal vooral gekeken worden naar het raamwerk moderne vreemde talen (BVE raad, 2001), de Nederlandse vertaling van het CEFR.

Het raamwerk 'Nederlands' en het raamwerk 'moderne vreemde talen' zijn geschreven naar aanleiding van een lage taalvaardigheid van studenten op het MBO (MBO raad, 2007). Een grote

zorg bestond over de ontwikkeling van de talige competenties van studenten op het MBO. Docenten gaven aan dat ze meer behoefte hadden aan een instrument waarmee ze de ontwikkeling van talige competenties van studenten zichtbaar konden maken en sturen. Ze wilden graag een richtlijn waaraan ze het taalvaardigheidniveau van studenten konden beoordelen en op basis waarvan leermaterialen en een taalportfolio ontwikkeld konden worden. Daarnaast was er ook nog een sterke behoefte aan een Nederlandse ondersteuning voor de kwaliteitsprofielen in het MBO (MBO raad, 2007).

Kenmerkend voor het Europese referentiekader is de opdeling in niveaus (common reference levels) die de taalvaardigheid van de taallerende aangeeft (zie voorbeeld Bijlage 1). Er is een raamwerk van zes levels omschreven. Dit is volgens de council of Europe (2001) een goede weerspiegeling van het leerproces van de gemiddelde Europese taallerende. Binnen het basisdocument zijn deze levels respectievelijk: breakthrough (A1), waystage (A2), threshold (B1), vantage (B2), effective operational proficiency (C1) en mastery (C2). Deze zes levels zijn terug te brengen naar drie niveaus van taalontwikkeling; A, beginnend taalgebruiker (basic user), B, onafhankelijke taalgebruiker (independent user) en C, vaardig taalgebruiker (proficient user), (Council of Europe, 2001). Een taalgebruiker op niveau C2 functioneert op academisch niveau. Dit is een niveau dat normaliter niet voorkomt op het VO of MBO. Deze wordt dan ook binnen het raamwerk 'moderne vreemde talen' niet omschreven (BVE raad, 2001).

Binnen de verschillende niveaus van het raamwerk voor moderne vreemde talen (BVE raad, 2001) wordt een onderscheid gemaakt in de verschillende taalvaardigheden: luisteren, lezen, gesprekken voeren, spreken en schrijven. Binnen deze vaardigheden worden competenties omschreven die studenten moeten behalen om het betreffende taalvaardigheidniveau te behalen. Om de competenties specifiek te kunnen omschrijven worden er rubrieken opgenomen binnen de verschillende taalvaardigheden. Deze zijn: kenmerken van de teksten, algemene beschrijving van de vaardigheid, subvaardigheden, kenmerken van de taakuitvoering, strategieën en kennis (BVE raad, 2001). Deze zorgen ervoor dat de competenties concreet en volgens de SMART-doelen worden omschreven. Dit betekent dat de competenties die gesteld worden, Specifiek, Meetbaar, Acceptabel, Realistisch en Tijdsgebonden moeten zijn. Door de uitgebreide omschrijving van de competenties voldoen ze aan de eisen die gesteld worden aan SMART doelen zoals hiervoor genoemd. Ze zijn specifiek en makkelijk meetbaar. Daarnaast zijn ze makkelijk te accepteren en te begrijpen voor iedereen die betrokken is bij het onderwijs. Doordat het CAP authentiek moet zijn, worden de doelen ook realistisch opgesteld. Als laatste punt moeten de competenties ook tijdgebonden zijn. Dit punt verschilt per leerling en hierover kan binnen de competenties geen uitspraak gedaan worden. Een voorbeeld van een taalcompetentie is de volgende:

Luisteren A1: Ik kan vertrouwde woorden en basiszinnen begrijpen die mezelf, mijn familie en directe concrete omgeving betreffen, wanneer de mensen langzaam en duidelijk spreken.

2.5.1 Voor- en nadelen van het CEFR

Aan het basisdocument van het CEFR zitten een aantal voor- en nadelen. Een nadeel is bijvoorbeeld volgens Little (2007) dat het CEFR erg actiegeoriënteerd is. Een actiegeoriënteerd doel betekent dat het gefocust is op een relevante inhoud en ervaringen, ook holistische activiteiten, zodat leerlingen strategieën kunnen ontwikkelen. Dit betekent echter wel dat er veel

gedragsuitkomsten beschreven worden (leeruitkomsten). North (2007) vindt dat een raamwerk voor taalonderwijs juist tegemoet zou moeten komen aan het beschrijven van het uitbreiden van mogelijkheden en competenties die gerelateerd zijn aan de ontwikkeling van het kind. Daarnaast zou het de opbouw hiervan moeten aanmoedigen, er zou dus meer een formatief dan een summatief doel centraal moeten staan.

Een ander nadeel van het CEFR is dat er nog niet optimaal aandacht besteedt wordt aan reflectie (Little, 2007). Het is volgens Little (2007) juist de bedoeling dat leerkrachten/ gebruikers van het CEFR over heel Europa met elkaar kunnen praten over hun onderwijs. Dit betekent dus dat er meer aandacht zou moeten zijn voor reflectie gericht op het gehele taalonderwijs.

Een voordeel van het CEFR is echter wel dat er aandacht is voor reflectie op het niveau van de student. Byrnes (2007) bevestigt dit in zijn artikel. Hij vindt dat het CEFR goed aandacht besteedt aan het zelfsturend leren van studenten binnen het curriculum. Little (2007) geeft hier aan dat er speciaal binnen het CEFR een 'self- assessment grid' is gemaakt waarin studenten hun ervaringen en kwalificaties kunnen documenteren. Little (2007) verwijst hier naar de 'I can- statements' binnen het CEFR, vooral gericht op de formatieve kant van het toetsen waaraan een student zijn level kan bepalen. Een leerling moet hierbij bijvoorbeeld kunnen aangeven of hij de handleiding van een technisch apparaat in het Engels kan lezen en begrijpen. Aan de hand van de antwoorden van de student kan bepaald worden op welk niveau hij zich bevindt. Vooral binnen het Europees taalportfolio is hier veel aandacht voor (Little, 2007).

Een ander voordeel van het CEFR is dat de beschrijving van de niveaus altijd positief is. Ze refereren nooit aan wat een leerling niet kan. Ze zijn helder, kort, onafhankelijk en permanent (Little, 2007). Aan de andere kant zijn er over dit punt ook weer klachten gekomen. North (2007) beschrijft in zijn artikel dat gebruikers van het CEFR vinden dat de omschrijvingen erg de 'algemene taal' weergeven. Beschrijvingen zijn dus erg algemeen. Hierdoor is het moeilijk ze te specificeren voor een bepaalde context. Daarnaast geven gebruikers aan dat de beschrijvingen heel ingewikkeld zijn opgeschreven en vaak bedoeld zijn voor lerenden van zestien jaar en ouder. Dit laatste gegeven bevestigt Alderson (2007) in zijn artikel waarin hij aangeeft dat het raamwerk niet geschikt is voor jonge studenten. Bij het vertalen van het CEFR kwam men erachter dat er een gebrek aan definities was, er was sprake van overlap en tweeslachtigheden, er werd niet consistent omgegaan met terminologie en er waren grote gaten tussen de verschillende niveaus. Hieruit kan geconcludeerd worden dat de beschrijvingen niet optimaal zijn uitgevoerd. North (2007) geeft bij de discussie over de beschrijvingen van de niveaus aan dat een raamwerk contextvrij moet zijn om generaliseerbare resultaten te verkrijgen. Tegelijkertijd moeten de beschrijvingen van de niveaus context- relevant zijn, gerelateerd of vertaalbaar naar iedere relevante context en geschikt voor de functies die gekoppeld zijn aan die context. Dit betekent praktisch dat de beschrijvingen a) gerelateerd moeten zijn aan de categorisering in de theorie van taalcompetenties, b) relevant moeten zijn voor de context van de lerende populatie ondanks dat deze context niet met zekerheid voorspeld kan worden en c) gebruiksvriendelijk moeten blijven en toegankelijk moeten zijn voor mensen in de praktijk. Aan deze voorwaarden is niet volledig tegemoet gekomen.

Ten slotte is een voordeel van het CEFR dat de competenties op de verschillende niveaus niet alleen zorgen voor een mix tussen *wat* en *hoe goed* maar ook voor een niveau van linguïstische competenties zoals woordenschat, woordenschat controle, grammatica en fonologische (klank) controle. Dit maakt het raamwerk compleet, maar zorgt tegelijkertijd ook voor

onduidelijkheden. Men kan zich dan afvragen of een leerling op niveau B2 ook alle gebieden van B2 moet beheersen of dat hij bijvoorbeeld op het niveau van woordenschat nog op niveau B1 zit. Over deze inschaling van studenten bestaat nog onduidelijkheid (Hulstijn, 2007).

2.5.2 Toetsen binnen het CEFR

Het CEFR is niet alleen ontworpen om de mate van communicatie af te lezen aan afgenomen toetsen/ examens, maar ook om deze toetsen/ examens te vergelijken met elkaar, van taal tot taal en van land tot land. Er wordt door Little (2007) aangegeven dat het CEFR alleen dient als een uitgangspunt. Het is geen op zichzelfstaande didactiek die zo toegepast kan worden. Het kan wel gebruikt worden als een aanzet tot toetsontwikkeling.

Binnen het basisdocument (Council of Europe, 2001) worden drie toepassingsmogelijkheden van het CEFR binnen assessment genoemd. Ten eerste kan het raamwerk toegepast worden voor de specificatie van de inhoud van toetsen en examens. Ten tweede kan het raamwerk gebruikt worden voor het vaststellen van de criteria voor het behalen van een toets of niveau. Ten derde kan het raamwerk ingezet worden voor het beschrijven van de levels van bekwaamheid binnen toetsen en examens, hierdoor kunnen deze vergeleken worden en gebruikt worden voor verschillende systemen van kwalificatie.

Binnen het raamwerk 'moderne vreemde talen' (BVE raad, 2001), wordt aangegeven dat de taalniveaus de mogelijkheid bieden om de voortgang in de taalvaardigheid van studenten af te meten aan externe, niet in de leerstof gelegen normen. Dit is nuttig omdat er dan een algemeen beeld ontstaat van talige competenties van studenten in de desbetreffende vreemde taal. Naast het geven van cijfers voor beoordeling, worden de cijfers ook gekoppeld aan de Europese referentieniveaus. Dit geeft dan een indicatie van het bereikte taalniveau dat in heel Europa begrepen wordt.

Het raamwerk 'moderne vreemde talen' (BVE raad, 2001) helpt docenten en toetsontwikkelaars om praktijksituaties te kiezen die illustratief zijn voor een bepaald niveau. Dit helpen bestaat uit het aangeven van de moeilijkheidsgraad (van bijvoorbeeld tekstkenmerken) en het verschaffen van criteria voor het geven van een voldoende prestatie. Ook beoordelingsmodellen kunnen uit het raamwerk afgeleid worden.

De discussie zoals deze ook in de literatuur naar voren komt wordt ook door het raamwerk 'moderne vreemde talen' bevestigd. Deze discussie gaat om het probleem dat praktijktoetsen, psychometrisch gezien, minder betrouwbaar zijn en veel meer afhankelijk zijn van het oordeel van een beoordelaar. Er wordt dus ook hier gezocht naar een combinatie van toetsen die deze betrouwbaarheid vergroot. Voor de talen worden steeds meer niveautoetsen ontwikkeld die gekoppeld zijn aan het Europese taalniveau.

In de toekomst vindt men via toetsbanken op internet een steeds betere indicatie van het Europese taalniveau waarop men functioneert binnen de verschillende talen. Een voorbeeld van een toetsbank is het DIALANG programma voor toetsing wat beschikbaar is via het internet. DIALANG is het eerste toetsprogramma dat gebaseerd is op het Europese raamwerk. Het bevat samengestelde en gevalideerde toetsen voor diverse taalvaardigheden en geeft uitgebreide feedback en advies om de taalvaardigheid van studenten te vergroten. Het biedt geen certificaten en is alleen bedoeld om de taalvaardigheid vast te stellen (DIALANG, n.d.). Naast een toetsprogramma is er ook een Europees taalportfolio online beschikbaar. Het is een instrument

waarmee alle ervaringen met het leren van talen gedocumenteerd kunnen worden. Per vaardigheid kan vastgesteld worden op welk niveau van het Europees referentiekader de student zich bevindt. Het taalportfolio bestaat uit drie delen, een talenpaspoort (hierin worden alle gegevens over de student verzameld, diploma's die gehaald zijn, gevolgd onderwijs, welke talen er zijn geleerd en welk niveau dit is, etc.), een taalbiografie (hierin staan taalervaringen binnen en buiten school) en een dossier (hierin staan alle bestanden opgeslagen, een archief) (Nationaal bureau voor moderne vreemde talen, 2006).

2.6 Omschrijving methode Taalblokken

Taalblokken is een innovatief systeem voor individuele taalondersteuning. Het is een flexibele methode die studenten de kans geeft om op ieder moment van de dag en op iedere plaats te werken aan zijn of haar taalcompetenties. Taalblokken is een methode die afgestemd is op de taalcompetenties en de CEF-niveaus die een student nodig heeft voor de gekozen beroepsrichting. De methode is er voor Nederlands, Engels en Duits, binnen de sectoren economie, economie uitgebreid met taaltaken voor horeca toerisme en recreatie, techniek en zorg en welzijn, voor de niveaus 2, 3 en 4. De database van taalblokken bestaat uit de volgende onderdelen:

- *Taalblokken*

Een blok bestaat uit interactieve (door de computer nagekeken) en statische (door de docent te beoordelen) oefeningen waarmee een leerling een vaardigheid kan oefenen die hem van het ene naar het andere niveau brengt. Elk blok van A1 t/m B1 kent drie componenten: zes lessen met elk acht oefeningen, drie taaltaken en een geïntegreerde opdracht. Omdat de onderdompeling in een taal op een hoger niveau belangrijk is, kent het B2-blok een andere opbouw: twaalf lessen met elk vier oefeningen en bij elke les een taaltaak. In een B2-blok komt geen geïntegreerde opdracht voor omdat de taaltaken al een geïntegreerd karakter hebben.

- *Taaltaken*

Taaltaken zijn opdrachten die gericht zijn op de productie van taal, waarbij een vaardigheid centraal staat. Taaltaken zijn altijd statische opdrachten. Taaltaken hebben als functie dat een student, na één of twee lessen geoefend te hebben met een vaardigheid, kan laten zien dat hij de vaardigheid kan toepassen. Om die reden worden de taaltaken aangeboden op het CEF- niveau van het betreffende blok. Taaltaken zijn ook thematisch te gebruiken. Een voorbeeld van een Taaltaak kan bijvoorbeeld zijn om een sollicitatiebrief te schrijven voor een bepaalde functie.

- *Toetsen*

Dit onderzoek richt zich op de methode Taalblokken en heeft hierbij specifiek gekeken naar de toetsen van Taalblokken. Hieronder is een overzicht van de verschillende toetsen die aanwezig zijn binnen Taalblokken weergegeven.

Bepaal je niveau toetsen

Als een student binnenkomt is het mogelijk om met behulp van de 'bepaal je niveau toets' per vaardigheid zijn niveau vast te stellen. Met deze toets kunnen de volgende CEF- niveaus vastgesteld worden

- voor niveau 2 → CEF- niveau A1, A2 of B1
- voor niveau 3-4 → CEF- niveau A2, B1 of B2

Bij elke vaardigheid zijn twee van deze toetsen beschikbaar.

Test jezelf- toetsen

Na de niveaubepaling kan een student aan de slag met de taalblokken, op het niveau dat bij hem past. Als een student een taalblok voor een bepaalde vaardigheid heeft doorgewerkt, kan hij een toets maken om te bekijken of hij het niveau van het blok bereikt heeft. Bij elke vaardigheid zijn vier van deze toetsen beschikbaar. Deze toetsen zijn formatief van aard.

Toetsen met beweringen en toetsen met bron

Er zijn bij zowel de 'bepaal je niveau toetsen' als bij de 'test jezelf toetsen' twee mogelijkheden, toetsen met beweringen en toetsen met bron.

- Toetsen met beweringen/ 'can do'- statements

Bij de toetsen die bestaan uit 'can do'- statements reflecteert de leerling op zijn eigen vaardigheden door op een aantal uitspraken te reageren met 'ja, dat kan ik' of 'nee, dat kan ik niet'.

De uitslag hiervan hangt sterk af van hoe iemand naar zichzelf kijkt. Het is daarom ook de bedoeling dat de docent de toets samen met de student bekijkt. De docent kan hierbij controleren of de student werkelijk deze competentie beheerst en anders wat er nog moet gebeuren om het wel te behalen. Bij 'bepaal je niveau toetsen' krijgen studenten 15 beweringen, bij 'test jezelf toetsen' krijgen studenten 10 beweringen. Daarnaast zijn het interactieve toetsen, de computer bepaalt de score. De toetsen zijn er voor alle CEF-niveaus en voor alle vaardigheden.

- Toetsen met bron

Bij de toetsen met bron krijgt de leerling vragen of opdrachten bij een bron (een audiofragment, een filmpje of een tekst), of, voor de productieve vaardigheden, in een context.

Toetsen voor de productieve vaardigheden

Voor de productieve vaardigheden, schrijven en gesprekken voeren, zijn er twee oefeningen per toets. Deze toetsen zijn niet interactief, de docent beoordeelt de prestatie. Voor deze beoordeling zijn beoordelingsmodellen toegevoegd. Hierin staat vanuit het raamwerk omschreven wat de leerling voor de desbetreffende (sub) vaardigheid en het CEF- niveau moet kunnen. Er wordt voor de vaardigheid schrijven een voorbeelduitwerking gegeven en voor de vaardigheid gesprekken voeren is er een houvast aan de hand van een gesprekskaart.

Toetsen voor de receptieve vaardigheid

Voor de receptieve vaardigheden, lezen en luisteren, zijn de toetsen interactief, dat wil zeggen: het systeem stelt de score vast. Er zijn twee of drie oefeningen per toets. De docent bekijkt het resultaat samen met de student.

Alle toetsen binnen Taalblokken zijn formatieve toetsen, dit betekent dat ze procesgericht zijn. Voor iedere toets binnen Taalblokken geldt dat de student het taalniveau beheerst bij een score van 80%. Bij interactieve toetsen krijgt de student bij een score van 77% dus ook te zien dat hij het niveau nog niet behaald heeft. De docent bepaalt vervolgens met de student samen wat deze nog moet doen om het niveau te behalen.

Docenten kunnen de toetsen in het leerarrangementsysteem beheren. Hierbij wordt een onderscheid gemaakt tussen assessments en toetsen. Hierbij zijn assessments verzamelingen van één of meer toetsen en kunnen dus toetsen op verschillende vaardigheden en zelfs op verschillende CEF- niveaus bevatten. Er wordt bij de toetsen van Taalblokken ook een onderscheid gemaakt tussen afgeschermd en openbare toetsen. Een afgeschermd toets is een toets die alleen ter beschikking staat van de docent. De docent kan zo'n toets in een assessment toewijzen aan een student of aan een groep studenten. Hij kan deze toets daarna ook weer afschermen, zodat deze niet meer beschikbaar is voor een student. Zo kan de toets beperkte tijd beschikbaar worden gemaakt.

Een openbare toets is een toets die altijd ter beschikking staat van een student. Het gaat daarbij om test jezelf -toetsen. De student kan daarmee aan het eind van een blok zelf testen of hij een niveaustap (bijvoorbeeld van A1 naar A2) heeft gemaakt

2.7 Doelstelling en vraagstelling

De doelstelling van dit onderzoek is om na te gaan of de toetsen Engels binnen de methode Taalblokken van Malmberg aan de eisen en voorwaarden van competentiegericht toetsen voldoen en of ze aansluiten bij de taal competentieniveaus (CEF- niveaus) van het CEFR.

De hoofdvraag in dit onderzoek is:

Voldoen de toetsen Engels binnen de methode Taalblokken van Malmberg aan de eisen en voorwaarden van competentie gericht toetsen en sluiten ze aan bij de taal competentieniveaus (CEF- niveaus) van het common Europeaan framework of reference?

Deelvragen binnen dit onderzoek zijn:

- *Wat is de mening van docenten over de kwaliteit van de eisen en voorwaarden van competentiegericht toetsen binnen de methode Taalblokken?*
- *Wat is de mening van studenten over de kwaliteit van de eisen en voorwaarden van competentiegericht toetsen binnen de methode Taalblokken?*
- *Op welke manier komen de kwaliteitscriteria van competentiegericht toetsen terug in de toetsen van de methode Taalblokken?*
- *Welke keuzes heeft Malmberg gemaakt bij het ontwerpen van de methode Taalblokken gericht op de kwaliteitseisen van het competentiegericht toetsen en het common Europeaan framework of reference?*

3. Methode

In dit hoofdstuk zal verder worden ingegaan op het onderzoeksontwerp, de onderzoeksgroep die deelnam aan het onderzoek en de instrumenten.

3.1 Onderzoeksontwerp

De aard van deze thesis is zowel kwalitatief als kwantitatief. De onderzoeksvragen zijn met zowel statistische als beschrijvende data beantwoord. Een combinatie van beide noemt men in de statistiek een mixed design (Robson, 2002).

De hoofdvraag binnen dit onderzoek: *'Voldoen de toetsen Engels binnen Taalblokken van Malmberg aan de eisen en voorwaarden van competentiegericht toetsen en sluiten ze aan bij de taalcompetentieniveaus van het common European framework of reference?'* is beantwoord door middel van interviews, vragenlijsten en een zelf ontworpen beoordelingsinstrument.

De eerste deelvraag: *'Wat is de mening van docenten over de kwaliteit van de eisen en voorwaarden van competentiegericht toetsen binnen de methode Taalblokken?'* is beantwoord met een interview dat is afgenomen bij docenten. De tweede deelvraag: *'Wat is de mening van studenten over de kwaliteit van de eisen en voorwaarden van competentiegericht toetsen binnen de methode Taalblokken?'* is beantwoord met een vragenlijst die is afgenomen bij studenten. De derde deelvraag: *'Op welke manier komen de kwaliteitscriteria van competentiegericht toetsen terug in de toetsen van de methode Taalblokken?'* is beantwoord met behulp van een ontwikkeld beoordelingsinstrument waarmee de toetsen van Taalblokken geëvalueerd zijn. De vierde deelvraag: *'Welke keuzes heeft Malmberg gemaakt bij het ontwerpen van de methode Taalblokken gericht op de kwaliteitseisen van het competentiegericht toetsen en het common European framework of reference?'* is beantwoord door het afnemen van een interview bij een medewerker van Malmberg. Zij is specialist op het gebied van Taalblokken.

3.2 Onderzoeksgroep

De onderzoeksgroep is samengesteld naar aanleiding van een lijst met scholen verkregen van Malmberg. Belangrijk hierbij was dat scholen aangeschreven werden die al met Taalblokken werkten en niet eerder aangeschreven waren voor een dergelijk onderzoek. Deze lijst bestond uit 23 scholen met 44 contactpersonen. Deze scholen zijn benaderd via de 44 contactpersonen. De doelstelling van dit onderzoek was om vijf docenten te interviewen en honderd studenten (twintig per school) een vragenlijst af te nemen. Op de mail waarmee docenten benaderd zijn hebben twee docenten van twee verschillende ROC's (A en B) gereageerd. Het aantal studenten is geselecteerd naar aanleiding van de bereidheid van de school om medewerking te verlenen. Docenten hebben de selectie voor dit onderzoek in handen gehad. Er werd een beroep gedaan op de aanwezige klassen en de bereidheid van deze klassen om mee te doen aan het onderzoek. Op ROC A waren veel studenten bereid de vragenlijst in te vullen (50 studenten), in tegenstelling tot ROC B (14 studenten).

Docenten

Twee docenten hebben medewerking verleend aan dit onderzoek. De docente van ROC A geeft lessen Engels binnen de opleiding administratie (alle richtingen behalve secretariael) op niveau twee en niveau vier. Daarnaast heeft er een docent gereageerd van ROC B. Hij geeft lessen Duits binnen de sector techniek op niveau twee, drie en vier. Ondanks dat zijn vakgebied op het taalgebied van Duits lag was ook hij goed in staat om de vragen in het interview naar behoren te beantwoorden. De methode is voor zowel Engels als Duits hetzelfde opgezet. Zijn mening hierover kon gelijk getrokken worden met Engels. Daarnaast was hij één van de weinige docenten die medewerking wilde verlenen.

Studenten

De studenten die bevroegd zijn waren gelinkt aan de docent op het desbetreffende ROC. Het aantal studenten verschilde per school. Vijftig studenten van het ROC A hebben een vragenlijst ingevuld. De leeftijd van deze studenten varieert van zestien tot 29. De gemiddelde leeftijd is 18,6 met een standaardafwijking van 2,05. De onderzoeksgroep bestaat uit 23 mannen en 27 vrouwen. Binnen deze vijftig studenten zijn er veertien uit het eerste en 36 uit het tweede leerjaar. Dertien studenten zitten op niveau twee en 37 op niveau vier. Negen studenten doen de opleiding juridisch medewerker op niveau vier, 41 studenten doen de opleiding administratie waarvan dertien op niveau twee en 28 op niveau vier.

Op het ROC B hebben veertien studenten een vragenlijst ingevuld. De leeftijd van deze studenten varieert van zeventien tot negentien. De gemiddelde leeftijd is 17,5 met een standaardafwijking van 0,76. De onderzoeksgroep bestaat uit dertien mannen en één vrouw. Binnen deze groep studenten zijn zeven uit het eerste en zeven uit het tweede leerjaar. Alle studenten zitten op niveau vier. Negen studenten doen de opleiding mobiliteit, drie studenten doen de opleiding technisch specialist/ werkplaatsmanager en twee studenten doen de opleiding commercieel bedrijfsleider. De respons op de vragenlijst was 100% voor beide groepen.

Medewerker Malmberg

Bij Malmberg is er een interview afgenomen bij de uitgever van Taalblokken. Zij is één jaar werkzaam als uitgever van Taalblokken. De hoofdtaak van een uitgever is het ontwikkelen van commercieel succesvolle educatieve producten binnen de kaders van een uitgeverij. Die ontwikkeling loopt van het eerste idee tot de uiteindelijke marktintroductie. De uitgever heeft een tweeledige rol. Enerzijds geef hij vanuit een marketingperspectief inhoudelijk sturing aan de totstandkoming van nieuwe producten. Anderzijds is een uitgever een procesbewaker en stuurt hij als zodanig op tijd, geld en kwaliteit.

Toetsen

Er zijn binnen dit onderzoek veertien toetsen geanalyseerd. Dit zijn alle aanwezige toetsen binnen het blok A2 > B1. Dit houdt in dat alle toetsen van alle vaardigheden geanalyseerd zijn. Vier toetsen voor gesprekken voeren, vier toetsen voor lezen, twee toetsen voor luisteren en vier toetsen voor schrijven. Hierbinnen zijn zes statische toetsen aanwezig en acht interactieve toetsen. Zie § 2.6 voor een uitleg over deze toetsen.

3.3 Instrumenten

Er zijn drie manieren voor dataverzameling gebruikt binnen deze thesis om antwoord te geven op de onderzoeksvraag. Bij docenten en bij een medewerker van Malmberg zijn interviews afgenomen, bij studenten zijn vragenlijsten afgenomen en voor de toetsen is gebruik gemaakt van een beoordelingsinstrument. Een criteria komt volledig naar voren als het bij alle vier de instrumenten positief scoort. Zie Tabel 3 voor een overzicht van de bevroegde criteria per instrument.

3.3.1 Interview bij docenten

Door middel van een interview bij docenten zijn de kwaliteitscriteria accepteerbaarheid, cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten, vergelijkbaarheid, tijd en kosten, onderwijskundige gevolgen, eerlijkheid, geschiktheid voor onderwijsdoelen, betekenisvolheid, herhaalbaarheid en transparantie onderzocht (zie Tabel 3). Er is gekozen voor interviews omdat het doel was om attitudes, opinies, gevoelens, gedachten en kennis te verzamelen (Baarda & De Goede, 2006; 't Hart, van Dijk, de Goede, Jansen & Teunissen, 1996). De vragen die gesteld werden aan de participanten zijn te kaderen onder de twaalf kwaliteitscriteria van Baartman en collega's (2007). Deze twaalf punten zijn erg abstract en daarom door de onderzoeker geoperationaliseerd in concrete punten en vragen. Het zijn open vragen die vrije antwoordmogelijkheden mogelijk maken. Het doel was om veel informatie te verzamelen. Door de open vragen was de interviewer in staat om door te vragen zodat er genuanceerde antwoorden naar boven komen. Er zijn negentien interviewvragen opgesteld (zie Bijlage 2). Een voorbeeld van een interviewvraag voor het criterium *accepteerbaarheid* is: "Vindt u de methode voldoende beroepsgericht? Voor zover u dat kan beoordelen?". Het interview heeft 'face to face' plaatsgevonden en kostte ongeveer 45 minuten tijd van de bevroegde docenten. Bij docent B liep het interview uit wat zorgde voor een aantal onbeantwoorde vragen, dit is terug te vinden in de resultatensectie, Tabel 4.

3.3.2 Interview Malmberg

Door middel van een interview met één medewerker van Malmberg zijn de kwaliteitscriteria tijd en kosten, geschiktheid voor onderwijsdoelen, ontwikkeling van zelfsturend leren en de effecten op het leren van studenten en herhaalbaarheid onderzocht (zie Tabel 3).

Bij de uitgever van Taalblokken van Malmberg is ook een interview afgenomen, net als bij de docenten. Naar aanleiding van de operationalisering van de kwaliteitscriteria van Baartman en collega's (2007) was het noodzakelijk om meer achtergrondinformatie te verkrijgen over de ontwikkeling van de methode Taalblokken. Een aantal kwaliteitscriteria zoals *Tijd en kosten* benadrukten het belang van deze informatie. Dit is de reden geweest om te rade te gaan bij de uitgever van Malmberg. Er zijn zestien vragen opgesteld die ongeveer 45 minuten van de geïnterviewde vroegen. Een voorbeeld van een vraag voor het criterium *tijd en kosten* is: "Evalueren jullie naar aanleiding van de ontwikkeling van de methode hoe deze tot stand is gekomen en wat er de volgende keer eventueel beter zou kunnen?".

3.3.3 Analyse interviews

De interviews zijn verwerkt door middel van de grounded theory (Glaser & Strauss, 1969). Deze theorie is vooral bedoeld voor de analyse van kwalitatieve data zoals een interview. Hierbij wordt de hoeveelheid informatie verminderd door het te schiften, samen te vatten en het in categorieën in te delen (Boeije, 2005). De data analyse vormt zich rond drie stappen, 'open coding', 'axial coding' en 'selective coding' (Glaser & Strauss, 1969).

Open coderen, alle gegevens die verzameld zijn, zijn gelezen en in fragmenten opgedeeld. De relevante informatie is vervolgens gelabeld en onderling vergeleken (Boeije, 2005). De gegevens hebben een label gekregen dat zo dicht mogelijk bij de tekst lag (Baarda & de Goede, 2006). *Axiaal coderen*, de codes van het open coderen zijn geordend en opnieuw gelabeld met een label dat refereert aan één van de twaalf kwaliteitscriteria van Baartman et al. (2007). Doordat de interviewvragen al waren opgesteld aan de hand van de kwaliteitscriteria waren de gegevens uit het interview snel te plaatsen binnen de kwaliteitscriteria. *Selectief coderen*, de bij het axiaal coderen verkregen labels zijn bij deze fase in een schema geordend.

3.3.4 Vragenlijsten studenten

Door middel van de vragenlijsten zijn de kwaliteitscriteria accepteerbaarheid, cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten onderzocht, tijd en kosten, onderwijskundige gevolgen, eerlijkheid en transparantie (zie Tabel 3).

De vragenlijst bestaat uit 23 beweringen. Het invullen van de vragenlijst kostte ongeveer twintig minuten tijd van de bevroegde studenten (zie Bijlage 3). Een voorbeeld van een bewering voor het criterium *accepteerbaarheid* is: "Ik vind het fijn dat we Taalblokken gebruiken in de klas". De criteria cognitieve complexiteit en de ontwikkeling van zelfsturend leren en de effecten op het leren van studenten zijn in de vragenlijst bevroegd met dezelfde items. Dit met de reden dat deze criteria grote overlap hebben en om deze reden samen zijn genomen. Een voorbeeld van een bewering is: "Ik let op de opmerkingen (feedback) die mijn docent mij geeft over het werk dat ik inlever of maak". Het criterium *tijd en kosten* is met de volgende bewering bevroegd: "Bekijk voorbeeld 2, dit is een toets voor Engels, op niveau A2 voor gesprekken voeren. Ik vind het redelijk dat ik 5 minuten de tijd krijg om deze opdracht uit te voeren (dit is zonder voorbereiding)". Een voorbeeld van een bewering voor het criterium *onderwijskundige gevolgen* is: "Door Taalblokken ben ik anders gaan leren voor Engels dan hiervoor toen we nog werkten met een boek en een werkboek. (Denk hierbij aan het leren van een toets, is je voorbereiding veranderd?)". Een voorbeeld van een bewering voor het criterium *eerlijkheid* is: "Ik krijg voldoende kansen binnen Taalblokken om te laten zien wat ik kan op het gebied van Engels". Een voorbeeld van een bewering voor het criterium *transparantie* is: "Ik vind dat de toetsen voor Engels goed te begrijpen zijn". Het ging hierbij vooral om de criteria waarbij de mening van studenten ook van belang is. In totaal zijn er in de vragenlijst zeven criteria gemeten met zes schalen.

Binnen de vragenlijst is er gekozen voor een zes puntsschaal van helemaal mee eens tot helemaal mee oneens. De reden voor de keuze voor een zes puntsschaal is dat de middencategorie ontbreekt. De studenten zijn hiermee tot een mening gedwongen. Hiermee is voorkomen dat de meerderheid van de studenten een neutraal antwoord zouden geven (Baarda & de Goede, 2001).

Over de vragenlijst is een betrouwbaarheidsanalyse uitgevoerd (n=64) (zie Tabel 2). Binnen de vragenlijst was sprake van zes schalen met ieder hun eigen aantal items. De Cronbach's alpha binnen de verschillende schalen ligt tussen de .24 en de .72. In deze analyse wordt een Cronbach's alpha van hoger dan .60 gezien als een voldoende schaal. De criteria cognitieve complexiteit/ontwikkeling van zelfsturend leren en de effecten op het leren van studenten en eerlijkheid hebben een lage Cronbach's alpha. Bij het criterium eerlijkheid is dit mogelijk te verklaren doordat dit criterium gemeten is met maar twee items. Over het criterium tijd en kosten is geen analyse uitgevoerd omdat dit criterium maar met één item gemeten is. De criteria accepteerbaarheid, onderwijskundige gevolgen en transparantie hebben een Cronbach's alpha hoger dan .60. Bij deze schalen zijn gemiddelde schaalscores berekend om zo een uitspraak te kunnen doen over bijvoorbeeld de mate van accepteerbaarheid, onderwijskundige gevolgen of de transparantie van de toetsen van Taalblokken in de perceptie van de studenten.

Tabel 2 Betrouwbaarheidsanalyse vragenlijst studenten

Criterium	Cronbach's alpha
Accepteerbaarheid (2 items)	.66
Cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten (5 items)	.57
Tijd en kosten (1 item)	x
Onderwijskundige gevolgen (6 items)	.71
Eerlijkheid (2 items)	.24
Transparantie (7 items)	.72

Om de validiteit van de vragenlijst te bepalen is er een factoranalyse uitgevoerd over alle items in de vragenlijst. De vragenlijst is opgezet met het doel om zes factoren te meten. Uit de factoranalyse zijn echter acht factoren naar voren gekomen. Items die hoger laden dan .40 zijn opgenomen binnen de factoren. Opvallend hierbij is dat de meeste items hoog laden op één factor (zie Bijlage 7). Dit betekent dat de vragenlijst hoofdzakelijk één factor meet. Voor de rest van het onderzoek betekent dit dat de resultaten voortkomend uit de vragenlijst mogelijk niet valide zijn en de resultaten bij meerdere criteria een verklaring kunnen geven.

3.3.5 Beoordelingsinstrument toetsen

Door middel van het beoordelingsinstrument (zie Bijlage 4) voor de analyse van de toetsen zijn de kwaliteitscriteria authenticiteit, cognitieve complexiteit, vergelijkbaarheid, geschiktheid voor onderwijsdoelen, ontwikkeling van zelfsturend leren en de effecten op het leren van studenten, betekenisvolheid en herhaalbaarheid onderzocht (zie Tabel 3).

Naar aanleiding van de operationalisering van de kwaliteitscriteria van Baartman en collega's (2007) is een aantal criteria op een rij gezet die getoetst konden worden met dit beoordelingsinstrument alleen kijkend naar de inhoud van de toetsen. Het beoordelingsinstrument heeft 25 items, de antwoordschalen die gebruikt zijn voor het beantwoorden van de verschillende items verschillen per vraag en variëren van een gesloten vraag (ja/nee) tot een meerkeuzevraag.

Bij iedere vraag is een opmerkingenveld toegevoegd om de beoordeling zo breed mogelijk te houden. Aspecten die opvielen bij de toetsen maar die buiten de vraag vielen kunnen zo toch meegenomen worden. Door middel van het opmerkingenveld werd ook voorkomen dat kostbare informatie verloren ging. Een voorbeeld van een gesloten vraag was: *Wordt de mening van studenten betrokken bij de beoordeling van het assessment? Staat dit beschreven in het beoordelingsmodel? Hebben studenten iets te zeggen bij de beoordeling?*. Een voorbeeld van een meerkeuzevraag is: *"Welke sociale processen spelen een rol tijdens de uitvoer?" - Samenwerken - Overleggen - Vergaderen - Beslissingen maken - Voordracht - Beleefdheid - Overig namelijk*". De keuze voor gesloten vragen lag in de praktische toepasbaarheid van het instrument. Door het invullen van gesloten vragen kost het de beoordelaar minder tijd en moeite en zal de drempel tot het gebruik van dit instrument lager blijven. Daarnaast zullen gesloten vragen bevorderend werken voor de objectiviteit van het instrument.

Het criterium authenticiteit neemt een groot deel van het beoordelingsinstrument in beslag. Dit met de reden dat authenticiteit alleen door de toetsen zelf te bestuderen gemeten kan worden. Dit is gedaan door de trap van Klarus (2001) en de kenmerken van Gulikers en collega's (2004) in het instrument te verwerken en ook aan de hand hiervan conclusies te trekken.

Het instrument kan ingezet worden bij het beoordelen van een CAP. Het model dat in deze thesis ontwikkeld is bevat niet alle criteria van Baartman en collega's (2007). Dit komt omdat niet alle criteria vertaald konden worden naar het instrument. Bij veel criteria was het noodzakelijk ook de mening van docenten en studenten te betrekken. Alleen kijkend naar een toets kunnen enkel de criteria authenticiteit, cognitieve complexiteit, vergelijkbaarheid, geschiktheid voor onderwijsdoelen, ontwikkeling van zelfsturend leren en de effecten op het leren van studenten, betekenisvolheid en herhaalbaarheid met dit instrument beoordeeld worden. Het is dus geen volledig dekkend beoordelingsinstrument. Het is een gemakkelijk hanteerbaar document dat digitaal of schriftelijk ingevuld kan worden. De vragen zijn concreet opgezet zodat ieder persoon ze gemakkelijk kan beantwoorden.

De objectiviteit van het beoordelingsinstrument is gemeten door drie andere personen het beoordelingsinstrument in de laten vullen voor drie toetsen, één beweringentoets, één toets voor lezen (interactief) en één toets voor gesprekken voeren (statisch). Over de hiermee verkregen data is een interbeoordelaars betrouwbaarheidsanalyse uitgevoerd. In Bijlage 5 zijn de overeenstemmingspercentages per item en over alle items samen weergegeven. De betrouwbaarheid wordt als voldoende beschouwd als het gemiddelde overeenstemmingspercentage over alle items boven de 70% ligt. De objectiviteit van het beoordelingsinstrument ligt op 88,58% met een minimum van 25% en een maximum van 100% (zie Bijlage 5). Het laagste percentage van de objectiviteit ligt bij item dertien. Dit item vraagt naar de verschillende soorten sociale processen die binnen een toets naar boven komen. Dit item valt binnen het criterium authenticiteit. De personen die het beoordelingsinstrument in hebben gevuld hebben hier allemaal een ander idee over gehad. Zo vond de ene persoon dat overleggen en discussiëren erg duidelijk naar voren kwam terwijl de andere persoon vond dat beslissingen maken meer naar voren kwam. Deze keuze heeft te maken met de persoonlijke benadering van iedere beoordelaar. Bij de items die onder de 70% scoren valt het op dat ze een hoge interpretatiewaarde hebben. Zo beoordelen bijvoorbeeld alle beoordelaars bij item elf een betekenisvolle situatie op basis van de taal verschillend. Het

overeenstemmingpercentage is hier dan ook maar 50%. Over het geheel genomen is het instrument echter voldoende objectief.

Tabel 3 Overzicht onderzochte kwaliteitscriteria

Interview docenten	Interview Malmberg	Vragenlijst	Beoordelingsinstrument
accepteerbaarheid	tijd en kosten	accepteerbaarheid	authenticiteit
cognitieve	geschiktheid voor	cognitieve	cognitieve complexiteit
complexiteit/ ontwikkeling van	onderwijsdoelen	complexiteit/ ontwikkelen van	vergelijkbaarheid
zelfsturend leren en	ontwikkeling van	zelfsturend leren en	geschiktheid voor
de effecten op het	zelfsturend leren en	de effecten op het	onderwijsdoelen
leren van studenten	de effecten op het	leren van studenten	ontwikkeling van
vergelijkbaarheid	leren van studenten	leren van studenten	zelfsturend leren en
tijd en kosten	herhaalbaarheid	tijd en kosten	de effecten op het
onderwijskundige		onderwijskundige	leren van studenten
gevolgen		gevolgen	betekenisvolheid
eerlijkheid		eerlijkheid	herhaalbaarheid
geschiktheid voor		transparantie	
onderwijsdoelen			
betekenisvolheid			
herhaalbaarheid			
transparantie			

4. Resultaten

De resultaten zullen beschreven worden aan de hand van de vier deelvragen van dit onderzoek. Binnen deze vier deelvragen zal per kwaliteitscriterium besproken worden wat de bevindingen zijn geweest.

4.1 Wat is de mening van docenten over de kwaliteit van de eisen en voorwaarden van competentiegericht toetsen binnen de methode Taalblokken?

Een samenvatting van de resultaten die voort zijn gekomen uit het interview bij docenten is verwerkt in Tabel 4.

4.1.1 Accepteerbaarheid

Beide docenten accepteren de doelen die gesteld worden in het raamwerk en in Taalblokken. Doordat de methode volledig aansluit bij het raamwerk kunnen de docenten goed zien aan welke kwalificaties de studenten moeten voldoen. Over de beroepsgerichtheid van de methode kunnen de docenten nog niet veel zeggen, omdat Taalblokken nog geen studenten af heeft geleverd. Wel zijn de docenten van mening dat de toetsen goed aansluiten. Studenten hebben een beter beeld van de beroepspraktijk in de ogen van de docenten.

4.1.2 Cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten

Er wordt een enkele keer feedback gegeven door docent A met de voorkeur voor mondelinge feedback. Docent B geeft regelmatig feedback met ook de voorkeur voor mondelinge feedback. Docent B koppelt de feedback aan de toetsen en het leerproces dat de studenten doormaken, docent A koppelt feedback alleen aan negatieve toetsresultaten. De docenten zijn beide niet tevreden over de feedback die de methode Taalblokken aan de studenten geeft. Deze is minimaal. De 'test jezelf toetsen' worden door beide docenten niet afgenomen. Deze zijn te tijdrovend voor docent B en op het ROC A wordt gewerkt met een Taalportfolio waardoor deze toetsen volgens docent A overbodig zijn. Studenten stellen geen leerdoelen op naar aanleiding van de gegeven feedback, dit is volgens docent A nog net één stap te ver. Studenten moeten eerst leren hoe ze moeten leren.

4.1.3 Vergelijkbaarheid

Op het ROC A werken ze nog met een toetsweek. Alle studenten krijgen op dezelfde tijd in dezelfde omgeving een toets. De vergelijkbaarheid is hier hoog. Op het ROC B bekijkt docent B wanneer een student toe is aan een toets. Formatieve toetsen worden dus per leerling op een ander moment afgenomen. Docent B bepaalt zelf wanneer dit is. De vergelijkbaarheid hangt hier af van de docent.

4.1.4 Tijd en kosten

Docent A is van mening dat er te weinig toetsen aanwezig zijn binnen Taalblokken. Hierbij moet wel rekening gehouden worden met de vele systeemfouten waar de school mee te maken heeft gehad zoals het niet opslaan van gegevens over toetsen et cetera. Docent B gebruikt veel van zijn eigen materiaal en doet geen uitspraak over de hoeveelheid toetsen binnen de methode. Beide

docenten vinden de toetsen efficiënt. Docent A heeft bijna geen voorbereidingstijd nodig voor de toetsen. Docent B geeft daarentegen aan dat hij wel iets langer bezig is met de voorbereiding dan in de methode aangegeven staat. Daarnaast hebben de studenten volgens de docenten voldoende tijd om de toets te maken kijkend naar de aangegeven tijd bij de toetsen.

4.1.5 Onderwijskundige gevolgen

De manier van onderwijs geven is bij beide docenten veranderd. Docent A geeft aan van een klassikale situatie volledig over te zijn gestapt naar een leerlinggestuurde situatie. Studenten starten en sluiten de les nog wel als groep om het groepsgevoel levend te houden. Docent B geeft ook aan deze verandering te hebben doorstaan. Hij besteedt één uur per week aan Taalblokken en één uur per week aan zijn eigen materiaal, alles leerlinggestuurd. De leerstijl van de studenten verandert volgens docent A door projecten en simulaties. Docent B geeft aan dat studenten steeds meer zelfstandig en doelgericht werken. Dit moet echter wel groeien, zo geven beide docenten aan.

4.1.6 Eerlijkheid

Docent A vindt zich voldoende competent om de beoordelingen uit te voeren. Studenten krijgen bij docent A ruimte om bezwaar te maken tegen de toetsen. Deze ruimte is echter wel beperkt. Daarnaast geeft docent A aan dat studenten nog niet altijd het nut van CGO inzien, waardoor ze meer bezwaren maken. Het is moeilijk voor de docent om dit aan studenten duidelijk te maken. Docent B is hierop niet bevraagd wegens tijdgebrek tijdens het interview.

4.1.7 Geschiktheid voor onderwijsdoelen

Alle vaardigheden (gesprekken voeren, lezen, schrijven en luisteren) komen volgens de docenten voldoende aan bod. De verhouding kennis, vaardigheden en attitude is voldoende aanwezig binnen de toetsen. Docent A geeft echter wel aan dat ze de toetsen erg kort vindt, waardoor het lastiger wordt voor docenten om een oordeel hierover te kunnen geven. De verhouding hangt volgens docent A ook af van de manier waarop de toets wordt afgenomen. Docenten moeten zich daarvan bewust zijn.

4.1.8 Betekenisvolheid

Er zijn nog geen studenten afgestudeerd met deze methode. De verwachting van beide docenten is dat de aansluiting met de beroepspraktijk beter zal zijn. Studenten hebben een beter beeld van wat het werk in gaat houden.

4.1.9 Herhaalbaarheid

Op het ROC A wordt geen gebruik gemaakt van de inzet van meerdere docenten bij een toets om de interbeoordelaarsbetrouwbaarheid te vergroten. Dit blijft achterwege om praktische redenen. Docent B is hierop niet bevraagd door tijdgebrek tijdens het interview.

4.1.10 Transparantie

De beoordelingscriteria worden door beide docenten als begrijpelijk genoeg ervaren. Ook naar de studenten toe vinden de docenten het duidelijk. Docent B geeft wel aan moeilijke woorden uit de beoordelingscriteria te halen.

Tabel 4 Schematische weergave resultaten docenten

criterium	Docent A	Docent B	Conclusie
1 Accepteerbaarheid	Accepteert de doelen Beroepsgerichtheid geen uitspraak, wel aansluiting goed	Accepteert de doelen Beroepsgerichtheid geen uitspraak, wel aansluiting goed	+
2 Cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten	Enkele keer feedback geven Mondelinge feedback voorkeur Feedback koppelen aan negatieve resultaten Niet tevreden over feedback van methode Werken met Taalportfolio dus test jezelf toetsen overbodig Leerdoelen opstellen één stap te ver	Regelmatig feedback Mondelinge feedback voorkeur Feedback koppelen aan leerproces Niet tevreden over feedback van methode Neemt geen test jezelf toetsen af	+/-
3 Vergelijkbaarheid	Toetsweek	Docent bepaalt toetsmoment	+/-
4 Tijd en kosten	Te weinig toetsen Veel systeemfouten Toetsen efficiënt Korte voorbereidingstijd voor toetsen Studenten krijgen voldoende tijd voor een toets	Veel eigen materiaal x Toetsen efficiënt Redelijke voorbereidingstijd toetsen Studenten krijgen voldoende tijd voor een toets	+
5 Onderwijskundige gevolgen	Van klassikaal naar leerlinggestuurd Projecten en simulaties, verandering leerstijl	Van klassikaal naar leerlinggestuurd Studenten zelfstandiger en doelgericht	+
6 Eerlijkheid	Docent voldoende competent Ruimte voor bezwaar studenten	X	+
7 Geschiktheid voor onderwijsdoelen	Vaardigheden voldoende Verhouding kennis, vaardigheden, attitude voldoende	Vaardigheden voldoende Verhouding kennis, vaardigheden, attitude voldoende x	+
8 Betekenisvolheid	Studenten hebben beter beeld beroepspraktijk	Studenten hebben beter beeld beroepspraktijk	+
9 Herhaalbaarheid	Geen interbeoordelaarsbetrouwbaarhei d	X	-
10 Transparantie	Beoordelingscriteria duidelijk docenten Voor studenten nog moeilijke woorden	Beoordelingscriteria duidelijk docenten Voor studenten nog moeilijke woorden	+

4.2 Wat is de mening van studenten over de kwaliteit van de eisen en voorwaarden van competentiegericht toetsen binnen de methode Taalblokken?

In Bijlage 6 zijn de gemiddelde scores, standaardafwijking, minimum en maximum scores van alle studenten weergegeven op de zes schalen die zijn onderzocht met de vragenlijst. Een samenvatting van de resultaten die voort zijn gekomen uit de vragenlijst zijn verwerkt in Tabel 11.

4.2.1 Accepteerbaarheid

Bij accepteerbaarheid ligt het gemiddelde over alle items op 2,87 met een standaardafwijking van 2,84. Over de accepteerbaarheid zeggen studenten dat ze de methode in zijn geheel nog niet fijn vinden maar ze zijn redelijk tevreden over het leerrendement van de toetsen. Zie tabel 5 voor de gegeven antwoorden.

Tabel 5 gegeven antwoorden accepteerbaarheid

Antwoordcategorie	Item 8	Item 23
Helemaal mee oneens (1)	6	27
Oneens (2)	12	11
Niet helemaal mee oneens (3)	13	8
Niet helemaal mee eens (4)	23	9
Mee eens (5)	8	5
Helemaal mee eens (6)	2	3

4.2.2 Cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten

Bij cognitieve complexiteit en de ontwikkeling van zelfsturend leren en de effecten op het leren van studenten ligt het gemiddelde voor item 17 op 2,56 met een standaardafwijking van 1,47. Item 19 scoort een gemiddelde van 4,11 met een standaardafwijking van 1,38. Item 20 scoort een gemiddelde van 3,03 met een standaardafwijking van 1,18 en item 21 scoort een gemiddelde van 2,95 met een standaardafwijking van 1,65. Studenten vinden het over het algemeen niet belangrijk om terug te kijken op een blok. De meeste studenten letten wel op de opmerkingen die de docent maakt over hun werk, hierdoor gaan veel studenten alleen niet anders te werk. Ze vinden het over het algemeen niet fijn als andere studenten naar hun werk kijken en zijn grotendeels niet tevreden over de feedback die de methode Taalblokken studenten aanreikt. Zie Tabel 6 voor de gegeven antwoorden.

Tabel 6 gegeven antwoorden cognitieve complexiteit

Antwoordcategorie	Item 17	Item 19	Item 20	Item 21	Item 22
Helemaal mee oneens (1)	19	3	7	17	18
Oneens (2)	18	6	14	17	10
Niet helemaal mee oneens (3)	8	9	21	16	10
Niet helemaal mee eens (4)	10	21	14	7	11
Mee eens (5)	6	13	8	6	10
Helemaal mee eens (6)	2	12	0	1	4

4.2.3 Tijd en kosten

Het gemiddelde over het enkele item van tijd en kosten ligt op 3,19 met een standaardafwijking van 1,53. Studenten vinden dat ze voldoende tijd krijgen om een toets uit te voeren. Zie tabel 7 voor de gegeven antwoorden.

Tabel 7 gegeven antwoorden tijd en kosten

Antwoordcategorie	Item 9
Helemaal mee oneens (1)	14
Oneens (2)	6
Niet helemaal mee oneens (3)	15
Niet helemaal mee eens (4)	16
Mee eens (5)	9
Helemaal mee eens (6)	4

4.2.4 Onderwijskundige gevolgen

Bij onderwijskundige gevolgen ligt het gemiddelde over alle items op 2,38 met een standaardafwijking van 1.45. Vanuit het perspectief van de studenten zijn de onderwijskundige gevolgen negatief. Studenten geven aan op een andere manier te zijn gaan leren. Opvallend is echter dat vanuit het perspectief van CGO, de leerstijl niet op een positieve manier verandert. Studenten kijken bijvoorbeeld niet terug op een blok, maken opdrachten niet nog een keer, oefenen niet samen met medestudenten en vragen geen extra materiaal aan hun docent bij slechte toetsresultaten. Hierbij is echter niet gevraagd naar de situatie voordat studenten begonnen met Taalblokken. Zie tabel 8 voor de gegeven antwoorden.

Tabel 8 gegeven antwoorden onderwijskundige gevolgen

Antwoordcategorie	Item 6	Item 7.1	Item 7.2	Item 7.3	Item 7.4	Item 7.5
Helemaal mee oneens (1)	10	30	36	39	34	30
Oneens (2)	7	13	12	10	11	7
Niet helemaal mee oneens (3)	9	11	8	6	8	10
Niet helemaal mee eens (4)	8	4	5	4	8	9
Mee eens (5)	7	5	2	4	1	7
Helemaal mee eens (6)	23	1	1	1	2	1

4.2.5 Eerlijkheid

Bij eerlijkheid ligt het gemiddelde voor item 14 op 3,00 met een standaardafwijking van 1,58. Voor item 18 ligt het gemiddelde op 4,83 met een standaardafwijking van 1,26. Studenten hebben het gevoel niet voldoende kans te krijgen om te laten zien wat ze kunnen. Opvallend zijn de hoge scores op item 18 wat weergeeft dat studenten wel het gevoel hebben dat ze gelijk behandeld worden tijdens het maken van een toets. Zie tabel 9 voor de gegeven antwoorden.

Tabel 9 gegeven antwoorden eerlijkheid

Antwoordcategorie	Item 14	Item 18
Helemaal mee oneens (1)	14	3
Oneens (2)	14	0
Niet helemaal mee oneens (3)	10	6
Niet helemaal mee eens (4)	12	8
Mee eens (5)	9	25
Helemaal mee eens (6)	4	21

4.2.6 Transparantie

Bij transparantie ligt het gemiddelde over alle items op 3,65 met een standaardafwijking van 1,46. Studenten zijn over het algemeen positief over de transparantie van de methode. Opvallend zijn de hoge scores op item 5, 10, 11 en 15. Studenten weten wat ze moeten leren, begrijpen de toetsen, vinden de toetsen goed uit te voeren en weten van wie ze een beoordeling krijgen. Wat hierbij ook opvalt, zijn de lage scores op item 12, 13 en 16. Studenten weten welke docent ze gaat beoordelen maar weten niet op welke punten ze beoordeeld worden en waaraan studenten dus moeten voldoen. Samenvattend geeft dit weer dat de toetsen helder en duidelijk zijn opgezet maar dat de transparantie van de beoordelingsmodellen nog niet optimaal is. Zie tabel 10 voor de gegeven antwoorden.

Tabel 10 gegeven antwoorden transparantie

Antwoordcategorie	Item 5	Item 10	Item 11	Item 12	Item 13	Item 15	Item 16
Helemaal mee oneens (1)	2	1	8	6	11	9	12
Oneens (2)	9	11	0	13	12	6	14
Niet helemaal mee oneens (3)	14	9	12	14	15	7	15
Niet helemaal mee eens (4)	15	15	10	15	12	8	10
Mee eens (5)	17	13	21	9	7	14	9
Helemaal mee eens (6)	7	12	10	4	5	17	2

4.2.7 Verschillen tussen deelnemende ROC's

Door middel van een onafhankelijke t- toets zijn de verschillen bekeken van de gegeven antwoorden op de verschillende items van de vragenlijst tussen het ROC A en het ROC B. Bij enkele items zijn significante verschillen gevonden: item 7.4 ($p = 0,03$), item 15 ($p = 0,03$), bij item 17 ($p = 0,01$) en bij item 21 ($p = 0,05$). Aangezien het maar weinig significante verschillen zijn bij items die niet samenhangen binnen een criterium wordt dit toegeschreven aan toeval.

Tabel 11 Schematische weergave resultaten vragenlijst

criterium	Studenten	Conclusie
1. Accepteerbaarheid	Methode geheel niet fijn Wel tevreden leerrendement toetsen	+/-
2. Cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten	Niet belangrijk terug te kijken op een blok Letten op opmerkingen docent Gaan niet anders te werk	+/-
3. Tijd en kosten	Voldoende tijd om een toets uit te voeren	+
4. Onderwijskundige gevolgen	Leren op een andere manier Leerstijl verandert niet op een postieve manier vanuit cgo Kijken niet terug op een blok Maken opdrachten niet nog een keer Oefenen niet met medestudenten Vragen geen extra materiaal aan docent	-
5. Eerlijkheid	Krijgen niet voldoende kans om te laten zien wat ze kunnen Worden gelijk behandeld	+
6. Transparantie	Weten wat ze moeten leren Begrijpen de toets Vinden de toets goed uit te voeren Weten van wie ze een beoordeling krijgen	+

4.3 Op welke manier komen de kwaliteitscriteria van competentiegericht toetsen terug in de toetsen van de methode Taalblokken?

Een samenvatting van de resultaten die voort zijn gekomen uit de het beoordelingsinstrument voor toetsen zijn verwerkt in Tabel 12.

4.3.1 Authenticiteit

De toetsen binnen de methode Taalblokken bestaan uit vier kennistoetsen, zes simulaties van de praktijk en vier toetsen die niet te plaatsen zijn op de trap van Klarus (2001). Bij zes toetsen wordt het assessment vooraf gegaan door een minimale authentieke instructie. Tien van de veertien toetsen, toetsen een betekenisvolle prestatie op het gebied van taal en de toekomstige beroepspraktijk. De mening van studenten wordt alleen bij de beweringentoetsen betrokken. Kennis, vaardigheden en attitude zijn bij drie van de veertien toetsen geïntegreerd aanwezig. De sociale processen die plaatsvinden binnen de toetsen zijn: overleggen, beslissingen maken, voordracht en beleefdheid. Bij zeven toetsen komen geen sociale processen voor. Bij tien toetsen wordt er een competentie getoetst die later in de beroepspraktijk ook van de student gevraagd kan worden.

4.3.2 Cognitieve complexiteit

De toetsen spelen geen van allen in op de verschillende denkprocessen die een student doorloopt. Wel worden de verschillende denkstappen bij zes van de veertien toetsen in de beoordelingsmodellen voor de docent aangegeven zodat deze hier bij de beoordeling rekening mee kan houden. Verder worden geen aanwijzingen gegeven hoe een docent een student hierop kan aansturen.

4.3.3 Vergelijkbaarheid

Acht van de veertien toetsen zijn gesloten (interactieve) toetsen, dit betekent dat hierbij de beoordeling voor iedere student hetzelfde is. Omdat de toetsen op de computer worden gemaakt is het gebruiksgemak voor de docent hoog, er zijn geen beoordelingscriteria of beoordelingsmodellen voor de docent. Bij zes van de veertien toetsen zijn beoordelingsmodellen aanwezig waarin de taak en de criteria worden omschreven. Bij geen enkel beoordelingsmodel wordt gesproken over werkcondities.

4.3.4 Geschiktheid voor onderwijsdoelen

De doelen die genoemd staan bij de beoordelingsmodellen voor de statische toetsen zijn letterlijk overgenomen uit het raamwerk. Bij de interactieve toetsen ontbreekt de doelomschrijving, hier kan dus geen uitspraak over gedaan worden. Bij drie toetsen zijn kennis, vaardigheden en attitude geïntegreerd. Bij tien toetsen is er sprake van kennistoetsing, bij alle veertien toetsen worden vaardigheden getoetst, bij drie toetsen is er ook sprake van attitude toetsing.

4.3.5 Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten

Bij de beweringentoetsen wordt van de docent verwacht dat deze feedback geeft aan de student, dit staat omschreven in de docentenhandleiding. Verder staat aangegeven dat docenten de advies 'tool' kunnen gebruiken om studenten feedback te geven. Bij de toetsen zelf wordt hier geen

uitspraak over gedaan. Vier van de veertien toetsen stimuleren het geven van feedback. Dit zijn de beweringentoetsen waarbij aan de student gevraagd wordt om na te denken over, of te reflecteren op de eigen competenties en daar in een gesprek met de docent over te praten.

4.3.6 Betekenisvolheid

Bij alle veertien toetsen wordt rekening gehouden met het referentiekader van de student. Bij een aantal toetsen is dit minimaal maar wel aanwezig.

4.3.7 Herhaalbaarheid

Bij geen van beide ROC's wordt er gebruik gemaakt van een tweede beoordelaar.

Tabel 12 Schematische weergave resultaten beoordelingsinstrument voor toetsen

criterium	Beoordelingsinstrument	Conclusie
1. Authenticiteit	Vier kennistoetsen Zes simulaties van de praktijk Vier niet te plaatsen toetsen Tien toetsen toetsen een betekenisvolle prestatie op het gebied van taal en de toekomstige beroepspraktijk Mening van studenten alleen bij beweringentoetsen Kennis, vaardigheden, attitude bij drie toetsen geïntegreerd Sociale processen (bij zeven toetsen): overleggen, beslissingen maken, voordracht en beleefdheid Tien toetsen toetsen een competentie die later in de beroepspraktijk ook gevraagd kan worden	+/-
2. Cognitieve complexiteit	Toetsen spelen niet in op denkprocessen Denkstappen wel bij zes toetsen verwerkt in beoordelingsmodel Geen aanwijzingen voor docent om student te sturen	+/-
3. Vergelijkbaarheid	Acht gesloten toetsen Bij zes toetsen is er een beoordelingsmodel	+
4. Geschiktheid voor onderwijsdoelen	Doelen beoordelingsmodellen letterlijk uit raamwerk Bij interactieve toetsen geen doelomschrijving Kennis, vaardigheden, attitude bij drie toetsen geïntegreerd Tien toetsen, kennis Veertien toetsen, vaardigheden Drie toetsen, attitude	+/-
5. Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten	Bij beweringentoetsen feedback door docent Advies tool gebruiken in methode Vier toetsen stimuleren het geven van feedback	+/-
6. Betekenisvolheid	Bij alle toetsen wordt rekening gehouden met referentiekader	+
7. Herhaalbaarheid	Geen tweede beoordelaar	-

4.4 Welke keuzes heeft Malmberg gemaakt bij het ontwerpen van de methode Taalblokken gericht op de kwaliteitseisen van het competentiegericht toetsen en het common Europeaan framework of reference?

Een samenvatting van de resultaten die voort zijn gekomen uit het interview bij de uitgever van Taalblokken is verwerkt in Tabel 13.

4.4.1 Tijd en kosten

De uitgever van Taalblokken geeft aan dat een educatieve uitgeverij rekening moet houden met een aantal kaders bij het ontwikkelen van een nieuwe methode. Een nieuwe methode moet namelijk voldoen aan de vraag van de markt, hier komt de uitgeverij achter door marktonderzoek te doen, gesprekken te voeren op scholen en literatuur te bekijken. Daarnaast komt een groot deel marketing en financiën aan de orde. Welke business modellen ga je gebruiken? Welke mensen heb je nodig? Wat is het budget? Waarom zou de klant het willen kopen? De uitgever geeft aan dat er bij de ontwikkeling van Taalblokken veel problemen zijn geweest met het systeem. Bij de ontwikkeling van een tweede versie van Taalblokken (versie 2.0) zal daar volgens de uitgever extra op gelet moeten worden. Naar aanleiding van de ontwikkeling van de methode is er een evaluatieronde geweest binnen Malmberg en in de markt. Zo kunnen er voor versie 2.0 onderdelen aangepast worden en anders aangepakt worden.

4.4.2 Geschiktheid voor onderwijsdoelen

De uitgever geeft aan dat er geen toetsmatrix aanwezig is voor de toetsen binnen Taalblokken. Er zijn nu vier toetsen per CEF- niveau per vaardigheid aanwezig. Dit zou volgens de uitgever voldoende dekkend moeten zijn. Verder zijn er zoveel mogelijk uitgangspunten van CGO betrokken binnen de methode, zo geeft de uitgever aan.

4.4.3 Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten

Reflecteren is volgens de uitgever voor MBO studenten en docenten erg moeilijk. Studenten hebben volgens haar niet het benodigde cognitief vermogen en inzicht om naar behoren te kunnen reflecteren. Daarnaast heeft de uitgever ervaren dat docenten ook nog moeite hebben met het aanleren van reflecteren aan studenten wat het voor de studenten moeilijker maakt. Er is bewust gekozen om de methode digitaal te maken omdat Malmberg als commercieel bedrijf tegemoet moet komen aan de wensen van de markt. Hier staat tegenover dat in een digitaal instrument het moeilijker is om reflectiemomenten in te bouwen.

4.4.4 Herhaalbaarheid

Voor de methode Taalblokken wordt er een implementatietraining aangeboden aan docenten vertelt de uitgever. Deze bestaat uit twee onderdelen: een knoppencursus waarin docenten kennis maken met de functies van de methode en een cursus voor taalbeleid binnen de school. Didactische onderwerpen kunnen hierbij aan bod komen, dit is echter geen vereiste. Bij de klantendag van Taalblokken worden didactische vraagstukken meestal wel besproken. Er is geen betrouwbaarheidsanalyse uitgevoerd over de toetsen van Taalblokken. Dit is wel een aandachtspunt voor versie 2.0, zo zegt de uitgever.

Tabel 13 Schematische weergave resultaten uitgever Taalblokken

criterium	Uitgever Taalblokken	Conclusie
1. Tijd en kosten	Methode moet voldoen aan vraag markt: - Marktonderzoek - Gesprekken voeren op scholen - Literatuur Marketing en financiën Business modellen Budget Personeel Bij ontwikkeling Taalblokken veel problemen met systeem Evaluatieronde na ontwikkeling Taalblokken binnen Malmberg	+
2. Geschiktheid voor onderwijsdoelen	Geen toetsmatrix Vier toetsen per CEF- niveau per vaardigheid voldoende	-
3. Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten	Reflecteren is voor MBO studenten en docenten moeilijk Docenten moeite aanleren reflecteren studenten Methode bewust digitaal gemaakt Bij digitaal instrument moeilijk reflectie inbouwen	-
4. Herhaalbaarheid	Implementatietraining - Knoppencursus - Taalbeleid Geen betrouwbaarheidsanalyse over toetsen van Taalblokken	+/-

5. Conclusie & Discussie

5.1 Korte samenvatting resultaten

Een schematische samenvatting van de conclusies die voort zijn gekomen uit het interview bij docenten, de vragenlijst, het beoordelingsinstrument en het interview bij de uitgever van Taalblokken is weergegeven in Tabel 14.

Tabel 14 Conclusies resultaten

Criteria	Interview docenten	Vragenlijst studenten	Beoordelings- instrument voor toetsen	Interview uitgever Taalblokken
1. Accepteerbaarheid	+	+/-	X	X
2. Authenticiteit	X	X	+/-	X
3. Cognitieve complexiteit	+/-	+/-	+/-	X
4. Vergelijkbaarheid	+/-	X	+	X
5. Tijd en kosten	+	+	X	+
6. Onderwijskundige gevolgen	+	-	X	X
7. Eerlijkheid	+	+	X	X
8. Geschiktheid voor onderwijsdoelen	+	X	+/-	-
9. Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten	+/- (samen met 3.)	+/- (samen met 3.)	+/-	-
10. Betekenisvolheid	+	X	+	X
11. Herhaalbaarheid	-	X	-	+/-
12. Transparantie	+	+	X	X

Door middel van drie instrumenten is er informatie verzameld over de kwaliteit van de methode Taalblokken die zich richt op het CEFR op basis van de twaalf kwaliteitscriteria van Baartman en collega's (2007).

Twee docenten zijn op tien kwaliteitscriteria bevraagd door middel van een interview. De methode voldoet volgens docenten aan accepteerbaarheid, tijd en kosten, onderwijskundige gevolgen, eerlijkheid, geschiktheid voor onderwijsdoelen, betekenisvolheid en transparantie. De criteria cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten en vergelijkbaarheid zijn nog niet optimaal. Docenten geven nog te weinig feedback en maken te weinig gebruik van de mogelijkheden binnen de methode tot feedback geven en reflectie. De vergelijkbaarheid verschilt per school evenals de onderwijskundige gevolgen. Dit heeft te

maken met de onderwijskundige visie van de school en de uitvoer hiervan. Herhaalbaarheid is voor beide docenten praktisch niet haalbaar en scoort hierdoor negatief.

Studenten (n=64) zijn op zes kwaliteitscriteria bevroegd. Er bleken een aantal significante verschillen te zijn tussen ROC A en ROC B. Deze verschillen hadden echter geen samenhang en worden daarom toegeschreven aan toeval. Studenten zijn positief over de tijd en kosten, eerlijkheid en de transparantie van de methode. Studenten zijn negatief over de onderwijskundige gevolgen van de methode. Studenten worden niet gestimuleerd om meer aandacht te besteden aan hun leerstijl. Daarnaast hebben ze het gevoel niets te leren. De criteria accepteerbaarheid en de cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten zijn volgens de studenten nog niet optimaal aanwezig binnen de methode. Studenten accepteren de methode in zijn geheel niet en hebben geen positieve verandering in hun leerstijl kijkend naar de uitgangspunten van het CGO.

De toetsen (n=14) zijn bekeken met een ontworpen beoordelingsinstrument. De toetsen voldoen aan vergelijkbaarheid en betekenisvolheid. De criteria authenticiteit, cognitieve complexiteit, geschiktheid voor onderwijsdoelen en ontwikkeling van zelfsturend leren en de effecten op het leren van studenten komen nog niet optimaal naar voren. De toetsen spelen niet in op de herhaalbaarheid van de toetsen. De toetsen binnen de methode Taalblokken bestaan uit vier kennistoetsen, zes simulaties van de praktijk en vier toetsen die niet te plaatsen zijn op de trap van Klarus (2001). Bij zes toetsen wordt het assessment voorafgegaan door een minimale authentieke instructie. Tien van de veertien toetsen, toetsen een betekenisvolle prestatie op het gebied van taal en de toekomstige beroepspraktijk. De mening van studenten wordt alleen bij de beweringstoetsen betrokken. Kennis, vaardigheden en attitude zijn bij drie van de veertien toetsen geïntegreerd aanwezig. De sociale processen die plaatsvinden binnen de toetsen zijn: overleggen, beslissingen maken, voordracht en beleefdheid. Bij zeven toetsen komen er geen sociale processen voor. Bij tien toetsen wordt een competentie getoetst die later in de beroepspraktijk ook van de student gevraagd kan worden. De statische toetsen sluiten volledig aan bij het raamwerk (Bve- raad, 2001). Vier toetsen stimuleren reflectie bij de studenten en het geven van feedback door de docenten en dus de ontwikkeling van zelfsturend leren en de effecten op het leren van studenten.

Een medewerker van Malmberg is bevroegd op vier kwaliteitscriteria. De methode is bewust digitaal gemaakt als wens van de markt. Tijd en kosten komt hierin naar voren. Deze staan in verhouding met de opbrengsten die het Malmberg oplevert. Hier staat tegenover dat de ontwikkeling van zelfstandig leren en de effecten op het leren van studenten niet naar voren komt. Er zijn beperkingen op het gebied van reflectie en feedback door de digitale aard van de methode. Er mist een betrouwbaarheidsanalyse en een toetsmatrix van de toetsen wat het moeilijk maakt over de herhaalbaarheid en de geschiktheid voor onderwijsdoelen uitspraken te doen. Er is echter wel een implementatietraining aanwezig die veel mogelijkheden biedt.

5.2 Verklaring resultaten

Het CAP van Baartman en collega's (2007) komt niet volledig terug in de methode Taalblokken van Malmberg. Hieronder worden de criteria besproken die niet volledig naar voren kwamen in de methode Taalblokken.

De accepteerbaarheid van de methode is niet optimaal. Alleen docenten zijn het eens met de doelen die gesteld worden. Studenten zouden het hier ook mee eens moeten zijn evenals de beroepspraktijk. Deze is in dit onderzoek niet bevraagd vanwege tijdsbeperkingen. Studenten vinden de methode niet fijn. Doordat alle ontwikkelingen binnen het MBO zo snel zijn verlopen is het moeilijk voor zowel docenten als studenten om dit te verwerken binnen hun lessen. Aan deze recente ontwikkelingen moet meer aandacht besteed worden in de vorm van projecten voor studenten en training voor docenten. Men heeft er echter de tijd en het geld niet voor. Studenten moeten in een redelijk korte tijd overstappen van een leerkrachtgestuurde leeromgeving naar een studentgestuurde leeromgeving. Studenten moeten zelfstandig worden, leren van een computerscherm en zelf hun ontwikkelingen bijhouden. De medewerker van Malmberg gaf aan dat studenten hier nog niet aan toe zijn. Door meer informatie te verschaffen over de doelen die gesteld worden in de methode en het nut van CGO zullen studenten meer geneigd zijn de methode te accepteren zodat dit criterium behaald wordt.

De authenticiteit van de methode is niet optimaal. Dit kan verklaard worden door de aard van de methode. Studenten worden veel getoetst achter een computer, dit is geen authentieke situatie. Wanneer studenten meer de beroepspraktijk in gestuurd zouden worden zou dit de authenticiteit van de toetsen bevorderen. De methode is echter bewust digitaal gemaakt zoals de uitgever aangeeft. Hier moeten afwegingen in gemaakt worden.

De cognitieve complexiteit van de methode is niet toereikend. Dit kan verklaard worden door de manier waarop de toetsen zijn vormgegeven. Deze houden geen rekening met de verschillende denkprocessen die een student moet doorlopen. Daarnaast geven docenten zoals eerder genoemd te weinig feedback naar aanleiding van de toetsen, waardoor de studenten zich ook moeilijker ontwikkelen. De handvaten voor deze stappen zijn echter wel weergegeven in de beoordelingsmodellen van de statische methode. Ook hier laat de uitvoer van de toetsen te wensen over.

De vergelijkbaarheid van de toetsen verschilt per ROC en wordt daarom als niet volledig weergegeven. Bij ROC A werken ze met een toetsweek wat de vergelijkbaarheid van de toetsen verhoogt. Op het ROC B bepaalt de docent wanneer de student een toets krijgt conform de richtlijnen van het CGO, dit bevordert echter niet de vergelijkbaarheid.

De onderwijskundige gevolgen ervaren studenten als negatief. Ook dit is te verklaren door de snelle veranderingen binnen het MBO. Studenten vinden het moeilijk hun leerstijl aan te passen en willen dit ook niet. Daarnaast hebben ze het gevoel hierbij te weinig uitgedaagd te worden door de methode. De methode is volgens de studenten te 'simpel'. Hier komt het 'backwash' effect naar voren (Alderson & Wall, 1993). Studenten leren minimaal voor de toetsen omdat ze niet uit worden gedaagd om er meer voor te doen. Dit kan ook verklaard worden door de aard van de methode. Door alle stappen te doorlopen zouden studenten vanzelf bij de eindstreep moeten komen, er is geen sprake meer van leertentamens of het maken van kennistoetsen. Studenten vinden het moeilijk om met competentiegericht toetsen om te gaan. Het is een manier van leren die ze niet gewend zijn. Ook hiervoor geldt dat dit zal moeten groeien.

De geschiktheid voor onderwijsdoelen is nog niet optimaal. Dit kan verklaard worden door het ontbreken van doelomschrijvingen bij de toetsen en het ontbreken van een toetsmatrix voor de toetsen. De medewerker van Malmberg geeft wel aan dat de toetsen opgezet zijn aan de hand van het raamwerk voor moderne vreemde talen (Bve- raad, 2001) en daarom voldoet aan de onderwijsdoelen. Deze uitspraak kan echter niet bevestigd worden.

De ontwikkeling van zelfsturend leren en de effecten op het leren van studenten komt bij meerdere criteria naar voren. Dit is al verklaard door de snelle verandering binnen het MBO. Studenten veranderen hun leerstijl wel maar niet altijd op een positieve manier. Daarnaast moeten studenten leren reflecteren wat moeilijk is gezien het cognitieve niveau van studenten. Ook maken docenten niet optimaal gebruik van de mogelijkheden tot reflecteren binnen de methode. De 'test- jezelf toetsen' worden door beide docenten bijvoorbeeld niet gebruikt. Zo blijft de ontwikkeling van zelfsturend leren uit bij studenten.

Er is weinig aandacht voor de herhaalbaarheid. Dit kan verklaard worden door de praktische inzetbaarheid van docenten. Baartman en collega's (2007) en Van der Vleuten en Schuwirtz (2005) geven aan dat beslissingen niet alleen af mogen hangen van een specifieke assessmentsituatie en beoordelaar. Dit is een ideaalbeeld wat in de praktijk niet haalbaar is. Hiervoor zijn te weinig docenten beschikbaar en staat er een te hoge tijdsdruk op de werkzaamheden van de docenten. Docenten geven dit zelf ook aan.

De kwaliteitscriteria eerlijkheid, tijd en kosten, betekenisvolheid en transparantie komen allen positief naar voren binnen de methode. Dit betekent dat de toetsen van de methode Taalblokken op deze kwaliteitscriteria voldoende scores kijkend naar de instrumenten waarmee deze criteria zijn gemeten.

5.3 Beperkingen onderzoek

De beperkingen van dit onderzoek hebben op de eerste plaats betrekking op de steekproefgrootte. De doelstelling van dit onderzoek was om vijf docenten te interviewen en honderd studenten (twintig per school). Deze doelstelling is niet behaald. Doordat de dataverzameling in de maand mei was kwam deze samen met de meivakantie en de tentamenweek. Na de meivakantie waren scholen voornamelijk bezig met de afsluiting van het schooljaar. Verder reageerde maar één docent op de e-mail die rondgestuurd was waardoor er nagebeld en nagemaild moest worden. Dit nam veel tijd in beslag en dat resulteerde in een kleine onderzoeksgroep. Voornamelijk de dataverzameling bij docenten was erg minimaal. Docenten hebben het druk, worden vaak gevraagd deel te nemen aan onderzoeken op school en willen vaak geen tijd vrij maken. Een gevolg van deze kleine steekproef is dat de resultaten die verkregen zijn uit het onderzoek mogelijk niet generaliseerbaar zijn. Ook de resultaten uit de vragenlijst die afgenomen is, zijn niet voldoende representatief. Deze is afgenomen bij twee scholen waarbij op de ene school vijftig studenten een vragenlijst hebben ingevuld en bij de andere school veertien studenten dit hebben gedaan. Dit is een onevenredige verdeling.

Het afnemen van interviews heeft op de tweede plaats ook een aantal beperkingen. De betrouwbaarheid van de informatie kan beïnvloed worden door het geven van sociaal wenselijke antwoorden (Baarda & de Goede, 2001). Dit probleem is ondervangen doordat de onderzoeker zo open mogelijk het interview in is gegaan. Docenten konden vrij praten zonder gevolgen. Daarnaast zijn mensen zich niet altijd bewust van hun eigen gedrag en de motieven voor hun gedrag. Dit is

ondervangen door specifiek door te vragen naar bepaalde concrete details. Wanneer een docent aangaf feedback te geven is er doorgevraagd naar het aantal keren dat hij feedback gaf, op welke momenten en op welke manier. Docenten moesten hierdoor dieper nadenken over hun motieven. Hier is ook bij het interview met de uitgever van Taalblokken rekening mee gehouden. Wat bij interviews ook een beperking kan zijn is de objectiviteit van de verwerking van de interviews. De interviews zijn door de onderzoeker alleen afgenomen en deze heeft hier een eigen interpretatie aan gegeven. Een andere onderzoeker zou misschien een andere betekenis verlenen aan bepaalde uitspraken.

Een andere factor die invloed heeft op de resultaten is de betrouwbaarheid van de vragenlijst. Bij de schalen van cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten, tijd en kosten en eerlijkheid is de betrouwbaarheid te laag. Deze criteria zijn met te weinig items getoetst om hier een betrouwbare uitspraak over te kunnen doen. De betreffende resultaten dienen dus voorzichtig geïnterpreteerd te worden. Door tijdsbeperking en mate van concentratie van MBO studenten was het niet mogelijk veel items op te nemen in de vragenlijst. Bij het opzetten van een grotere vragenlijst met meer respondenten was de betrouwbaarheid mogelijk hoger geweest. Daarnaast waren bepaalde items binnen de vragenlijst mogelijk niet representatief omdat ze bevraagd zijn met een specifiek toetsvoorbeeld. Hier is bijvoorbeeld een toets voor schrijven weergegeven voor niveau A2. Bij het nemen van een andere voorbeeldtoets hadden studenten misschien anders geantwoord. Het opnemen van alle toetsen was niet haalbaar waardoor hier een keuze gemaakt moest worden. Ten slotte bleek uit de factoranalyse van de vragenlijst dat bijna alle items gemeten werden met één factor. Dit betekent dat de validiteit van de vragenlijst niet optimaal is. Doordat er veel overlap bestaat tussen de verschillende criteria kunnen vragen uit de vragenlijst van één criteria resultaten van een ander criteria verklaren. Dit zorgt ervoor dat er geen eenduidige conclusies getrokken kunnen worden over de criteria afzonderlijk.

Baartman (2008) geeft in haar proefschrift aan dat veel criteria met elkaar samenhangen. Het is bijvoorbeeld vaak aangenomen dat een authentieke en meer transparante assessment het leren zal verbeteren (Frederiksen & Collins, 1989; Linn et al., 1991) en dat het hebben van toegang tot de criteria voor beoordeling noodzakelijk is om de eerlijkheid vast te stellen. Een mix van verschillende assessment methoden bevordert de eerlijkheid. De tijd die geïnvesteerd wordt in een assessment zou moeten resulteren tot positieve onderwijskundige gevolgen (Hambleton, 1996). De verschillende kwaliteitscriteria kunnen dus waarschijnlijk gegroepeerd worden binnen een kleiner aantal noemers (Baartman, 2008). Er is daarom volgens Baartman (2008) meer onderzoek nodig om te onderzoeken of het aantal kwaliteitscriteria gereduceerd kunnen worden en of dit gedaan kan worden kijkend naar de samenhang tussen de criteria. Dit is ook weergegeven in het theoretische kader. Binnen dit onderzoek is er in Tabel 1 een poging gedaan de samenhang tussen de verschillende criteria weer te geven.

Een beperking bij het ontworpen toetsinstrument is de objectiviteit van de beoordeling. De interbeoordelaarsbetrouwbaarheid is hoog (88,58%). Dit betekent echter niet dat de objectiviteit van het instrument optimaal is. Items die laag scoren op de objectiviteitmeting hebben nog een hoge interpretatiewaarde. Iedere beoordelaar heeft een andere mening over bepaalde onderwerpen. Zo zal de ene beoordelaar een instructie wel authentiek vinden terwijl een andere beoordelaar deze misschien als niet authentiek zou beoordelen. Door een uitgebreide beschrijving

te geven van de begrippen zal deze interpretatie lager worden en zal de objectiviteit van het instrument verbeteren. Daarnaast is het aan te raden een handleiding bij het instrument te voegen met een uitleg van de verschillende doelen van de toetsen en de manier waarop het instrument gebruikt moet worden.

Als laatste beperking zijn niet alle criteria getoetst met zowel de interviews, de vragenlijst en het beoordelingsinstrument voor de toetsen. Dit betekent dat niet alle criteria volledig onderzocht zijn op alle instrumenten. Er kan hierdoor geen uitspraak gedaan worden over het volledig voorkomen van bepaalde criteria. Beter zou zijn alle criteria te verwerken binnen alle instrumenten zodat er hierover wel uitspraken gedaan kunnen worden.

5.4 Mogelijkheden voor verder onderzoek

Er kan meer onderzoek gedaan worden naar het in deze thesis ontworpen instrument. Het is een praktisch instrument voor scholen en instellingen om een assessment programma te beoordelen. Een voorwaarde hiervoor is om de andere kwaliteitscriteria binnen het instrument te betrekken en het instrument concreter te maken, door het opnemen van een bijhorende handleiding. Bij meer tijd en welwillendheid van scholen zou het onderzoek nogmaals uitgevoerd kunnen worden met een grotere onderzoeksgroep. De resultaten die daaruit voortkomen, zullen wel generaliseerbaar zijn en een belangrijkere betekenis hebben voor de wetenschap en de praktijk. Uitspraken die dan gedaan worden zullen een grotere invloed hebben. Naar aanleiding van een vervolgstudie kan het beoordelingsinstrument verder ontwikkeld worden zodat het een instrument kan worden wat inzetbaar is voor het toetsen van assessmentprogramma's binnen het MBO.

Het zou interessant zijn om dit onderzoek over een aantal jaren nogmaals te voeren. Hierbij kan dan bekeken worden hoe de ontwikkelingen zijn geweest binnen met MBO. Het CGO en toetsen staan nu nog in de kinderschoenen. Over een aantal jaren is de implementatie van deze onderwijsvorm al verder gevorderd. Het zou erg zinvol zijn om die ontwikkeling weer te geven in het kader van de kwaliteitscriteria van Baartman en collega's (2007).

7. Literatuurlijst

Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, 91, 659-663.

Alderson, J. C., & Wall, D. (1993). Does Washback exist? *Applied Linguistics*, 14, 115-129.

Baarda, D. B., & de Goede, M. P. M. (2006). *Basisboek methoden en technieken*. Houten: Wolters-Noordhoff Groningen.

Baartman, L. K. J. (2008). *Assessing the assessment*. Utrecht: Netherlands organisation for scientific research.

Baartman, L. K. J., Prins, F. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2007). Determining the quality of competence assessment programs: a self- evaluation procedure. *Studies in Educational Evaluation*, 33, 258-281.

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2006). Teachers' opinions on quality criteria for competency assessment programs. *Teaching and Teachers Education*, 23, 857-867.

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2006). The Wheel of competency assessment: presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation*, 32, 153-170.

Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., et al., (2006). Position paper, A learning integrated assessment system. *Educational Research Review*, 1, 61-67.

Bloom, B. S., & Nemetz, A. (1957). Taxonomy of educational objectives: the classification of educational goals, handbook I: cognitive domain. *Journal of Higher Education*, 28 (5), 290-292.

Boeije, H. (2005). *Analyseren in kwalitatief onderzoek*. Hoofddorp: Boomonderwijs.

Boekaerts, M. (2002). Bringing about change in the classroom: strengths and weakness of the self-regulated learning approach- EARLI presidential address, 2001. *Learning and Instruction*, 12, 589-604.

Broekkamp, H., & van Hout-Wolters, B. H. A. M. (2007). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review*, 19 (4), 401-428.

BVE raad. (2001). *Raamwerk moderne vreemde talen in het secundair beroepsonderwijs*. 's-Hertogenbosch: CINOP.

- Byrnes, H. (2007). Developing national language education policies: Reflections on the CEFR. *Modern Language Journal*, 91 (4), 679–685.
- Carver, C. S., & Shreier, M. F. (2000). On the structure of behavioral self-regulation. In Boekaerts, M., Pintrick, P. R., Zeidner, M. (2000). *Handbook of Self-regulation* (pp 41-84). San Diego, CA: Academic press.
- Competentie- en taalgericht toetsen*, (n.d.). Gevonden op 3 maart 2008, op <http://www.taalgerichtvakonderwijs.nl/literatuur/algemeen/00012/>
- Council of Europe. (2001). *Common European Framework of Reference*. Strasbourg.
- Dekker, J. , & Sanders, P. (2008). *Kwaliteit van beoordeling in de praktijk*. Ede: Kenniscentrum handel.
- DIALANG*. (n.d.). Gevonden op 17 maart 2008, op <http://www.dialang.org/dutch/index.htm>
- Dierick, S., & Dochy, F. (2001). New lines in edumetrics: new forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27, 307-329.
- Eraut, M. (1994). *Developing professional knowledge and competence*. Londen: The falmer press.
- Studion, (2007). *Factoranalyse*, gevonden op 21 juli 2008 op <http://studion.fss.uu.nl/Bouwstenenonline/3a7factoranalyse.doc>.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18, 27-32.
- Glaser, G., & Strauss, A. L. (1969). The discovery of grounded theory: Strategies for qualitative research. *The British Journal of Sociology*, 20 (2), 227-228.
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five- dimensional framework for authentic assessment. *ETR&D*, 52 (3), 67-86.
- Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91, 663–667.
- Interview Claartje Bom, uitgever cluster MBO, Taalblokken, Malmberg, Den Bosch, 5 juni 2008.
- Interview Erkan Cevher, docent Duits en muziek, Albeda College, Rotterdam, 13 juni 2008.
- Interview Margreet Roelse- Minnaar, docente Engels en coördinator taal, ROC Zeeland, Middelburg, 14 mei, 2008.

- Johnsen, S. K. (1997). Assessment beyond definitions. *Peabody Journal of Education*, 72 (3&4), 136-152.
- Klarus, R. (2001). *Instrumenten voor competentie gericht toetsen en beoordelen*. Gevonden op 11 september 2003, op www.kenniscentrumevc.nl/site/documenten/KC3/pdf.
- Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17, 48-60.
- Linn, R.L., Baker, J., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20 (8), 15-21.
- Little, D. (2007). The common European framework of reference for languages: perspectives on the making of supranational language educational policy. *The Modern Language Journal*, 91, 645-655.
- MBO raad. (2007). *Raamwerk Nederlands, Nederlands in (v)MBO- opleiding, beroep en maatschappij*. 's-Hertogenbosch: CINOP.
- Mulder, M., Wesselink, R., Biemans, H., Nieuwenhuis, L., & Poell, R. (2003) *Competentiegericht beroepsonderwijs*. Houten: Wolters- Noordhoff.
- Nationaal bureau voor moderne vreemde talen (2006). *Europees taalportfolio*. Gevonden op 17 maart 2008, op <http://www.europeestaalportfolio.nl/TaalPortfolio/show.do?ctx=10010,10023&anav=10016>.
- North, B. (2007). The CEFR illustrative descriptor scales. *The Modern Language Journal*, 91, 656-659.
- Taalbeheersing voorwaarde voor toetreding arbeidsmarkt, persbericht*, (11 oktober 2007). Gevonden op 22 april 2008, op taalinMBO.kennisnet.nl.
- Robson, C. (2002). *Real world research*. United Kingdom: Blackwell publishing.
- Sociaal- economische raad. (2001). *Advies koers BVE 01-01*. Den Haag: SER.
- Stoof, A., Martens, R. L., & Merriënboer J. J. G. (2000). *What is competence?* Leiden: Open universiteit Nederland.
- Stuurgroep competentiegericht beroepsonderwijs. (2007). *Leren, loopbaan en burgerschap*. 's Hertogenbosch: CINOP.
- 't Hart, H., van Dijk, J., de Goede, M., Jansen, W., & Teunissen J. (1996). *Onderzoeksmethoden*. Amsterdam: Boom.

- Uhlenbeck, A. M., Verloop, N., & Beijaard, D. (2002). Requirements for an assessment procedure for beginning teachers: Implications from recent theories on teaching and assessment. *Teachers College Records, 104* (2), 242-272.
- Van der Maesen de Sombreff, P., & Schakel, L. (1999). Wat zijn competenties niet? *Opleiding en Ontwikkeling, 12*, 11-17.
- Van der Vleuten, C. P. M. (1996). The assessment of professional competence: developments, research and practical complications. *Advances in Health Education, 1*, 41-67.
- Van der Vleuten, C. P. M., Scherpbier, A. J. J. A., Dolmans, D. H. J. M., Schuwirth, L. W. T., Verwijnen, G. M., & Wolfhagen, H. A. P. (2000). Clerkship assessment assessed. *Medical Teacher, 22* (6), 592-600.
- Van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: from methods to programmes. *Medical Education, 39*, 309-317.
- Westera, W. (2001). Competences in education: a confusion of tongues. *J. Curriculum Studies, 33* (1), 75-88.
- White, R. W. (1959) Motivation reconsidered: the concept of competence. *Psychological Review, 66* (3), 297-233.
- Winn, F., & Bittner, A. (2005). Extended framework for older worker competence: Effects of cognitive complexity on performance. *International Congress Series, 1280*, 35-40.

Bijlagen

Bijlage 1 CEF niveaus

	Luisteren	Lezen	Gesprekken voeren	Spreken	Schrijven
C1	Kan uitgebreide betogen over abstracte en complexe onderwerpen volgen, ook buiten het eigen interesse- of vakgebied. Begrijpt veel idioom en spreektaal-uitdrukkingen.	Kan tot in detail lange, complexe teksten begrijpen, waaronder specialistische artikelen en lange technische instructies op het eigen vakterrein of bedoeld voor de geïnteresseerde leek, mits moeilijke passages herlezen kunnen worden.	Kan zich vloeiend en spontaan uitdrukken. Kan de taal flexibel en effectief gebruiken voor sociale en voor professionele doeleinden.	Kan gedetailleerde en precieze beschrijvingen geven van en formele presentaties houden over complexe onderwerpen	Kan duidelijke, goedgestructureerde teksten over complexe onderwerpen in werk, opleiding en privé-leven adequaat en accuraat schrijven.
B2	Begrijpt meer complexe informatie over onderwerpen uit het dagelijks leven en de wereld van opleiding en beroep in voldoende mate om de draad van het betoog te volgen.	Begrijpt een breed scala aan teksten op het eigen vak- of interessegebied. Begrijpt teksten over specialistische onderwerpen als hij voldoende kennis heeft van het desbetreffende onderwerp.	Kan op een effectieve wijze deelnemen aan (semi-)formele en informele gesprekken over onderwerpen van praktische, sociale en beroepsmatige aard. Kan in een discussie een mening geven en die met argumenten onderbouwen.	Kan een duidelijk, gedetailleerd betoog houden over een breed scala van onderwerpen uit het eigen interesse- of werkgebied.	Kan duidelijke, gedetailleerde tekst schrijven over een breed scala van onderwerpen in het dagelijks leven, in het beroepsleven en in opleidingen.
B1	Begrijpt feitelijke informatie over veel voorkomende	Kan feitelijke teksten over onderwerpen uit de eigen werk- of leefomgeving	Kan met redelijk gemak deelnemen aan gesprekken over	Kan een eenvoudige uiteenzetting geven over vertrouwde onderwerpen	Kan eenvoudige samenhangende tekst schrijven over vertrouwde

	onderwerpen uit dagelijks leven en werk.	lezen met een redelijke mate van begrip.	onderwerpen uit het dagelijks leven, gericht op het onderhouden van sociaal contact en het regelen van zaken.	uit de eigen leef- en werkomgeving.	onderwerpen in het dagelijks leven, op opleiding of werk.
A2	Begrijpt de belangrijkste punten van korte en eenvoudige boodschappen en aankondigingen als er langzaam en duidelijk gesproken wordt.	Kan korte en eenvoudige teksten lezen over concrete onderwerpen uit de eigen werk- of leefomgeving.	Kan korte gesprekken voeren in standaardsituaties. Kan vragen stellen en beantwoorden en ideeën en informatie uitwisselen over vertrouwde onderwerpen in voorspelbare, alledaagse situaties.	Kan in eenvoudige bewoordingen een beschrijving geven van mensen, leef- en werkomstandigheden, dagelijkse routines enzovoort.	Kan korte, informele briefjes schrijven en boodschappen noteren in telegramstijl.
A1	Begrijpt vertrouwde uitdrukkingen en eenvoudige zinnen als er heel langzaam en duidelijk gesproken wordt.	Kan vertrouwde namen, woorden en zeer eenvoudige zinnen begrijpen in standaardteksten als mededelingen, posters, catalogi.	Kan deelnemen aan eenvoudige gesprekken in situaties die veel voorkomen in het dagelijks leven en op de werkplek.	Kan zich uitdrukken in losse woorden en in korte, eenvoudige zinnen over concrete zaken uit de eigen leef- of werkomgeving.	Kan losse woorden zoals personalia in formulieren invullen en een kort tekstje schrijven zoals vakantiegroeten op een ansichtkaart.

Bijlage 2 Interview met docenten

Nr.	Vraag	CAP
1.	Docenten zijn bekend met de doelen die gesteld worden in het CEF. Ik leg docenten uit dat ik weet dat ze op de hoogte zijn van het CEF. Ik hou de doelen er nog wel even naast ter ondersteuning. Het zal gemakkelijker zijn om zo de vragen te beantwoorden. Ik leg de docenten de doelen voor van het CEFR. Ik geef de docenten even de tijd om ze door te nemen. Ik stel ze de volgende vraag:	...
1a.	Vindt u dat de doelen die hier omschreven worden voor de verschillende vaardigheden compleet genoeg zijn? Vindt u ze acceptabel?	Accepteerbaarheid
2.	Komen volgens u alle vijf de vaardigheden (lezen, luisteren, spreken, gesprekken voeren en schrijven) voldoende aan bod binnen de methode?	Geschiktheid voor onderwijsdoelen
3.	Vindt u de methode voldoende beroepsgericht? Voor zover u dat kan beoordelen?	Accepteerbaarheid
4.	Bij deze vraag wil ik een kort inleidend stukje vertellen over competentie gericht onderwijs en dat het erg belangrijk is dat je een goede verhouding hebt tussen kennis, vaardigheden en attitude. Dit is dus de kennis die een leerling bezit, de handelingen die hij hiermee uit kan voeren en de houding die hij heeft tegenover deze kennis en handelingen. Deze moeten geïntegreerd aan bod komen binnen (bijna) iedere toets die gegeven wordt.	...
4a.	Zijn deze drie elementen (kennis, vaardigheden, attitude) volgens u aanwezig binnen Taalblokken? Hoe is volgens u de verhouding tussen kennis, vaardigheden en attitude binnen de toetsen van Taalblokken	Geschiktheid voor onderwijsdoelen
5.	Op welke manier geeft u de studenten feedback op opdrachten of toetsen? Denk hierbij aan mondeling of schriftelijk. In de vorm van extra opdrachten, tips, aanvullingen, etc. (let u op denkstappen/ en kijkt u later nog terug op wat de leerlingen hebben gedaan met de feedback?)	Cognitieve complexiteit & ontwikkeling van zelfsturend leren en de effecten op het leren van studenten
6.	Wat vindt u van de feedback die leerlingen krijgen in de methode bij het fout beantwoorden van een vraag?	Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten
7.	Wat is uw mening over de 'test jezelf' toetsen met de 'can- do-	Ontwikkeling van

	statements'? Geven ze voldoende indicatie van het niveau waarop een leerling zich bevindt? Heeft de leerling voldoende reflectiemogelijkheden om zelf een reële schatting te maken?	zelfsturend leren en de effecten op het leren van studenten
8.	Laat u aan de hand van uw feedback de studenten leerdoelen opstellen aan het begin van een blok? En hoe toetst u deze?	Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten
9.	Wanneer bepaald een docent of een student een toets moet maken? Is dit moment voor iedere student hetzelfde? Zijn studenten bij het maken van een toets altijd in dezelfde toetsomgeving?	Vergelijkbaarheid
10.	Vindt u dat er voldoende toetsmomenten aanwezig zijn in de methode?	Kosten en efficiëntie
11.	Ik leg de docenten de toets voor van gesprekken voeren A1 > A2, toets 307. Ik leg de docenten uit dat dit een statische toets is uit taalblokken die ze toe moeten kennen aan de studenten (hij is dus niet openbaar). Daarnaast leg ik de beoordelingscriteria voor die de docent moet aanhouden bij het beoordelen van de student aan de hand van deze toets	...
11a.	Vindt u de beoordelingscriteria voor de toets van gesprekken voeren A1 > A2, toets 307 begrijpelijk genoeg?	Transparantie
11b.	Vindt u dat de beoordelingscriteria zoals ze hier gesteld zijn voor deze toets voldoende aansluiten bij de doelen die gesteld worden in het CEFR? (de docent heeft deze doelen nog voor zich en kan daar eventueel naar terugkijken)	Accepteerbaarheid
11c.	Wat vindt u specifiek binnen deze toets van de verhouding kennis, vaardigheden en attitude?	Geschikt voor onderwijsdoelen
12.	Ik leg de docent nogmaals de toets gesprekken voeren A1 > A2, toets 307 voor.	...
12a.	Hoeveel tijd heeft u nodig om deze toets voor te bereiden? Er staat 5 minuten voor het uiteindelijke gesprek, vindt u dit voldoende? Is de voorbereiding die u in een toets steekt een redelijke tijdsbesteding in verhouding met hetgeen wat het u oplevert?	Kosten en efficiëntie
13.	Heeft u het gevoel dat u voldoende competent bent om een dergelijke beoordelingen uit te voeren? (deze zijn natuurlijk specifiek gericht op de CEF- niveaus en zijn anders dan de gestandaardiseerde beoordelingscriteria)	Eerlijkheid
14.	Is uw stijl van lesgeven veranderd na de invoer van taalblokken?	Onderwijskundige gevolgen

15.	Ziet u verandering in de leerstijl van uw studenten na de invoer van taalblokken?	Onderwijskundige gevolgen
16.	Is er ruimte voor leerlingen om bezwaar te maken tegen een toets? (bijvoorbeeld als ze vinden dat er te hoge eisen gesteld worden of het niet eens zijn met de manier waarop de toets wordt afgenomen)	Eerlijkheid
17.	Om een beoordeling zo betrouwbaar mogelijk te houden is het aan te raden meerdere beoordelaars in te zetten bij het afnemen van een toets. Wat vindt u hiervan? (Ik kan er denk ik vanuit gaan dat docenten niet werken met een interbeoordelaarsbetrouwbaarheid!	Herhaalbaarheid
18.	Op hoeveel toetsmomenten baseert u een eindoordeel over een student? (per vaardigheid, per niveau)	Herhaalbaarheid
19.	Vindt u de beoordelingen die gegeven worden aan studenten zinvol met betrekking tot de toekomstige beroepspraktijk? (komen de beoordelingen de beroepspraktijk tegemoet?)	Betekenisvol

Bijlage 3 Vragenlijst studenten

Nr.	Vraag	Schaal	CAP
	LEREN		
5.	Ik weet wat ik moet leren voor Engels met Taalblokken.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Transparantie
6.	Door Taalblokken ben ik anders gaan leren voor Engels dan hiervoor toen we nog werkten met een boek en een werkboek. (Denk hierbij aan het leren van een toets, is je voorbereiding veranderd?)	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Onderwijskundige gevolgen
7.	Ik bereid me voor op een toets schrijven van Engels door <ul style="list-style-type: none"> - Opdrachten uit Taalblokken nog een keer te maken. - Opdrachten die ik al gemaakt heb nog een keer te bekijken en van mijn fouten te leren. - Samen met medestudenten oefenen (vb. ik maak een tekst en mijn medestudenten kijken deze na) - Ik vraag extra oefenmateriaal aan mijn docent. - Ik leer alles wat ik moet leren voor mijn toets uit mijn hoofd. - Overig namelijk _____ 	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Onderwijskundige gevolgen
8.	Bekijk voorbeeld 1, dit is een toets voor schrijven op niveau A2. Je moet voor deze toets aantekeningen maken bij een gesprek. Door deze toets leer ik beter te schrijven voor Engels	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Accepteerbaarheid
	BEOORDELING/ VOORTGANG		
9.	Bekijk voorbeeld 2, dit is een toets voor Engels, op niveau A2 voor gesprekken voeren. Ik vind het redelijk dat ik 5 minuten de tijd krijg om deze opdracht uit te voeren (dit is zonder voorbereiding).	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Tijd en kosten
10.	Ik vind dat de toetsen voor Engels goed te begrijpen zijn.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Transparantie
11.	Ik vind dat de toetsen voor Engels goed uit te voeren zijn.	helemaal mee oneens 1 2 3 4	Transparantie

		5 6 helemaal mee eens	
12.	Ik weet waaraan ik moet voldoen voordat ik een toets mag maken van mijn docent voor Engels bij Taalblokken.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Transparantie
13.	Ik weet waarop ik word beoordeeld bij Taalblokken, Engels. (Bijvoorbeeld als ik een toets moet maken voor luisteren.)	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Transparantie
14.	Ik krijg voldoende kansen binnen Taalblokken om te laten zien wat ik kan op het gebied van Engels.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Eerlijkheid
15.	Ik weet van wie ik een beoordeling krijg voor een toets voor Engels.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Transparantie
16.	Ik weet waar mijn docent op let bij het nakijken van een toets van Engels en het geven van een beoordeling.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Transparantie
17.	Aan het einde van een blok van Engels vind ik het belangrijk om terug te kijken op alles wat ik heb geleerd, wat er al erg goed ging en waar ik nog beter op moet oefenen.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Cognitieve complexiteit/ Ontwikkelen van zelfsturend leren en de effecten op het leren van studenten
18.	Iedere student wordt volgens mij gelijk behandeld bij het maken van een toets voor Engels.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Eerlijkheid
	FEEDBACK		
19.	Ik let op de opmerkingen (feedback) die mijn docent mij geeft over het werk dat ik inlever of maak.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Cognitieve complexiteit/ Ontwikkelen van zelfsturend leren en de effecten op het leren van studenten
20.	Door deze opmerkingen van mijn docent ga ik anders	helemaal mee	Cognitieve

	te werk binnen taalblokken.	oneens 1 2 3 4 5 6 helemaal mee eens	complexiteit/ Ontwikkelen van zelfsturend leren en de effecten op het leren van studenten
21.	Ik vind het fijn als andere studenten mijn werk nakijken/ bekijken en mij feedback geven.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Cognitieve complexiteit/ Ontwikkelen van zelfsturend leren en de effecten op het leren van studenten
22.	Ik vind de extra uitleg die taalblokken mij aanreikt als ik een fout antwoord geef duidelijk genoeg.	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Cognitieve complexiteit/ Ontwikkelen van zelfsturend leren en de effecten op het leren van studenten
	TAALBLOKKEN ALGEMEEN		
23.	Ik vind het fijn dat we Taalblokken gebruiken in de klas	helemaal mee oneens 1 2 3 4 5 6 helemaal mee eens	Accepteerbaarheid

Bijlage 4 Beoordelingsinstrument voor toetsen

BEOORDELINGSINSTRUMENT BEKIJKEN TOETSEN		
ALGEMEEN		
1.	Toetsomschrijving: <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>	
2.	Binnen welke sector valt deze toets?	<input type="radio"/> Economie <input type="radio"/> Economie- HTR <input type="radio"/> Techniek <input type="radio"/> Zorg en Welzijn <input type="radio"/> Basistaal traject
3.	Binnen welk niveau valt deze toets?	<input type="radio"/> Niveau 2 <input type="radio"/> Niveau 3-4
4.	Binnen welk blok wordt deze toets gegeven?	<input type="radio"/> 0 > A1 <input type="radio"/> A1 > A2 <input type="radio"/> A2 > B1 <input type="radio"/> B1 > B2 <input type="radio"/> B2 > C1
5.	Wat is het doel van de toets? <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>	
6.	Is de toets interactief of statisch?	<input type="radio"/> Interactief <input type="radio"/> Statisch

AUTHENTICITEIT		
7.	Op welke traptrede van de trap van Klarus is de toets in te delen? (Klarus, 2001)	
a.	<i>Portfolio- inventarisatie</i>	
	Wordt er gevraagd naar verzamelde opdrachten en verslagen in het portfolio?	<input type="radio"/> Ja <input type="radio"/> Nee
7a. Opmerkingen:		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
b.	<i>Kennistoets</i>	
	Wordt de toepassing van kennis getoetst los van de context?	<input type="radio"/> Ja <input type="radio"/> Nee
7b. Opmerkingen:		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
c.	<i>Beoordelingsverslagen</i>	
	Is de toets een verslag dat gemaakt moet worden?	<input type="radio"/> Ja <input type="radio"/> Nee
	Wordt er rekening gehouden met de volgende punten?	<input type="radio"/> Authenticiteit <input type="radio"/> Actualiteitswaarde <input type="radio"/> Relevantie <input type="radio"/> Kwantiteit <input type="radio"/> Variatie

7c. Opmerkingen:

d.	<i>Schriftelijke beoordeling volledige handeling</i>	
	Wordt er gevraagd om handelingen uit de beroepspraktijk op papier weer te geven?	<input type="radio"/> Ja <input type="radio"/> Nee

7d. Opmerkingen:

e.	<i>Criterium gericht interview</i>	
	Wordt de competentie van de student bevraagd in een gesprek?	<input type="radio"/> Ja <input type="radio"/> Nee
	Wordt dit gesprek ingericht met het STAR model (situatie, taken, actie, resultaten)	<input type="radio"/> Ja <input type="radio"/> Nee

7^e. Opmerkingen:

f.	<i>Portfolio beoordeling</i>	
	Wordt het portfolio zoals verzameld ook beoordeeld?	<input type="radio"/> Ja <input type="radio"/> Nee
7f. Opmerkingen:		
g.	<i>Simulatie praktijk</i>	
	Is de toets een simulatie van de beroepspraktijk?	<input type="radio"/> Ja <input type="radio"/> Nee
7g. Opmerkingen:		
h.	<i>Authentieke beoordeling</i>	
	Wordt de competentie getoetst in de beroepspraktijk?	<input type="radio"/> Ja <input type="radio"/> Nee
7h. Opmerkingen:		

8.	Wordt het assessment vooraf gegaan door een authentieke instructie? (alleen bij statische toetsen)	<input type="radio"/> Ja <input type="radio"/> Nee
8. Opmerkingen: <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>		
9.	Kunnen studenten hun competentie laten zien door een waardevolle en betekenisvolle prestatie?	<input type="radio"/> Ja <input type="radio"/> Nee
9. Opmerkingen: <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>		
10.	Is de mening van studenten belangrijk met betrekking tot de invloed van het assessment op leren?	<input type="radio"/> Ja <input type="radio"/> Nee
10. Opmerkingen: <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>		
11.	Is de taal betekenisvol en relevant? (de opdracht moet zoveel mogelijk identiek zijn aan de opdracht die in de praktijk uitgevoerd moet worden en er moet rekening gehouden worden met de integratie van kennis, vaardigheden en attitude)	<input type="radio"/> Ja <input type="radio"/> Nee

11. Opmerkingen:

12.	Reflecteert de omgeving waarin de praktijkopdracht uitgevoerd wordt hoe kennis, vaardigheden en attitude gebruikt worden in de praktijk?	<input type="radio"/> Ja <input type="radio"/> Nee
-----	--	---

12. Opmerkingen:

13.	Welke sociale processen spelen een rol tijdens de uitvoer?	<input type="radio"/> Samenwerken <input type="radio"/> Overleggen <input type="radio"/> Vergaderen <input type="radio"/> Beslissingen maken <input type="radio"/> Overig namelijk <input type="radio"/> _____ <input type="radio"/> _____ <input type="radio"/> _____ <input type="radio"/> _____
-----	--	--

13. Opmerkingen:

14.	Is het resultaat van de toets een product wat later in de beroepspraktijk ook van een student gevraagd kan worden?	<input type="radio"/> Ja <input type="radio"/> Nee
-----	--	---

14. Opmerkingen:

COGNITIEVE COMPLEXITEIT

15.	Speelt de toets in op de verschillende denkprocessen die een student doorloopt?	<input type="radio"/> Tips/ Aanwijzingen in de vorm van feedback <input type="radio"/> Extra informatie (taalwijzer) <input type="radio"/> Hardop denken stimuleren
-----	---	---

15. Opmerkingen:

16.	Houdt de beoordelaar rekening met de verschillende denkprocessen die een student doorloopt?	
	Staat hier in het beoordelingsmodel iets over geschreven?	<input type="radio"/> Ja <input type="radio"/> Nee
	Staat er bij de beoordelingscriteria dat de docent moet aangeven hoe een student de volgende keer het beter kan doen en welke stappen hij dan moet doorlopen?	<input type="radio"/> Ja <input type="radio"/> Nee
16. Opmerkingen:		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
VERGELIJKBAARHEID		
17.	Zijn er antwoordmodellen aanwezig voor de toetsen die ervoor zorgen dat iedere student gelijk wordt beoordeeld?	<input type="radio"/> Ja <input type="radio"/> Nee
17. Opmerkingen:		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
<hr/>		
18.	Worden de werkomstandigheden voor de studenten beschreven?	<input type="radio"/> Ja <input type="radio"/> Nee

18. Opmerkingen:

GESCHIKTHEID VOOR ONDERWIJSDOELEN

19.	Komen de doelen die gesteld worden in het raamwerk voor moderne vreemde talen terug in de toetsen?	
	Doelen:	
	-	<input type="radio"/> Ja <input type="radio"/> Nee
	-	<input type="radio"/> Ja <input type="radio"/> Nee

19. Opmerkingen:

20.	Hoe is de verhouding kennis, vaardigheden en attitude binnen deze toets? Zijn ze ook geïntegreerd?	<input type="radio"/> Kennis <input type="radio"/> Vaardigheden <input type="radio"/> Attitude <input type="radio"/> Geïntegreerd? <input type="radio"/> Ja <input type="radio"/> Nee
-----	--	--

20. Opmerkingen:

ONTWIKKELING VAN ZELFSTUREND LEREN

21.	Op welke manier wordt er bij de toetsen feedback gegeven aan studenten?	<input type="radio"/> Mondeling <input type="radio"/> Schriftelijk
-----	---	---

21. Opmerkingen:

22.	Stimuleert de gegeven feedback zelfsturend leren?	<input type="radio"/> Ja <input type="radio"/> Nee
-----	---	---

22. Opmerkingen:

23.	Stimuleren de toetsen het doen van reflectie?	<input type="radio"/> Ja <input type="radio"/> Nee
-----	---	---

23. Opmerkingen:

BETEKENISVOLHEID

24.	Is er in de toetsen rekening gehouden met het referentiekader van de studenten? Deze zorgt er namelijk voor dat de toetsen meer betekenis krijgen.	<input type="radio"/> Ja <input type="radio"/> Nee
-----	--	---

24. Opmerkingen:

HERHAALBAARHEID

25.	Zijn de toetsen betrouwbaar? (kan ik dit meten?)	<input type="radio"/> Ja <input type="radio"/> Nee
-----	--	---

25. Opmerkingen:

Bijlage 5 Overeenstemmingpercentages objectiviteitmeting beoordelingsinstrument

Item	kwaliteitscriteria	Overeenstemmingspercentage Toets 1	Overeenstemmingspercentage Toets 4	Overeenstemmingspercentage Toets 5	Gemiddeld overeenstemmingspercentage
1	Algemeen	100	100	100	100
2	Algemeen	100	100	100	100
3	Algemeen	100	100	100	100
4	Algemeen	100	100	100	100
5	Algemeen	100	100	100	100
6	Algemeen	100	100	100	100
7a	Authenticiteit	100	100	100	100
7b	Authenticiteit	100	75	100	91,67
7c	Authenticiteit	100	100	100	100
7d	Authenticiteit	100	100	100	100
7 ^e	Authenticiteit	75	100	100	91,67
7f	Authenticiteit	100	100	100	100
7g	Authenticiteit	100	100	100	100
7h	Authenticiteit	100	100	100	100
8	Authenticiteit	100	100	100	100
9	Authenticiteit	100	100	0	66,67
10	Authenticiteit	100	100	100	100
11	Authenticiteit	75	75	0	50
12	Authenticiteit	75	50	100	75
13	authenticiteit	0	0	75	25
14	authenticiteit	75	50	75	66,67
15	Cognitieve complexiteit	100	100	100	100
16a	Cognitieve complexiteit	75	100	100	91,67
16b	Cognitieve complexiteit	100	100	100	100
17	Vergelijkbaarheid	100	100	100	100
18	Vergelijkbaarheid	100	100	100	100
19	Geschiktheid voor onderwijsdoelen	100	100	100	100
20	Geschiktheid voor onderwijsdoelen	100	75	100	91,67
21	Ontwikkeling van zelfsturend leren	100	0	100	66,67
22	Ontwikkeling van zelfsturend leren	50	100	100	83,33
23	Ontwikkeling van zelfsturend leren	100	100	100	100
24	Betekenisvolheid	75	100	100	91,67
25	Herhaalbaarheid	100	100	100	100
	Totaal	88,89%	83,33%	90,74%	88,58%

Bijlage 6 Resultaten vragenlijst studenten

criterium	M	SD	Min	Max
1. Accepteerbaarheid	2,87	2,84	1	6
- item 8	3,33	1,27	1	6
- item 23	2,41	1,57	1	6
2. Cognitieve complexiteit/ ontwikkeling van zelfsturend leren en de effecten op het leren van studenten			1	5,8
- item 17	2,56	1,47	1	6
- item 19	4,11	1,38	1	6
- item 20	3,03	1,18	1	5
- item 21	2,55	1,33	1	6
- item 22	2,95	1,65	1	6
3. Tijd en kosten				
- item 9	3,19	1,53	1	6
4. Onderwijskundige gevolgen	2,38	1,45	1	6
- item 6	4	1,89	1	6
- item 7.1	2,13	1,36	1	6
- item 7.2	1,88	1,24	1	6
- item 7.3	1,86	1,33	1	6
- item 7.4	2,02	1,35	1	6
- item 7.5	2,36	1,53	1	6
5. Eerlijkheid				
- item 14	3	1,58	1	6
- item 18	4,83	1,26	1	6
6. Transparantie	3,65	1,46	1,14	6
- item 5	3,89	1,33	1	6
- item 10	4,05	1,43	1	6
- item 11	4,21	1,31	2	6
- item 12	3,33	1,40	1	6
- item 13	3,11	1,52	1	6
- item 15	4,03	1,80	1	6
- item 16	2,94	1,44	1	6

Bijlage 7 Tabel uitkomst factoranalyse

Componenten Matrix								
	Componenten							
	1	2	3	4	5	6	7	8
Item 5 transparantie			.623					
Item 6 onderwijskundige gevolgen								.593
Item 7.1 onderwijskundige gevolgen	.583				-.413			
Item 7.2 Onderwijskundige gevolgen	.595							
Item 7.3 Onderwijskundige gevolgen	.657						-.475	
Item 7.4 Onderwijskundige gevolgen	.532	-.436						
Item 7.5 Onderwijskundige gevolgen	.451							
Item 8 Accepteerbaarheid	.636							
Item 9 Tijd en kosten			.487	.406				
Item 10 Transparantie		.778						
Item 11 Transparantie		.729						
Item 12 Transparantie	.712							
Item 13 Transparantie	.649							
Item 14 Eerlijkheid	.515	.413	-.442					
Item 15 Transparantie	.491			-.536				
Item 16 Transparantie					.494			
Item 17 Cognitieve complexiteit/ Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten	.449							-.403
Item 18 Eerlijkheid				-.682				
Item 19 Cognitieve complexiteit/ Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten	.426	-.459						
Item 20 Cognitieve complexiteit/ Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten	.616							
Item 21 Cognitieve complexiteit/ Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten							-.646	
Item 22 Cognitieve complexiteit/ Ontwikkeling van zelfsturend leren en de effecten op het leren van studenten						-.597		
Item 23 Accepteerbaarheid	.499				-.408			