

**Praktijkbeoordelingen bij godsdienstlerarenopleiding:
Welke percepties zijn er over de edumetrische kwaliteit?**

Janneke Catharina (Rianne) Annot

3112071

Universiteit Utrecht

Masterthesis Onderwijskundig Ontwerp en Advisering

Opdrachtgever: Jan Marten Praamsma, Christelijke Hogeschool Ede

Eerste beoordelaar: Havva Jongen

Tweede beoordelaar: Marieke van der Schaaf

Datum: augustus 2010

Inhoudsopgave

Samenvatting	2
1. Inleiding	3
1.1 Assessmentcultuur	3
1.2 Praktijkbeoordeling.....	3
1.3 Kwaliteitscriteria.....	5
1.4 Competentie assessment wiel.....	7
Startcriterium	7
Voorwaardelijke criteria	8
Aanvullende criteria	8
Omgevende criteria.....	8
1.5 Kwaliteit van praktijkbeoordelingen.....	9
2. Methode.....	11
2.1 Deelnemers	11
2.2 Instrumenten	12
2.3 Design	13
2.4 Procedure	13
Dataverzameling voorbereiden.....	14
Data verzamelen	14
Data verwerken.....	14
2.5 Analyse	15
3. Resultaten	16
3.1 Perceptie van GL-docenten.....	16
Perceptie startcriterium.....	16
Perceptie voorwaardelijke criteria	17
Perceptie aanvullende criteria.....	18
Perceptie omgevende criteria.....	20
Perceptie totale kwaliteit.....	21
3.2 Perceptie van studenten.....	24
Perceptie startcriterium.....	24
Perceptie voorwaardelijke criteria	25
Perceptie aanvullende criteria.....	27
Perceptie omgevende criteria.....	28
Perceptie totale kwaliteit.....	29
3.3 Perceptie van stagebegeleiders.....	31
3.4 Overeenkomsten en verschillen van percepties van GL-docenten en studenten.....	31
Belangrijkheid: overeenkomsten en verschillen	32
Toepasbaarheid: overeenkomsten en verschillen	33
Tevredenheid: overeenkomsten en verschillen.....	34
4. Conclusie	34
4.1 Perceptie van GL-docenten.....	34
4.2 Perceptie van studenten.....	35
4.3 Overeenkomsten en verschillen percepties docenten en studenten.....	36
4.4 Verwachtingen	36
5. Discussie.....	37
5.1 Deelnemers	37
5.2 Kwaliteitsmeter	37
5.3 Praktijkbeoordelingen	38
Dankbetuiging	39
Referenties.....	39
Bijlage 1. <i>Overzicht kwaliteitscriteria en indicatoren</i>	Error! Bookmark not defined.
Bijlage 2. <i>Kwaliteitsmeter</i>	Error! Bookmark not defined.

Praktijkbeoordelingen bij godsdienstlerarenopleiding: Welke percepties zijn er over de edumetrische kwaliteit?

J.C. Annot

Samenvatting

Door middel van praktijkbeoordelingen bij de lerarenopleiding Godsdienst / Levensbeschouwing aan de Christelijke Hogeschool Ede worden elk jaar de bekwaamheden van de studenten vastgesteld. Deze beoordelingen zijn niet eerder beoordeeld op kwaliteit. Een oorzaak hiervan is dat de praktijkbeoordelingen tot de nieuwe assessmentvormen behoren waarbij andere kwaliteitscriteria gelden, namelijk edumetrische, dan bij oude assessmentvormen, zoals schriftelijke toetsen. Door middel van dit onderzoek is de edumetrische kwaliteit van de praktijkbeoordelingen vastgesteld aan de hand van de twaalf kwaliteitscriteria van Baartman (2008). De docenten en studenten die betrokken zijn bij de genoemde praktijkbeoordelingen, hebben door middel van het invullen van een kwaliteitsmeter duidelijk gemaakt wat hun perceptie is over de kwaliteit van de praktijkbeoordeling. Hieruit bleek dat alle kwaliteitscriteria belangrijk en toepasbaar zijn bij de praktijkbeoordelingen. Ook werd duidelijk dat de docenten en studenten gemiddeld tevreden zijn over de kwaliteit van de praktijkbeoordelingen.

Trefwoorden: assessment, competentie, godsdienstleraar, kwaliteit, kwaliteitsmeter, perceptie, praktijkbeoordeling

1. Inleiding

1.1 Assessmentcultuur

Het onderwijs is de laatste decennia gekenmerkt door de verschuiving van testcultuur naar assessmentcultuur (Birenbaum, 2003; Falchikov, 2005; Gulikers, Bastiaens, Kirschner & Kester, 2006; Sambell, McDowell & Brown, 1997). De testcultuur kenmerkte zich door schriftelijke toetsen, focus op product en scheiding tussen kennis en kunde (Falchikov, 2005). De assessmentcultuur biedt kennis en kunde tegelijk aan en heeft als doelen: integratie van assessment, onderwijs en leren; authentieke assessmentvormen; actieve studenten; en focus op zowel proces als product (Sambell et al., 1997). Met deze doelen is gestreefd naar een betere aansluiting van het onderwijs op de arbeidsmarkt (Birenbaum, 2003).

De huidige assessmentcultuur in het onderwijs kent een grote variëteit in assessmentvormen waarmee de competenties van studenten, bestaande uit kennis, vaardigheden en houdingen, in kaart gebracht kunnen worden (Falchikov, 2005). Met een assessment wordt de geschiktheid van een persoon voor een bepaalde taak of functie op een authentieke manier vastgesteld (Falchikov, 2005). Voorbeelden van assessmentvormen zijn: portfolio, presentatie, simulatie, self-assessment, peer-assessment en co-assessment (Dochy & Struyven, 2002). De laatst genoemde vormen geven aan dat de beoordeling met behulp van een assessmentvorm niet altijd door een docent gedaan wordt, maar ook door medestudenten of studenten zélf uitgevoerd kan worden. De authenticiteit van een assessment is afhankelijk van de taak, de fysieke context, de sociale context, de methode, en de resultaatcriteria (Gulikers et al., 2006).

1.2 Praktijkbeoordeling

Praktijkbeoordelingen bij de lerarenopleiding Godsdienst / Levensbeschouwing [GL] aan de Christelijke Hogeschool in Ede [CHE] worden gerekend tot de authentieke assessmentvormen (Van Peer & Van Beek, 2009). De praktijkbeoordelingen zijn gekoppeld aan stages die de studenten elk jaar uitvoeren op een middelbare school. Tijdens deze stages observeren de studenten lessen van hun stagebegeleider. Ook geven de studenten een aantal lessen en schrijven ze een verslag. Elk volgend

leerjaar is de stage ingewikkelder doordat meer van de student verwacht wordt op het gebied van zelfstandigheid en lesgeven. Alle stageperiodes worden afgesloten met een praktijkbeoordeling. De stages in het tweede en derde studiejaar kennen ook een tussenbeoordeling. De eisen met betrekking tot de stages en praktijkbeoordelingen staan in de *Pocket praktijkstage GL* (Van Peer en Van Beek, 2009). Het doel van de praktijkbeoordelingen is het komen tot een goed oordeel over de onderwijsbekwaamheid van de studenten. De tussenbeoordelingen zijn formatief van aard terwijl de eindbeoordelingen zowel een formatieve als summatieve functie hebben. De gewenste bekwaamheid wordt, zowel in de opdracht als in de beoordeling, beschreven aan de hand van 4 kerntaken en de competenties voor onderwijsgevend. Tabel 1 laat zien hoe de samenhang tussen de competenties en kerntaken is vastgesteld door Van Peer en Van Beek (2009). De bekwaamheden van de kerntaken worden afgelezen aan de hand van de beoordelingen van de competenties.

Tabel 1. *Competenties in samenhang met kerntaken van GL* (Van Peer & Van Beek, 2009, p.6)

Competenties	Kerntaken			
	Onderwijsleer- processen verzorgen	Leerling- begeleiding	Communiceren binnen en buiten school	Nascholing en beziging
1. interpersoonlijk	X			
2. pedagogisch	X	X		
3a. vakdidactisch	X			
3b. vakinhoudelijk	X			
4. organisatorisch	X			
5. omgang met collega's		X	X	
6. omgang met omgeving		X	X	
7. reflectie en ontwikkeling	X			X

Elke praktijkbeoordeling komt tot stand met inbreng van de student en de stagebegeleider (Van Peer & Van Beek, 2009). Een docent van de CHE geeft het eindoordeel, voornamelijk gebaseerd op de inbreng van de student en stagebegeleider. De beoordelingen worden gegeven aan de hand van beoordelingsformulieren. Op deze formulieren staan de competenties genoemd met daarbij indicatoren. Deze indicatoren verschillen per stagejaar, zodat het bekwaamheidsniveau per competentie aangepast is aan het niveau dat in de stage aan bod komt. Per indicator wordt, door de stagebegeleider en student, op het formulier aangegeven in welke mate de student bekwaam is. De docent van de CHE geeft aan de hand hiervan een totaaloordeel. De formatieve beoordelingen geven

weer waar de student al voldoende bekwaam in is en bij welke competentie en kerntaken nog verbetering vereist is. De summatieve beoordeling geeft aan of de student wel of niet bekwaam is om door te gaan naar de volgende stage. Indien de student beoordeeld wordt als zijnde onvoldoende bekwaam, doet de docent van de CHE een voorstel tot herkansing aan de examencommissie. Deze commissie beslist uiteindelijk welke herkansingsmaatregel uitgevoerd moet worden door de student. Over de betrouwbaarheid en validiteit van de praktijkbeoordelingen zijn geen gegevens beschikbaar, omdat de kwaliteit niet beoordeeld is.

1.3 Kwaliteitscriteria

De kwaliteit van assessmentvormen is lastig vast te stellen (Dochy & Struyven, 2002; Gipps, 1994). Dit omdat de assessmentcultuur en het competentiegerichte onderwijs andere eisen stellen aan kwaliteit dan de testcultuur deed. In de testcultuur werd kwaliteit psychometrisch bepaald door betrouwbaarheid en validiteit. Bij betrouwbaarheid ging het om herhaalbaarheid en vergelijkbaarheid van de toetsen (Baartman, et al., 2007; Baartman, 2008). Validiteit ging over: wordt er gemeten wat er gemeten moet worden (Sambell et al., 1997). Tegenwoordig is validiteit ook gedefinieerd als: het testen van de hypothese dat er een causaal verband is tussen het te meten construct en de resultaten van de beoordeling (Borsboom & Mellenbergh, 2007).

In de wetenschappelijke literatuur is een discussie gevoerd over de inhoud en reikwijdte van validiteit (Baartman, 2008; Borsboom & Mellenbergh, 2004, 2007; Linn, Baker & Dunbar, 1991; Messick, 1995). Linn en collega's (1991) maakten onderscheid tussen acht dimensies van validiteit, namelijk: consequenties, eerlijkheid, transfer, inhoud, cognitieve complexiteit, betekenisvolheid, begrijpelijkheid, en kosten en efficiëntie. Kosten en efficiëntie zijn onderdeel van validiteit omdat het daarbij gaat om de toepasbaarheid en uitvoerbaarheid van het geheel. Messick (1992, 1995) ging echter uit van constructvaliditeit waarmee het onderliggende construct gemeten wordt en hield rekening met ongewenste invloed van andere variabelen (Borsboom & Mellenbergh, 2004). Daarbij schaarde Messick (1992, 1995) zes aspecten onder constructvaliditeit, namelijk: inhoud, substantieel, structureel, extern, generaliseerbaar en consequentieel (Messick, 1995). Construct validiteit meet of er echt competenties en complete leerdoelen gemeten worden en is daarom belangrijk bij het vaststellen

van kwaliteit bij assessmentvormen. Er kunnen echter ook irrelevante constructen gemeten worden, dan is er sprake van constructirrelevante variatie (Messick, 1995). Ook kan het voorkomen dat de assessment slechts een deel van een competentie meet, dan is er sprake van construct ondervertegenwoordiging. Baartman (2008) voegde de dimensies en aspecten van Linn en collega's (1991) en Messick (1995) samen en liet zien dat deze dimensies en aspecten vrijwel met elkaar overeenkomen, zie tabel 2.

De huidige assessmentvormen zijn gericht op het meten van kennis, vaardigheden en houdingen, ook wel competenties genoemd (Diederick & Dochy, 2001). Hierbij zijn en blijven betrouwbaarheid en validiteit belangrijk, maar zijn niet meer voldoende om de kwaliteit van assessments vast te stellen (Borsboom & Mellenbergh, 2004). Nieuwe, edumetrische, criteria zoals transparantie, impact en volledigheid zijn toegevoegd aan de oorspronkelijke, psychometrische, kwaliteitscriteria om de kwaliteit van assessments vast te stellen (Linn et al., 1991). Hiermee wordt overgegaan van psychometrische kwaliteit naar edumetrische kwaliteit (Baartman, 2008).

Baartman (2008) gaf in haar proefschrift een nieuw en compleet raamwerk om de kwaliteit van assessments in competentiegericht onderwijs te beoordelen. Dit raamwerk, het competentie assessment wiel, is ontworpen mede aan de hand van de zes construct validiteitaspecten van Messick (1995) en de acht validiteitdimensies van Linn en collega's (1991), zie tabel 2. De betrouwbaarheid is terug te vinden in de criteria vergelijkbaarheid en herhaalbaarheid (Baartman 2008).

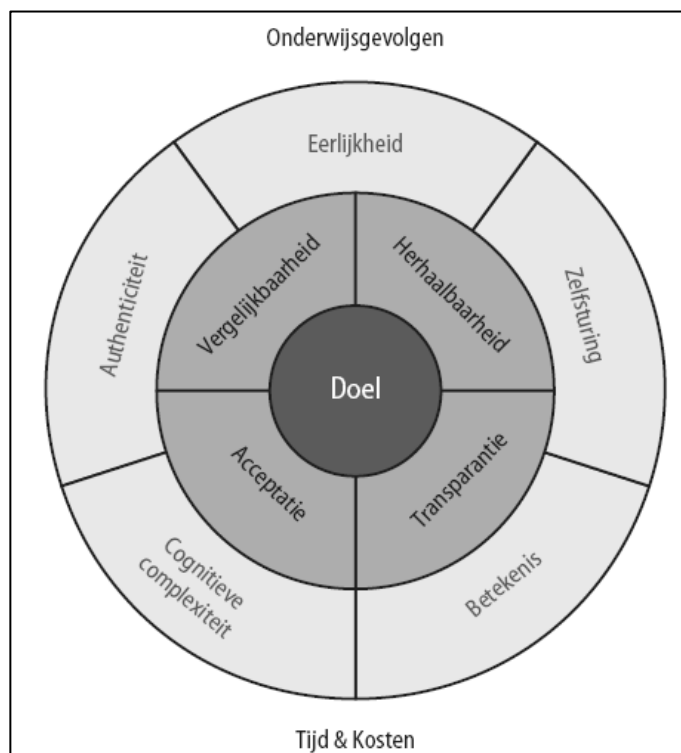
Tabel 2. *Vergelijking van competentie assessment wiel met validiteitdimensies en -aspecten*

Competentie assessment wiel (Baartman, 2008)	Validiteitdimensies (Linn et al, 1991)	Constructvaliditeitaspecten (Messick, 1995)
Onderwijsgevolgen	X	X
Eerlijkheid	X	X
Zelfsturing		
Betekenis	X	X
Cognitieve complexiteit	X	X
Authenticiteit		X
Vergelijkbaarheid	X	X
Herhaalbaarheid	X	X
Transparantie	X	X
Acceptatie		
Doel		
Kosten en efficiëntie	X	

De kwaliteitscriteria *zelfsturing*, *acceptatie* en *doel* zijn aan het competentie assessment wiel toegevoegd, omdat het belangrijke componenten zijn van het competentiegerichte onderwijs (Baartman, 2008). In paragraaf 1.4 wordt uitgelegd hoe het competentie assessment wiel opgebouwd is en wat de inhoud is van de verschillende kwaliteitscriteria.

1.4 Competentie assessment wiel

Het competentie assessment wiel in figuur 1 laat zien hoe de 12 kwaliteitscriteria onlosmakelijk met elkaar verbonden zijn (Baartman, Prins, Kirschner & Van der Vleuten, 2007). In het midden van het competentie assessment wiel staat het startcriterium *doel*. De binnenste cirkel van figuur 1 toont voorwaardelijke basiscriteria voor de aanvullende kwaliteitscriteria in de buitenste cirkel. De omgeving om het wiel toont de omgevende criteria. Door met het complete wiel rekening te houden, is het volgens Baartman (2008) mogelijk de kwaliteit van een assessment vast te stellen.



Figuur 1. Competentie Assessment wiel (Baartman et al, 2007, p.21)

Startcriterium

Het Competentie assessment wiel van Baartman (2008) heeft als startcriterium *geschikt voor onderwijsdoelen*. Dit wordt afgekort als *doel*. Bij het startcriterium gaat het om de afstemming van de

doelen van het onderwijs op de gebruikte assessmentvormen. Bij competentiegericht onderwijs is het belangrijk dat zowel kennis als vaardigheden en houdingen in de onderwijsdoelen en assessmentvormen naar voren komen (Sambell et al., 1997).

Voorwaardelijke criteria

De voorwaardelijke criteria van het competentie assessment wiel, zijn *herhaalbaarheid*, *transparantie*, *acceptatie* en *vergelijkbaarheid* (Baartman, 2008). Bij *herhaalbaarheid* is het belangrijk dat de oordelen over studenten gebaseerd zijn op meerdere verschillende bewijzen en beoordelaars. Het criterium *transparantie* gaat over de duidelijkheid en begrijpelijkheid van de assessmentvormen voor alle betrokkenen. Bij criterium *acceptatie* gaat het over de mate waarin de betrokkenen zich kunnen vinden in de opzet en uitvoering van de assessmentvormen. Met het criterium *vergelijkbaarheid* wordt rekening gehouden met de consistente opzet en uitvoering van de assessmentvormen.

Aanvullende criteria

Eerlijkheid, *ontwikkeling van zelfsturend leren*, *betekenisvolheid*, *cognitieve complexiteit* en *authenticiteit* vormen samen de vijf aanvullende criteria van het competentie assessment wiel (Baartman, 2008). Door middel van het criterium *eerlijkheid* wordt gemeten of alle studenten een eerlijke kans krijgen om hun competenties te tonen. De *ontwikkeling van zelfsturend leren*, of *zelfsturing*, gaat over de stimulans tot zelfreflectie en het formuleren van eigen leerdoelen door de studenten. Het criterium *betekenisvolheid* gaat over de feedback en leerervaring die de assessmentvormen met zich meebrengen. Bij *cognitieve complexiteit* wordt gemeten of de assessmentvorm niet alleen gedrag, maar ook onderliggende kennis test. Ook wordt daarbij op zowel het proces als het product van de assessmentvorm gelet. Het criterium *authenticiteit* houdt rekening met de mate waarin de assessmentvorm lijkt op de beroepspraktijk.

Omgevende criteria

De omgevende criteria, *onderwijsgevolgen* en *tijd en kosten*, zijn de factoren die op de achtergrond meespelen bij het beoordelen van competenties. Het criterium *onderwijsgevolgen* is belangrijk, omdat niet alle assessmentvormen de gewenste positieve invloed hebben op leren en instructie.

Het criterium *tijd en kosten* gaat om de praktische mogelijkheid tot uitvoering van de assessmentvormen.

1.5 Kwaliteit van praktijkbeoordelingen

Het competentie assessment wiel van Baartman (2008) is stevig gefundeerd in relevante wetenschappelijke theorieën en onderzoeken en biedt een goede mogelijkheid om de edumetrische kwaliteit van nieuwe assessmentvormen vast te stellen. Hierbij is geen onderscheid te maken tussen psychometrische en edumetrische kwaliteit, omdat de definities van het competentie assessment wiel gericht zijn op de edumetrie (Baartman, 2008). Aangezien assessment de hoeksteen van het onderwijs is, dienen assessmentvormen van goede kwaliteit te zijn en aan te sluiten bij de rest van het onderwijs (Diederick & Dochy, 2001). De kwaliteit, in het bijzonder de edumetrische kwaliteit, van de praktijkbeoordelingen bij GL is niet eerder onderzocht. Ook is niet vastgesteld of de docenten van de CHE het met elkaar eens zijn over de praktijkbeoordelingen bij GL. Voor de accreditatie van de opleiding is het echter van groot belang dat duidelijk is wat de edumetrische kwaliteit van de praktijkbeoordelingen is.

Volgens Baartman en collega's (2007) is het wetenschappelijk relevant de kwaliteit van een assessment vast te laten stellen vanuit verschillende perspectieven. De percepties van de diverse betrokkenen, zoals docenten en studenten kunnen, vanwege het verschil van perspectief en functie, belangrijke verschillen en overeenkomsten vertonen. Door wetenschappelijk te onderzoeken wat de percepties van docenten, studenten en stagebegeleiders zijn over de praktijkbeoordelingen bij GL, kan vastgesteld worden in hoeverre de percepties van de docenten, studenten en stagebegeleiders met elkaar in overeenstemming zijn en wat de edumetrische kwaliteit is van de praktijkbeoordelingen bij GL.

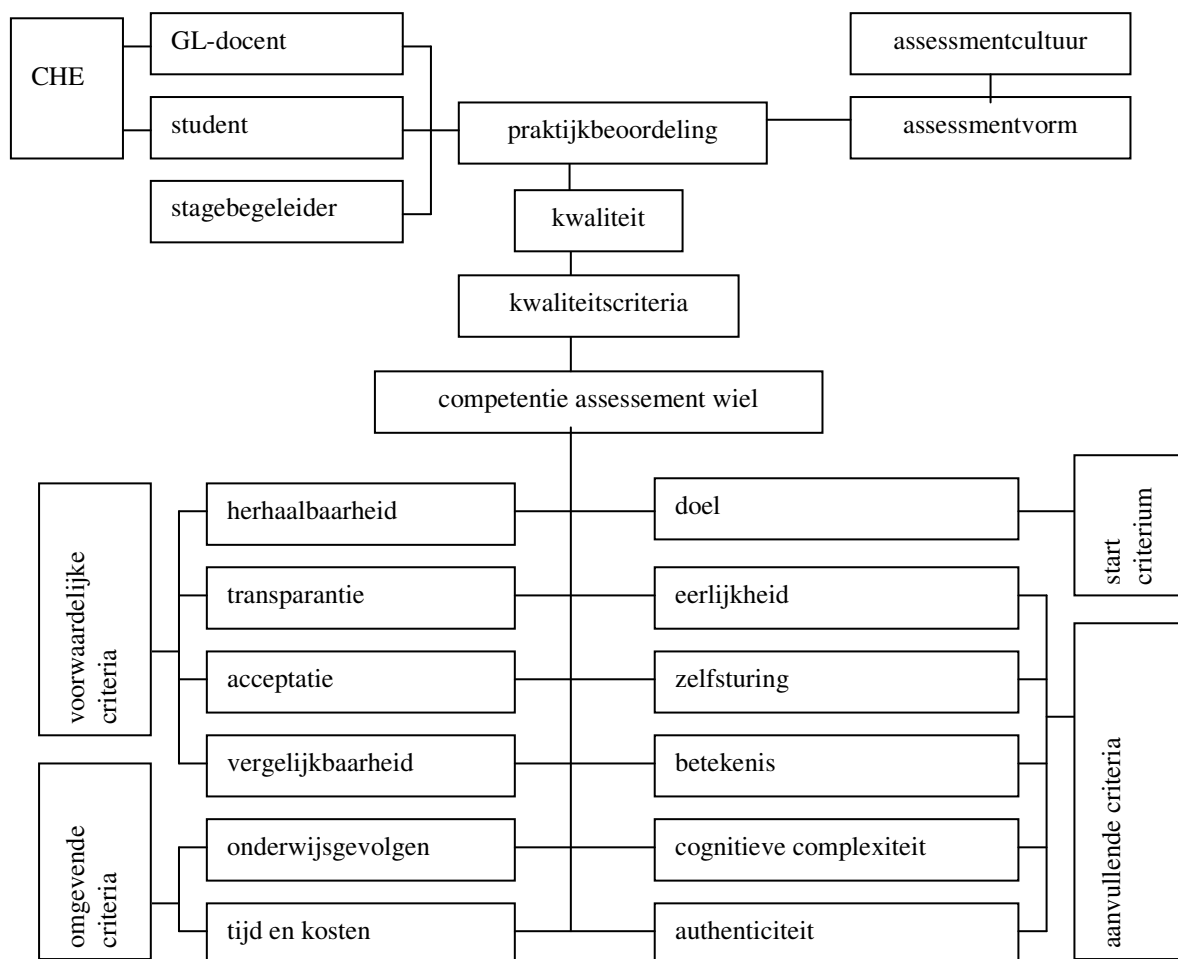
Hieruit volgt het wetenschappelijke en maatschappelijk relevante doel van dit onderzoek: het vaststellen van de edumetrische kwaliteit van de praktijkbeoordelingen van GL aan de hand van de percepties van docenten, studenten en stagebegeleiders. De centrale vraag is dan ook: *Wat is volgens docenten, studenten en stagebegeleiders de edumetrische kwaliteit van de praktijkbeoordelingen bij de*

lerarenopleiding Godsdienst / Levensbeschouwing op de Christelijke Hogeschool Ede? Het antwoord op deze centrale vraag wordt afgeleid van de antwoorden op de deelvragen. De deelvragen zijn:

1. *Wat is de perceptie van docenten van de edumetrische kwaliteit van de praktijkbeoordelingen?*
2. *Wat is de perceptie van studenten van edumetrische kwaliteit van de praktijkbeoordelingen?*
3. *Wat is de perceptie van stagebegeleiders van de edumetrische kwaliteit van de praktijkbeoordelingen?*
4. *In hoeverre komen de percepties van docenten, studenten en stagebegeleiders met elkaar overeen en wat zijn de belangrijkste verschillen?*

De docenten van de CHE hebben mondeling enkele verwachtingen weergegeven. Deze docenten waren betrokken bij de ontwikkeling van de praktijkbeoordelingen en hebben daarom naar hun eigen verwachting een eenduidig en overwegend positief oordeel over de kwaliteit van praktijkbeoordelingen. Baartman (2008) gaf aan dat docenten alle criteria van het competentie assessment wiel belangrijk vinden. De verwachting is dat in dit onderzoek de docenten ook alle criteria belangrijk vinden. Door gebrek aan informatie is het niet mogelijk verwachtingen op te stellen van de afzonderlijke kwaliteitscriteriagroepen. De docenten van de CHE spraken de verwachting uit dat studenten en stagebegeleiders ook positief zullen oordelen over de kwaliteit van de praktijkbeoordelingen, omdat er weinig klachten geuit worden over de beoordelingen.

Het conceptuele model laat de samenhang zien tussen alle relevante begrippen van dit onderzoek, zie figuur 2. De praktijkbeoordeling is een assessmentvorm uit de assessmentcultuur. Deze beoordeling komt tot stand door inbreng van docenten en studenten van de opleiding GL van de CHE en door stagebegeleiders van de stagescholen. De kwaliteit van de praktijkbeoordeling wordt in dit onderzoek vastgesteld met behulp van de kwaliteitscriteria van het competentie assessment wiel. Deze kwaliteitscriteria zijn gegroepeerd als startcriterium, voorwaardelijke criteria, aanvullende criteria en omgevende criteria. Zo hangen alle begrippen uit figuur 2 met elkaar samen.



Figuur 2. Conceptueel model

2. Methode

2.1 Deelnemers

De deelnemers van dit onderzoek zijn allen betrokken bij de praktijkbeoordelingen van GL op de CHE. Het betreft de 5 betrokken docenten van de CHE, de hele populatie, deelnemersgroep 1. Deelnemersgroep 2 bestaat uit de tweedejaars studenten van de lerarenopleiding GL van de CHE, 27 personen in totaal, dit is ongeveer 25% van de gehele GL-studenten populatie. Door praktische overwegingen konden er niet meer studenten betrokken worden bij dit onderzoek. De studenten hebben in februari een tussenbeoordeling ontvangen. In juni kregen zij een eindbeoordeling. Ook zijn 15 stagebegeleiders bij het onderzoek betrokken, de hele populatie, deelnemersgroep 3. Op deze manier zijn alle partijen die betrokken zijn bij de praktijkbeoordeling in het onderzoek meegenomen.

2.2 Instrumenten

Voor het verkrijgen van de benodigde data is gebruik gemaakt van de gevalideerde CAP-kwaliteitsmeter van Baartman (2008), waarbij CAP staat voor Competentie Assessment Programma. Deze kwaliteitsmeter is een uitwerking van het competentie assessment wiel en bevat per kwaliteitscriterium gemiddeld 5 indicatoren, zoals uitgewerkt in bijlage 1 (Baartman, n.d.). De CAP-kwaliteitsmeter van Baartman (2008) is op een paar onderdelen aangepast, zodat de kwaliteitsmeter toepasbaar was bij dit onderzoek. De indicatoren zijn ongewijzigd overgenomen. De term *CAP* is veranderd in *tussentijdse praktijkbeoordeling*, omdat dit duidelijker was voor de respondenten. Het woord *tussentijdse* is toegevoegd, omdat de studenten nog geen uiteindelijke beoordeling hadden ontvangen op het moment van afname van de kwaliteitsmeter. De gewijzigde kwaliteitsmeter is digitaal gemaakt met hulp van de website www.thesistools.nl (Van Rixtel, n.d.). Voor deze website is gekozen omdat het de enige snel toegankelijke en toepasbare manier was om de kwaliteitsmeter digitaal te maken. Met de website van Van Rixtel (n.d.) was het niet mogelijk de schaalverdelingen van Baartman (n.d.) aan te houden. Daarom is de oorspronkelijk 100-puntsschaal met als uitersten *helemaal niet van toepassing* en *helemaal wel van toepassing* gewijzigd in een 10-puntsschaal. Door deze verandering konden de respondenten minder nauwkeurig, maar wel meer bewust aangeven in welke mate de indicator van toepassing is (Baartman, 2008). De vraag of de mate van toepassing voldoende is, kon met ja of nee beantwoord worden, zie figuur 3. Dit is identiek aan de oorspronkelijke CAP-kwaliteitsmeter (Baartman, n.d.). Ook de mogelijkheid om aan te geven of de indicator al dan niet bekend was, is ongewijzigd overgenomen.

In welke mate is dit van toepassing op de tussentijdse praktijkbeoordeling? Vindt u dit voldoende?

	helemaal niet	helemaal wel	ja	nee	onbekend
Alle te bewijzen competenties worden beoordeeld in de tussentijdse praktijkbeoordeling	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In de tussentijdse praktijkbeoordeling wordt voldoende aandacht besteed aan kennis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figuur 3. Voorbeeld van indicatoren en daarbij gestelde vragen van de kwaliteitsmeter

In de CAP-kwaliteitsmeter is bij elke indicator de mogelijkheid gegeven uitleg te geven over de gegeven antwoorden. Deze mogelijkheid is in dit onderzoek beperkt tot het geven van uitleg per criterium. Dit is veranderd om de lengte van de kwaliteitsmeter in te perken en overzichtelijk te houden. Verder werd per criterium gevraagd naar de mate van belangrijkheid van dat criterium voor de tussentijdse praktijkbeoordeling. Deze vraag kon beantwoord worden op een 8-puntsschaal met als uitersten *helemaal niet belangrijk* en *heel belangrijk*. De kwaliteitsmeter werd afgesloten met open vragen naar de sterke en zwakke punten van de praktijkbeoordeling en 8-puntsschaalvragen naar de totale kwaliteit van de praktijkbeoordeling en de moeilijkheid van het evalueren, zie figuur 4. Bijlage 2 bevat de complete gewijzigde versie van de kwaliteitsmeter zoals deze gebruikt is in dit onderzoek.

In hoeverre vindt u de totale kwaliteit van uw praktijkbeoordeling voldoende?

helemaal niet voldoende ○○○○○○○○ helemaal voldoende

Wat vond u van het evalueren van de praktijkbeoordeling?

helemaal niet moeilijk ○○○○○○○○ heel moeilijk

Figuur 4. Afsluitende vragen van de kwaliteitsmeter

2.3 Design

Dit onderzoek is beschrijvend van aard en te typeren als survey-onderzoek (Stokking, 1998). Dit type onderzoek is geschikt omdat het gaat om empirisch onderzoek aan de hand van een steekproef. De vraagstelling van dit onderzoek geeft aan dat er niet direct aanleiding is om causale verbanden te onderzoeken. Ook daarom is een survey-onderzoek de juiste methode. De onderzoeksvraag is beantwoord met hulp van gegevens die verzameld zijn door middel van een vragenlijst die kwaliteitsmeter genoemd wordt.

2.4 Procedure

De procedure van dit onderzoek bestond uit 3 stappen, namelijk:

1. dataverzameling voorbereiden;
2. data verzamelen;
3. data verwerken.

Dataverzameling voorbereiden

Het voorbereiden van de dataverzameling bestond uit het aanpassen van de CAP-kwaliteitsmeter van Baartman (2008), zoals beschreven is in paragraaf 2.2. Ook is tijdens het voorbereiden van de dataverzameling een brief samengesteld met daarin uitleg over het onderzoek en het gebruik van de kwaliteitsmeter. Deze brief bevatte voor elke deelnemer een persoonlijke toegangscode voor de kwaliteitsmeter. Voorafgaand aan de dataverzameling is de brief digitaal verzonden aan alle deelnemers.

Data verzamelen

Nadat de voorbereidingen voor de dataverzameling klaar waren, is de kwaliteitsmeter digitaal afgenomen bij alle deelnemers. Alle deelnemers kregen een persoonlijke mail met daarin uitleg over de kwaliteitsmeter en het onderzoek. Ook zijn er toegangscode's in vermeld. De deelnemers kregen 14 dagen de tijd om de kwaliteitsmeter in te vullen. Om de respons zo hoog mogelijk te maken, zijn er na 7 en 13 dagen e-mails ter herinnering verzonden naar de deelnemers. De antwoorden die gegeven werden op de vragen in de kwaliteitsmeter vormden samen de data voor dit onderzoek.

Vanwege de zeer geringe respons op de digitale afname, is de kwaliteitsmeter naderhand schriftelijk afgenomen bij de studenten en stagebegeleiders. De GL-docenten hebben allemaal de digitale kwaliteitsmeter ingevuld, daarom zijn zij niet schriftelijk benaderd.

Data verwerken

De digitale kwaliteitsmeter is door 5 personen helemaal ingevuld. Deze respondenten zijn allen docent aan de CHE, zij vormen de gehele docentenpopulatie van deelnemersgroep 1.

De schriftelijke kwaliteitsmeter is door 11 van de 27 studenten (40%), deelnemersgroep 2, ingevuld. De stagebegeleiders, deelnemersgroep 3, hebben geen van allen de digitale of schriftelijke versie van de kwaliteitsmeter ingevuld. Enkele stagebegeleiders gaven aan geen tijd en zin te hebben om de uitgebreide kwaliteitsmeter in te vullen. Van andere stagebegeleiders werd geen enkele respons ontvangen. Van deze deelnemersgroep zijn daardoor geen gegevens beschikbaar voor de analyse.

De analyse van de data van deelnemersgroep 1 en 2 is uitgevoerd zoals staat weergegeven in paragraaf 2.5. Door het geringe aantal deelnemers was het niet relevant uitgebreide statistische toetsen uit te voeren.

2.5 Analyse

De gegevens die verkregen zijn door de kwaliteitsmeter, bestond uit kwantitatieve en kwalitatieve gegevens. Van de kwantitatieve gegevens zijn de algemene beschrijvende statistieken berekend, zoals gemiddelde en standaardafwijking. Dit is berekend per indicator, per kwaliteitscriterium en per criteriagroep. Door middel van de genoemde analysemethoden is vastgesteld welke perceptie de docenten en studenten hebben van elk kwaliteitscriterium. De verkregen kwalitatieve gegevens zijn gecategoriseerd per kwaliteitscriterium. De kwalitatieve gegevens zijn gebruikt om de kwantitatieve uitkomsten te onderbouwen.

De gegevens zijn onder meer verzameld aan de hand van 8-puntschalen en 10-puntschalen. Deze schalen hadden als uitersten 1 *helemaal niet* en 8 dan wel 10 *helemaal wel*. De gemiddelden van deze schalen zijn in 4 categorieën verdeeld. Bij de 8-puntsschaal geldt: gemiddeld 1.0-2.5 is *helemaal niet*, 2.6-4.5 is *niet*, 4.6-6.5 is *wel*, 6.6-8.0 is *helemaal wel*. Bij de 10-puntsschaal geldt: gemiddeld 1.0-2.9 is *helemaal niet*, 3.0-5.5 is *niet*, 5.6-7.9 is *wel*, 8.0-10.0 is *helemaal wel*.

Bij de kwaliteitsmeter werd per indicator gevraagd naar de bekendheid van de indicator. De respondenten konden dit aangeven door bij *bekend* al dan niet een kruisje te zetten. Wanneer de respondenten geen kruisje hebben gezet, kan er echter sprake zijn van onbekendheid of missing values. Hierdoor zijn er geen betrouwbare resultaten weer te geven over de bekendheid van indicatoren en kwaliteitscriteria. Hiervan zullen geen resultaten beschreven worden.

De uitkomsten van de verschillende analyses zijn als resultaten gerapporteerd op de volgorde die de vraagstelling aangeeft. Per deelnemersgroep is eerst beschreven hoe de respondenten de kwaliteit van het startcriterium beoordelen. Vervolgens zijn de voorwaardelijke, aanvullende en omgevende criteria aan bod gekomen. Tot slot is de totale perceptie van de kwaliteit van de tussentijdse praktijkbeoordelingen beschreven. Bij alle criteria zijn de standaardgegevens, zoals gemiddelde scores en standaardafwijkingen, weergegeven, evenals relevante resultaten en opvallende uitkomsten. Van

alle criteriagroepen zijn zoveel als mogelijk op identieke wijze de resultaten weergegeven, zodat de structuur duidelijk en overzichtelijk is.

3. Resultaten

3.1 Perceptie van GL-docenten

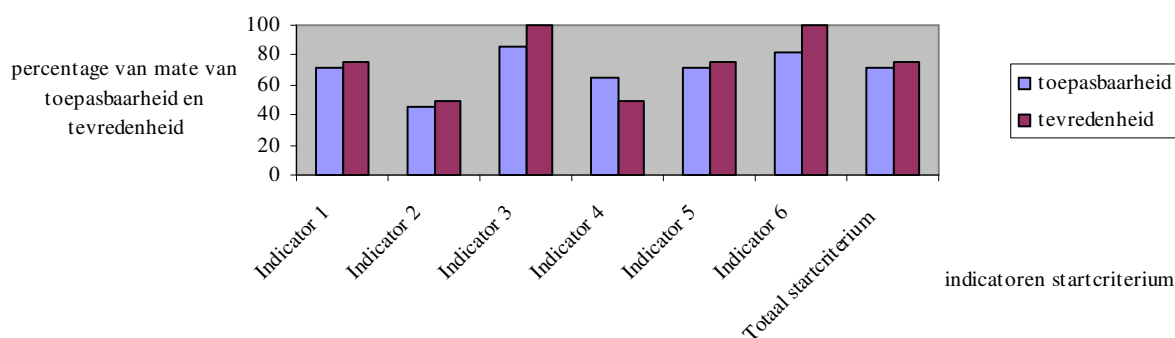
De 5 GL-docenten die betrokken zijn bij de praktijkbeoordelingen, hebben de kwaliteitsmeter van Baartman (2008) volledig ingevuld. Hierbij is per kwaliteitscriterium gevraagd in hoeverre elk criterium belangrijk is bij de praktijkbeoordelingen. Per indicator is gevraagd naar de mate van toepasbaarheid en of dat voldoende dan wel onvoldoende is. Eén GL-docent heeft nergens aangegeven of de indicatoren voldoende of onvoldoende toepasbaar zijn, andere GL-docenten hebben incidenteel geen antwoord gegeven. Hierdoor is sprake van missing values. De resultaten van de ingevulde kwaliteitsmeter zijn beschreven op de volgorde van de kwaliteitsmeter, namelijk:

- Perceptie van startcriterium
- Perceptie van voorwaardelijke criteria
- Perceptie van aanvullende criteria
- Perceptie van omgevende criteria
- Perceptie van totale kwaliteit

Perceptie startcriterium

Het startcriterium *geschikt voor onderwijsdoelen* uit het competentie assessment wiel van Baartman (2008) is in de kwaliteitsmeter uitgesplitst in 6 indicatoren, zie bijlage 1. De GL-docenten vinden deze indicatoren heel belangrijk, met een gemiddelde van 7.0 en standaarddeviatie van .6 op een 8-puntsschaal, en over het algemeen ook toepasbaar, met een gemiddelde van 7.1 en standaarddeviatie van 2.1 op een 10-puntsschaal. Alleen de tweede indicator *voldoende aandacht besteed aan kennis* wordt aangemerkt als niet toepasbaar, met een score van 4.6 bij een standaarddeviatie van 2.4 op een 10-puntsschaal. Het meest toepasbaar is de derde indicator *Uitvoeringsvaardigheden komen voldoende voor* met een gemiddelde van 8.6 en standaarddeviatie van 0.5 op een 10-puntsschaal. De mate van

tevredenheid, uitgedrukt in percentage respondenten dat *ja* antwoordde op de vraag of de toepasbaarheid voldoende is, loopt vrijwel gelijk met de mate van toepasbaarheid, zie figuur 5.

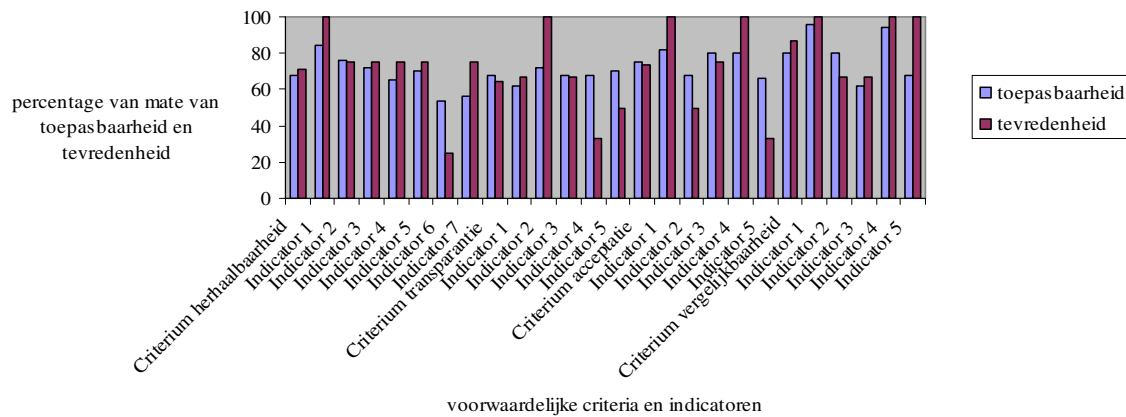


Figuur 5. Perceptiegegevens van toepasbaarheid en tevredenheid bij startcriterium ($n = 5$)

In de kwalitatieve data gaven de docenten unaniem aan dat hun perceptie van het startcriterium gebaseerd is op de beoordelingsformulieren en de daarin genoemde competenties en indicatoren.

Perceptie voorwaardelijke criteria

De vier voorwaardelijke criteria *herhaalbaarheid*, *transparantie*, *acceptatie* en *vergelijkbaarheid* zijn volgens de GL-docenten allemaal heel belangrijk, met een gemiddelde van 6.8 en standaarddeviatie van 0.8 op een 8-puntsschaal, en toepasbaar, met gemiddelde van 7.2 en standaarddeviatie van 2.2 op een 10-puntsschaal. Het kwaliteitscriterium *transparantie* is het meest belangrijk, met een gemiddelde van 7.2 en standaarddeviatie van 0.4 op een 8-puntsschaal, en het kwaliteitscriterium *herhaalbaarheid* is in verhouding het minst belangrijk van deze vier criteria, met een gemiddelde van 6.4 en een standaarddeviatie van 1.2 op een 8-puntsschaal. De *herhaalbaarheid* is ook het minst toepasbaar, met gemiddelde van 6.7 en standaarddeviatie van 2.3. *Transparantie* is niet het meest toepasbaar, met gemiddelde van 6.8 en standaarddeviatie van 1.7. Dat is de *vergelijkbaarheid* met gemiddelde toepasbaarheid van 8.0 en standaarddeviatie van 2.5. De mate van tevredenheid, uitgedrukt in percentage respondenten dat *ja* antwoordde op de vraag of de toepasbaarheid voldoende is, is meer gevarieerd dan de mate van toepasbaarheid, maar de tevredenheid en toepasbaarheid lopen wel vrij parallel aan elkaar, zie figuur 6.



Figuur 6. Perceptiegegevens van toepasbaarheid en tevredenheid bij voorwaardelijke criteria ($n = 5$)

De mate van tevredenheid is bij *herhaalbaarheid* indicator 6 *gelijkwaardig overleg tussen beoordelaars* onvoldoende, zie figuur 6 en bijlage 1, terwijl uit de kwalitatieve data blijkt dat er gesprekken plaatsvinden tussen de beoordelaars. Daarbij is de beoordeling een samenspel tussen de beoordelaars.

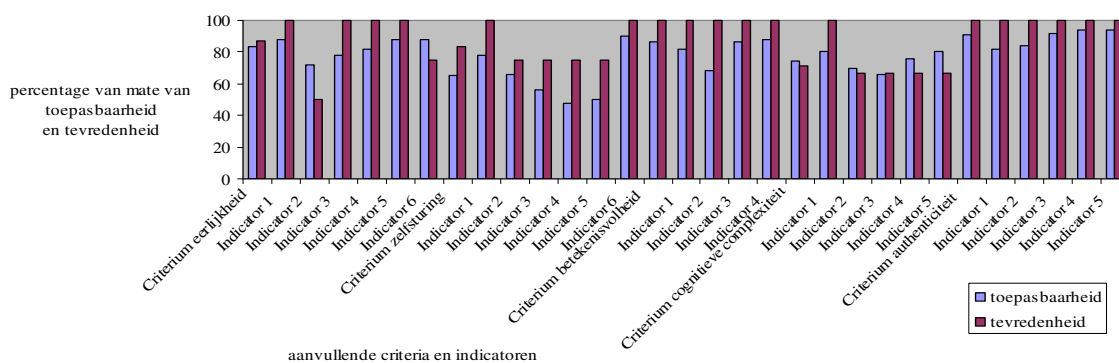
Bij indicator 4 van *transparantie, de docenten kunnen zich vinden in de tussentijdse praktijkbeoordeling*, zijn de respondenten ook onvoldoende tevreden over de toepasbaarheid, zie figuur 6. In de kwalitatieve data geven de docenten aan dat het voor studenten en praktijkbegeleiders niet altijd duidelijk is aan welke normering en criteria ze moeten voldoen. Ook zien studenten niet overal het nut van in.

De mate van tevredenheid bij *acceptatie* is onvoldoende bij indicator 5 *het beroepenveld kent en begrijpt de inhoud, criteria en uitvoering van de tussentijdse praktijkbeoordeling*, zie figuur 6. Uit de kwalitatieve data blijkt dat het werkveld nog niet altijd op de hoogte is van het competentieleren en dat er daarom meer aandacht nodig is voor het scholen van praktijkbegeleiders.

Perceptie aanvullende criteria

De vijf aanvullende criteria, *eerlijkheid, zelfsturing, betekenisvolheid, cognitieve complexiteit* en *authenticiteit* zijn in de perceptie van de docenten heel belangrijk, met een gemiddelde van 6.9 en standaarddeviatie van 0.8 op een 8-puntsschaal. De aanvullende criteria zijn toepasbaar, met een gemiddelde van 7.9 en standaarddeviatie van 1.8 op een 10-puntsschaal.

Van de aanvullende criteria wordt de *betekenisvolheid* het meest belangrijk gevonden, met een gemiddelde van 7.4 en standaarddeviatie van .5 op een 8-puntsschaal. Het criterium *zelfsturing* wordt het minst belangrijk gevonden, met gemiddelde van 6.0 en standaarddeviatie van 1.1 op een 8-puntsschaal. Ook de toepasbaarheid van *zelfsturing* is gering, met een gemiddelde van 6.5 en standaarddeviatie van 2.4. Het criterium *authenticiteit* is heel toepasbaar, met een gemiddelde van 9.1 en standaarddeviatie van 1.0 op een 10-puntsschaal. De mate van tevredenheid, uitgedrukt in percentage respondenten dat *ja* antwoordde op de vraag of de toepasbaarheid voldoende is, is minder gevarieerd dan de mate van toepasbaarheid, maar de tevredenheid en toepasbaarheid lopen wel vrij parallel aan elkaar, zie figuur 7.



Figuur 7. Perceptiegegevens van toepasbaarheid en tevredenheid bij aanvullende criteria ($n = 5$)

In figuur 7 is te zien dat de tevredenheid bij indicator 2 *de beoordelaars zijn competent* van criterium *eerlijkheid* vrij laag is. Uit de kwalitatieve data blijkt dat de respondenten van mening zijn dat de beoordelaars meer ingezet kunnen worden op hun eigen competenties, bijvoorbeeld vakinhoudelijke docenten over de vakinhoudelijke competenties.

Bij het criterium *zelfsturing* is de toepasbaarheid opvallend laag bij indicator 4 en 5, zie figuur 7. Deze indicatoren gaan respectievelijk over *studenten beoordelen zichzelf en elkaar* en *studenten hebben eigen inbreng bij de keuze van beoordelingsopdrachten*, zie bijlage 1. De tevredenheid van de respondenten is bij deze indicatoren niet laag, zie figuur 7. In de kwalitatieve data geven de respondenten aan dat de studenten alleen zichzelf beoordelen en dat beoordelen van anderen niet belangrijk is bij de tussentijdse praktijkbeoordelingen. Slechts 2 respondenten geven uitleg over de

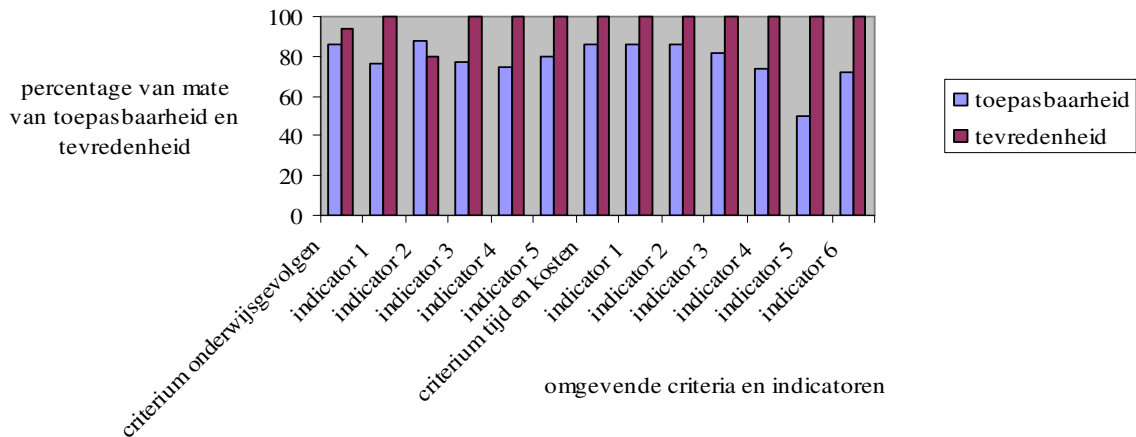
lage toepasbaarheid van indicator 5. De ene respondent geeft hierbij aan dat studenten nu geen eigen inbreng hebben bij het kiezen van beoordelingsopdrachten, maar dat dit wel wenselijk is. De andere respondent vindt dat de norm, en daarmee de keuze, van de beoordelingsopdrachten bij de docent en stagebegeleider hoort te liggen.

Opvallend is dat de respondenten unaniem tevreden zijn over de toepasbaarheid van de criteria *betekenisvolheid* en *authenticiteit*, zie figuur 5. In de kwalitatieve data geven zij bij *betekenisvolheid* aan dat de belangrijkste betekenis van de tussentijdse praktijkbeoordeling zit op het punt van de vervolgstappen die de student leert maken naar aanleiding van het beoordelingsgesprek. Bij het criterium *authenticiteit* wordt vermeld dat de stages altijd plaatsvinden in een zeer realistische en authentieke omgeving. Ook wordt hierbij opgemerkt dat de competenties zijn gebaseerd op het landelijke opleidingsprofiel.

Bij het criterium *cognitieve complexiteit* is, bij vrijwel alle indicatoren, de toepasbaarheid groter dan de tevredenheid. Uit de kwalitatieve data blijkt dat de respondenten vinden dat er vooral registreerbaar gedrag gemeten wordt en minder de achterliggende denkprocessen. Het reflecteren door de studenten wordt als heel belangrijk ervaren, maar sommige studenten moeten hierin bijgeschoold worden omdat ze te taakgericht denken.

Perceptie omgevende criteria

De omgevende criteria *onderwijsgevolgen* en *tijd en kosten* zijn door de respondenten beoordeeld als belangrijk, met een gemiddelde van 6.0 en standaarddeviatie van 1.5 op een 8-puntsschaal. Deze criteria zijn ook toepasbaar, met een gemiddelde van 7.7 en standaarddeviatie van 1.9 op een 10-puntsschaal. Het criterium *onderwijsgevolgen* is volgens de perceptie van de respondenten heel belangrijk, met een gemiddelde van 7.0 en standaarddeviatie van .0 op een 8-puntsschaal, en heel toepasbaar, met een gemiddelde van 8.0 en een standaarddeviatie van 1.5 op een 10-puntsschaal. Het criterium *tijd en kosten* is minder belangrijk, met een gemiddelde van 5.2 en standaarddeviatie van 1.3 op een 8-puntsschaal, maar wel toepasbaar, met een gemiddelde van 7.5 en standaarddeviatie van 2.1 op een 10-puntsschaal. De respondenten zijn aanzienlijk tevreden over de toepasbaarheid van de omgevende criteria, zie figuur 8.



Figuur 8. Perceptiegegevens van toepasbaarheid en tevredenheid bij omgevende criteria ($n = 5$)

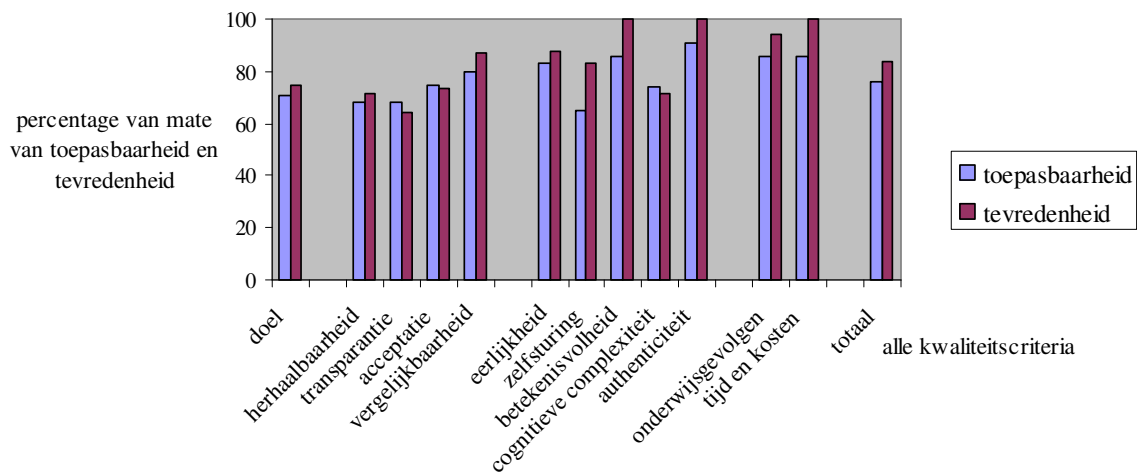
In figuur 8 valt op dat van criterium *onderwijsgevolgen* indicator 2 *de tussentijdse praktijkbeoordeling roept gewenst leergedrag op* een hoge mate van toepasbaarheid heeft met een lagere mate van tevredenheid. Uit de kwalitatieve data blijkt dat de ontevredenheid van 1 respondent voortkomt uit het feit dat hij de indicator onduidelijk vindt. De andere respondenten geven geen uitleg bij deze indicator, maar beoordelen de toepasbaarheid hoog, met een gemiddelde 7.6 en standaarddeviatie van 1.2.

Verder is opvallend dat alle respondenten tevreden zijn over het criterium *tijd en kosten* met alle bijbehorende indicatoren, terwijl de toepasbaarheid heel verschillend is per indicator, zie figuur 8. Indicator 5 *voorafgaand aan de implementatie van de tussentijdse praktijkbeoordeling wordt een inschatting gemaakt van de tijd en kosten* wordt als niet toepasbaar beoordeeld door de docenten, met een gemiddelde van 5.0 en standaarddeviatie van 3.0. De hoge standaarddeviatie duidt op een grote verscheidenheid van perceptie tussen de respondenten. Uit de kwalitatieve data komt naar voren dat er nauwelijks kosten zijn en dat alles over het algemeen vrij goed in te plannen is, al hebben stagebegeleiders wel vaak te maken met een gebrek aan uren.

Perceptie totale kwaliteit

De totale mate van toepasbaarheid en tevredenheid zijn berekend aan de hand van de gegevens van de afzonderlijke kwaliteitscriteria, zie figuur 9 en tabel 5. De totale toepasbaarheid heeft een gemiddelde van 7.6 bij een standaarddeviatie van 2.0 op een 10-puntsschaal. Alle criteria zijn toepasbaar tot heel toepasbaar. Figuur 9 laat zien dat de mate van toepasbaarheid en tevredenheid in verhouding bij de

omgevende criteria het hoogst zijn. Het *startcriterium* en de *voorwaardelijke criteria* scoren in verhouding vrij laag, terwijl de *aanvullende criteria* een zeer verdeeld beeld geven, zie figuur 9. Bij de criteria *transparantie*, *acceptatie* en *cognitieve complexiteit* is de mate van toepasbaarheid groter dan de mate van tevredenheid. Bij de andere criteria is de mate van tevredenheid groter dan de mate van toepasbaarheid, zie figuur 9.



Figuur 9. Perceptiegegevens van toepasbaarheid en tevredenheid bij alle criteria ($n = 5$)

De respondenten vinden alle criteria belangrijk tot heel belangrijk, met een gemiddelde van 6.7 en een standaarddeviatie van 1.0 op een 8-puntsschaal. Het meest belangrijk is het criterium *betekenisvolheid*, met een gemiddelde van 7.4 en standaarddeviatie van 0.5 op een 8-puntsschaal. Het minst belangrijk is het criterium *tijd en kosten*, met een gemiddelde van 5.2 en standaarddeviatie van 1.3 op een 8-puntsschaal. Bij dit criterium is de standaarddeviatie het hoogst, wat aangeeft dat de percepties van de respondenten hier het minst overeen stemmen, zie tabel 5.

Tabel 5. Gemiddelde en standaarddeviatie van belangrijkheid en toepasbaarheid per criterium

Kwaliteitscriteria Categorie	Criterium	Belangrijkheid		Toepasbaarheid	
		Gemiddeld op 8-puntsschaal	Standaard- deviatie	Gemiddeld op 10-puntsschaal	Standaard- deviatie
Start	Doel	7.0	0.6	7.0	2.1
Voorwaardelijk	Herhaalbaarheid	6.4	1.2	6.7	2.3
	Transparantie	7.2	0.4	6.7	1.7
	Acceptatie	6.8	0.8	7.6	1.7
	Vergelijkbaarheid	6.8	0.4	8.0	2.5
	Totaal	6.8	0.8	7.2	2.2
Aanvullend	Eerlijkheid	6.8	0.4	8.3	1.6
	Zelfsturing	6.0	1.1	6.5	2.4
	Betekenisvolheid	7.4	0.5	8.6	0.7
	Cognitieve complexiteit	7.0	0.6	7.4	1.5
	Authenticiteit	7.2	0.4	9.1	1.0
Omgevend	Totaal	6.9	0.8	7.9	1.8
	Onderwijsgevolgen	7.0	0.0	8.0	1.5
	Tijd en kosten	5.2	1.3	7.5	1.2
Totaal	Totaal	6.0	1.5	7.7	1.9

n = 5

Op de afsluitende vraag *In hoeverre vindt u de totale kwaliteit van de tussentijdse praktijkbeoordeling voldoende*, is door de respondenten op een 8-puntsschaal een gemiddelde aangegeven van 5.8 met een standaarddeviatie van 1.5. Dit gemiddelde past in de categorie *wel voldoende* en is daarmee niet *helemaal voldoende*. Op kwalitatieve wijze is gevraagd naar de sterke en zwakke punten van de praktijkbeoordeling. Als zwakke punten werden door de docenten verschillende dingen genoemd, namelijk:

- Het is nog teveel een zaak vanuit de opleidingsschool en te weinig een gezamenlijke verantwoordelijkheid
- Het werken met gedragsindicatoren die teveel versnippering van het beeld geven en te vaak niet van toepassing zijn
- De beperkte keuze van de indicatoren en de opsomming van de indicatoren
- De student zou meer aan de voorbereiding kunnen doen
- De praktijksituatie is soms te smal alleen gebaseerd op lessen

De sterke punten van de praktijkbeoordeling zijn volgens de docenten:

- Directe betrokkenheid op student en zijn leerproces
- De gerichtheid op competenties, maar ook de vrijheid om zaken aan te vullen waarin de student excelleert

- Integraal, uitgebreid en authentiek
- De feedback op de student vanuit de open vragen op het beoordelingsformulier
- Het proceskarakter; het meet waar je staat en wat nog moet.

3.2 Perceptie van studenten

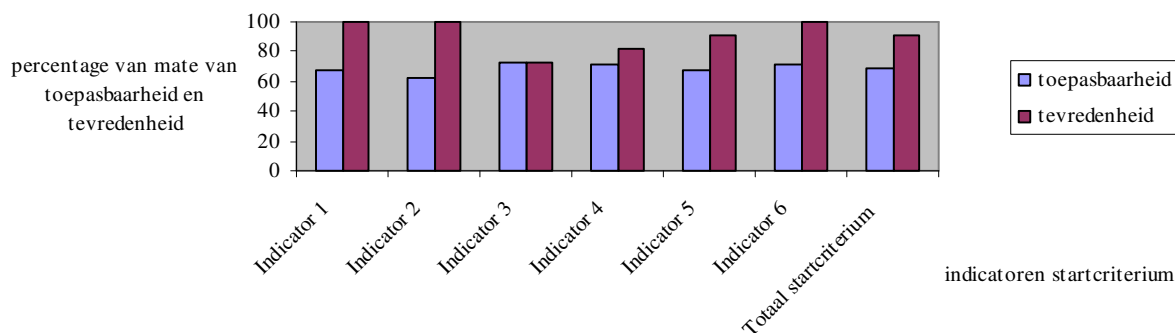
Van de 27 ondervraagde tweedejaarsstudenten hebben 11 studenten de kwaliteitsmeter van Baartman (2008) volledig ingevuld. Hierbij is per kwaliteitscriterium gevraagd in hoeverre elk criterium belangrijk is bij de praktijkbeoordelingen. Per indicator is gevraagd naar de mate van toepasbaarheid en of dat voldoende dan wel onvoldoende is. Eén student heeft nergens aangegeven of de indicatoren voldoende of onvoldoende toepasbaar zijn, andere studenten hebben incidenteel geen antwoord gegeven. Hierdoor is sprake van missing values. De resultaten van de ingevulde kwaliteitsmeter zijn beschreven op de volgorde van de kwaliteitsmeter, namelijk:

- Perceptie van startcriterium
- Perceptie van voorwaardelijke criteria
- Perceptie van aanvullende criteria
- Perceptie van omgevende criteria
- Perceptie van totale kwaliteit

Perceptie startcriterium

Het startcriterium *geschikt voor onderwijsdoelen* uit het competentie assessment wiel van Baartman (2008) is in de kwaliteitsmeter uitgesplitst in 6 indicatoren, zie bijlage 1. De studenten vinden deze indicatoren belangrijk, met een gemiddelde van 5.9 en standaarddeviatie van 1.5 op een 8-puntsschaal, en over het algemeen ook toepasbaar, met een gemiddelde van 6.9 en standaarddeviatie van 1.7 op een 10-puntsschaal. Het minst toepasbaar is de tweede indicator *voldoende aandacht besteed aan kennis*, met een score van 6.2 bij een standaarddeviatie van 2.4 op een 10-puntsschaal. Het meest toepasbaar is de derde indicator *Uitvoeringsvaardigheden komen voldoende voor* met een gemiddelde van 7.3 en standaarddeviatie van 1.3 op een 10-puntsschaal. Opvallend is dat de mate van tevredenheid,

uitgedrukt in percentage respondenten dat *ja* antwoordde op de vraag of de toepasbaarheid voldoende is, niet gelijk loopt met de mate van toepasbaarheid, zie figuur 10.

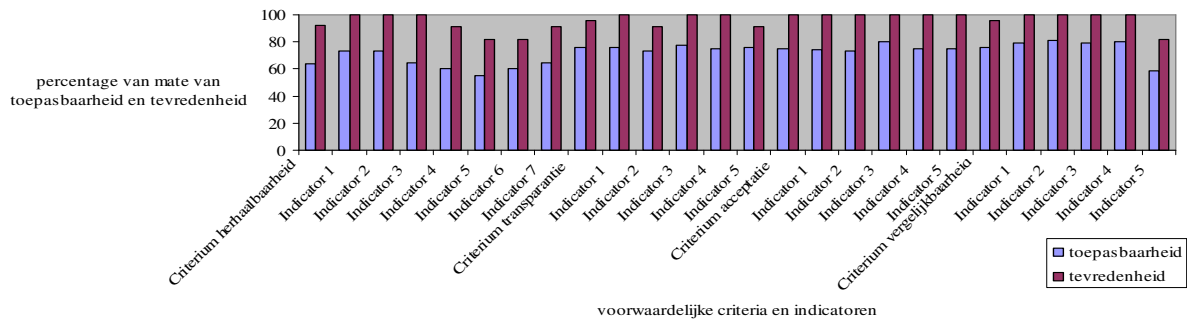


Figuur 10. Perceptie van studenten van toepasbaarheid en tevredenheid bij startcriterium ($n = 11$)

In de kwalitatieve data geven de studenten unaniem aan dat hun perceptie van het startcriterium gebaseerd is op de beoordelingsformulieren waarin alle competenties aan bod komen en in de praktijk getoetst worden.

Perceptie voorwaardelijke criteria

De vier voorwaardelijke criteria *herhaalbaarheid*, *transparantie*, *acceptatie* en *vergelijkbaarheid* zijn volgens de studenten allemaal belangrijk, met een gemiddelde van 5.9 en een standaarddeviatie van 1.3 op een 8-puntsschaal, en toepasbaar, met een gemiddelde van 7.2 en standaarddeviatie van 1.7 op een 10-puntsschaal. Het kwaliteitscriterium *acceptatie* is de belangrijkste, met een gemiddelde van 6.2 en standaarddeviatie van 1.0 op een 8-puntsschaal, en het kwaliteitscriterium *herhaalbaarheid* is in verhouding het minst belangrijk van deze vier criteria, met een gemiddelde van 5.7 en een standaarddeviatie van 1.5 op een 8-puntsschaal. De *herhaalbaarheid* is ook het minst toepasbaar, met gemiddelde van 6.4 en standaarddeviatie van 1.8. De mate van tevredenheid, uitgedrukt in percentage respondenten dat *ja* antwoordde op de vraag of de toepasbaarheid voldoende is, is minder gevarieerd dan de mate van toepasbaarheid, maar de tevredenheid en toepasbaarheid lopen wel vrij parallel aan elkaar, zie figuur 11.



Figuur 11. Perceptiegegevens van studenten van toepasbaarheid en tevredenheid bij voorwaardelijke criteria ($n = 11$)

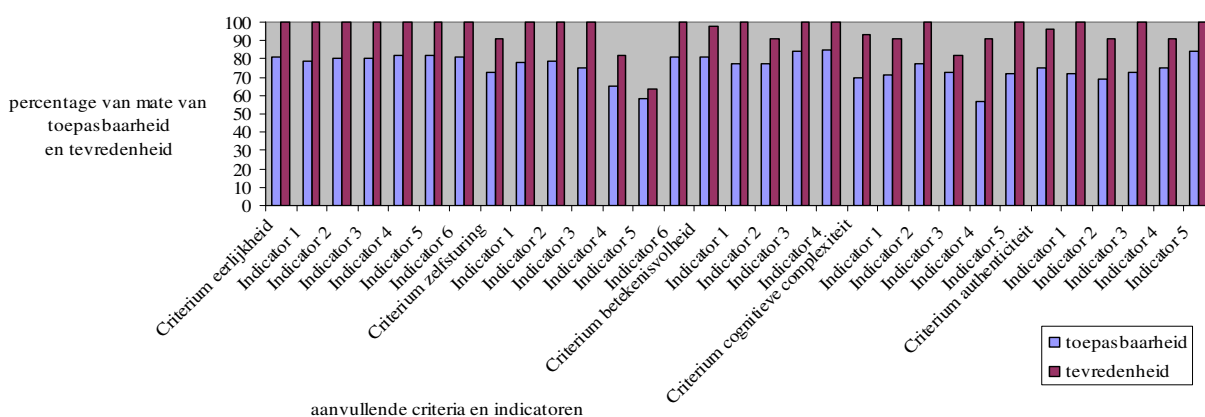
De mate van toepasbaarheid is bij *herhaalbaarheid* indicator 5 de beoordelaars hebben een *verschillende achtergrond* onvoldoende, zie figuur 11, terwijl uit de kwalitatieve data blijkt dat elke student een beoordelaar heeft vanuit de praktijk, de stagebegeleider, en een beoordelaar vanuit de opleiding, een GL-docent.

De mate van toepasbaarheid bij *acceptatie* is zeer hoog bij indicator 3 de studenten kennen en begrijpen de uit te voeren *beoordelingsopdrachten*, zie figuur 11 en bijlage 1. Uit de kwalitatieve data blijkt dat de studenten de opdrachten helder vinden. Tegelijkertijd geven de studenten aan dat de stagebegeleiders de beoordelingsopdrachten niet altijd duidelijk vinden en soms zelfs niets weten van de beoordelingsopdrachten.

Bij *vergelijkbaarheid* zijn de studenten zeer tevreden over de hoge mate van toepasbaarheid van indicator 2, de *beoordelingscriteria en normering* zijn voor alle studenten *vergelijkbaar*, en indicator 4, de *beoordelingsopdrachten* zijn voor alle studenten *hetzelfde*, zie figuur 11. Over de mindere toepasbaarheid van de *werkomstandigheden tijdens de beoordeling* zijn voor alle studenten *vergelijkbaar*, indicator 5, zijn de studenten aanzienlijk minder tevreden. Uit de kwalitatieve data blijkt dat elke student een andere werkomgeving en stagebegeleider heeft, waardoor de werkomstandigheden niet goed vergelijkbaar zijn. Volgens de studenten zijn bepaalde praktijkopdrachten meer toepasbaar op protestants christelijk onderwijs dan op reformatorisch onderwijs.

Perceptie aanvullende criteria

De vijf aanvullende criteria, *eerlijkheid*, *zelfsturing*, *betekenisvolheid*, *cognitieve complexiteit* en *authenticiteit* zijn in de perceptie van de studenten belangrijk, met een gemiddelde van 6.4 en standaarddeviatie van 1.2 op een 8-puntsschaal. De aanvullende criteria zijn toepasbaar, met een gemiddelde van 7.6 en standaarddeviatie van 1.7 op een 10-puntsschaal. Van de aanvullende criteria wordt de *authenticiteit* het meest belangrijk gevonden, met een gemiddelde van 6.8 en standaarddeviatie van .8 op een 8-puntsschaal. Opvallend is dat *authenticiteit* het minst toepasbaar is, met een gemiddelde van 7.0 en standaarddeviatie van 1.7. Het criterium *eerlijkheid* wordt het minst belangrijk gevonden, met gemiddelde van 6.3 en standaarddeviatie van 1.4 op een 8-puntsschaal. De mate van tevredenheid, uitgedrukt in percentage respondenten dat *ja* antwoordde op de vraag of de toepasbaarheid voldoende is, is minder gevarieerd dan de mate van toepasbaarheid, maar de tevredenheid en toepasbaarheid lopen wel vrij parallel aan elkaar, zie figuur 12.



Figuur 12. Perceptiegegevens van studenten van toepasbaarheid en tevredenheid bij aanvullende criteria ($n = 11$)

In figuur 12 is te zien dat de toepasbaarheid bij criterium *eerlijkheid* bij alle indicatoren bijzonder hoog is. Uit de kwalitatieve data blijkt dat de studenten van mening zijn dat er duidelijkheid, openheid en goede communicatie is op dit terrein.

Bij het criterium *zelfsturing* is de toepasbaarheid en tevredenheid vrij laag bij indicator 5, zie figuur 12. Deze indicator gaat over *studenten hebben eigen inbreng bij de keuze van beoordelingsopdrachten*, zie bijlage 1. In de kwalitatieve data geven de studenten aan dat ze geen

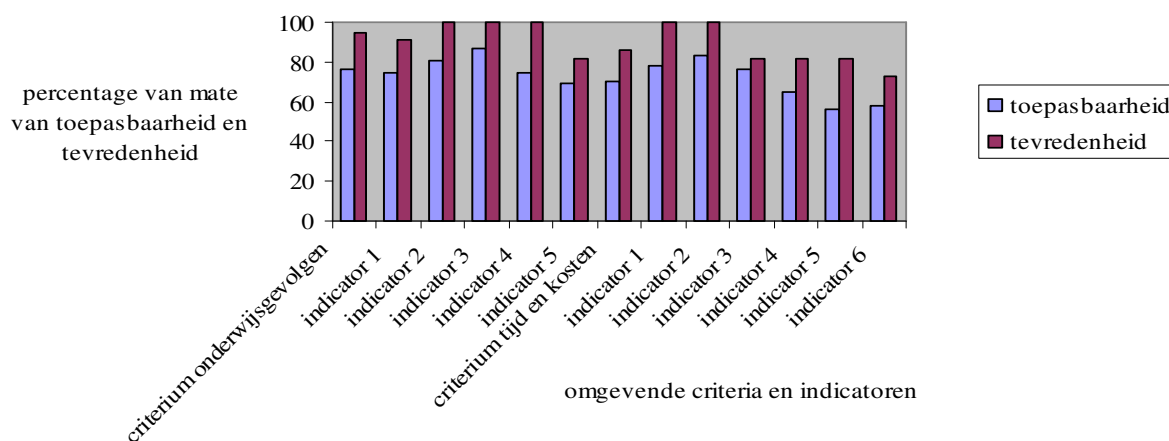
eigen inbreng hebben bij het kiezen van opdrachten, maar dat ze het wel wenselijk vinden om opdrachten te kunnen kiezen die bij hun persoonlijke doelen en competenties passen.

Bij de criteria *betekenisvolheid* en *authenticiteit* geven de studenten aan dat alle indicatoren toepasbaar tot heel toepasbaar zijn, zie figuur 12. In de kwalitatieve data geven zij bij *betekenisvolheid* aan dat de tussentijdse praktijkbeoordeling een leermoment is waarbij goede feedback gegeven wordt. Bij het criterium *authenticiteit* wordt vermeld dat de stages al aanvoelen als een echt beroep.

Bij het criterium *cognitieve complexiteit* is de toepasbaarheid van indicator 4, *de beoordeling vindt plaats in een nieuwe, complexe situatie*, vrij laag. De tevredenheid is hierbij wel hoog. De studenten geven niet aan waarom de toepasbaarheid laag is.

Perceptie omgevende criteria

De omgevende criteria *onderwijsgevolgen* en *tijd en kosten* zijn door de studenten beoordeeld als belangrijk, met een gemiddelde van 6.0 en standaarddeviatie van 1.2 op een 8-puntsschaal. Deze criteria zijn ook toepasbaar, met een gemiddelde van 7.2 en standaarddeviatie van 4.4 op een 10-puntsschaal. Deze standaarddeviatie is opvallend hoog, wat aangeeft dat de studenten over de toepasbaarheid van de omgevende criteria heel verschillend denken. Het criterium *onderwijsgevolgen* is volgens de perceptie van de studenten belangrijk, met een gemiddelde van 5.9 en standaarddeviatie van 1.1 op een 8-puntsschaal, en toepasbaar, met een gemiddelde van 7.6 en een standaarddeviatie van 2.0 op een 10-puntsschaal. Het criterium *tijd en kosten* is ook belangrijk, met een gemiddelde van 6.2 en standaarddeviatie van 1.3 op een 8-puntsschaal en toepasbaar, met een gemiddelde van 7.0 en standaarddeviatie van 2.1 op een 10-puntsschaal. De mate van tevredenheid, uitgedrukt in percentage respondenten dat *ja* antwoordde op de vraag of de toepasbaarheid voldoende is, loopt vrij parallel aan de mate van toepasbaarheid, zie figuur 13.

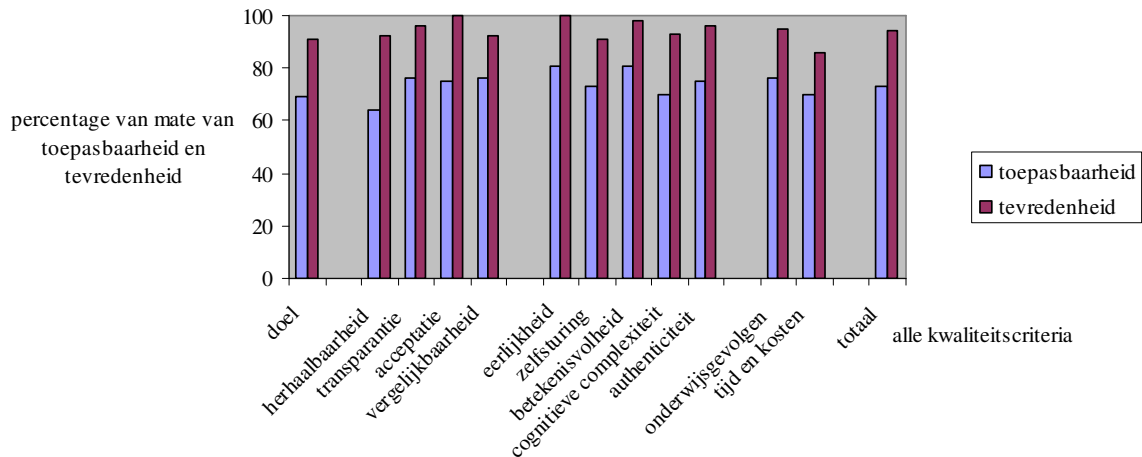


Figuur 13. Perceptiegegevens van studenten van toepasbaarheid en tevredenheid bij omgevende criteria ($n = 11$)

In figuur 13 valt op dat zowel de mate van toepasbaarheid als de mate van tevredenheid veel varieert tussen de verschillende indicatoren van de omgevende criteria. In de kwalitatieve data geven de studenten aan dat ze niet goed weten wat ze moeten antwoorden bij de indicatoren, omdat ze geen goede kijk hebben op de *onderwijsgevolgen* en *tijd en kosten*.

Perceptie totale kwaliteit

De totale mate van toepasbaarheid en tevredenheid zijn berekend aan de hand van de gegevens van de afzonderlijke kwaliteitscriteria, zie figuur 14 en tabel 6. De totale toepasbaarheid heeft een gemiddelde van 7.3 bij een standaarddeviatie van 3.6 op een 10-puntsschaal. Alle criteria zijn toepasbaar tot heel toepasbaar. Figuur 14 laat zien dat de mate van toepasbaarheid en tevredenheid in verhouding niet veel van elkaar verschillen. De toepasbaarheid is bij alle criteria voldoende beoordeeld. Bij alle criteria is de mate van tevredenheid groter dan de mate van toepasbaarheid, zie figuur 14.



Figuur 14. Perceptie van studenten van toepasbaarheid en tevredenheid bij alle criteria ($n = 11$)

De studenten vinden alle criteria belangrijk tot heel belangrijk, met een gemiddelde van 6.1 en een standaarddeviatie van 1.3 op een 8-puntsschaal, zie tabel 6. Het meest belangrijk is het criterium *authenticiteit*, met een gemiddelde van 6.8 en standaarddeviatie van 0.8 op een 8-puntsschaal. Het minst belangrijk is het criterium *herhaalbaarheid*, met een gemiddelde van 5.7 en standaarddeviatie van 1.5 op een 8-puntsschaal.

Tabel 6. Gemiddelde en standaarddeviatie van belangrijkheid en toepasbaarheid per criterium van studentenperceptie

Kwaliteitscriteria	Belangrijkheid	Toepasbaarheid			
Categorie	Gemiddeld op 8-puntsschaal	Gemiddeld op 10-puntsschaal			
Criterium	Standaarddeviatie	Standaarddeviatie			
Start	Doel	5.9	1.5	6.9	1.7
Voorwaardelijk	Herhaalbaarheid	5.7	1.5	6.4	1.8
	Transparantie	6.0	1.3	7.6	1.3
	Acceptatie	6.2	1.0	7.5	1.7
	Vergelijkbaarheid	5.7	1.3	7.6	2.0
	Totaal	5.9	1.3	7.2	1.7
Aanvullend	Eerlijkheid	6.3	1.4	7.1	1.4
	Zelfsturing	6.4	1.1	8.1	1.6
	Betekenisvolheid	6.5	1.2	8.1	2.0
	Cognitieve complexiteit	5.8	1.3	7.0	1.7
	Authenticiteit	6.8	0.8	7.5	1.6
Omgevend	Totaal	6.4	1.2	7.6	1.7
	Onderwijsgevolgen	5.9	1.1	7.6	2.0
	Tijd en kosten	6.2	1.3	7.0	2.1
Totaal	Totaal	6.0	1.2	7.2	4.4
Totaal		6.1	1.3	7.3	3.6

$n = 11$

Op de afsluitende vraag *In hoeverre vindt u de totale kwaliteit van de tussentijdse praktijkbeoordeling voldoende*, is door de respondenten op een 8-puntsschaal een gemiddelde aangegeven van 5.8 met een standaarddeviatie van 1.5. Dit gemiddelde past in de categorie *wel voldoende* en is daarmee niet *helemaal voldoende*. Op kwalitatieve wijze is gevraagd naar de sterke en zwakke punten van de praktijkbeoordeling. Als zwakke punten werden door de studenten verschillende dingen genoemd, namelijk:

- Het beoordelingsformulier kan te letterlijk genomen worden
- Vage omschrijvingen van competenties en lastig uitvoerbare eisen
- Het kost heel veel tijd
- Het gaat teveel over vaardigheden en te weinig over het vertalen van de kennis naar de praktijksituatie
- Het zijn alleen momentopnamen

De sterke punten van de praktijkbeoordeling zijn volgens de studenten:

- door de beoordeling krijg je een goed beeld van waar je staat, wat je al kunt en waar je nog aan moet werken
- de concrete feedback op je ontwikkeling
- dat de CHE langskomt, getuigt van betrokkenheid
- de toetsituatie is identiek aan de beroepssituatie
- alle competenties komen aan bod

3.3 Perceptie van stagebegeleiders

Geen van de 15 stagebegeleiders heeft de kwaliteitsmeter ingevuld. Hierdoor is het niet mogelijk resultaten weer te geven van de percepties van de stagebegeleiders.

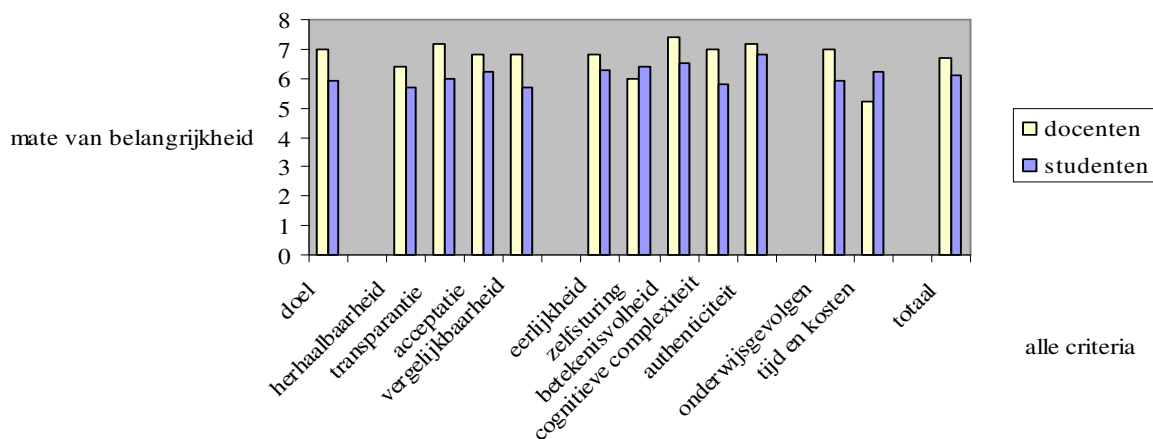
3.4 Overeenkomsten en verschillen van percepties van GL-docenten en studenten

De percepties van de GL-docenten hebben meer overeenkomsten dan verschillen met de percepties van de studenten. De meest opvallende overeenkomsten en verschillen worden in deze paragraaf

beschreven, waarbij onderscheid gemaakt wordt tussen belangrijkheid, toepasbaarheid en tevredenheid.

Belangrijkheid: overeenkomsten en verschillen

De mate van belangrijkheid is bij bijna alle criteria hoger bij de docenten dan bij de studenten, zie figuur 15 en tabel 7. Alleen de criteria *zelfsturing* en *tijd en kosten* vinden de studenten belangrijker dan de docenten. De verschillen tussen de percepties van docenten en studenten zijn niet opvallend groot. Op de afsluitende vraag *in hoeverre vindt u de totale kwaliteit van uw praktijkbeoordeling voldoende?* Gaven zowel de docenten als de studenten gemiddeld een 5.8 op een 8-puntsschaal met een standaarddeviatie van 1.5. Deze 5.8 is lager dan de 6.7, van de docenten, en 6.1, van de studenten, die voortvloeien uit de gemiddelden van de criteria.



Figuur 15. Mate van belangrijkheid per criterium van docentenperceptie en studentenperceptie

($n^{\text{docenten}} = 5$, $n^{\text{studenten}} = 11$)

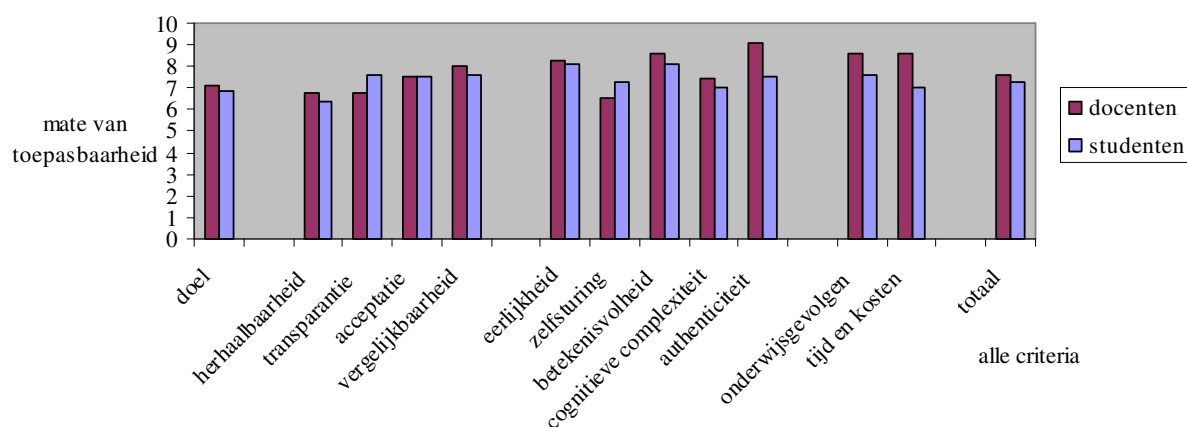
Tabel 7. Gemiddelde en standaarddeviatie van belangrijkheid en toepasbaarheid per criterium van docentenperceptie en studentenperceptie

Kwaliteitscriteria Categorie	Criterium	Belangrijkheid				Toepasbaarheid			
		Gemiddeld op 8-puntsschaal		Standaard- deviatie		Gemiddeld op 10-puntsschaal		Standaard- deviatie	
		Docent	Student	Docent	Student	Docent	Student	Docent	student
Start	Doel	7.0	5.9	0.6	1.5	7.0	6.9	2.1	1.7
Voorwaardelijk	Herhaalbaarheid	6.4	5.7	1.2	1.5	6.7	6.4	2.3	1.8
	Transparantie	7.2	6.0	0.4	1.3	6.7	7.6	1.7	1.3
	Acceptatie	6.8	6.2	0.8	1.0	7.6	7.5	1.7	1.7
	Vergelijkbaarheid	6.8	5.7	0.4	1.3	8.0	7.6	2.5	2.0
	Totaal	6.8	5.9	0.8	1.3	7.2	7.2	2.2	1.7
Aanvullend	Eerlijkheid	6.8	6.3	0.4	1.4	8.3	7.1	1.6	1.4
	Zelfsturing	6.0	6.4	1.1	1.1	6.5	8.1	2.4	1.6
	Betekenisvolheid	7.4	6.5	0.5	1.2	8.6	8.1	0.7	2.0
	Cognitieve complexiteit	7.0	5.8	0.6	1.3	7.4	7.0	1.5	1.7
	Authenticiteit	7.2	6.8	0.4	0.8	9.1	7.5	1.0	1.6
Omgevend	Totaal	6.9	6.4	0.8	1.2	7.9	7.6	1.8	1.7
	Onderwijsgevolgen	7.0	5.9	0.0	1.1	8.0	7.6	1.5	2.0
	Tijd en kosten	5.2	6.2	1.3	1.3	7.5	7.0	1.2	2.1
Totaal	6.0	6.0	1.5	1.2	7.7	7.2	1.9	4.4	
Totaal		6.7	6.1	1.0	1.3	7.6	7.3	2.0	3.6

$n^{\text{docenten}} = 5, n^{\text{studenten}} = 11$

Toepasbaarheid: overeenkomsten en verschillen

De mate van toepasbaarheid is bij bijna alle criteria hoger bij de docenten dan bij de studenten, zie figuur 16 en tabel 7. Alleen het criterium *transparantie* vinden de studenten meer toepasbaar dan de docenten. De verschillen tussen de percepties van docenten en studenten zijn niet opvallend groot.

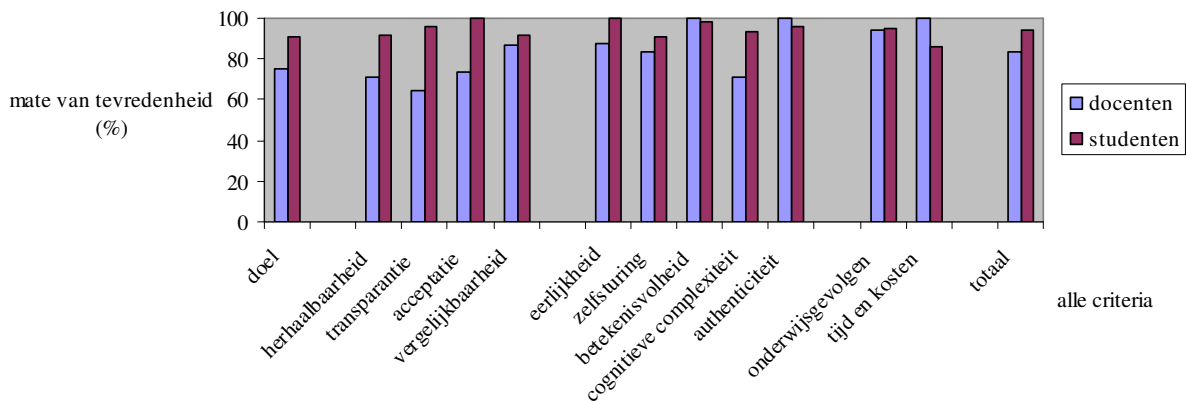


Figuur 16. Mate van toepasbaarheid per criterium van docentenperceptie en studentenperceptie

($n^{\text{docenten}} = 5, n^{\text{studenten}} = 11$)

Tevredenheid: overeenkomsten en verschillen

De mate van tevredenheid is bij bijna alle criteria lager bij de docenten dan bij de studenten, zie figuur 17. Alleen over de toepasbaarheid van de criteria *betekenisvolheid*, *authenticiteit* en *tijd en kosten* zijn de docenten meer tevreden dan de studenten. De verschillen tussen de percepties van docenten en studenten zijn bij de voorwaardelijke criteria vrij groot.



Figuur 17. Mate van tevredenheid per criterium van docentenperceptie en studentenperceptie ($n^{\text{docenten}} = 5$, $n^{\text{studenten}} = 11$)

4. Conclusie

4.1 Perceptie van GL-docenten

De perceptie van docenten van de kwaliteit van het startcriterium met betrekking tot de praktijkbeoordelingen is gematigd positief. De docenten vinden het criterium heel belangrijk, maar enkele indicatoren zijn minder of niet toepasbaar. De tevredenheid loopt parallel met de mate van toepasbaarheid.

De docenten vinden alle voorwaardelijke criteria heel belangrijk, maar niet allemaal heel toepasbaar. Over de toepasbaarheid van de *vergelijkbaarheid* zijn de docenten tevreden. De *transparantie* is wel toepasbaar, maar de docenten zijn niet tevreden over de mate van toepasbaarheid. Volgens de docenten moet de tussentijdse praktijkbeoordeling transparanter worden door meer duidelijkheid te geven richting de stagebegeleiders en studenten.

De aanvullende criteria worden door de docenten heel belangrijk gevonden, maar de mate van toepasbaarheid en tevredenheid varieert per criterium. De criteria *eerlijkheid*, *betekenisvolheid* en

authenticiteit zijn goed toepasbaar en de docenten zijn daar tevreden over. Het criterium *zelfsturing* is minder toepasbaar en ook niet heel belangrijk volgens de docenten. Toch zijn de docenten tevreden over de mate van toepasbaarheid van *zelfsturing*, omdat dit criterium niet echt bij de tussentijdse praktijkbeoordeling past, maar wel bij andere beoordelingen naar voren komt. Bij het criterium *cognitieve complexiteit* zijn de docenten ontevreden over de mate waarin de kennis beoordeeld wordt. De perceptie van de docenten is dat de tussentijdse praktijkbeoordeling teveel gericht is op het beoordelen van zichtbaar gedrag.

Bij de omgevende criteria is de perceptie van de docenten dat het criterium *onderwijsgevolgen* belangrijker en meer toepasbaar is dan het criterium *tijd en kosten*. Over beide criteria zijn de docenten tevreden.

Alle criteria samenvoegend, is duidelijk dat de docenten alle criteria belangrijk en toepasbaar vinden. Het criterium *betekenisvolheid* is, in de perceptie van de docenten, het meest belangrijk en het criterium *tijd en kosten* is het minst belangrijk. De toepasbaarheid is bij het criterium *zelfsturing* het laagst en bij het criterium *authenticiteit* het hoogst. De mate van tevredenheid van de toepasbaarheid is bij *transparantie* het laagst, maar ook bij *cognitieve complexiteit* en *acceptatie* gering. Over de toepasbaarheid van de criteria *betekenisvolheid*, *authenticiteit* en *tijd en kosten* zijn de docenten zeer tevreden.

4.2 Perceptie van studenten

De perceptie van studenten van de kwaliteit van het startcriterium met betrekking tot de praktijkbeoordelingen is positief. De studenten vinden het criterium belangrijk, en alle indicatoren zijn toepasbaar. De tevredenheid loopt niet parallel met de mate van toepasbaarheid.

De studenten vinden alle voorwaardelijke criteria belangrijk en toepasbaar. De studenten zijn tevreden over de mate van toepasbaarheid van de indicatoren.

De aanvullende criteria worden door de studenten belangrijk tot heel belangrijk gevonden. Ook de mate van toepasbaarheid en tevredenheid zijn hoog tot heel hoog. De criteria *zelfsturing*, *cognitieve complexiteit* en *authenticiteit* zijn goed toepasbaar en de studenten zijn daar redelijk tot heel tevreden over. De criteria *eerlijkheid* en *betekenisvolheid* worden door de studenten als heel toepasbaar

aangemerkt. De studenten zijn hier heel tevreden mee, maar ze vinden deze zeer toepasbare criteria niet heel belangrijk.

De perceptie van de studenten bij de omgevende criteria is dat het criterium *tijd en kosten* belangrijker, maar minder toepasbaar is dan *onderwijsgevolgen*. Over de toepasbaarheid van *tijd en kosten* zijn de studenten niet bij alle indicatoren tevreden. De studenten gaven bij de omgevende criteria aan, geen volledig beeld te hebben van de praktijk op dit gebied en daardoor niet heel betrouwbaar te kunnen antwoorden.

Alle criteria samenvoegend, is duidelijk dat de studenten alle criteria belangrijk en toepasbaar vinden. Het criterium *authenticiteit* is, in de perceptie van de studenten, het meest belangrijk en de criteria *herhaalbaarheid* en *vergelijkbaarheid* zijn het minst belangrijk. De toepasbaarheid is bij het criterium *herhaalbaarheid* het laagst en bij de criteria *eerlijkheid* en *betekenisvolheid* het hoogst. De mate van tevredenheid van de toepasbaarheid is bij alle criteria hoog.

4.3 Overeenkomsten en verschillen percepties docenten en studenten

De centrale vraag van dit onderzoek was: *Wat is volgens docenten, studenten en stagebegeleiders de edumetrische kwaliteit van de praktijkbeoordelingen bij de lerarenopleiding Godsdienst / Levensbeschouwing op de Christelijke Hogeschool Ede?* Deze vraag is niet compleet te beantwoorden doordat de stagebegeleiders geen medewerking wilden verlenen. De percepties van de docenten en de studenten geven aan dat de edumetrische kwaliteit van de praktijkbeoordelingen bij de lerarenopleiding Godsdienst / Levensbeschouwing op de CHE, rekening houdend met de kwaliteitscriteria van het competentie assessment wiel van Baartman (2008), goed is. De percepties van de docenten en studenten komen bijna geheel met elkaar overeen. De studenten zijn meer tevreden dan de docenten over de toepasbaarheid van de criteria.

4.4 Verwachtingen

De verwachting van de docenten voorafgaand aan dit onderzoek, was dat zij een eenduidig oordeel zouden hebben over de edumetrische kwaliteit van de praktijkbeoordelingen. Uit de resultaten blijkt, dat de percepties van de docenten globaal genomen inderdaad redelijk eenduidig zijn. Bij enkele

criteria en indicatoren zijn de standaarddeviaties echter vrij hoog, waaruit blijkt dat de percepties van de docenten daar niet eenduidig zijn. Dit is voornamelijk het geval bij de mate van toepasbaarheid van het startcriterium, de voorwaardelijke en omgevende criteria. In de kwalitatieve data kwam vaak wel een eenduidig oordeel naar voren. De verschillen tussen de docenten hoeven niet per definitie te betekenen dat zij het niet met elkaar eens zijn. De oorzaak kan ook gevonden worden in de interpretatie van de indicatoren en in de percepties over de kwaliteitsmeter.

De docenten geven aan dat alle kwaliteitscriteria belangrijk tot heel belangrijk zijn. Dit is een bevestiging van de kwaliteit van de kwaliteitsmeter en onderzoek van Baartman (2008). Ook bevestigt dit de verwachting dat de docenten alle criteria belangrijk zouden vinden.

5. Discussie

5.1 Deelnemers

In dit onderzoek is de edumetrische kwaliteit van de praktijkbeoordelingen van de lerarenopleiding Godsdienst / Levensbeschouwing op de CHE beoordeeld aan de hand van de percepties van de betrokken docenten en studenten. Dit is gedaan met behulp van, een aanpaste versie van, de kwaliteitsmeter van Baartman (2008), zie paragraaf 2.2. Stagebegeleiders bleken niet bereid de uitgebreide kwaliteitsmeter in te vullen waardoor het niet mogelijk was hun percepties mee te nemen in het onderzoek. Hierdoor kan een scheef beeld ontstaan zijn van de edumetrische kwaliteit van de praktijkbeoordelingen.

Bij vervolgonderzoek is het aan te bevelen vooraf stagebegeleiders te peilen over manieren waarop zij mee willen en kunnen werken, want voor een volledig beeld van de edumetrische kwaliteit van praktijkbeoordelingen is het noodzakelijk dat ook de stagebegeleiders hun perceptie duidelijk maken.

5.2 Kwaliteitsmeter

De kwaliteitsmeter van Baartman (2008) waarmee de edumetrische kwaliteit van assessmentvormen bepaald kan worden, is voor dit onderzoek aangepast, zoals beschreven in paragraaf 2.2. Deze aanpassingen kunnen er onbedoeld voor gezorgd hebben dat de resultaten minder betrouwbaar en

valide zijn dan bij de officiële gevalideerde kwaliteitsmeter van Baartman (2008). Door de aanpassingen is er een fout opgetreden bij de vragen naar de bekendheid van de indicatoren. Bij de resultaten was niet duidelijk of er sprake was van niet-ingevulde vragen of van onbekende indicatoren. Daarom is de bekendheid van de indicatoren niet meegenomen in de resultaten en conclusies.

De respondenten hebben aangegeven dat alle criteria belangrijk zijn. Er is echter niet gevraagd of deze kwaliteitscriteria de enige belangrijke criteria zijn. Hierdoor is niet met zekerheid te zeggen of de respondenten de kwaliteitsmeter compleet vinden. Voor verder onderzoek is aan te bevelen de kwaliteitsmeter zo aan te passen dat respondenten de criteria zelf kunnen aanvullen met aandachtspunten die zij belangrijk vinden.

In dit onderzoek is gevraagd naar percepties over de kwaliteit van *tussentijdse* praktijkbeoordelingen omdat de eindbeoordelingen nog niet gegeven waren. De tussenbeoordelingen zijn formatief en kunnen in kwaliteit afwijken van de summatieve eindbeoordelingen. Verder onderzoek is nodig om aan te tonen of er daadwerkelijk verschil is tussen de percepties over de edumetrische kwaliteit van de tussentijdse en eindbeoordelingen.

Verschillende respondenten gaven aan dat de kwaliteitsmeter te veel indicatoren en vragen bevat. Hierdoor kregen de respondenten de neiging vragen over te slaan of antwoorden in te vullen zonder nauwkeurig te lezen. Om respondenten gemotiveerd en betrokken te houden bij het invullen van de kwaliteitsmeter is het aan te bevelen bij vervolgonderzoek de kwaliteitsmeter op een betrouwbare en valide manier in te korten.

5.3 *Praktijkbeoordelingen*

Uit dit onderzoek is gebleken dat de docenten en studenten een positieve perceptie hebben van de edumetrische kwaliteit van de praktijkbeoordelingen. Ze zijn tevreden over de toepasbaarheid van alle criteria. Toch gaven de docenten en studenten enkele verbeterpunten aan.

Het beoordelingsformulier kan verbeterd worden door meer duidelijke beschrijvingen van de competenties en een betere schaal voor de normering. Verder is het, volgens de docenten, nodig dat de beoordeling een gezamenlijke verantwoordelijkheid wordt, terwijl de verantwoordelijkheid nu nog bij de docenten ligt. De studenten sluiten hierbij aan door een verzoek om meer zelfsturing.

Dankbetuiging

Dit onderzoek had niet uitgevoerd kunnen worden zonder de enthousiaste hulp van de GL-docenten en studenten. Een bijzonder woord van dank gaat uit naar Jan Marten Praamsma die de opdracht gaf en altijd bereid was mee te denken en helpen. Liesbeth Baartman wordt hartelijk bedankt voor het beschikbaar stellen van de kwaliteitsmeter. Ook Havva Jongen en Marieke van der Schaaf mogen in dit dankwoord niet ongenoemd blijven. Zij hebben keer op keer feedback gegeven, waardoor dit onderzoek tot een goed resultaat gebracht kon worden.

Referenties

- Baartman, L.K.J. (2008). *'Assessing the assessment' Development and use of Quality criteria for Competence Assessment Programmes*. Print Partners Ipskamp.
- Baartman, L.K.J. (n.d.). *CAP-kwaliteitsmeter*. Gevonden op 19 februari 2010, op <http://www.cap-kwaliteit.nl/menu.php>.
- Baartman, L.K.J., Prins, F.J., Kirschner, P.A., & Vleuten, C.P.M. van der (2007). Kwaliteitsmeting van Competentie Assessment Programma's via zelfevaluatie. *Onderwijsinnovatie, 1*, 17 – 26.
- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F.J.R.C. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp.13 – 36). Dordrecht: Kluwer Academic Publishers.
- Diederick, S., & Dochy, F. (2001). New Lines in edumetrics: new forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation, 27*, 307 – 329.
- Dochy, F., & Struyven, K. (2002). *Assessment: betekenis en assessmentvormen*. In F. Dochy, L. Heylen, en H. van de Mosselaer (Eds.). *Assessment in onderwijs*. Utrecht: Lemma.
- Falchikov, N. (2005). *Improving assessment through student involvement*. London: RoutledgeFalmer.
- Gipps, C. (1994). *Beyond testing*. New York: RoutledgeFalmer.
- Gulikers, J.T.M., Bastiaens, T.J., Kirschner, P.A., & Kester, L. (2006). Relations between student perceptions of assessment authenticity, study approaches and learning outcome. *Studies in Educational Evaluation, 32*, 381 – 400.

- Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: expectations and validation criteria. *Educational Researcher*, 23, 13 – 23.
- Messick, S. (1992). *The interplay of evidence and consequences in the validation of performance assessments*. New Jersey: Educational Testing Service Princeton.
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741 – 749.
- Peer, S. van & Beek, H.M. van (2009). *Pocket praktijkstage GL. Lesgeven: beleven of overleven?!*. Ede: Theologie Christelijke Hogeschool Ede.
- Sambell, K., McDowell, L., & Brown, S. (1997). "But is it fair?" An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23, 349 – 371.
- Stokking, K.M. (1998). *Bouwstenen voor onderzoek in onderwijs en opleiding*. Gevonden op 8 februari 2010, op <http://studion.fss.uu.nl/Bouwstenenonline/Bouwstenenonline.html>.
- Rixtel, J. van (n.d.). *Thesistools*. Gevonden op 30 maart 2010, op <http://www.thesistools.nl>.