

Het beoordelingsproces van afstudeeronderzoek bij de opleiding **Huidtherapie**

□ Froukje Jellema

Dit artikel gaat in op het ontwikkelen van een beoordelingsprocedure voor het afstudeeronderzoek bij de hbo-opleiding Huidtherapie. Het team van deze opleiding heeft een beoordelingsprocedure en rubrics ontwikkeld die zowel voor summatieve als voor formatieve toetsing gebruikt kunnen worden. Alle docenten van het team die betrokken zijn bij de beoordeling van het afstudeerrapport, stemmen twee keer per jaar met elkaar af, door hetzelfde afstudeerrapport te beoordelen en te bespreken. Juist het bespreken lijkt goed te werken en te leiden tot meer overeenstemming tussen beoordelaars.

Het afstudeeronderzoek bij de opleiding Huidtherapie

De opleiding Huidtherapie is onderdeel van de Faculteit voor Gezondheid, Voeding & Sport van de Haagse Hogeschool. Huidtherapie is een uniek vak dat alleen in Nederland en Australië als specialisme voorkomt. Onderzoek is een belangrijk onderdeel van de opleiding. In dit artikel gaan we eerst in op drie problemen bij de opleiding, namelijk dat bij het afstudeeronderzoek de beoordelaarsovereenstemming laag was, de beoordelingsprocedure tijdrovend en dat studenten de beoordeling niet transparant vonden. Daarna presenteren we onze oplossingen.

Context

De student voert gelijk in het eerste onderwijsblok

van jaar 1 een klein onderzoek uit. Gedurende de studie worden de opdrachten steeds complexer en moeten ze in een grotere mate van zelfstandigheid worden uitgevoerd. Iedere student rondt de opleiding af met een individueel onderzoek, bestaande uit literatuuronderzoek en praktijkonderzoek, waarin de onderzoekscompetenties op het eindniveau worden getoetst.

Studenten worden bij hun afstudeeronderzoek in groepjes van acht begeleid door een docent. Deze geeft feedback over opzet en aanpak. Studenten worden daarnaast gestimuleerd elkaar feedback te geven. Het onderzoek resulteert in een afstudeerrapport, dat beoordeeld wordt door twee andere docenten dan de begeleider, op de onderdelen 'vorm en taalgebruik' en 'onderzoek'. Beide onderdelen bestaan uit een aantal criteria met zes-puntschalen, die worden omgezet in een cijfer per onderdeel en een definitief eindcijfer op tien-puntschalen. De afstudeerbegeleider beoordeelt het onderdeel 'proces'. Bij het eindcijfer telt het onderdeel 'onderzoek' mee voor 70% en 'vorm en taalgebruik' en 'proces' elk voor 15%. Het rapport is voldoende wanneer het eindcijfer een 5,5 of hoger is en zowel het onderdeel 'vorm- en taalgebruik' als 'onderzoek' voldoende is.

Beoordeling kritisch onder de loep

In studiejaar 2012-2013 was er een drietal redenen om de beoordeling van het afstudeerrapport kritisch te bekijken. Ten eerste zaten beoordelaars niet altijd op één lijn. Het was onduidelijk wanneer je als beoordelaar welke score gaf.

Een ander punt was de procedure om te komen tot het eindoordeel. De twee beoordelaars beoordeelden los van elkaar en gingen vervolgens met elkaar in gesprek om tot een gezamenlijke beoordeling te komen. Dit was tijdrovend. Wanneer het eindcijfer uitkwam tussen de 4,5 en de 6,5, werd er een derde beoordelaar bij gevraagd. Dit kwam vaak voor en vroeg veel docentinzet. Bovendien was de vraag of dit leidt tot een kwalitatief betere beoordeling.



De student voert gelijk in het eerste onderwijsblok een klein onderzoek uit

Ten slotte kwam het signaal van studenten dat zij de beoordeling niet transparant vonden. Beoordelingscriteria waren globaal, met formuleringen als 'taalgebruik', 'structuur' en 'onderzoeksmethode'. Dit gaf studenten weinig steun. Het was bijvoorbeeld onduidelijk wat de opleiding verwachtte met betrekking tot de structuur en wat de student moest doen om een goede prestatie te leveren. De beoordelingscriteria waren daardoor ook moeilijk bruikbaar voor zelf- en peerfeedback.

De opleiding vroeg zich af op welke manier de beoordeling met betrekking tot deze drie punten verbeterd zou kunnen worden.

Aanpak

In het team werd besproken of het criterium voor het betrekken van een derde beoordelaar juist was. De regel was dat de derde beoordelaar ingezet moest worden als er afwijkingen waren rond de cesuur, omdat er dan veel voor de student vanaf hangt. Vanuit beoordelingsperspectief kan beargumenteerd worden dat het criterium voor de inzet van de derde beoordelaar niet moet zijn of de beoordeling rond

de cesuur zit maar de mate waarin beoordelaars het oneens zijn.

Een andere vraag was of de mondelinge afstemming tussen beoordelaars iets toevoegde. De afstudeercoördinator heeft de beoordeling van tien afstudeerrapporten (de eerste tien uit het bestand met scripties van cohort 2009) bekeken om te kijken of de mondelinge afstemming tussen twee beoordelaars leidde tot een andere beoordeling. De ene beoordelaar kan namelijk met goede argumenten de andere beoordelaar tot echt andere inzichten laten komen. In dat geval voegt een dergelijke afstemming iets toe aan de beoordeling, ook al is het tijdrovend. Uit dit onderzoek (zie tabel 1) bleek dat het gesprek echter niets opleverde ten opzichte van het middelen van de twee beoordelingen, aangezien het cijfer na de afstemming vrijwel altijd een score dichtbij het gemiddelde was. En dat terwijl het gesprek veel tijd kostte. Het team vond het overigens wel belangrijk om met elkaar af te stemmen, maar niet op deze manier. De overeenstemming tussen beoordelaars was vaak laag. Wellicht kwam dit door onduidelijkheid van de criteria en van wanneer je als beoordelaar nou welke score gaf. Besloten werd om samen rubrics te

Tabel 1. Scores bij afstemming beoordelaars versus middelen, cohort 2009

Student	Beoordelaar 1	Beoordelaar 2	Cijfer na afstemming	Gemiddelde van score 1 en 2
1	7,4	6,5	7,0	7,0
2	7,5	7,8	7,5	7,7
3	7,6	8,2	7,9	7,9
4	4,6	4,9	4,9	4,8
5	7,2	6,6	6,9	6,9
6	8,5	7,2	7,7	7,9
7	7,6	7,9	7,7	7,8
8	6,3	7,3	6,5	6,9
9	6,9	6,8	6,8	6,9
10	6,2	5,2	5,6	5,7

Studenten vonden de beoordeling niet transparant



ontwikkelen. Deze kunnen behalve voor summatieve beoordeling ook gebruikt worden voor formatieve beoordeling: de docent kan gerichte feedback geven en studenten kunnen de rubrics gebruiken voor zelf- en peerfeedback. Ze leiden tot een hogere transparantie. De criteria werden 'verdeeld' op basis van expertise: de communicatiedocenten dachten na over vorm- en taalcriteria, onderzoeksdocenten over onderzoekscriteria en de studieloopbaanbegeleiders over procescriteria.

Welke oplossing is gevonden?

De nieuwe oplossing heeft een aantal elementen:

a) Nieuwe procedure

De beoordelaars gaan niet meer met elkaar in gesprek, maar hun beoordelingen worden gemiddeld. Er komt een derde beoordelaar bij wanneer de scores voor het eindcijfer (gemeten op een tien-puntschaal) van de twee beoordelaars meer dan één punt van elkaar afwijken. Het eindcijfer is in dat geval het gemiddelde van drie oordelen.

b) Formulieren

Er is een Excelformulier gemaakt waarin de scores automatisch worden gemiddeld, waardoor de beoordeling gebruiksvriendelijker is. Het formulier met de gemiddelde scores wordt aan de student gemaild, met feedback van beide beoordelaars. Vanuit het oogpunt van transparantie was het voorstel van de afstudeercoördinator om niet alleen de gemiddelde scores van de twee beoordelaars terug te koppelen maar ook de scores van de afzonderlijke beoordelaars. Zodoende zou de student precies zien welke beoordelaar welke score en feedback heeft gegeven. In het docententeam riep een dergelijke individuele terugkoppeling van scores op dat moment nog te veel weerstand op. Docenten vermoedden dat er teveel discussie ontstaat als studenten inzage hebben in de scores van individuele docenten.

c) Rubrics

Voor ieder beoordelingscriterium zijn door docenten voorsnog de scores nul (slecht), drie (voldoende) en vijf (uitstekend) ingevuld. Vanwege gebrek aan tijd is ervoor gekozen om in eerste instantie de uiterste

scores in te vullen, zodat duidelijk is wat absoluut onvoldoende en juist heel goed is. Daarnaast is de score die de cesuur vertegenwoordigt ingevuld, zodat duidelijk is wat de student minimaal moet laten zien om nog een voldoende te hebben voor een criterium. Docenten hebben deze criteria zo concreet mogelijk geformuleerd, zodat het voor hen en studenten duidelijk is wat verwacht wordt (zie tabel 2 en 3 voor een voorbeeld). Op een later moment kunnen ook de overige scores nog ingevuld worden, indien de huidige opzet te weinig houvast zou geven.

d) Afstemming in het team

Twee keer per jaar vindt een teamafstemming plaats. Doel is om zoveel mogelijk overeenstemming te krijgen in de beoordelingen en de interpretatie van de criteria. Iedereen beoordeelt hetzelfde eindrapport met behulp van de rubrics. De afstudeercoördinator doet vooraf een analyse, zodat zichtbaar is bij welke criteria de oordelen uiteenlopen. Iedere beoordelaar ziet bovendien hoe hij of zij scoort ten opzichte van de anderen. Deze analyse is input voor een overleg van anderhalf uur waarbij iedereen aanwezig is. Hieronder een voorbeeld van de afstemming van mei 2014. De beoordeling op het eindcijfer liep uiteen van 7,4 tot 8,1. Volgens de beoordelingsprocedure van de opleiding mogen beoordelaars maximaal één punt verschillen dus dit was een goed resultaat. De deelcijfers voor 'vorm en taalgebruik' en 'onderzoek' liepen meer uiteen, respectievelijk 2,4 en 1,3 waarbij met name één docent een stuk lager zat dan de rest.

Allereerst heeft iedereen de resultaten gekregen en kunnen bekijken. Vervolgens is een aantal opvallende uitkomsten besproken. Aan de hand van onderstaande voorbeelden is te zien hoe bekeken werd op welke manier de criteria en/of de rubrics aangepast konden worden:

- Voorbeeld 1. Discussie/aanbevelingen/eindconclusie: het is onduidelijk wat het verschil is tussen 'eindconclusie' en 'conclusie' (een ander criterium). Besloten is om 'eindconclusie' uit de omschrijving te halen en bij 'discussie' in de rubrics op te nemen dat de student moet aangeven wat de waarde is van het onderzoek. Besloten is om het onderdeel 'discussie/aanbevelingen' in de rubrics te splitsen.



Studenten kunnen de rubrics gebruiken voor zelf- en peerfeedback.

Tabel 2. Voorbeeld rubrics voor één van de onderzoekscriteria

Onderzoek	0	3	5
Discussie en aanbevelingen	<u>Discussie</u> : er wordt niets gezegd over beperkingen van het onderzoek en het onderzoek wordt niet in de context van eerder onderzoek geplaatst. Er wordt niet aangegeven wat de waarde van het onderzoek is, gezien de beperkingen van het onderzoek.	<u>Discussie</u> : er wordt aangegeven wat de beperkingen van dit onderzoek zijn. Daarnaast wordt het onderzoek in de context van eerder onderzoek geplaatst waarbij gebruik wordt gemaakt van relevante bronnen (welke eerder in het rapport zijn aangehaald). Er wordt aangegeven wat de waarde is van het onderzoek, gezien de beperkingen van het onderzoek.	<u>Discussie</u> : beschrijft duidelijk en objectief de beperkingen van dit onderzoek. Daarnaast wordt het onderzoek in de context van eerder onderzoek geplaatst en wordt evt. een verklaring gegeven voor gevonden verschillen. Hierbij wordt gebruik gemaakt van relevante en nieuwe bronnen die nog niet in de rest van het rapport naar voren komen. Er wordt aangegeven wat de waarde is van het onderzoek, gezien de beperkingen van het onderzoek en er is een duidelijke eindconclusie.
	<u>Aanbevelingen</u> : het rapport bevat geen aanbevelingen.	<u>Aanbevelingen</u> : het rapport bevat onderbouwde en concrete aanbevelingen voor nader onderzoek OF inhoudelijke aanbevelingen voor de opdrachtgever/praktijk (1 van de twee mist) OF het rapport bevat aanbeveling voor zowel onderzoek als opdrachtgever/praktijk, maar deze zijn onvolledig onderbouwd/concreet.	<u>Aanbevelingen</u> : het rapport bevat zowel aanbevelingen voor nader onderzoek alsmede inhoudelijke aanbevelingen voor de opdrachtgever/praktijk. Aanbevelingen zijn onderbouwd en concreet.

Voorbeeld 2. Procescriteria: het team vond deze criteria moeilijk meetbaar. Het is lastig om te beoordelen wanneer het proces goed is: moet een student vaak aan de bel trekken of juist zelfstandig werken? Ook is het proces niet altijd zichtbaar, waarbij de ene beoordelaar 'dan maar' vijf punten gaf en de andere drie. Deze criteria worden dus onbetrouwbaar beoordeeld. Besloten is om procescriteria niet langer bij het rapport te beoordelen maar bij het eindgesprek.

Feedback op het proces is vanzelfsprekend wel belangrijk tijdens de begeleiding.

Nu het team twee keer per jaar een afstemming doet, zijn beoordelaars meer op één lijn gaan beoordelen. Bij de beoordelingsronde in juni 2014 was voor ruim 40 afstudeerrapporten geen enkele keer een derde beoordelaar nodig en in juni 2015 slechts één keer. Deels komt dit door de rubrics, maar ook de afstemming heeft meerwaarde. De rubrics zijn bijvoorbeeld

*Nu het team twee keer per jaar een afstemming doet,
zitten beoordelaars meer op één lijn*



Tabel 3. Voorbeeld rubrics voor één van de vorm- en taalcriteria

Vorm en taalgebruik	0	3	5
Vakspecifieke termen	Vakspecifieke termen: de tekst bevat afwisselend jargon van de onderzoekende huidtherapeut en "lekentaal": de beoogde doelgroep is daarmee voor de lezer onduidelijk. Met name onderzoeksterminologie wordt inconsequent en/of fout gehanteerd.	Vakspecifieke termen: de tekst bevat het gewenste jargon van de huidtherapeut; dat van de onderzoeker wordt nog niet consequent en overal goed gehanteerd.	Vakspecifieke termen: de tekst bevat het gewenste jargon van de onderzoekende huidtherapeut. Er is een grote mate van academisch woordgebruik en dit wordt consequent gehanteerd.

ook bij een externe afstemming gebruikt, waarbij drie deskundigen met behulp van de rubrics hetzelfde rapport beoordeelden. Daar lagen de oordelen erg uiteen. Het zou dus kunnen zijn dat rubrics alleen onvoldoende zijn en dat het juist gaat om het gesprek erover.

e) Gebruik in het onderwijs

De afstudeerder weet door de rubrics precies wat de opleiding goed onderzoek vindt en wat verwacht wordt. De rubrics worden ook gebruikt tijdens de groepsbegeleiding, om elkaar feedback te geven. Inmiddels worden ze in eerdere studie jaren gebruikt om studenten een beeld te geven van waar ze gedurende vier jaar naar toewerken.

Verdere ontwikkeling van de beoordeling van het afstudeeronderzoek

De betrokkenen in de opleiding vinden dat de nieuwe rubrics goed werken. Vanzelfsprekend staan deze niet voor altijd vast, maar wordt ze continu geëvalueerd en indien nodig aangescherpt. Het team blijft de afstemming doen, zodat ook nieuwe docenten in het afstemmingsproces worden meegenomen. Studenten worden op dit moment niet bij de afstemming betrokken. Reden hiervoor is dat het de taak van

de docent is om de summatieve beoordeling te doen. Studenten dienen de rubrics te gebruiken voor zelf- en peerfeedback. Het zou interessant zijn om onder studenten te evalueren of zij de rubrics daadwerkelijk gebruiken en of ze de rubrics voldoende bruikbaar en transparant vinden.

Dit jaar kijkt het team kritisch naar de beoordeling van het eindgesprek. Wanneer het rapport is goedgekeurd, geeft de student een presentatie en wordt een eindgesprek gevoerd met twee docenten. Dit is een losstaand studieonderdeel, waarvoor de student een aparte beoordeling en studiepunten krijgt. Ook hier waren tot vorig jaar alleen globale beoordelingscriteria en hadden zowel docenten als studenten weinig houvast. Inmiddels heeft het team rubrics ontwikkeld en de bedoeling is om dit jaar een afstemming in het team te doen waarin alle docenten hetzelfde gesprek beoordelen. Ook daar verwachten we toegevoegde waarde van teamafstemming.

☐ Froukje Jellema is hogeschoolhoofddocent bij de Faculteit voor Gezondheid, Voeding en Sport en afstudeercoördinator van de opleiding Huidtherapie van de Haagse Hogeschool. Zij is als onderzoeker verbonden aan het lectoraat Duurzame Talentontwikkeling. Email: F.A.Jellema@hhs.nl