

Dit artikel is gepubliceerd in Tijdschrift voor Hoger Onderwijs, Jaargang 33, Nr. 2, 2015

Beoordelingsmodel voor eindwerkstukken voor een Faculteit Economie & Management in het hbo

Daan Andriessen

Dr. D. G. Andriessen (daan.andriessen@hu.nl) is werkzaam bij het lectoraat Methodologie van Praktijkgericht Onderzoek van Hogeschool Utrecht

Irene van der Marel

Ir. I. van der Marel (Irene.vandermarel@hu.nl) is werkzaam bij de Faculteit Economie en Management en het lectoraat Methodologie van Praktijkgericht Onderzoek van Hogeschool Utrecht

Beoordelingsmodel voor eindwerkstukken voor een Faculteit Economie & Management in het hbo

Samenvatting

De laatste jaren is er in het hbo veel aandacht voor de legitimering van de diploma's. Het rapport "Vreemde ogen dwingen" heeft een impuls gegeven aan het verbeteren van de afstudeerprocessen bij vrijwel alle opleidingen. Onderdeel van de verbetering is het verbeteren van de modellen waarmee eindwerken worden beoordeeld. Deze modellen moeten bijdragen aan een valide, betrouwbaar en voor de student inzichtelijke beoordeling. Belangrijke vraag daarbij is hoe zo'n model kan worden vormgegeven en of een beoordelingsmodel specifiek is voor een opleiding of hetzelfde kan zijn voor meerdere opleidingen. Dit artikel beschrijft een project aan de Hogeschool Utrecht waarin een gezamenlijk beoordelingsmodel voor afstuderen in het economiedomein is ontwikkeld. Het bleek mogelijk een gezamenlijk beoordelingsformulier te ontwerpen waarin ruimte is voor het toetsen van opleidings specifieke eindkwalificaties. Het artikel beschrijft de manier waarop het beoordelingsmodel is ontwikkeld en het resultaat van deze ontwikkeling en geeft een aantal succesfactoren aan.

Towards a joint scoring guide for the assessment of bachelor theses in higher professional education

Summary

In recent years much emphasis has been placed in Dutch higher professional education on the validity of the diploma's. The committee report "Vreemde ogen dwingen" has been a strong impulse for improving the graduation process. As part of this improvement process institutions are developing new models for assessing bachelor theses of students. Good assessment models contribute to the validity, reliability and transparency of the assessment scores. Key questions are how to develop a good assessment model and to what extent assessment models can be similar across different institutes. This article describes a research into a project at HU Utrecht University of Applied Sciences in which a joint assessment model was developed that also covers institute specific competences. The article describes the process that was used to develop the model, the end results and provides criteria for success.

Trefwoorden

Hbo; afstuderen; beoordelingsformulier; eindkwalificaties; beoordelingsmodel

Keywords

Higher professional education; assessment form; competencies; assessment model

Inleiding

Op 10 juli 2010 berichtte de Volkskrant over onrechtmatigheden bij het toekennen van diploma's bij Hogeschool Inholland. Sindsdien ligt het afstuderen in het hbo onder een vergrootglas en is er veel ondernomen om de kwaliteit ervan te verbeteren. Een van de aandachtspunten is de kwaliteit van de beoordelingen van afstudeerprestaties. Beoordelingsmodellen spelen hierin een belangrijke rol. Dit artikel beschrijft op welke manier een beoordelingsmodel voor het beoordelen van afstudeerwerken kan worden ontwikkeld, aan de hand van een casus bij de Faculteit Economie & Management van de Hogeschool Utrecht.

De perikelen bij Inholland, en later ook bij een aantal andere hogescholen, hebben ertoe geleid dat er grote aandacht is gekomen voor de externe legitimering van examens in het hbo. In opdracht van de Vereniging Hogescholen is hier over door de Commissie Bruijn het rapport "Vreemde ogen dwingen" geschreven (Commissie externe validering examenkwaliteit hoger beroepsonderwijs, 2012). De commissie heeft onder andere geadviseerd aan de Vereniging om te komen tot een landelijk protocol voor de beoordeling van eindwerkstukken. Een expertgroep heeft dit advies uitgewerkt (Expertgroep Protocol, 2014) en geconcludeerd dat één landelijk beoordelingsmodel voor het beoordelen van afstudeerprestaties niet mogelijk is, maar dat het wel belangrijk is dat opleidingen goede beoordelingsmodellen in het afstuderen hanteren. Dit draagt bij aan de validiteit, betrouwbaarheid en inzichtelijkheid van de beoordeling (Sluijsmans, Andriessen, Sporken, Dijkstra, & Vonk, 2014).

Veel opleidingen in het hbo zijn op dit moment bezig om hun beoordelingsmodellen aan te scherpen. Daarbij rijzen allerlei vragen: aan welke eisen moeten beoordelingsmodellen voldoen? Hoe maak je ze inhoudelijk goed en tegelijkertijd ook hanteerbaar? Hoe gedetailleerd moeten ze zijn? In welke mate kunnen beoordelingsmodellen voor meerdere opleidingen hetzelfde zijn?

In dit artikel komen deze vragen aan bod. Ze worden geïllustreerd met een casus van de Hogeschool Utrecht. De Faculteit Economie & Management (FEM) heeft in 2014 gewerkt aan een gemeenschappelijk beoordelingsmodel voor alle 13 opleidingen van de faculteit. Achtereenvolgens gaan we in op de rol van beoordelingsmodellen in afstudeerprogramma's. Vervolgens wordt de aanpak van de FEM beschreven, de eisen die in deze casus aan het beoordelingsmodel zijn gesteld, en het resultaat waar die aanpak toe heeft geleid. Op basis van die ervaringen wordt een aantal geleerde lessen en aanbevelingen geformuleerd voor andere opleidingen in het hbo.

De rol en vorm van beoordelingsmodellen

Tijdens het afstuderen wordt vastgesteld of de student beroepsbekwaam is. De eisen daarvoor zijn vastgelegd in de eindkwalificaties van de opleiding. In het hbo is er in toenemende mate sprake van competentiegericht onderwijs: onderwijs gericht op het vermogen om theoretische en praktische kennis te integreren en te leren van ervaring (Atkins, 1995). Deze verandering vraagt ook een verschuiving van het toetsen richting het geïntegreerd toetsen van kennis, vaardigheden en houding (Baartman, 2008).

Omdat in competentiegericht onderwijs veelal een scala aan competenties moet worden getoetst bestaat het afstudeerprogramma van een opleiding vaak uit meerdere toetsonderdelen (Dijkstra, Van der Vleuten, & Schuwirth, 2010; Schuwirth et al., 2002; Southgate et al., 2001). In het hbo zijn die onderdelen over het algemeen performance assessments (Straetmans, 2006) oftewel beroepsopdrachten; dat wil zeggen opdrachten waarin onder natuurgetrouwe omstandigheden een realistische beroepstaak wordt uitgevoerd.

De prestatie die de student binnen de beroepstaak moet leveren wordt gebruikt om te toetsen of de student over de vereiste competenties beschikt. Dit wordt beoordeeld op basis van de kwaliteit van de geleverde prestatie waarin zowel het proces als het product wordt betrokken. Bij iedere beroepsopdracht levert de student een of meerdere prestaties. In het afstu-

deurprogramma van een opleiding zijn dat vaak een beroepsproduct, een verantwoordingsverslag, een presentatie en een verdediging.

Deze prestaties moeten worden beoordeeld door een deskundige examiner. Beoordelen is zelden objectief te maken (Expertgroep Protocol, 2014; Straetmans, 2006) en de kwaliteit van de beoordeling is sterk afhankelijk van de kwaliteit van de examiner (McMullan et al., 2003). Om de examiner bij het beoordelen te ondersteunen werken opleidingen met beoordelingsmodellen (Van Berkel, 2012). In navolging van de Expertgroep Protocol definiëren wij een beoordelingsmodel als “Een model dat de kwaliteit specificiert waarop bij een afstudeerprestatie product en proces beoordeeld worden, middels prestatiecriteria en beoordelingsschalen en beslisregels.”

Competenties moeten geïntegreerd worden beoordeeld. Onder andere Van der Vleuten & Schuwirth (2005), Oosterheert, Eldik, & Kral (2007) en Van Berkel (2012) noemen dit ‘holistisch beoordelen’. Daarmee wordt bedoeld dat kennis, vaardigheden en houding niet afzonderlijk worden geëvalueerd maar op een geïntegreerde wijze zodat niet alleen de expliciete maar ook de impliciete expertise wordt beoordeeld. Deze terminologie kan verwarrend zijn omdat bij het geïntegreerd beoordelen van competenties zowel een holistisch als een analytisch beoordelingsmodel kan worden gebruikt.

Bij het maken van een beoordelingsmodel moet onder andere worden gekozen voor een holistische of een analytische beoordelingsschaal (Kuhlemeier, 2002; Mertler, 2004). Bij een holistisch beoordelingsmodel geeft de examiner één algemeen oordeel over de kwaliteit van een prestatie. Bij een analytisch beoordelingsmodel worden verschillende dimensies van kwaliteit afzonderlijk beoordeeld waarna de oordelen worden gecombineerd.

Over welk model de voorkeur geniet wordt in de literatuur verschillend gedacht. Mertler(2004) stelt dat een holistisch model de voorkeur heeft bij een toets van complexe vaardigheden waarbij er niet één juist antwoord is en er een overall oordeel moet worden gegeven op meerdere kwaliteitsdimensies. Kuhlemeier (2002) stelt echter dat een holistisch model juist geschikt is bij het beoordelen op één algemene dimensie bij relatief eenvoudige vaardigheden. Volgens hem is een analytisch model juist heel geschikt bij complexe vaardigheden. Straetmans (2006) is ook geen voorstander van holistisch beoordelingen. Deze lopen het gevaar dat ze teveel mening en te weinig feitelijk zijn en ook zijn ze weinig informatief voor studenten. Straetmans erkent wel dat er grenzen zijn aan analytisch beoordelen. Het in detail voorschrijven van wat er beoordeeld moet worden kan tot problemen leiden omdat de kwaliteit van het geheel van een prestatie niet altijd een optelling is van de kwaliteit van de onderdelen: “Objectiveren moet. Maar er moet voor gewaakt worden dat, omwille van de objectiviteit, de beoordeling zich toespitst op de gemakkelijk te objectiveren zaken die noch elk op zich noch gezamenlijk representatief zijn voor de betreffende competentie” (p. 26).

Er moet dus een optimum worden gevonden in de mate van gedetailleerdheid van het beoordelingsmodel (Kuhlemeier, 2002). In onze visie is hier sprake van een continuüm: hoe minder oordelen er worden gevraagd hoe holistischer en hoe meer oordelen hoe analytischer het beoordelingsmodel. In de praktijk van het hbo blijkt dat vaak een punt van discussie onder examinatoren. Als gevolg van bovengenoemde perikelen is er grote druk gekomen op de verantwoording van de zak- en slaagbeslissingen in het hbo. Het gevolg is dat er steeds gedetailleerdere beoordelingsmodellen worden ontwikkeld. Sommige examinatoren hebben daar behoefte aan omdat ze onzeker zijn geworden over hun oordeelsvermogen, anderen vinden dat onwenselijk omdat het gevaar bestaat dat de kwaliteit van het geheel van de prestatie uit het oog wordt verloren.

De casus

De Faculteit Economie & Management (FEM) van de Hogeschool Utrecht verzorgt 13 bachelor opleidingen die vallen binnen het thema Economie en Management. In de afgelopen

twee jaar is een aantal van de opleidingen geaccrediteerd. Andere opleidingen staan aan de vooravond van een accreditatie. Enkele opleidingen aan de FEM zijn in een hersteltraject terechtgekomen door het onvoldoende kunnen aantonen van de gerealiseerde eindkwalificaties. Uit deze hersteltrajecten is de noodzaak tot verbeteren gebleken.

Omdat toetsing en gerealiseerde eindkwalificaties het probleem vormden is gestart met het verbeteren van de beoordeling van de eindkwalificaties. Als eerste stap is gekozen om te werken aan een het verbeteren van het beoordelingsmodel voor het beoordelen van het eindwerkstuk. Het eindwerkstuk is bij alle opleiding onderdeel van de eindopdracht binnen het afstudeerprogramma. De eindopdracht betreft het schrijven, presenteren en verdedigen van een advies aan een opdrachtgever. Bij sommige opleidingen kan de student ook een businessplan voor een eigen onderneming maken. Het management van de faculteit wilde graag toe naar één beoordelingsmodel voor alle opleidingen omdat dit het uitwisselen van examinatoren makkelijker maakt en er dan maar één keer in een ontwikkeltraject geïnvesteerd hoeft te worden. Ook beschikken niet alle opleidingen over de vereiste expertise.

Er is een ontwikkelgroep opgericht voor het maken van een nieuw beoordelingsmodel. De ontwikkelgroep bestond uit een afgevaardigde van iedere opleiding, een projectleider die niet specifiek aan een opleiding verbonden is en twee adviseurs van het lectoraat Methodologie van Praktijkgericht Onderzoek van de hogeschool. Bij de start van het traject is eerst besproken aan welke eisen het nieuwe beoordelingsmodel zou moeten voldoen. De NVAO stelt dat de toetsen en de beoordeling van een opleiding valide, betrouwbaar en voor de student inzichtelijk moeten zijn (NVAO, 2014). Op basis hiervan zijn de volgende eisen bepaald:

1. Uit de eis van validiteit volgt dat de prestatiecriteria in het beoordelingsmodel navolgbaar zijn afgeleid van de te toetsen eindkwalificaties. Deze eis heeft tot gevolg dat beoordelingsmodellen sterk opleidingsgebonden zijn. Iedere opleiding heeft immers andere eindkwalificaties.
2. Het beoordelingsmodel is bedoeld voor het beoordelen van een schriftelijk eindwerkstuk. Bij het beoordelen hiervan hebben examinatoren behoefte aan criteria die gaan over de kwaliteit van schriftelijke producten. Hoewel een oordeel moet worden gegeven over eigenschappen van de student, moeten deze eigenschappen worden geoperationaliseerd in termen van eisen die te stellen zijn aan het schriftelijke werkstuk. Hierdoor wordt voorkomen dat iedere examinerator deze operationalisatie naar eigen inzicht doet. Dit komt de betrouwbaarheid van de beoordeling ten goede.
3. Voor een goede inzichtelijkheid van de beoordeling is het van belang dat het eindoordeel gebaseerd is op een beoordeling van meerdere beoordelingsdimensies. Dit betekent de keuze voor een analytisch beoordelingsmodel. Wat betreft de mate van gedetailleerdheid van het beoordelingsmodel werd de eis gesteld dat het model voldoende prestatiecriteria per beoordelingsdimensie moet hebben om examinatoren houvast te geven. Tegelijkertijd moet het model ook de mogelijkheid geven om naar het grotere geheel te blijven kijken. Dat betekent dat niet alle criteria afzonderlijk worden gescoord en opgeteld, maar dat de examinerator per beoordelingsdimensie een oordeel vormt, gebruikmakend van de criteria.
4. De betrouwbaarheid wordt ook bevorderd wanneer de schaalpunten goed van elkaar te onderscheiden zijn. Dit betekent niet teveel schaalpunten. Uiteindelijk is er gekozen voor vier schaalpunten: onvoldoende – ontwikkelend – geaccepteerd – best practice.
5. De inzichtelijkheid van de beoordeling kan verder worden bevorderd door de examinatoren per beoordelingsdimensie kort te laten verantwoorden op welke manier zij tot hun oordeel zijn gekomen.
6. Tot slot moet het beoordelingsmodel navolgbare beslisregels bevatten om te komen tot een beslissing.

Het ontwikkelproces

Het proces waarlangs het beoordelingsmodel tot stand is gekomen is te reconstrueren aan de hand van de stappen van Allen & Tanner (2006): 1) opstellen van een lijst van eindkwalificaties die in de eindopdracht worden getoetst; 2) beschrijven aan de hand van welke beoordelingsdimensies bepaald kan worden of de studenten deze eindkwalificaties beheersen; 3) opstellen van een lijst met prestatiecriteria per beoordelingsdimensie; 4) beschrijven van score per criteria op het hoogste niveau; 5) beschrijven van de overige niveaus van beheersing en het bepalen van de beslisregels; en 6) verzamelen van feedback van examinatoren en het uitproberen op prestaties van studenten.

Stap 1: Om tot een lijst van eindkwalificaties te komen is onderzoek gedaan naar de eindkwalificaties en afstudeerprogramma's van alle opleidingen en de beoordelingsmodellen die daarin werden gebruikt. Hieruit bleek dat er grote verschillen zijn in eindkwalificaties tussen de 13 opleidingen. Er bleken slechts twee eindkwalificaties voor *alle* opleidingen gelijk te zijn namelijk 'onderzoekend vermogen' en 'communicatie'. De eindkwalificatie 'communicatie' bleek overigens qua beschrijving verschillend. Hierin kwamen vooral de eisen op het gebied van schriftelijke communicatie vrij sterk overeen. De afstudeerprogramma's lijken wel sterk op elkaar. Iedere opleiding wordt afgesloten met eenzelfde beroepsopdracht, het maken van een advies voor een opdrachtgever. Op basis van deze analyse werd besloten dat een gemeenschappelijk beoordelingsmodel voor alle eindkwalificaties niet mogelijk is. In plaats daarvan is ervoor gekozen de focus te richten op het onderzoekend vermogen en de schriftelijke communicatie.

Stap 2: Voor het vaststellen van de beoordelingsdimensies is gekeken naar een beoordelingsmodel voor een vergelijkbare taak (Allen & Tanner, 2006), namelijk naar een rubric die vanaf 2010 in de eerstegraads lerarenopleiding van het toenmalige IVLOS en later COLUU van de Universiteit Utrecht in gebruik is om studenten kwaliteitscriteria te bieden voor het plannen, uitvoeren en communiceren van hun praktijk gericht onderzoek. Hij is ontwikkeld door Maaïke Endedijk en Frans Kranenburg en wordt gebruikt bij de beoordeling van het uiteindelijke artikel binnen de opleiding over het praktijkgerichte onderzoek. Hij is mede gebaseerd op de kwaliteitscriteria voor onderzoek van Oost (Oost & Markenhof, 2002; Oost, 2002a, 2002b).

Op basis hiervan zijn de volgende beoordelingsdimensies bepaald. Deze volgen de stappen in een onderzoek:

1. Kwaliteit van de verkenning van de kwestie
2. Kwaliteit van het literatuurverslag en het theoretisch kader
3. Kwaliteit van methodologie
4. Kwaliteit van analyse
5. Kwaliteit van conclusie en aanbevelingen
6. Kwaliteit van rapportagetechniek en taalgebruik

Daarnaast is een beoordelingsdimensie 'Aanpak kiezen' toegevoegd. Met deze dimensie wordt het proces beoordeeld waarlangs het eindwerkstuk tot stand is gekomen, alsmede de kwaliteit van de reflectie daar op.

Stap 3: Per beoordelingsdimensie zijn prestatiecriteria bepaald. Ook hiervoor is gekeken naar de rubrics van het COLUU maar ook naar de kwaliteitscriteria van Andriessen & Van Weert (2008). Het bleek handig om iedere beoordelingsdimensie uit te splitsen in subdimensies en hierbij een aantal criteria te benoemen. Zo werd dimensie 1 uiteengelegd in:

- 1.1 **Kwaliteit beschrijving situatie en aanleiding**, met als criteria: *volledig en probleemdefinitie aanwezig*;
- 1.2 **Kwaliteit beschrijving doelstellingen van opdrachtgever en onderzoek**, met als criteria: *precies, kritische blik, onderbouwd en relevant*;

1.3 **Kwaliteit vraagstelling**, met als criteria: *precies geformuleerd, ingeperkt, onderbouwd, gedefinieerd, voortkomend uit aanleiding/doelstellingen, zonder tegenstrijdigheden, relevant en gerelateerd aan kennisgebied.*

Deze stap leverde een totaal van 65 criteria op, verdeeld over 23 subdimensies.

Stap 4: De ontwikkelgroep koos ervoor om het voorbeeld van het COLUU te volgen en het beoordelingsmodel ook als rubrics vorm te geven. Een rubrics is een matrix die voor een specifieke taak en set van criteria de niveaus van kwaliteit specificceert (Allen & Tanner, 2006). Het uitsplitsen en beschrijven van de verschillende niveaus in een rubrics draagt bij aan de interne consistentie en de betrouwbaarheid van de beoordeling (Jonsson & Svingby, 2007). Per subdimensie werd het hoogste niveau van kwaliteit beschreven. Zo werd voor subdimensie 1.1 Kwaliteit beschrijving situatie en aanleiding het hoogste niveau ('best practice') omschreven als: *"De bestudeerde situatie en aanleiding tot het onderzochte probleem zijn dusdanig volledig beschreven dat de student komt tot nieuwe inzichten over de doelstelling van het onderzoek. Er is een duidelijke probleemdefinitie gegeven."*

Stap 5: De ontwikkelgroep koos voor vier niveaus van ontwikkeling met bijbehorende schaalpunten: 'onvoldoende', 'ontwikkelen', 'geaccepteerd' en 'best practice'. Voor elke subdimensie en voor elk niveau is de rubrics vervolgens uitgeschreven. Er is voor gekozen om dit uitschrijven niet op het niveau van prestatiecriteria te doen, want dat zou hebben geleid tot een rubrics waarin de examiner 78 oordelen had moeten geven. In plaats daarvan is de rubrics uitgeschreven voor de 19 subdimensies.

Vervolgens zijn de volgende beslisregels afgesproken: In de eerste plaats is bepaald dat de examiner een oordeel geeft voor iedere beoordelingsdimensies en niet per subdimensie. Dit dwingt examinatoren om in hun oordeelsvorming geïntegreerd naar de scores op de subdimensies te kijken, vanuit het idee dat de kwaliteit van het geheel meer kan zijn dan de som der delen. In de tweede plaats is afgesproken dat het eindoordeel wordt bepaald door te kijken naar de individuele scores op de dimensies. Beslisregels hierbij zijn:

- het eindoordeel is onvoldoende wanneer een van de dimensies als onvoldoende is beoordeeld.
- Het werkstuk is een bespreekgeval tussen de examinatoren als minimaal één dimensie een score op 'ontwikkelen' heeft en geen scores heeft op 'onvoldoende'.
- Het werkstuk is voldoende als geen enkele dimensie een score heeft op 'onvoldoende' of 'ontwikkelen'.

Stap 6: In de volgende fase van het project hebben de teams van de verschillende opleidingen aan de FEM het prototype getest in zogenaamde kalibreersessies. De opmerkingen uit de kalibreersessies werden verzameld en verwerkt. De projectleider van de ontwikkelgroep heeft het formulier vervolgens voorgelegd aan het managementteam, waarna het formulier is vastgesteld.

Het resultaat

Het uiteindelijke beoordelingsmodel is bedoeld voor het beoordelen van de werkstukken die studenten in het laatste deel van jaar vier maken. Hierbij is de beroepsopdracht het geven van een advies aan een opdrachtgever (of het maken van een businessplan voor een eigen bedrijf). Alle opleidingen toetsen met deze beroepsopdracht in ieder geval het onderzoekend vermogen van de student. Daarnaast wordt via deze opdracht nog een aantal andere eindkwalificaties getoetst. Welke dat zijn verschilt per opleiding. Verschillende eindkwalificaties leiden in de beroepsopdrachten met name tot verschillende eisen ten aanzien van de analyse. Om die reden bevat het beoordelingsmodel ruimte voor een eigen invulling van de beoordelingsdimensie 'Analyse'. Deze kan worden gebruikt om specifieke eindkwalificaties van een opleiding te toetsen. Tijdens de analyse tonen studenten eindkwalificaties die vaak opleidingsspecifiek zijn zoals een financiële, strategische of marketinganalyse. Daarbij worden

vaak ook vakspecifieke analysemodellen gebruikt. De prestatiecriteria binnen beoordelingsdimensie 4 zijn dan ook in principe per opleiding verschillend.

De opleidings specifieke eindkwalificaties die in dimensie 4 worden getoetst bleken bij veel opleidingen de regulatieve cyclus te volgen: probleemstelling, diagnose, ontwerp, implementatie en evaluatie (Van Strien, 1997). Deze opleidingen hebben die stappen als subdimensies genomen. Hoewel beoordelingsdimensie 4 de ruimte geeft aan iedere opleiding om eigen prestatiecriteria te formuleren, hebben de meeste opleidingen deze dimensie opvallend genoeg niet opleidings specifiek, maar met een aantal opleidingen gezamenlijk ingevuld, bijvoorbeeld door te verwijzen naar zogenaamde Bachelor of Business Administration kernvakgebieden zoals finance of strategic management (Vereniging Hogescholen, 2012).

Iedere beoordelingsdimensie is onderverdeeld in een aantal subdimensies. Vaak verwijzen die naar concrete onderdelen van het werkstuk. Binnen de beoordelingsdimensie methodologie is bijvoorbeeld de beschrijving van de wijze van dataverzameling een onderdeel. Per subdimensie heeft de ontwikkelgroep een aantal prestatiecriteria geformuleerd en deze zijn uitgewerkt in de schaalpunten: 'onvoldoende', 'ontwikkelen', 'geaccepteerd' en 'best practice'. Ter illustratie is in tabel 1 de volledige rubric opgenomen van beoordelingsdimensie 1: Verkennen kwestie. Het volledige beoordelingsformulier is opvraagbaar bij de auteurs.

----- tabel 1 ongeveer hier-----

Toepassen van het model

Het beoordelingsmodel is bij de meeste opleidingen toegepast tijdens het afstuderen in studiejaar 2013-2014, alsmede tijdens kalibreersessies die op regelmatige basis worden gehouden binnen de faculteit. De ontwikkelgroep heeft in het najaar van 2014 de bevindingen van de opleidingen verzameld. De opleidingen vinden het formulier over het algemeen goed bruikbaar. Wel zien ze mogelijkheden tot verbetering:

1. De opleidings specifieke eindkwalificaties zijn nog onvoldoende in het formulier opgenomen. Mede daardoor krijgt het onderzoekend vermogen te veel de nadruk en is het onvoldoende ondersteunend aan het beroepsproduct.
2. In de praktijk wordt de mogelijkheid op het formulier om per dimensie een toelichting te schrijven nog te weinig gebruikt. Hierdoor krijgen studenten weinig feedback en is de beoordeling nog niet volledig transparant.
3. Het formulier is lang en wordt als behoorlijk analytisch ervaren.
4. De benaming 'best practice' voor het hoogste schaalpunt vormt een drempel om dit schaalpunt te gebruiken.
5. Op sommige punten worden de eisen als hoog ervaren voor de bachelor, bijvoorbeeld de eisen op het gebied van het gebruik van literatuur.

De komende tijd gaat de ontwikkelgroep aan de slag met het adresseren van deze punten. Verbeteringen worden ingevoerd in schooljaar 2015 - 2016.

Succesfactoren

Het ontwikkelen van een beoordelingsmodel voor alle opleidingen binnen de faculteit was zeer succesvol. Het ontwikkelteam dat het model heeft gemaakt heeft in januari 2015 de prijs ontvangen voor het beste onderwijsteam van de faculteit. Terugkijkend op het proces zijn de volgende succesfactoren aan te wijzen: Allereerst zat er een groep docenten in het ontwikkelteam met kennis van zaken op het gebied van afstuderen en onderzoek. De docenten werden aangesproken op hun expertise. Ten tweede waren opleidingen in het team vertegenwoordigd. De deelnemers gaven aan dat men blij was met de projectmatige brede aanpak. Het was vanaf de eerste bijeenkomst merkbaar dat er een grote bereidheid was tot het samen werken aan een beoordelingsmodel. Tegelijkertijd kwamen in de bijeenkomsten ook

de verschillen tussen de opleidingen naar voren. Er bleek ook behoefte aan behoud van autonomie als opleiding. Derde succesfactor is dat er ruimte was voor de eigen uniciteit als opleiding. Deze ruimte werd gevonden door het invoegen van een beoordelingsdimensie die naar eigen inzicht kan worden ingevuld.

Bij het ontwikkelen van het model was het merkbaar dat er verschil was in ervaring met het ontwikkelen van dergelijke beoordelingsmodellen tussen de leden van de verschillende opleidingen. De groepsleden uit opleidingen die recent waren geaccrediteerd waren vaardiger dan die uit andere opleidingen. Door het meegeven van opdrachten en het laten geven van presentaties aan en door de leden van de ontwikkelgroep, konden we bereiken dat iedereen snel op hetzelfde niveau kon meedenken.

Een vijfde belangrijke succesfactor voor de kwaliteit van het proces is permanente aanwezigheid van de groepsleden bij de bijeenkomsten. De samenstelling van de groep wisselde weleens, waardoor soms onderwerpen terug kwamen op de agenda waarover al een besluit was genomen.

Ten zesde is het van belang het ontwerp voor het beoordelingsmodel goed te testen. Zoals hierboven beschreven zijn er kalibreersessies gehouden met afstudeerbegeleiders van de afzonderlijke opleidingen met als doel om het beoordelingsmodel te testen. Deze sessies hebben een goede bijdrage geleverd aan de kwaliteit van het beoordelingsmodel en het draagvlak onder de afstudeerbegeleiders.

Tot slot is het belangrijk dat er een permanent proces van verbetering wordt ingericht. De ontwikkelgroep binnen de faculteit is blijven bestaan en werkt aan een stapsgewijze verbetering van het model op basis van feedback uit de opleidingen.

Conclusie

Het verhogen van de kwaliteit van beoordelingsmodellen is een belangrijke stap in de kwaliteitsverbetering van het hbo-onderwijs en daarmee voor het herstellen van het vertrouwen van de maatschappij in het hbo-diploma. Het ontwikkelen van een goed beoordelingsmodel kost veel tijd. Indien mogelijk is het daarom verstandig om dit gezamenlijk te doen. Uitgangspunt daarbij moet zijn: gemeenschappelijk waar het kan en verschillend waar dat moet. Harmonisering tussen verschillende opleidingen van een faculteit is mogelijk, mits deze opleidingen eenzelfde beroepsopdracht hebben en er enige gemeenschappelijkheid is in de eindkwalificaties die daarin worden getoetst. Het is belangrijk dat er ruimte blijft voor valide prestatiecriteria die zijn afgeleid uit de specifieke eindkwalificaties van de opleiding. Deelname van vertegenwoordigers uit alle betrokken opleidingen met verstand van zaken is een belangrijke voorwaarde voor succes. Ook het testen van het concept-beoordelingsmodel in alle opleidingen is een belangrijke stap die de kwaliteit en het draagvlak voor het beoordelingsmodel vergroot.

Een goed beoordelingsmodel is een noodzakelijke maar nog niet voldoende voorwaarde voor valide, betrouwbare en voor de student inzichtelijke beoordeling van eindwerkstukken. Daarnaast moeten de afstudeerbegeleiders en –beoordelaars worden getraind en dienen regelmatig kalibreersessies te worden gehouden waarin beoordelaars gezamenlijk eindwerken beoordelen om te komen tot normvinding en gezamenlijke interpretatie van het beoordelingsmodel.

Referenties

Allen, D., & Tanner, K. (2006). Approaches to Biology Teaching and Learning Rubrics : Tools for Making Learning Goals and Evaluation Criteria Explicit for Both Teachers and Learners. *CBE Life Sciences Education*, 5, 197–203.

- Andriessen, D. G., & Van Weert, T. (2008). Kwaliteitszorg voor praktijkgericht onderzoek aan hogescholen. *Onderzoek van Onderwijs*, 37(1), 27–31.
- Atkins, M. (1995). What should we be assessing. In P. Knight (Ed.), *Assessment for learning in higher education* (pp. 25–33). London: Kogan Page.
- Baartman, L. (2008). “Assessing the assessment”: Development and use of quality criteria for Competence Assessment Programmes. <http://igitur-archive.library.uu.nl/dissertations/2008-0423-200627/UUindex.html>
- Commissie externe validering examenkwaliteit hoger beroepsonderwijs. (2012). *Vreemde ogen dwingen* (pp. 1–80). Den Haag. www.vereniginghogescholen.nl
- Dijkstra, J., Van der Vleuten, C., & Schuwirth, L. W. (2010). A new framework for designing programmes of assessment. *Advances in Health Science Education*, 15, 379–939.
- Expertgroep Protocol. (2014). *Beoordelen is mensenwerk*. Den Haag. www.vereniginghogescholen.nl
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
- Kuhlemeier, H. (2002). *Beoordelingsschalen in praktijktoetsen : hoe ontwikkel en gebruik je ze?* (pp. 1–29). <http://toetswijzer.kennisnet.nl/html/praktijktoetsen/praktijktoetsen3.pdf>
- McMullan, M., Endacott, R., Gray, M. a., Jasper, M., Miller, C. M. L., Scholes, J., & Webb, C. (2003). Portfolios and assessment of competence: a review of the literature. *Journal of Advanced Nursing*, 41(3), 283–294.
- Mertler, C. A. (2004). Designing Scoring Rubrics for Your Classroom. *Practical Assessment, Research & Evaluation*, 7(25), 1–9. https://resources.oncourse.iu.edu/access/content/user/mikuleck/Filemanager_Public_Files/EFL_Assessment/Unit_3/Metler_Designing_scoring_rubrics_for_your_classroom.pdf
- NVAO. (2014). *Beoordelingskaders accreditatiestelsel hoger onderwijs*. http://www.nvao.net/page/downloads/Beoordelingskaders_NL_19_december_2014.pdf
- Oost, H. (2002a). *Een onderzoek rapporteren*. Baarn: HBUitgevers.
- Oost, H. (2002b). *Een onderzoek uitvoeren*. Baarn: HBUitgevers.
- Oost, H., & Markenhof, A. (2002). *Een onderzoek voorbereiden*. Baarn: HBUitgevers.
- Oosterheert, I., Eldik, S. van, & Kral, M. (2007). *Het digitaal portfolio als instrument voor summatieve competentiebeoordeling* (pp. 0–34). Nijmegen: SURF / Hogeschool Arnhem Nijmegen. https://www.han.nl/onderzoek/kennismaken/kenniscentrum-kwaliteit-van-leren/lectoraat/leren-met-ict/publicaties/_attachments/gehele_rapport_dpf_instrument_scb.pdf
- Schuwirth, L. W. T., Southgate, L., Page, G. G., Paget, N. S., Lescop, J. M. J., Lew, S. R., ... Baron-Maldonado, M. (2002). When enough is enough: a conceptual basis for fair and defensible practice performance assessment. *Medical Education*, 36(10), 925–930.

- Sluismans, D., Andriessen, D., Sporken, S., Dijkstra, A., & Vonk, F. (2014). Goed beslissen over beroepsbekwaamheid in het hbo. *Onderwijsinnovatie*, (3), 17–26.
- Southgate, L., Cox, J., David, T., Hatch, D., Howes, A., Johnson, N., ... Turner, J. (2001). The assessment of poorly performing doctors: the development of the assessment programmes for the General Medical Council's Performance Procedures. *Medical Education*, 35(s1), 2–8.
- Straetmans, G. J. J. M. (2006). *Bekwaam beoordelen en beslissen* (p. 28). Enschede: Saxion Hogescholen.
- Van Berkel, A. (2012). Kritische reflectie op competentietoetsen in het hbo. *Onderwijsinnovatie*, 2, 17–26.
- Van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: from methods to programmes. *Medical Education*, 39(3), 309–17.
- Van Strien, P. J. (1997). Towards a methodology of psychological practice, the regulative cycle. *Theory and Psychology*, 7(5), 683–700.
- Vereniging Hogescholen. (2012). *Standaard Bachelor of Business Administration* (pp. 3–4). www.vereniginghogescholen.nl

Tabel 1 Rubrics voor het verkennen van een kwestie

1. Verkennen kwestie				
<p>1.1. Situatie en aanleiding</p> <p><i>Volledig</i> <i>Probleemdefinitie aanwezig</i></p>	<p>1.1 De situatie bij de opdrachtgever en de aanleiding tot het probleem zijn niet omschreven en er is geen duidelijke probleemdefinitie gegeven.</p>	<p>1.1 De situatie bij de opdrachtgever en de aanleiding tot het probleem zijn ten dele omschreven en de probleemdefinitie is niet geheel duidelijk.</p>	<p>1.1 De situatie bij de opdrachtgever en de aanleiding tot het probleem zijn grotendeels omschreven en er is een duidelijke probleemdefinitie gegeven.</p>	<p>1.1. De bestudeerde situatie en aanleiding tot het onderzochte probleem zijn dusdanig volledig beschreven dat de student komt tot nieuwe inzichten over de doelstelling van het onderzoek. Er is een duidelijke probleemdefinitie gegeven.</p>
<p>1.2. Doelstellingen van opdrachtgever en onderzoek</p> <p><i>Precies</i> <i>Kritische blik</i> <i>Onderbouwd</i> <i>Relevant voor opdrachtgever</i></p>	<p>1.2 De doelstelling van het onderzoek is niet of vaag omschreven en/of niet onderbouwd. Er is geen blijk gegeven van een kritische blik ten opzichte van de verzamelde en/of door de opdrachtgever aangereikte informatie. De relevantie voor de opdrachtgever is niet (zichtbaar) gewaarborgd.</p>	<p>1.2. De doelstelling van het onderzoek is niet precies omschreven, maar de kern is zichtbaar. Er is enige onderbouwing. De student geeft nauwelijks blijk van een kritische blik ten opzichte van de verzamelde en/of door de opdrachtgever aangereikte informatie. De waarborging van de relevantie voor de opdrachtgever is niet volledig zichtbaar.</p>	<p>1.2. De doelstelling van het onderzoek is precies geformuleerd en onderbouwd. Er is blijk gegeven van een kritische blik ten opzichte van de verzamelde en door de opdrachtgever aangereikte informatie. De relevantie voor de opdrachtgever blijft gewaarborgd.</p>	<p>1.2. De doelstelling van het onderzoek is precies geformuleerd en onderbouwd. Er is blijk gegeven van een kritische blik ten opzichte van de verzamelde en door de opdrachtgever aangereikte informatie. De relevantie voor de opdrachtgever blijft gewaarborgd.</p> <p>De doelstelling is bovengemiddeld complex.</p>

<p>1.3. Vraagstelling</p> <p><i>Precies geformuleerd</i> <i>Ingeperkt</i> <i>Onderbouwd</i> <i>Gedefinieerd</i> <i>Voortkomend uit aanleiding/doelstellingen</i> <i>Zonder tegenstrijdigheden</i> <i>Relevant</i> <i>Gerelateerd aan kennisgebied</i></p>	<p>1.3 De onderzoeksvraag is omslachtig geformuleerd en/of omvat meer dan 1 vraag en/of er is geen sprake van inperking door uitwerking in deelvragen. Beantwoorden van de deelvragen leidt niet of gedeeltelijk tot beantwoorden van de hoofdvraag. Gebruikte termen worden niet gedefinieerd.</p> <p>De vraagstelling (hoofd en deelvragen) vloeit niet of niet zichtbaar voort uit aanleiding en doelstellingen en er zijn tegenstrijdigheden. De vraagstelling is niet onderbouwd met referenties. Praktijkrelevantie en plaats in het kennisgebied zijn niet zichtbaar gemaakt.</p>	<p>1.3 De onderzoeksvraag is niet precies geformuleerd of niet goed ingeperkt door uitwerking in deelvragen die onduidelijk zijn toegelicht. Beantwoorden van de deelvragen leidt niet zichtbaar tot beantwoorden van de hoofdvraag. Gebruikte termen worden gedefinieerd met enige ruimte voor interpretatie.</p> <p>Er is niet helemaal duidelijk gemaakt hoe de vraagstelling (hoofd en deelvragen) voortvloeit uit aanleiding en doelstellingen en/of er zijn tegenstrijdigheden. De vraagstelling is niet of onduidelijk onderbouwd met referenties waarbij praktijkrelevantie en plaats in het kennisgebied meer aandacht verdienen.</p>	<p>1.3 De onderzoeksvraag is precies geformuleerd en ingeperkt door uitwerking in deelvragen die zijn toegelicht. Beantwoorden van de deelvragen leidt tot beantwoorden van de hoofdvraag. Gebruikte termen worden gedefinieerd.</p> <p>De vraagstelling (hoofd en deelvragen) vloeit logisch voort uit aanleiding en doelstellingen en er zijn geen tegenstrijdigheden. De vraagstelling is onderbouwd met referenties waarbij praktijkrelevantie en plaats in het kennisgebied zichtbaar zijn gemaakt.</p>	<p>1.3 De onderzoeksvraag is precies geformuleerd en goed ingeperkt door uitwerking in deelvragen die zijn toegelicht aan de hand van eerder onderzoek. Beantwoorden van de deelvragen leidt tot beantwoorden van de hoofdvraag. Gebruikte termen worden gedefinieerd en geven inzicht in wat de student eronder verstaat.</p> <p>De vraagstelling (hoofd en deelvragen) vloeit logisch voort uit aanleiding en doelstellingen en er zijn geen tegenstrijdigheden. De vraagstelling is onderbouwd met referenties waarbij praktijkrelevantie en plaats in het kennisgebied zichtbaar zijn gemaakt.</p> <p>De vraagstelling is bovengemiddeld complex.</p>
<p>Beoordeling standaard 1:</p>	<p>Onvoldoende (<5,0)</p>	<p>Ontwikkelend (5,0-5,9)</p>	<p>Geaccepteerd (6,0-7,9)</p>	<p>Best practice (8,0-10,0)</p>
<p>Toelichting:</p>				

|