

Beoordeling, antwoordmodellen en beoordelaarsverschillen

Van iedere toets, project of afstudeerverslag moet van tevoren vastgelegd zijn hoe de student beoordeeld wordt. Het antwoordmodel op een regulier tentamen wordt na afloop van het tentamen openbaar gemaakt en het beoordelingsmodel voor een afstudeerscriptie of een project is al voor de start van de opdracht bekend bij de student, zodat hij weet wat er van hem verwacht wordt.

Antwoordmodellen

Bij het opstellen van het antwoordmodel zijn een aantal punten belangrijk:

- **De goede antwoorden formuleren:** In het geval van een meerkeuzetoets is dit niet moeilijk, maar let er wel goed op dat er steeds slechts één antwoord goed is en niet meerdere. Waak er ook voor dat wanneer je schuift in de volgorde van de alternatieven in de vraag, dat je ook de daarbij behorende sleutel aanpast, want ineens is A niet meer het goede antwoord, maar B.
- **Wat te doen bij een gedeeltelijk goed antwoord:** Bij open vragen of casusvragen is het heel belangrijk om zo nauwkeurig mogelijk vast te leggen wanneer de student nog een gedeelte van het te behalen aantal punten kan krijgen en wanneer niet. Ook moet je vastleggen hoe je met doorreken- of doorredeneerfouten om gaat, zeker wanneer de studenten met het antwoord van de vorige vraag door moeten in een volgende vraag. Lopen ze dan bij een fout antwoord op de eerste vraag het aantal punten van beide vragen mis of sta je het toe om met een fout antwoord door te rekenen, zodat de student toch de volle punten krijgt als hij de goede berekening uitvoert.
- **Becijfering:** Maak voor de student inzichtelijk hoe tot het cijfer wordt gekomen. Hoeveel punten staan er voor iedere (deel)vraag, hoeveel punten zijn er in totaal te verdienen en bij hoeveel punten is de student geslaagd (cesuur)?
- **Uittesten antwoordmodel nieuw tentamen:** Het kan nuttig zijn om de 'werking' van het antwoordmodel bij de beoordelaars te testen door een proefbeoordeling uit te voeren. Eerst wordt dan een steekproef van bijvoorbeeld vijf tentamenuitwerkingen nagekeken door de beoordelaars: elke beoordelaar kijkt dezelfde vijf tentamens na. Gezamenlijk wordt daarna gekeken of het beoordelingsmodel goed werkt: komt elke beoordelaar op ongeveer dezelfde score? Zo nee, op welke plekken moet het model worden aangescherpt om te zorgen dat alle beoordelaars op een vergelijkbare manier beoordelen?

Beoordelingsmodellen voor praktische opdrachten

Ook bij praktische opdrachten moet van tevoren zijn vastgelegd waar je de student op gaat beoordelen. De beoordelingscriteria op dit beoordelingsformulier moeten aan de volgende voorwaarden voldoen:

- De criteria moeten terug te herleiden zijn tot de leerdoelen.
- De criteria moeten concreet en waarneembaar zijn
- Bedenk of je wil dat elk element een eigen weging heeft of dat je meerdere elementen in één beoordelingspunt stopt. Bijv: De onderzoeksmaterialen waren goed gekozen en op de juiste manier ingezet. Dit impliceert dat de goede materialen bepalen per se voorspelt dat ze ook op de goede manier worden gebruikt. Dit hoeft echter niet zo te zijn en dan is dit punt niet eenvoudig met één voldoende of onvoldoende te beoordelen, dus dit beoordelingspunt moet in twee punten gesplitst worden.
Echter, als je deze punten splitst, krijgt elk punt afzonderlijk meer gewicht in de eindcijferbepaling. Als je dit niet wilt, dan kun je ze samengenomen houden en in je beoordeling beide aspecten meenemen.
- Het moet duidelijk zijn wanneer een criterium met een voldoende gescoord wordt en wanneer onvoldoende.

- Naast beoordelingscriteria moeten er ook afspraken gemaakt worden over eventuele weging van onderdelen en de cesuur.
- Bepaal een schaal waarop je wil beoordelen. In essentie beoordeel je alleen op voldoende/onvoldoende, maar deze twee kwalificaties worden doorgaans verder uitgewerkt in zeer onvoldoende, onvoldoende, voldoende, ruim voldoende, goed, uitstekend etc.

Beoordelaarsvalkuilen

Met uitzondering van meerkeuzetoetsen wordt de docent bij het beoordelen van tentamenwerk van studenten altijd in meer of mindere mate beïnvloed door (onterechte) externe factoren. Hieronder staat een lijstje met deze zogenaamde *beoordelaarsvalkuilen*.

De belangrijkste beoordelaarsvalkuilen worden vermeden door het vooraf opstellen van heldere normantwoorden en de tentamens te laten nakijken door beoordelaars die de studenten niet kennen (niet het onderwijs hebben verzorgd), bij voorkeur anoniem en na te kijken per vraag in plaats van per heel tentamen.

- **Mildheid en strengheid:** beoordelaars zijn mensen en mensen zijn nou eenmaal streng of mild in hun oordeel. Dit effect kun je tegengaan door van tevoren duidelijke en heldere normantwoorden te formuleren, zodat de ruimte om streng of mild te oordelen zo klein mogelijk blijft.
- **Horn-effect:** de beoordelaar laat negatieve eigenschappen van de student onterecht meewegen in de beoordeling. Bijvoorbeeld het slechte handschrift van een student, de reputatie van de student, sympathie/antipathie voor de student of stereotypering: de beoordelaar beoordeelt het wiskunde-tentamen van een meisje milder dan dat van een jongen “omdat meisjes minder goed in wiskunde zijn”. Voor een groot gedeelte kunnen deze effecten worden voorkomen door het anonimiseren van de antwoorden voordat er wordt nagekeken.
- **Halo-effect:** de beoordelaar gaat er (onterecht) van uit dat de aanwezigheid van een bepaalde kwaliteit iets zegt over de aanwezigheid van andere kwaliteiten. Bijvoorbeeld een docent die twee vakken geeft en de student bij een goede score op vak 1 coulanter beoordeelt bij het beoordelen van de fouten in een tentamen van vak 2.
- **Sequentie-effect:** de beoordelaar heeft de neiging om een redelijk gemaakt tentamen hoger te waarderen na een reeks zwakker gemaakte tentamens dan wanneer hetzelfde tentamen na een reeks zeer goed gemaakte tentamens was gekomen. Het sequentie-effect wordt beperkt door het tentamen per vraag na te kijken in plaats van per student.
- **Signifisch effect:** als van tevoren niet duidelijk is hoe er moet worden beoordeeld, bestaat het risico dat de beoordelaars ieder aan andere onderdelen van de beoordeling een ander gewicht hangen. De ene beoordelaar vindt lay-out belangrijker terwijl de andere taalgebruik belangrijker vindt.
- **Pygmalion/self-fulfilling prophecy:** een sterke mate van (on)geloof in iemand, waardoor dat (on)geloof realiteit wordt. Als de beoordelaar vanaf het begin geen vertrouwen heeft in de kwaliteit van een student, dan kan dit weerspiegeld worden in de beoordeling, terwijl de student misschien positiever was beoordeeld door een beoordelaar die de student niet kent.
- **Net-als-ik-effect/projectie:** de beoordelaar beoordeelt een student positiever omdat hij gelijkenissen met zichzelf in de student herkent. “Ik was vroeger ook zo in de klas, en met mij is het ook goed gekomen.”
- **Contaminatie-effect:** de beoordelaar gebruikt de beoordeling voor oneigenlijke doelen, bijvoorbeeld als maatstaf voor zijn eigen onderwijs. Lage tentamenresultaten kunnen namelijk indirect ook iets zeggen over de kwaliteit van het onderwijs en dus van de docent/beoordelaar.
- **Leniency-effect:** de beoordelaar geeft uit toegeeflijkheid relatief te positieve oordelen.

- **Strictness-effect/centrale tendentie:** de beoordelaar wil niet te scheutig zijn in positieve of negatieve oordelen en probeert daarom gemiddelde scores toe te wijzen. Dit effect wordt verminderd door het vooraf opstellen van heldere normantwoorden.