

## Een beoordelingsmodel voor het mondeling tentamen literatuur

De betrouwbaarheid en validiteit van het mondeling tentamen literatuur zijn vaak moeilijk aan te tonen. In mijn praktijkonderzoek voor mijn master Nederlands heb ik me gericht op het ontwerpen van een beoordelingsmodel waarmee het mondeling tentamen literatuur betrouwbaar en valide wordt.

Het is bij mondelinge toetsen net zo belangrijk als bij schriftelijke toetsen dat ze betrouwbaar en valide zijn (Bax en Van Berkel 2006). In de praktijk blijkt dat lastig te realiseren. Daarnaast wordt de laatste jaren op steeds meer scholen met RTTI (of een soortgelijk systeem) gewerkt. Zo ook op het Farelcollege, waar RTTI een vast onderdeel is van proefwerken en schoolexamens. Naast het vergroten van de betrouwbaarheid en validiteit was het een uitdaging om RTTI toe te passen in het mondeling tentamen literatuur. Om dat te bewerkstelligen, heb ik hiervoor een beoordelingsschema

ontworpen, waarin de eisen voor validiteit en betrouwbaarheid gewaarborgd zijn. Dat schema kan ingezet worden in de voorbereiding en afname van elk mondeling tentamen.

Validiteit bestaat uit meerdere elementen:

1. inhoudsvaliditeit;
2. cognitieve validiteit;
3. de weging van de verschillende onderdelen ten opzichte van elkaar.

Een toets is inhoudsvalide als deze een goede afspiegeling is van de voorbereide stof. De vragen moeten representatief zijn voor de behandelde stof en er moet een goede spreiding zijn over de verschillende onderdelen (Wools 2011). Bij het ontwerpen van het model is uitgegaan van de manier waarop een schriftelijke toets geconstrueerd wordt. Dat betekent dat de leerdoelen uit de eindtermen zijn omgezet in toetstermen. Toetstermen omschrijven welke inhoud en welk gedrag een leerling moet laten zien op een toets. Op basis van de toetstermen worden vragen geformuleerd. Bij één toetsterm moeten verschillende vragen gesteld kunnen worden (Teelen 2011). De toetstermen zijn uiteindelijk opgenomen in het beoordelingsmodel, zodat elke leerling aan dezelfde toetstermen voldoet, maar de docent de vrijheid heeft om verschillende vragen te stellen. Het model is verder onderverdeeld in de te toetsen domeinen. In het geval van het mondeling tentamen literatuur gaat het om ‘literatuurgeschiedenis’, ‘verhaalanalyse’ en ‘persoonlijke ontwikkeling’.

Een toets is cognitief valide als op alle cognitieve niveaus is getoetst (Wools 2011). De toetstermen moeten op een evenwichtige manier verdeeld kunnen worden over de verschillende cognitieve niveaus, passend bij het niveau van havo-5 of vwo-6 (Drost & Verra 2009). Om cognitieve validiteit te kunnen waarborgen, is gebruikgemaakt van RTTI als taxonomie. Zoals eerder gezegd, wordt er steeds meer met RTTI gewerkt als taxonomie. RTTI onderscheidt voor het middelbaar onderwijs vier cognitieve niveaus:

1. reproductie;
2. toepassing 1;
3. toepassing 2;
4. inzicht.

Reproductievragen zijn erop gericht te testen of een leerling de stof heeft geleerd. Toepassing 1-vragen zijn erop gericht te onderzoeken of een leerling het geleerde kan toepassen in een bekende situatie. Toepassing 2-vragen zijn erop gericht te onderzoeken of een leerling het geleerde kan toepassen in een nieuwe situatie. En inzichtvragen onderzoeken of een leerling zelf bekende en nieuwe zaken met elkaar in verband kan brengen (Drost & Verra 2009). In het examenjaar is het de bedoeling dat bij toetsen 30% van de te behalen punten uit R/T1-vragen komen en 70% uit T2/I-vragen.

RTTI is in het model alleen als taxonomie gebruikt om cognitieve validiteit te bewerkstelligen. De bijbehorende rekenmodule met behulp waarvan gedifferentieerd kan worden en aangepast vervolgonderwijs ingericht kan worden, is niet gebruikt, omdat het mondeling tentamen literatuur een toets met een summatieve functie is: een leerling slaagt of zakt; er is dus geen vervolgonderwijs (Bax & Van Berkel 2006). In het beoordelingsmodel zijn alle toetstermen met RTTI gemarkeerd en is een richtlijn voor de juiste RTTI-verdeling opgenomen. Daarbij is rekening gehouden met het feit dat elk domein toetstermen bevat van verschillende RTTI-niveaus, want de docent moet de verdeling over de cognitieve niveaus af kunnen stemmen op de literatuurlijst van de leerling.

Om een goede weging van de verschillende onderdelen ten opzichte van elkaar mogelijk te maken, is ervoor gekozen om elke bevroegde toetssterm hetzelfde aantal punten op te laten leveren. De toetstermen wegen dus allemaal even zwaar. Dat betekent dat de toetstermen allemaal (ongeveer) even groot zijn. Hierdoor kan een juiste verdeling in vragen worden aangebracht om een juiste RTTI-verhouding te bewerkstelligen. Bij elke toetssterm kan op een vooraf vastgesteld aantal prestatieniveaus gescoord worden. In het model is gekozen voor vijf prestatieniveaus, want dat aantal doet recht aan de verschillende typen antwoorden die leerlingen geven.

Een toets is betrouwbaar als de toets ook echt iets toetst: leerlingen halen bij dezelfde prestatie op een ander moment hetzelfde cijfer (Teelen 2011). Betrouwbaarheid is de basis voor een valide toets (Bax & Van Berkel 2006). De eisen voor inhoudsvaliditeit dragen ook bij aan de betrouwbaarheid van de toets. Binnen de gegeven tijd worden zoveel mogelijk vragen gesteld, verspreid over zoveel mogelijk boeken. Naarmate er meer vragen verspreid worden gesteld, wordt een toets meer betrouwbaar (Bax en Van Berkel 2006), want leerlingen krijgen voldoende kansen om te laten zien en horen wat ze weten en kunnen.

Het feit dat er op vijf vooraf vastgestelde verschillende prestatieniveaus gescoord kan worden, vergroot ook de betrouwbaarheid. Op vijf niveaus is het mogelijk om specifieke kenmerken aan de beschrijvingen toe te voegen die de verschillen in de beoordeling tussen docenten onderling verkleinen. De kans wordt daardoor groter dat verschillende docenten aan een antwoord dezelfde score toekennen. Doordat de beschrijving van de verschillende prestatieniveaus en puntentoekenning zijn opgenomen in het model, kan de beoordeling van iedere toetssterm worden aangegeven door een simpel kruisje. Over het te geven aantal punten hoeft niet te worden nagedacht.

Het model is uiteindelijk een schema geworden dat bestaat uit 8 kolommen, waarin RTTI, toetstermen, schrijfruimte en prestatieniveaus zijn opgenomen (zie de beknopte weergave onderaan). Door dit schema te gebruiken, wordt de aanpak van het mondeling tentamen literatuur meer uniform en kunnen validiteit en betrouwbaarheid

worden vergroot.

In de workshop zal een uitgebreid voorbeeld van het model nader worden toegelicht en wordt onder andere stilgestaan bij de volgende vragen:

- Op welke manier is gewerkt met het model?
- Wat levert het op?
- Worden de betrouwbaarheid en validiteit van de toets met behulp van dit model echt vergroot?

1	2	3	4	5	6	7	8
RTTI	Toetstermen per domein	Schrijfruimte	Onvol.	Matig	Vold.	Vold.+	Uitst.

Figuur 1: Beknopte weergave beoordelingsmodel mondeling tentamen literatuur.

## Referenties

Bax, A. & H. van Berkel (2006). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.

Drost, M.E. & P. Verra (2009). *Toetsen op school. Handboek voor docenten. Sturen van leerprocessen met RTTI*. Bodegraven: docentplus.nl.

Teelen B.V. (2011). *Toetsontwikkeling in de praktijk. Hoe maak ik goede vragen en toetsen?* Wilp: Teelen B.V.

Wools, S. (2011). "De validiteit van toetsscores". In: *Toetsen op school*. Arnhem: Cito.