

Betrouwbaar beoordelen

[Theorie](#)

[Betrouwbaar beoordelen](#)

[Checklist](#)

[DOEN](#)

[Verdieping](#)

Theorie

Betrouwbaar beoordelen

Veel docenten worstelen met de vraag hoe ze toetsen zo betrouwbaar mogelijk kunnen beoordelen. Het 'sec' nakijken en beoordelen van de antwoorden op de MC-toetsen levert geen problemen op, omdat voor die toetsen de correcte antwoorden bekend zijn en de regels voor vaststelling van de cesuur zijn vastgelegd.

Het wordt lastiger bij open vragen en bij toetsen met een nog opener karakter, zoals rapporten en stageverslagen of bijvoorbeeld presentaties of gesprekstechnieken. Hoe kun je er als docent voor zorgen dat je zo betrouwbaar mogelijk beoordeelt?

Mogelijke problemen bij het nakijken

Normverschuiving

De beoordelingsnorm vervlakt of verscherpt ten opzichte van het gemiddelde prestatieniveau van de beoordeelenden.

Bijvoorbeeld: een beoordelaar wordt meer coulant als het percentage onvoldoendes (te) hoog dreigt te worden, of strenger als de gemiddelde kwaliteit tijdens het beoordelen hoger blijkt dan verwacht.

Strategie ter voorkoming: Om ervoor te zorgen dat niet steeds de studenten waarvan de vragen steeds als eerste worden nagekeken, de dupe worden van de normverschuiving bij beoordelaars, is het raadzaam de volgorde van antwoordvellen van studenten regelmatig wijzigen.

Sequentie-effect

Er is een nawerking van voorafgaande beoordelingen van dezelfde schrijfp opdracht.

Bijvoorbeeld: een beoordelaar heeft de neiging om na enkele zwakke schrijfproducten een volgend redelijk product een relatief hoger cijfer te geven dan wanneer dit product volgt op enkele zeer goede schrijfproducten (zie ook Meuffels, 1992).

Strategie ter voorkoming: Het sequentie-effect wordt beperkt door het beoordelen van de antwoorden per vraag (in plaats van per student). Ook het wijzigen van de volgorde van antwoordvellen tijdens de beoordeling is hiervoor een oplossing.

Halo-effect

Een eigenschap van de tekst of van de schrijver heeft een storende uitstraling (halo) op de beoordeling, ook als die eigenschap niet relevant is.

Bijvoorbeeld: een goede tekst wordt laag beoordeeld omdat hij er slordig uitziet, of een matige tekst krijgt een hoog cijfer omdat de schrijver bekend staat als hardwerkend en slim (zie ook Nisbett & DeCamp Wilson, 1977).

Strategie ter voorkoming: Het halo-effect wordt beperkt door de toets anoniem na te kijken. Studenten wordt gevraagd niet hun naam op de toets te zetten, maar hun ID-nummer. Ook kan dit effect worden voorkomen door het inschakelen van een tweede beoordelaar, onafhankelijk van de eigen docent.

Contaminatie-effect

De beoordelaar gebruikt de beoordeling voor oneigenlijke doelen.

Bijvoorbeeld: een laag cijfer als 'straf' voor iemand die altijd te laat komt, een hoog cijfer voor een student die aanmoediging behoeft, of voor een student die met zijn onderzoeksresultaten aansluit bij de theorie van zijn beoordelaar (zie ook Meuffels, 1991).

Strategie ter voorkoming: Het contaminatie-effect is moeilijk oplosbaar zolang er geen scheiding is aangebracht tussen doceren/begeleiden en beoordelen. Intercollegiaal uitwisselen van beoordelingsactiviteiten (docent A beoordeelt de toets van studenten van docent B, en andersom) biedt een oplossing.

Persoonlijke tendens

De individuele normen en gewoontes van de beoordelaar hebben een verhullende uitwerking op de beoordeling.

Bijvoorbeeld: een docent geeft op een schaal van tien nooit hoger dan acht of nooit lager dan een vier.

Strategie ter voorkoming: Het is bijzonder lastig om strategieën te ontwikkelen ter voorkoming van een persoonlijke tendens, omdat meestal persoonlijk (impliciete) overtuigingen ten grondslag liggen aan de tendens. Intercollegiaal uitwisselen van beoordelingsactiviteiten kan mogelijk een persoonlijke tendens aan het licht brengt en daarmee ter discussie worden gesteld.

Strategieën ter voorkoming zijn ook helder uitgewerkt in Berkel & Bax, 2006. Lees hoofdstuk 7.4 “Oorzaken voor verschillen in beoordeling”, hoofdstuk 7.5 “Oplossingen voor beoordelaarsverschillen” en hoofdstuk 7.6 “Modelantwoord” (blz. 121 tm 124).

Inter- en intrabeoordelaarsbetrouwheid

Het gaat hier om interbeoordelaarsbetrouwbaarheid (hoe betrouwbaar is de beoordeling door één docent die alle studentprestaties beoordeelt) en de intrabeoordelaarsbetrouwbaarheid (hoe betrouwbaar zijn de beoordelingen als verschillende docenten ieder een deel van de studentprestaties beoordelen).

Onderzoek naar deze vormen van betrouwbaarheid heeft tot op heden geen beoordelingsmethodiek opgeleverd die voor open vraag toetsen 100% betrouwbaarheid oplevert. Wel zijn er veel methodes om de inter- en intrabeoordelaarsbetrouwbaarheid zo hoog mogelijk te maken.

Checklist

Normverschuiving

Om ervoor te zorgen dat niet steeds de studenten waarvan de vragen steeds als eerste worden nagekeken, de dupe worden van de normverschuiving bij beoordelaars, is het raadzaam de volgorde van antwoordvellen van studenten regelmatig wijzigen.

Sequentie-effect

Het sequentie-effect wordt beperkt door het beoordelen van de antwoorden per vraag (in plaats van per student). Ook het wijzigen van de volgorde van antwoordvellen tijdens de beoordeling is hiervoor een oplossing.

Halo effect

Het halo-effect wordt beperkt door de toets anoniem na te kijken. Studenten wordt gevraagd niet hun naam op de toets te zetten, maar hun ID-nummer of het inschakelen van een tweede beoordelaar, onafhankelijk van de eigen docent.

Contaminatie-effect

Het contaminatie-effect is moeilijk oplosbaar zolang er geen scheiding is aangebracht tussen doceren en beoordelen. Intercollegiaal uitwisselen van beoordelingsactiviteiten (docent A beoordeelt de toets van studenten van docent B, en andersom) biedt een oplossing.

Persoonlijke tendens

Het is bijzonder lastig om strategieën te ontwikkelen ter voorkoming van een persoonlijke tendens, omdat meestal persoonlijk (impliciete) overtuigingen ten grondslag liggen aan de tendens. Intercollegiaal uitwisselen van beoordelingsactiviteiten kan een oplossing bieden.

DOEN

Interbeoordelaarsbetrouwbaarheid

Wanneer je een toets met open vragen hebt afgenomen, kun je met een collega afstemmen dat jullie beiden een aantal dezelfde toetsen nakijken. Achteraf kun je bepalen wat de interbeoordelaarsbetrouwbaarheid is (of het percentage overeenstemming per vraag). Dat geeft inzicht in de mate waarop verschillende beoordelaars eenzelfde toets vergelijkbaar beoordelen. Vervolgens is het interessant om de casussen waar jullie qua beoordeling het meest verschillen te bespreken om zo de beoordelingscriteria meer helder te krijgen.

Verdieping

Erkens, T. Toetsen met open vragen (http://toetswijzer.kennisnet.nl/html/literatuur/openvragen_toetsenho.pdf).

Meuffels, B. (1992). Sequentie-effecten. Tijdschrift voor Taalbeheersing, 14(1), 30-43.

Meuffels, B. (1991). Contaminatie-effecten bij het beoordelen van opstellen. Tijdschrift voor Taalbeheersing, 13(1), 15-29.

Milius, J. (2007). Schriftelijk Tentamineren: een draaiboek voor docenten in het hoger onderwijs. Universiteit Utrecht (Ivlos): Utrecht.

Nisbett, R.E., & DeCamp Wilson, T. (1977). The halo effect: evidence for unconscious alteration of judgments. Journal of Personality and Social Psychology, 35(4), 250-256.

Inter- en intrabeoordelaarsbetrouwheid

“Beoordelingslijsten gebruiken zonder afstemming?” Blog van René van Kralingen
<http://www.gazine.eu/article-detail/beoordelingslijsten-gebruiken-zonder-afstemming/82>

“Toetsen in het Hoger Onderwijs”. Hoofdstukken 7.4, 7.5 en 7.6. Henk van Berkel en Anneke Bax. Bohn Stafleu van Loghum. Houten, 2006.

“Toetstechnische begrippenlijst”. Cito.
<http://www.cito.nl/static/oenw/ttb/beglist1.htm#top>