

Beoordelaarseffecten, waar moet je op letten?

written by Teun Kuipers

24 oktober 2023



Om te zien en laten zien waar leerlingen en studenten staan in hun ontwikkeling, beoordelen we. Dit doen we aan het einde van een (leer)periode, zoals bij een proefwerk of tentamen, maar ook doorlopend in de lessen, bijvoorbeeld met formatief handelen. Bij een digitale meerkeuzetoets hoef je zelf weinig te beoordelen, maar bij een product, presentatie of open vragen is dat anders. Dit beoordelen doe je het best zo objectief mogelijk... maar kunnen we dat wel? Welke factoren spelen een rol en hoe kun je daar rekening mee houden? Een artikel met inzichten en praktische tips.

Wat zijn 'beoordelaarseffecten'?

Het goed [beoordelen van prestaties](#) van studenten levert veel inzichten op, zowel voor de studenten als voor de docenten. Een goede beoordeling is echter niet vanzelfsprekend; het proces van beoordelen is een complex en foutgevoelig cognitief proces (Cronbach, 1990).

Niet relevante factoren in en buiten het werk kunnen het beoordelingsgedrag van de beoordelaar, en daarmee de resultaten, beïnvloeden; de beoordeling weerspiegelt dan niet alleen het werk van de student, maar ook van de beoordelaar. Deze invloeden noemen we beoordelaarseffect.

De mate waarin beoordelaarseffecten kunnen optreden is sterk afhankelijk van het werk; [opdrachten zoals een werkstuk of essay](#) zijn meer open van aard en bieden de student meer ruimte om iets van zijn persoonlijkheid mee te nemen in de uitwerking. De kans op een fout in de beoordeling veroorzaakt door een beoordelaarseffect is bij dit soort opdrachten groter dan bij opdrachten die bijvoorbeeld meer gesloten van aard zijn, zoals invulvragen. Beoordelaarseffecten kunnen echter altijd optreden en het is bijna onmogelijk ze helemaal te voorkomen.

4 veel voorkomende beoordelaarseffecten

1. Het soepelheidseffect (Het leniency-effect)

Het soepelheidseffect is de neiging om studenten (herhaaldelijk) te voorzien van hogere beoordelingen dan passend bij hun prestaties en kan om uiteenlopende redenen optreden. Tegenovergesteld aan dit effect is er de strengheidstendens: de beoordelaar heeft dan juist de neiging om zijn studenten te voorzien van lagere beoordelingen dan passend bij hun prestaties.

In de situatie waarbij verschillende beoordelaars de prestaties van studenten beoordelen kan het gebeuren dat dit verschil in strengheid leidt tot oneerlijke beoordelingsresultaten. Een hoog resultaat is dan niet het gevolg van de prestatie maar een vorm van geluk hebben met de beoordelaar.

2. Inconsistentie

Beoordelaarsinconsistentie is de neiging van een beoordelaar om een of meer beoordelingsschalen op een bepaalde manier toe te passen die niet gelijk is aan de manier waarop de andere beoordelaars dezelfde schalen toepassen. (zie ook: *signifisch-effect*)

Inconsistentie ontstaat vaak door het ontbreken van beoordelingscriteria of een gebrek aan inzicht van de beoordelaar in de beoordelingscriteria.

3. Het halo-effect

Bij het halo-effect laat de beoordelaar zich beïnvloeden door een kenmerk van de student of een aspect van zijn werk die los staat van de aspecten die beoordeeld worden. Denk bijvoorbeeld aan een klik met de student of een positieve indruk van de opmaak van een werkstuk. Het oordeel over het aspect dat niet relevant is voor de beoordeling straalt vervolgens uit naar het aspect dat men eigenlijk wil beoordelen, wat leidt tot een onbetrouwbare beoordeling. De beoordelaar kan zich ook laten beïnvloeden door een minder positief kenmerk van de student of een minder positief aspect van het werk. Het werk wordt dan ondergewaardeerd. Dit noemen we het *horn-effect*.

Het halo-effect treedt op als de beoordelaar niet of in mindere mate over het vermogen beschikt om de verschillende aspecten van het werk te onderscheiden of om daar onderscheid in te maken (Linn & Gronlund, 2000).

4. Restriction-of-range

Het Restriction-of-range effect heeft als gevolg dat de beoordelingen van een beoordelaar zich clusteren rond één bepaald uiterste van de waarderingschaal. Zo kan een beoordelaar onder invloed van het restriction-of-range effect overwegend mild beoordelen (mildheid), overwegend streng beoordelen (strengheid) of overwegend gemiddeld beoordelen. Dit laatste wordt omschreven als *central tendency* (Linn & Gronlund, 2000).

Beoordelaarseffecten voorkomen

Beoordelaarseffecten beïnvloeden de betrouwbaarheid van beoordelingen en hierdoor zijn de scores die voortkomen uit de beoordelingen niet langer objectief. De beoordelingen leveren nog steeds inzichten op maar deze geven een vertekend beeld van de competenties van de studenten. Het mag duidelijk zijn dat [beslissingen nemen op basis van vertekende inzichten](#) verre van wenselijk is en het is dan ook belangrijk om beoordelaarseffecten zoveel mogelijk te voorkomen zodat fouten in de beoordeling worden geminimaliseerd.

Echter, omdat we deze fouten onbewust maken is het lastig om ze volledig te vermijden. Zeker als we snel beslissingen moeten nemen of tijdsdruk ervaren kan het gebeuren dat ons brein ons oplossingsstrategieën (heuristieken) aanreikt, waardoor ons beoordelingsvermogen wordt beïnvloedt en er fouten in zowel de beoordeling als de scores ontstaan. Je kan de kans op het ontstaan van deze fouten echter wel verkleinen.

Totaaloverzicht beoordelaarseffecten

Op de eerste plaats is het belangrijk dat je als beoordelaar je bewust bent van het bestaan van beoordelaarseffecten. In dit artikel hebben we de meest voorkomende effecten beschreven. Wil je een meer uitgebreid overzicht, [klik dan hier](#).

Procedures voor meer objectieve beoordeling

Daarbij kan je dan ook de procedures van nakijken en scoren dusdanig inrichten dat de kans op beoordelaarseffecten wordt verkleind. Denk bijvoorbeeld aan het anoniem nakijken van het werk of het afzonderlijk van elkaar beoordelen en scoren van het werk. Wil je een meer uitgebreide opsomming van adviezen, [klik dan hier](#).

Hopelijk heb je wat aan deze tips en inzichten. Succes met beoordelen!

Literatuur

Van Berkel, H. (2017). Toetsen in het hoger onderwijs. A. Bax, & D. Joosten-ten Brinke (Eds.). Bohn Stafleu van Loghum.

Iramaneerat C., Yudkowsky R. (2006). How good are our raters? Chicago: University of Illinois.

Cronbach, L. J. (1990). Essentials of psychological testing (5th ed.). New York: Harper and Row.

Linn, R. L., & Gronlund, N. E. (2000). Measurement and assessment in teaching (8th ed.). Upper Saddle River, NJ: Prentice-Hall.

Kahneman, D. (2011). Thinking, fast and slow (42th ed.). New York: Farrar, Straus and Giroux.

Wilbrink, B. (2023). Persoonlijke communicatie.

Dit artikel is afkomstig van [Vernieuwenderwijs.nl](https://www.vernieuwenderwijs.nl).

Bekijk de meeste recente versie van ons artikel op onze website.

