

WHITEPAPER

GEBRUIKEN VAN BEOORDELINGSSCHALEN VOOR OPDRACHTEN

BEOORDELINGSSCHALEN VOOR CHECKLISTS EN RUBRICS,
HOE GEBRUIK JE ZE?

24 VOORBEELDEN

BTB
TOETSEN
OP NIVEAU



maart 2017

Harry Molkenboer, toetsdeskundige
bureau@toetsen-beoordelen.nl

06 53 67 47 03

Inhoudsopgave

1. Wat is een beoordelingsschaal	5
2. Uitgangspunten voor beoordelingsschalen	5
2.1. Aantal beoordelingsniveaus	6
2.2. Even of een oneven aantal beoordelingsniveaus	6
2.3. Volgorde van de beoordelingsniveaus	6
2.4. Aanduidingen van de beoordelingsniveaus	7
3. Voorbeelden van beoordelingsschalen voor checklists en rubrics	7
4. Enkele opmerkingen bij de verschillende beoordelingsschalen	9
Referenties	10



1. Wat is een beoordelingsschaal

Beoordelingsschalen vormen de basis om prestaties van studenten in niveaus in te delen, om daarmee de prestaties te scoren. Een beoordelingsschaal kan niet los worden gezien van beoordelingscriteria (wat wordt beoordeeld). Zie ook het WhitePaper over het “Gebruiken van beoordelingsvormen voor opdrachten”.

Toetswijzer van Cito beschrijft een beoordelingsschaal als volgt: “Een beoordelingsschaal is een beoordelingsinstrument waarmee op een glijdende schaal kan worden aangegeven in welke mate kennis, vaardigheden of houdingen bij een leerling aanwezig zijn. De glijdende schaal bestaat uit meerdere punten die een bepaalde positie of rangorde aangeven.”

2. Uitgangspunten voor beoordelingsschalen

Beoordelingsschalen worden binnen toetsing vooral gebruikt bij het beoordelen van opdrachten zoals een presentatie, werkstuk, stageverslag, afstudeerscriptie, gedrags-opdracht en dergelijke door middel van een checklist of een rubric. Beoordelingsschalen bestaan uit twee of meer beoordelingsniveaus met een aanduiding per niveau waarmee beoordelingscriteria binnen checklists en rubrics worden gescoord (zie ook het WhitePaper “Gebruiken van beoordelingsvormen voor opdrachten”).

De opbouw van een beoordelingsschaal is afhankelijk van het doel van de checklist of de rubric. Dit doel wordt medebepaald door:

- de aard van de opdracht (formatief, summatief, gedrags-, product- of procesopdracht);
- een toetsspecificatieplan met de uitgangspunten voor de opdracht;
- een toetsmatrijs (bij opdrachten wordt dit een opdrachtenontwerpschema genoemd) met een passende taxonomie (cognitief, affectief en/of psychomotorisch), leerdoelen en beoordelingscriteria;
- de aard van de te geven feedback (alleen gezakt/geslaagd, globale feedback of meer specifieke feedback).

Om de beoordelingsschalen goed te kunnen gebruiken zijn de volgende aspecten van belang:

- het aantal beoordelingsniveaus;
- een even of een oneven aantal beoordelingsniveaus;
- de volgorde van de beoordelingsniveaus;
- de aanduidingen van de beoordelingsniveaus.

2.1. Aantal beoordelingsniveaus

Uit hoeveel beoordelingsniveaus een beoordelingsschaal bestaat is onder meer afhankelijk van (Sanders, 2011):

- Complexiteit van het beoordelingsobject: hoe complexer, hoe meer niveaus.
- Grootte van de prestatieverschillen tussen de studenten: hoe groter de verschillen tussen studenten, hoe meer niveaus nodig zijn om tussen studenten te kunnen differentiëren.
- Tijd die de beoordelaar in de examensituatie aan de beoordeling kan besteden: hoe minder tijd, hoe minder schaalpunten.

Sanders (2011) adviseert om **bij formatieve toetsen** uit te gaan van minimaal vier tot maximaal zeven beoordelingsniveaus. Drie of minder niveaus zijn te weinig om ontwikkeling goed te onderscheiden. Bij meer dan zeven niveaus lukt het beoordelaars niet meer goed om prestaties van elkaar te onderscheiden. Straetmans (2016) zegt over het aantal beoordelingsniveaus **bij summatieve toetsen** het volgende: "Naarmate er meer prestatieniveaus onderscheiden moeten worden, nemen de problemen om die prestatieniveaus zo helder en objectief mogelijk van elkaar te onderscheiden toe. De literatuur is niet eenduidig over het aantal prestatieniveaus dat onderscheiden moet worden, maar veel auteurs houden het op drie tot maximaal zes prestatieniveaus."

Soms zijn twee niveaus voldoende. Bijvoorbeeld om aan te geven of een student onvoldoende of voldoende heeft gescoord of gezakt of geslaagd is. Soms zijn tien niveaus gewenst. Denk bijvoorbeeld aan de examens binnen het voortgezet onderwijs. Binnen checklists en rubrics voor opdrachten blijken in de praktijk drie tot vijf niveaus veel voor te komen, waarbij vier niveaus dominant is. Bij weinig niveaus, zoals bij twee of drie niveaus, wordt het lastiger om gedifferentieerd feedback op de beoordeling te geven.

2.2. Even of een oneven aantal beoordelingsniveaus

Bij een even aantal beoordelingsniveaus ontbreekt een beoordelingsmidden. De beoordelaar moet dan per beoordelingscriterium kiezen tussen een beoordeling die zich aan de zijde van onvoldoende of aan de zijde van voldoende bevindt. Bij een oneven aantal niveaus is een dergelijk midden er wel en kan dus een min of meer neutraal oordeel worden gegeven (niet slecht, niet goed). De literatuur is niet eenduidig wat aan te bevelen is. Maak zelf bewust de keuze voor een even of een oneven aantal niveaus, afhankelijk van het doel en de aard van de te beoordelen prestaties.

2.3. Volgorde van de beoordelingsniveaus

De volgorde van de niveaus kan van laag naar hoog of van hoog naar laag lopen, naar gelang wat al qua volgorde in de organisatie wordt gebruikt of waar de voorkeur naar uit gaat. Een volgorde van laag naar hoog is het meest natuurlijk en komt het meeste voor.



2.4. Aanduidingen van de beoordelingsniveaus

De aanduidingen van de beoordelingsniveaus en het aantal beoordelingsniveaus moeten aansluiten op de aard van de beoordeling. Een belangrijk aspect hierbij is de nauwkeurigheid van de beoordeling. Hoe nauwkeuriger beoordeeld wordt, hoe meer niveaus nodig zijn en hoe belangrijker het wordt om de niveaus goed van elkaar te scheiden door duidelijke aanduidingen. Daarvoor kan het nodig zijn een toelichting te geven op de aanduidingen.

Aanduidingen beoordelingsniveaus	De student heeft het betreffende beoordelingscriterium naar de mening van de inhoudsdeskundige en getrainde beoordelaar:
Zeer onvoldoende	niet laten zien of bijzonder slecht uitgevoerd
Onvoldoende	wel uitgevoerd, maar op een niveau dat niet voldoet aan de standaard
Voldoende	uitgevoerd op een niveau dat voldoet aan de standaard
Goed	uitgevoerd op een niveau dat hoger is dan de standaard

Tabel 1: Voorbeeld van een toelichting op aanduidingen binnen een beoordelingsniveau

Bij een rubric vormen de indicatoren de toelichting op de aanduidingen (zie het WhitePaper "Gebruiken van beoordelingsvormen voor opdrachten").

3. Voorbeelden van beoordelingsschalen voor checklists en rubrics

Hierna wordt een lijst met 24 voorbeelden van beoordelingsschalen gegeven. Voor het overzicht zijn de schalen gegroepeerd met de dikke contouren. De lijst is niet uitputtend en de niveauaanduidingen dienen slechts als voorbeeld. De voorbeelden zijn uit diverse bronnen verzameld, waaronder uit Van Berkel, Bax en Joosten-ten Brinke (2013).

De verticale groene lijn | in de lijst met voorbeelden geeft de grenswaarde tussen zakken en slagen weer (cesuur) als de schaal de score op een summatieve toets weer zou geven. Waar de verticale groene lijn ontbreekt is geen sprake van zakken of slagen, is grenswaarde niet aan te geven of nog nader te bepalen. Zoals uit de lijst blijkt hoeft de grenswaarde niet altijd in het midden te liggen.

1.	Niet voldaan					Voldaan				
2.	Onvoldoende					Voldoende				
3.	Gezakt					Geslaagd				
4.	Beginner					Gevorderde				
5.	Beginnend			Gevorderd			Expert			
6.	Laag			Gemiddeld			Hoog			
7.	Nog niet vaardig			Vaardig			Zeer vaardig			
8.	Onvoldoende			Voldoende			Goed			
9.	Onder verwacht niveau			Op verwacht niveau			Boven verwacht niveau			
10.	Onvoldoende		Matig		Voldoende		Goed			
11.	Zeer onvoldoende		Onvoldoende		Voldoende		Goed			
12.	Onvoldoende		Voldoende		Goed		Uitstekend			
13.	Beginnend		Gevorderd		Geoefend		Startbekwaam			
14.	Het begin is er		Nog wat te ontwikkelen		Goed op weg		Op niveau			
15.	--		-		+		++			
16.	--		-	-+		+	++			
17.	1		2		3		4			
18.	1	2		3		4		5		
19.	1	2	3	4	5	6				
20.	1	2	3	4	5	6	7			
21.	Onder verwacht niveau			Op verwacht niveau			Boven verwacht niveau			
	4	5	6	7	8	9	10			
22.	1	2	3	4	5	6	7	8	9	10
23.	Zeer slecht	Slecht	Zeer sterk onvoldoende	Sterk onvoldoende	Onvoldoende	Voldoende	Ruim voldoende	Goed	Zeer goed	Uitmuntend
24.	Zeer slecht	Slecht	Matig	Onvoldoende	Bijna voldoende	Voldoende	Ruim voldoende	Goed	Zeer goed	Uitstekend

Tabel 2: Voorbeelden van beoordelingsschalen



4. Enkele opmerkingen bij de verschillende beoordelingsschalen

1. Beoordelingsschaal op een hoog abstract niveau. Het nadeel is dat weinig gedifferentieerd kan worden tussen de niveaus en tussen studenten. Gerichtte feedback is lastig. Toegepast bij bijvoorbeeld het rijexamen.
2. **en 3.** Idem, maar nu met andere aanduidingen.
4. Weer met een andere aanduiding, maar door de aard van de aanduidingen lijkt de schaal gebruikt te worden voor een formatief doel, namelijk inschalen of een student nog beginnend of gevorderd is, om daarop zijn studieprogramma (verder) af te stemmen. Let goed op het woordgebruik voor de aanduidingen.
5. Idem als 4., maar nu met drie niveaus. Ook bij drie ranges zijn nog weinig mogelijkheden om studenten van elkaar te onderscheiden. Bij een formatieve toets hoeft dit een minder groot probleem te zijn dan bij een summatieve toets.
6. Idem als 5, maar nu met ander aanduidingen. Let op dat de betekenis van de aanduidingen voor beoordelaars duidelijk en ook eenduidig is. Wat is laag, midden of hoog? Zorg voor een toelichting wat onder laag, midden en hoog moet worden verstaan.
7. Deze schaal lijkt voor een summatieve toets bedoeld te zijn. Bedenk dat aan de zakkant van de cesuur één scorerange voorkomt en aan de slaagkant twee scoreranges. Denk hieraan bij het vaststellen van de cesuur.
8. Idem als 7. maar met meer voorkomende aanduidingen. Merk op dat de range van 'Goed' de student kan uitdagen meer te doen dan alleen een voldoende halen.
9. Idem als 7. en 8. maar specifiek voor een bepaalde situatie.
10. Met vier niveaus kunnen beoordelingen onderscheidend worden gemaakt en kan ook een duidelijk onderscheid in niveaus worden aangegeven. Sanders (2011) geeft aan om minimaal vier en maximaal zeven niveaus te gebruiken. Merk nog op dat de volgorde van de aanduidingen 'Onvoldoende' en 'Matig' hier omgekeerd is ten opzichte van de beoordelingsschaal genoemd bij 21. Nogmaals, zorg bij de aanduidingen voor eenduidigheid.
11. Idem als 10, maar door de benaming van 'Zeer onvoldoende' in plaats van 'Onvoldoende' wordt aangegeven dat een dergelijke beoordeling zwaar onder de maat is. Om consistent te blijven is 'Matig' aangepast in 'Onvoldoende'.
12. Bedenk dat aan de zakkant van de cesuur één scorerange voorkomt en aan de slaagkant drie scoreranges. Denk hieraan bij het vaststellen van de cesuur. Merk op dat de studenten met deze schaal worden uitgedaagd tot ook het leveren van een 'Uitstekende' prestatie.
13. Zonder de context te kennen is dit een wat onduidelijk schaal. Hij lijkt summatief bedoeld, maar waar ligt dan de cesuur, tussen 'Beginnend' en 'Gevorderd' of tussen 'Gevorderd' en 'Geoefend'? Of is de schaal formatief bedoeld?
14. Deze schaal lijkt te zijn bedoeld voor een formatieve toets.
15. Een schaal met een veel voorkomende aanduidingen, maar minder gebruikelijk bij toetsen. Apart van de schaal zal een verklaring voor de aanduidingen nodig zijn zoals bijvoorbeeld 'Onvoldoende', 'Matig', 'Voldoende' en 'Goed'. Doordat de '-+' ontbreekt is aannemelijk dat een eventuele cesuur ligt tussen '-' en '+'.



16. Idem als 13. maar nu als een vijfpuntsschaal. Zonder de context te kennen is niet duidelijk waar een eventuele cesuur ligt.
17. **tot en met 20:** oplopende schalen in cijfers van 4 tot en met 7. Dit aantal niveaus (4 tot en met 7) maakt volgens Sanders (2011) onderscheid tussen de beoordelingsniveaus mogelijk. De betekenis van de cijfers dient ergens weergegeven te worden, zodat duidelijk is waarvoor een niveau staat. Waar de cesuur ligt als de beoordelingsschaal ook voor een toets wordt gebruikt, dient onderbouwd en eenduidig weergegeven te worden.
21. Combinatie van twee beoordelingsschalen in één (zie 8. in combinatie met een 7-puntschaal, gerelateerd aan de bekende 10-puntsschaal, maar zonder de lage cijfers 1, 2 en 3).
22. De standaard tien niveaus die in het onderwijs worden gebruikt. Het nadeel van deze schaal voor het beoordelen van opdrachten is dat de niveaus 1, 2 en 3 en 9 en 10 veelal weinig worden gebruikt en de mate van detaillering groot is. Het voordeel van deze schaal is dat iedereen een goed gevoel heeft bij de waarde van de niveaus.
23. Vertaling van de cijferniveaus van 22. in tekst volgens een Nederlandse wet uit 1929 (Sanders, 2011).
24. Vertaling van de cijferniveaus van 22. in tekst die volgens Wikipedia (z.j.) in Nederland op scholen wordt gebruikt.

Referenties

Berkel, H.J.M. van, Bax, A.E. & Joosten-ten Brinke, D (2014). Toetsen in het hoger onderwijs. Houten: Bohn Stafleu van Loghum.

Sanders, P. (2011). Toetsen op school. Arnhem: Cito

Straetmans, G.J.J.M. (2016). Procedure voor het systematisch ontwikkelen van theorie- en praktijktoetsen, onderdeel van de hand-out binnen een college van de Masteropleiding Toetsdeskundige.

Wikipedia. <https://nl.wikipedia.org/wiki/Schoolcijfer>. Geraadpleegd op 26 december 2016.

