**Ontwikkelen en beoordelen van rubrics**

**Van**: Een handreiking van het Kennispunt Onderwijs & Examinering.

**Datum**: April 2018

**Contact**: [info@kennispuntmbo.nl](mailto:info@kennispuntmbo.nl)

|  |  |
| --- | --- |
| **Plaats in de PE** | **Procesgebied inkopen/construeren en vaststellen** |
| Datum publicatie | April 2018 |
| Versie | 1.1 |
| Omschrijving | Handreiking voor ontwikkelaars en vaststellers van examens waarbij sprake is van het gebruik van rubrics. De handreiking helpt hen om tot goede rubrics te komen en/of om te beoordelen of de rubrics ‘goed’ zijn. |
| Wijzigingen |  |
| Eigenaar | Kennispunt Onderwijs & Examinering |
| Contact |  |
| Gerelateerde servicedocumenten | Niet van toepassing |

**Inleiding**

***Wat is een rubric?***

Een rubric - ook wel scoring rubric - is een uitgeschreven voorschrift voor het waarderen van een prestatie, meestal gedrag of een product. Een rubric bestaat uit twee elementen: 1) de criteria waarop een prestatie moet worden beoordeeld en 2) de verschillende niveaus van de criteria. Rubrics worden vaak gebruikt om de beoordeling van lastig te beoordelen zaken objectiever (en concreter) te maken.

De rubric heeft dus de vorm van een vooraf ingevulde tabel. Zie figuur 1 voor een voorbeeld.

|  |  |  |  |
| --- | --- | --- | --- |
| **Beoordelingseenheid:……** | | | |
| **Criterium** | **Niveau X** | **Niveau Y** | **Niveau Z** |
| <beschrijving van criterium 1> | <beschrijving van de prestatie bij criterium 1 op niveau X> | <beschrijving van de prestatie bij criterium 1 op niveau Y> | <beschrijving van de prestatie bij criterium 1 op niveau Z> |
| <beschrijving van criterium 2> | <beschrijving van de prestatie bij criterium 2 op niveau X> |  |  |
|  |  |  |  |

*Fig.1: Uiterlijk van een eenvoudige rubric*

Er bestaan verschillende soorten rubrics. Er zijn twee dimensies waarop rubrics kunnen verschillen:

1. Mate van detaillering van het oordeel: holistisch versus analytisch. Een holistische rubric is erop gericht om een prestatie in één score uit te drukken, door alle onderliggende criteria tegelijkertijd te beoordelen. Een analytische rubric beoordeelt een prestatie afzonderlijk aan de hand van meerdere beoordelingscriteria.
2. Doel: generieke of meer (taak)specifieke beoordeling: taak specifieke rubrics bevatten criteria die speciaal geformuleerd zijn voor de beoordeling van een bepaalde afgebakende taak. Een generieke rubric bevat meer een algemene beschrijving en is daarmee toepasbaar voor meerdere prestaties.[[1]](#footnote-1). Denk bij een taakspecifieke beoordeling bijvoorbeeld aan een specifieke vakvaardigheid, en bij een meer generieke beoordeling een (verzameling van) werkprocessen.

Rubrics worden toegepast in het primair en voortgezet onderwijs, maar je komt ze steeds vaker tegen in het middelbaar en hoger beroepsonderwijs. Ook universiteiten maken gebruik van rubrics. Rubrics worden soms ook in het bedrijfsleven gebruikt als HR-instrument bij het beoordelen en inschalen van medewerkers.

***Rubrics gebruiken voor examens versus rubrics gebruiken voor ontwikkelingsgerichte toetsen***

Waarom zou je rubrics gebruiken in je exameninstrument? Hiervoor vinden we in de literatuur en praktijk, de onderstaande redenen:

* Een rubric kan de objectiviteit van de beoordeling verhogen, doordat er duidelijke, vrij gedetailleerde onderliggende criteria geformuleerd zijn: dit helpt om meer consistentie te creëren tussen beoordelingen, wanneer prestaties door verschillende beoordelaars worden beoordeeld.
* Een rubric verhoogt de transparantie doordat duidelijk is wat er precies van de student verwacht wordt.
* De rubric help bij het objectiveren/ kwantificeren van zaken die subjectief zijn. Het gaat er bijvoorbeeld niet om of een examenproduct mooi is, maar wat maakt dat het mooi is.
* Rubrics kunnen een hulpmiddel zijn om op een onderbouwde en transparante manier tot een oordeel te komen over de prestatie van studenten. De ‘score’ per criterium bestaat immers niet alleen uit bijvoorbeeld een O/V/G, maar ook uit de bijbehorende (vooraf geformuleerde) beschrijving van die O, V of G.
* Een beoordeling door middel van een rubric geeft de student meer informatie over de onderbouwing van het oordeel dan alleen een cijfer of score.

Een groot deel van de eerder genoemde argumenten geldt ook voor het ontwikkelingsgericht toetsen met behulp van rubrics. Aanvullend gelden nog de volgende redenen om te werken met rubrics voor ontwikkelingsgericht toetsen:

* Een rubric maakt voor studenten duidelijk wat er van hen verwacht wordt in verschillende fases, wat het einddoel is (bij voorkeur de meest rechtse kolom) en waar hij/zij staat ten opzichte van het einddoel.
* Een rubric helpt de voortgang te bewaken. Het geeft niet alleen docenten inzicht, maar kan ook studenten en hun ouders inzicht bieden in de voortgang van de student.
* Met gebruik van een rubric kan de docent de student goede feedback geven over de waargenomen prestatie en wat een volgende stap is in zijn/haar ontwikkeling.
* Een rubric is een goed hulpmiddel voor de student zelf om leerervaringen te evalueren en reflectie te stimuleren.

Daarnaast geeft het werken met generieke rubrics informatie die interessant is voor curriculumontwikkeling:

* Wanneer een rubric vakoverstijgend wordt ingezet kan daarmee voor zowel docenten als studenten inzichtelijk worden gemaakt hoe hun ontwikkeling door het hele curriculum is, naast hun ontwikkeling binnen het vak.
* Ook kan het gebruiken van generieke rubrics indicaties geven aan welke onderdelen van het curriculum meer aandacht besteed zou moeten/kunnen worden, wanneer studenten op bepaalde onderdelen systematisch laag scoren.

***Aandachtspunten bij het gebruik van rubrics***

Naast de genoemde voordelen, kent het werken met rubrics ook nadelen:

* Sommige gedrags- of houdingsaspecten zijn nu eenmaal moeilijk meetbaar of kwantificeerbaar te maken omdat er geen objectieve of observeerbare maat voor is. Ter illustratie: denk aan aspecten zoals ‘Bejegent de patiënt op een prettige manier’ of ‘bewaakt nauwgezet de grenzen van zijn rol’. Dat los je niet op met een rubric.
* Het kost vaak veel tijd om goede rubrics te ontwikkelen. Het is vaak een intensief en iteratief proces van ontwikkelen, uitproberen, evalueren, bijstellen, et cetera. Ontwikkelaars geven vaak aan dat het lastig is om te komen tot een rubric waar zij en hun studenten helemaal tevreden over zijn.
* Door elk onderdeel tot in de details te beschrijven zal een deel van de studenten zich berekend opstellen: zij proberen proces te laten zien wat er gevraagd wordt en laten dan mogelijk ‘een kunstje’ zien.

Tot slot: Een gedetailleerd exameninstrument kan de aandacht afleiden van twee andere factoren die essentieel zijn bij de kwaliteit van het beoordelen: het examenproces zelf en de onafhankelijkheid en deskundigheid van de beoordelaars. Met andere woorden: met het gebruik van goede rubrics ben je er nog niet.

***Enkele ervaringen***

Uit onderzoek[[2]](#footnote-2) komt naar voren dat docenten het gebruik van een rubric verschillend ervaren:

* Positieve ervaringen zijn dat het gebruiken van een rubric hen meer het gevoel geeft dat ze objectieve beoordelingen geven en dat daardoor beoordelingen van verschillende beoordelaars beter vergelijkbaar zijn.
* Het inzetten van rubrics stuit soms ook op weerstand. Meestal hangt die weerstand samen met de overtuiging dat het gebruiken van een rubric meer tijd kost dan het beoordelen op de reeds bestaande manier en/of dat assessoren het niet eens zijn met wat er in de rubric staat en het gevoel hebben dat ze daarmee geen valide oordeel kunnen geven over de prestatie van een student (zie ook validiteit).
* Sommige beoordelaars vinden het prettig om juist zo holistisch en algemeen mogelijk te beoordelen. Zij kunnen een rubric ervaren als veel tekst en veel werk.

Ervaringen van studenten met rubrics zijn over het algemeen positiever. Studenten geven aan dat zij het gebruik van een rubric zien als een belangrijke meerwaarde. Zij geven aan dat het hen helpt bij het monitoren en plannen van hun werk, het reflecteren op hun eigen prestatie en het geven en ontvangen van feedback. Daarnaast hebben zij het gevoel beter te presteren wanneer een rubric wordt ingezet. Sommige onderzoeksresultaten suggereren dat het inzetten van een rubric bijdraagt aan betere prestaties van studenten en dieper leren.

Wel is daarbij van belang dat de rubric tijdig beschikbaar is, dat wil zeggen vóórdat studenten iets moeten inleveren of laten zien. Op die manier kan de rubric studenten inzicht geven in wat er verwacht wordt en waar ze dus nog aan moeten werken.

***Kwaliteitseisen rubrics***

Het gebruik van rubrics kent zoals gezegd een aantal voordelen voor de validiteit en betrouwbaarheid van de examinering. Dit gaat echter alleen op als de rubric ‘goed’ is. Straetmans[[3]](#footnote-3) geeft ons in zijn artikel een aantal eisen aan de ontwikkeling van een rubric:

* De te beoordelen prestatie is adequaat geoperationaliseerd;
* De beschrijvingen zijn geformuleerd in observeerbare en objectieve termen;
* Betrokkenen accepteren de inhoud van de rubric.

Om de kwaliteit van rubrics te analyseren hebben Arter en Chappuis (2006) een ‘rubric for rubrics’ ontwikkeld. Deze hebben Johan van Strien en Desirée Joosten-ten Brinke vertaald naar het Nederlands en op een aantal punten aangescherpt (zie tabel 1). De rubric omvat drie hoofdcriteria:

* Validiteit. Dit betreft de mate waarin de inhoud een goede afspiegeling geeft van wat er nodig is om een goede prestatie te leveren en van wat er van studenten verwacht mag worden gezien hun niveau.
* Kwaliteit van de criteria. Dit betreft het aantal, de relevantie van en de samenhang tussen de criteria.
* Kwaliteit van de indicatoren. Dit betreft de invulling van de beoordelingsniveaus van de criteria. Zijn deze concreet en parallel geformuleerd?

De onderstaande tabel[[4]](#footnote-4) is dus in feite een rubric, waarmee de kwaliteit van een andere rubric kan worden beoordeeld.

|  | **Goed (2 pt)** | **Voldoende (1 pt)** | **Onvoldoende (0 pt)** |
| --- | --- | --- | --- |
| **Validiteit** | | | |
| Dekkings-  graad | De inhoud weerspiegelt correct wat er van studenten realistisch gezien mag worden verwacht gezien hun niveau en de leerdoelen. De inhoud van de rubric geeft de best mogelijke weergave van wat er nodig is om een goede prestatie te leveren op de vaardigheid of het product dat wordt beoordeeld. | De inhoud weerspiegelt wat er van studenten realistisch gezien mag worden verwacht gezien hun niveau en de leerdoelen, al geldt dit niet voor alle criteria. De inhoud van de rubric geeft grotendeels een goede weergave van wat er nodig is om een goede prestatie te leveren, maar er zijn punten waarop verbetering nodig is | De inhoud weerspiegelt niet wat er van studenten mag worden verwacht gezien hun niveau en de leerdoelen. De inhoud van de rubric geeft geen goede weergave van wat er nodig is om een goede prestatie te leveren, waardoor niet duidelijk is wat de rubric beoogt te meten. |
| Prestatie-  niveau | Studenten mogen verondersteld worden in staat te zijn het hoogste niveau te bereiken op alle criteria. Om elk criterium op ‘voldoende’ niveau te beheersen moeten studenten een acceptabele prestatie laten zien. | Studenten mogen verondersteld worden in staat te zijn het hoogste niveau te bereiken op alle criteria. Op enkele criteria kan met een zwakke prestatie reeds een voldoende of hoger worden behaald. | De eisen die worden gesteld om het hoogste niveau te bereiken zijn onrealistisch en/of studenten kunnen door een zwakke prestatie op meerdere criteria reeds een voldoende of hoger halen. |
| **Kwaliteit van de criteria** | | | |
| Aantal, samenhang, opbouw | De rubric omvat precies het juiste aantal criteria, zodat de complexiteit van de te leveren prestatie wordt weerspiegeld. De criteria kennen een goede samenhang en logische opbouw. | Het aantal criteria vereist enige aanpassing, zoals het opsplitsen van een afzonderlijk criterium in twee aparte criteria, of het samenvoegen van twee criteria. De criteria kennen verder een goede samenhang, maar de opbouw is niet altijd logisch | De rubric bestaat uit een lange lijst criteria die niet of nauwelijks met elkaar lijken samen te hangen en waarin geen keuzes lijken te zijn gemaakt. De opbouw is niet logisch |
| Weging | Het is duidelijk of de rubric bedoeld is voor formatieve en/of summatieve doeleinden. Indien summatieve doeleinden, dan is in de rubric expliciet aangegeven hoe zwaar elk criterium meetelt, waardoor is af te leiden welke criteria belangrijker zijn dan andere (eventueel in begeleidende tekst). Alle criteria krijgen de nadruk die zij verdienen. | Het is duidelijk of de rubric bedoeld is voor formatieve en/of summatieve doeleinden. Indien summatieve doeleinden, dan is in de rubric weliswaar expliciet (eventueel in begeleidende tekst) aangegeven hoe zwaar elk criterium meetelt en welke criteria belangrijker zijn dan andere, maar sommige criteria krijgen te veel of te weinig nadruk. | Het is niet duidelijk of de rubric bedoeld is voor formatieve en/of summatieve doeleinden. In de rubric wordt niet expliciet (ook niet in begeleidende tekst) aangegeven hoe zwaar elk criterium meetelt en welke criteria belangrijker zijn dan andere. De criteria krijgen hierdoor niet de nadruk die zij verdienen. |
| Onafhankelijk-heid | De criteria zijn onafhankelijk van elkaar, waardoor zij verschillende aspecten meten. Er is geen overlap tussen verschillende indicatoren. Wat hoort bij de ene indicator komt nergens anders terug. | De criteria zijn grotendeels onafhankelijk. Er is enige overlap tussen verschillende indicatoren, maar de rubric is desondanks in staat om verschillende aspecten te meten. | De criteria zijn niet onafhankelijk. Er is veel overlap tussen verschillende indicatoren. Verschillende indicatoren omvatten hetzelfde, waardoor de criteria niet verschillende aspecten meten. |
| Onderscheidend vermogen | Het aantal beoordelingsniveaus per criterium is logisch en toereikend. Er zijn voldoende niveaus om adequaat onderscheid te kunnen maken tussen goede en minder goede studenten en om voortgang te kunnen meten. | Het aantal beoordelings-niveaus per criterium is niet overal toereikend om adequaat onderscheid te kunnen maken tussen goede en minder goede studenten en om voortgang te kunnen meten, maar het aantal is redelijk eenvoudig aan te passen door één niveau toe te voegen dan wel twee niveaus samen te voegen. | Het aantal beoordelingsniveaus is veel te groot of juist te klein om zinvol en betrouwbaar onderscheid te kunnen maken tussen studenten. Grote aanpassingen zijn nodig. |
| **Kwaliteit van de indicatoren** | | | |
| Aanwezig | Elk beoordelingsniveau is voorzien van indicatoren. | Niet alle beoordelingsniveaus zijn voorzien van indicatoren. | Geen van de beoordelings-niveaus is voorzien van indicatoren. |
| Transparantie en parallelliteit | De indicatoren zijn gedetailleerd en concreet genoeg zodat duidelijk is wat de student moet laten zien om een goede prestatie te leveren (ze zijn kwalitatief ipv kwantitatief) en de beoordelaar een juiste beoordeling kan geven.  Er is een logische opbouw van indicatoren over de niveaus (parallelliteit) | Er is een poging ondernomen om indicatoren gedetailleerd en concreet te formuleren, maar sommige bevatten nog enkele vage beschrijvingen. Ze zijn wel grotendeels kwalitatief (en niet kwantitatief)  Er is veelal sprake van parallelliteit, maar bij een aantal criteria is de opbouw niet logisch. | De indicatoren zijn vaag en weinig concreet beschreven, waardoor niet duidelijk wordt wat er van de student wordt verwacht. Of er wordt alleen onderscheid gemaakt tussen beoordelingsniveaus met woorden als ‘zeer’, ‘erg’ en ‘enige’, of ze zijn volledig kwantitatief.  De parallelliteit ontbreekt meestal. |

Samenvattende Invultabel voor het beoordelen van de kwaliteit van rubric ….X…….:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Kwaliteitseisen Rubric** | **Goed** | **Voldoende** | **Onvoldoende** | **Verbeteractie** |
| Validiteit: Dekkingsgraad |  |  |  |  |
| Validiteit: Prestatieniveau |  |  |  |  |
| Kwaliteit van de criteria: Aantal, samenhang, opbouw |  |  |  |  |
| Kwaliteit van de criteria: Weging |  |  |  |  |
| Kwaliteit van de criteria: Onafhankelijkheid |  |  |  |  |
| Kwaliteit van de criteria: Onderscheidend vermogen |  |  |  |  |
| Kwaliteit van de indicatoren: Aanwezig |  |  |  |  |
| Kwaliteit van de indicatoren: Transparantie en parallelliteit |  |  |  |  |

**Stappenplan: het ontwikkelen van een rubric**

Een rubric bestaat uit drie onderdelen:

* beoordelingseenheid (bv. kerntaak, werkproces, resultaat, beroepsproduct)
* de beoordelingscriteria (op basis van omschrijving bij het werkproces, gedragscriteria, kennis of vaardigheid)
* de niveaubeschrijving van de criteria.

Het ontwikkelproces bestaat globaal uit 6 stappen:

*Stap 1: Bepaal de aard van de rubric (zie paragraaf 1). Dit geeft richting aan de verdere uitwerking van de rubric.*

* Moet de rubric holistisch of analytisch zijn?
* Hoe generiek of (taak-)specifiek moet de beoordeling zijn?

*Stap 2: Beoordelingseenheid: wat moet worden beoordeeld?*

* Onderzoek de beoordelingseenheid. Kijk naar het kwalificatiedossier of keuzedeel. Betreft het een kerntaak, werkproces, vaardigheid resultaat of beroepsproduct?
* Bepaal de eenheid die beoordeeld wordt. Is dit één werkproces/vaardigheid of één product, een clustering van werkprocessen of wellicht nóg breder en algemener binnen de opleiding (zoals bij generieke rubrics wordt bedoeld).

*Stap 3: Beoordelingscriteria*

* Welke omschrijvingen uit het kwalificatiedossier / kwalificatiedossier zijn bruikbaar, welke gedragscriteria en welke kennis of vaardigheden? Deze kunnen soms letterlijk worden overgenomen, of met aangepaste formulering. Voor het beperken van het aantal criteria is het aan te raden inhoudelijk gelijke indicatoren bij elkaar voegen.
* Is het nodig om aanvullende beoordelingscriteria te formuleren? Formuleer deze zoveel mogelijk in concreet waarneembaar gedrag. Waaraan kan men zien in hoeverre de student de doelen bereikt heeft?
* Beperk je bij voorkeur tot ongeveer 6 criteria.
* Plaats deze beoordelingscriteria in een logische volgorde van uitvoering.

*Stap 4: Niveaus*

* Bepaal de schaal waarop de beoordelingseenheid per criterium worden gewaardeerd (gescoord). Dat is niet per se dezelfde als de schaal waarop het eindoordeel wordt uitgedrukt. Globaal gezien komen drie typen waarderingen voor:
* Uitdrukking van de kwaliteit van de prestatie op het betreffende criterium. Bijvoorbeeld:
* onvoldoende – voldoende – goed
* Uitdrukking van de mate van gevorderdheid van degene die op dat curriculum een dergelijke prestatie levert. Bijvoorbeeld:
* beginner – gevorderde – expert
* Neutrale schaal: meer punten = beter van kwaliteit of verder gevorderd
* 1 – 2 – 3 – 4
* Bepaal per criterium de niveaus die een student kan behalen, waarbij in de meest rechterkolom het hoogste niveau wordt opgenomen en links het laagste niveau plaatst.
* Werk dan eerst de ‘voldoende’ uit. Dit is de minimumprestatie. Hoe ziet, voor elk criterium, een prestatie eruit die nét voldoende is? Bepaal vervolgens ook 'onvoldoende' en 'goed' (of andere en/of meer waarderingen).
* Beschrijf het prestatie zo kernachtig mogelijk. De niveaus moeten duidelijk van elkaar te onderscheiden zijn en een kwalitatieve beschrijving geven. Vermijd verwijzingen naar andere criteria (dus niet: beter dan...).

*Stap 5: Weging en beoordeling*

* Controleer of de uitwerking van de niveaus niet te uitgebreid is.
* Bepaal cruciale criteria.
* Bepaal of de criteria onderling even belangrijk zijn, of laat hier een weging op los.
* Bepaal hoe de totaalscore berekend/bepaald wordt.
* Controleer de rubric aan de hand van de checklist kwaliteitseisen rubrics (uit het vorige hoofdstuk).

*Stap 6: Afstemming en ingebruikname*

* Leg de rubric voor aan diverse personen, bijvoorbeeld collega-ontwikkelaars, mensen uit het beroepenveld en/of studenten.
* Bespreek de bevindingen en stel waar nodig bij.
* Een rubric is nooit echt af. Als het goed (genoeg) voelt, ga ermee aan de slag.
* Evalueer na een jaar gebruik en stel waar nodig bij.

**Geraadpleegde literatuur**

* Van de Moosdijk, N.N.G. Valide, Betrouwbaar en Transparant Toetsen van Competenties met Rubrics. Open Universiteit Nederland, Heerlen, 2017.
* Van Strien, Johan & Joosten-ten Brink, Desirée. Het beoordelen van de kwaliteit van rubrics. In: Examens - februari 2016 - nr 1.
* Straetmans, Gerard J.J.M. Gaan rubrics ons helpen om beter te beoordelen? In: Examens - november 2015 - nr 4.
* Huisman, Willibrord. Richten en toetsen met rubrics. Radboud Universiteit Nijmegen. 2015. http://www.willibrordhuisman.nl/onderwijsadvies/65/
* Stichting leerplanontwikkeling. Rubrics als beoordelingsinstrument voor vaardigheden. Enschede, 2006. http://www.slo.nl/downloads/rubricsals.pdf
* <https://toetsing.sites.uu.nl/modules/rubrics/rubrics-theorie/>
* <https://www.vernieuwenderwijs.nl/rubrics-klas-zo-ga-er-mee-aan-slag/>
* <https://score.hva.nl/docent/toetscyclus/Paginas/Beoordelen.aspx>
* histoforum.net/webquests/Rubrics\_of\_Rubrieken.doc
* <https://leerling2020.nl/wp-content/uploads/2017/04/Stappenplan.pdf>

1. Straetmans, Gerard J.J.M. (2015). Gaan rubrics ons helpen om beter te beoordelen? [↑](#footnote-ref-1)
2. <https://toetsing.sites.uu.nl/modules/rubrics/rubrics-theorie/> [↑](#footnote-ref-2)
3. Straetmans, Gerard J.J.M. (2015). Gaan rubrics ons helpen om beter te beoordelen? [↑](#footnote-ref-3)
4. Van Strien, Johan & Joosten-ten Brink, Desirée (2016). Het beoordelen van de kwaliteit van rubrics. In: Examens - februari 2016 - nr 1. [↑](#footnote-ref-4)