

WHITEPAPER

GEBRUIKEN VAN BEOORDELINGSVORMEN VOOR OPDRACHTEN

BEOORDELINGSVORMEN VOOR CHECKLISTS EN RUBRICS,
HOE GEBRUIK JE ZE?

5 VOORBEELDEN

BTB
TOETSEN
OP NIVEAU



maart 2017

Harry Molkenboer, toetsdeskundige

bureau@toetsen-beoordelen.nl

06 53 67 47 03

Inhoudsopgave

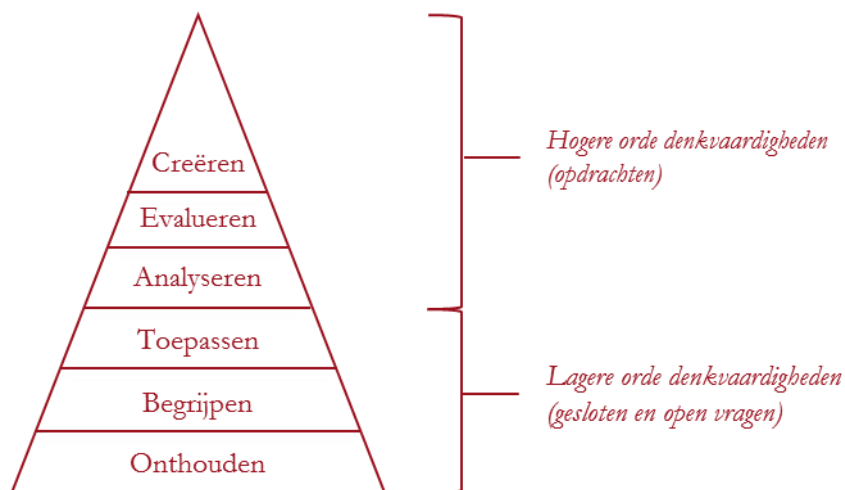
1. Inleiding	5
2. Goede opdrachten beginnen met een goede basis.....	7
3. Ontwikkelen en beoordelen van opdrachten lopen bij opdrachten in elkaar over	9
4. Schema voor het ontwikkelen en beoordelen van opdrachten	10
4.1. Opdrachtbeschrijving	11
4.2. Beoordelingsvoorschrift	11
4.3. Beoordelingscriteria en wat daar nog bij komt kijken	11
5. Vijf voorbeelden van drie hoofdvormen voor het beoordelen van opdrachten	12
5.1. Voorbeeld 1 van 5: Specifieke checklist met twee beoordelingsniveaus	15
5.2. Voorbeeld 2 van 5: Specifieke checklist met meerdere beoordelingsniveaus	17
5.3. Voorbeeld 3 van 5: Globale checklist met meerdere beoordelingsniveaus	18
5.4. Voorbeeld 4 van 5: Analytische rubric.....	20
5.5. Voorbeeld 5 van 5: Holistische rubric	22
6. Overwegingen bij een checklist en een rubric.....	22
6.1. Overwegingen bij een specifieke checklist	22
6.2. Overwegingen bij een globale checklist	23
6.3. Overwegingen bij een rubric.....	24
7. Kiezen van een beoordelingsvorm	25
Referenties	26



1. Inleiding

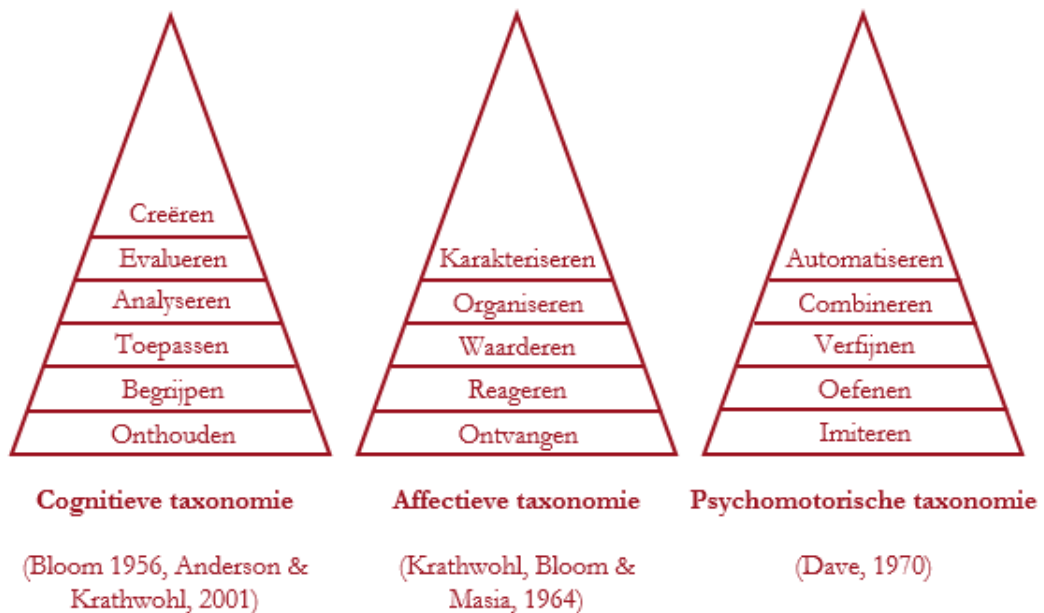
Aanleiding voor dit whitepaper is de situatie dat veel docenten in het beroepsonderwijs (middelbaar en hoger) onzeker zijn over het ontwikkelen en beoordelen van opdrachten zoals een presentatie, werkstuk, stageverslag, afstudeerscriptie, gedragsopdracht en dergelijke. Dat is jammer, omdat als een overzicht bestaat in wat erbij komt kijken, het niet ingewikkeld is.

Opgaven worden meestal toegepast voor het meten prestaties van hogere orde denkvaardigheden.



Figuur 1: Overzicht toetstaken gerelateerd aan de taxonomie van Bloom (1956) en Anderson & Krathwohl (2001)

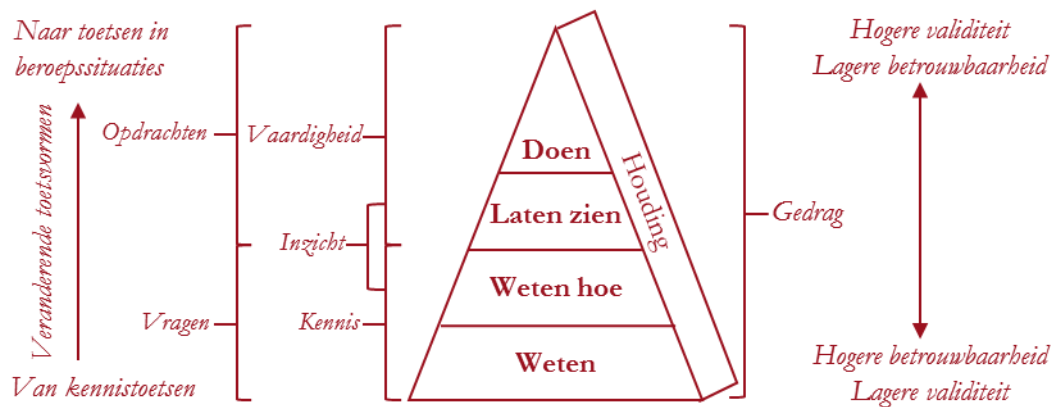
Opgaven zullen echter meestal niet alleen cognitief van aard zijn. Bij opdrachten worden kennis, inzichten, vaardigheden en competenties vaak integratief met elkaar getoetst. Dit betekent dat naast denkvaardigheden ook affectieve en psychomotorische vaardigheden aan de orde zijn. Hiervoor zijn goed toegankelijke en bruikbare taxonomieën beschikbaar (Bloom, Krathwohl, Masia & Dave, 1956, 1964, 1970, 2001).



Figuur 2: Overzicht gedragsniveaus binnen een cognitieve, affectieve en een psychomotorische taxonomie

De Piramide van Miller (1990) is van betekenis voor de keuze van de toetsvormen. Met behulp van de piramide kan ook zichtbaar worden gemaakt dat toetsen met lagere orde denkvaardigheden over het algemeen een lagere validiteit kennen maar betrouwbaarder zijn. En dat toetsen gericht op hogere orde beroepssituaties meer valide maar minder betrouwbaar zijn. Toetsen in of in gesimuleerde beroepssituaties liggen dicht bij de praktijk (meer valide), maar meten vaak maar een of een beperkt aantal beroepsbekwaamheden (minder betrouwbaar). Met toetsen met gesloten vragen gericht op kennis en inzicht kunnen in een korte tijd veel vragen worden aangeboden waardoor een breed spectrum van een kennisdomein kan worden bevraagd (meer betrouwbaar), maar dat alleen het meten kennis over het algemeen verder van de beroepspraktijk afstaat (minder valide) (zie figuur 3)

Omdat praktijkopdrachten in het algemeen minder betrouwbare toetsscores opleveren is het zaak om meetfouten, die elke toets altijd heeft, zoveel mogelijk te verkleinen. Dit is mede mogelijk door opdrachten zorgvuldig te ontwikkelen, waarbij de beoordeling al bij de ontwikkeling zo wordt georganiseerd dat de intra- en interbeoordelaarsbetrouwbaarheid zo groot mogelijk worden gemaakt. Bij opdrachten (maar ook bij open vragen) krijgt de beoordeling van opdrachten bij de ontwikkeling van de opdracht dan ook zeer nadrukkelijk de aandacht.



Figuur 3: De piramide van Miller (1990) in relatie tot toetsvormen en validiteit en betrouwbaarheid van toetsen

2. Goede opdrachten beginnen met een goede basis

Het beoordelen van gesloten en open vragen levert meestal geen of weinig problemen op. Het beoordelen van gesloten vragen is niet meer dan het turven van de goede antwoorden. Het beoordelen van open vragen wordt al wat lastiger. Maar met een goed antwoordmodel, een goede beoordelingsinstructie, gedetailleerde beoordelingscriteria en inhoudelijk deskundige beoordelaars is dit goed te doen.

Het ontwikkelen en beoordelen van opdrachten wordt over het algemeen als lastig ervaren. Dat is jammer want als de voorbereidingen goed zijn uitgevoerd, is ook het ontwikkelen en beoordelen van opdrachten niet heel erg complex. Tot de voorbereidingen behoren het ontwikkelen van een toetsspecificatieplan met de uitgangspunten voor de opdracht, het formuleren van concrete, eenduidige en meetbare leerdoelen en het ontwikkelen van een toetsmatrijs met een passende taxonomie. Bij opdrachten wordt de toetsmatrijs een opdrachtenontwerpschema genoemd. In dit schema vormen de leerdoelen de basis voor een of meer opdrachten. In het schema worden de leerdoelen niet uitgewerkt tot concrete, eenduidige en meetbare sub-leerdoelen, maar gelijk tot (globale) beoordelingscriteria.

Het is verstandig om de beoordelingscriteria gekoppeld aan het leerdoel uit te werken bij de ontwikkeling van het opdrachtenontwerpschema en niet, zoals veel voorkomt, pas bij de ontwikkeling van de opdracht. Ontwikkelaars van opdrachten laten zich bij de beoordelingscriteria dan teveel leiden door de opdracht zelf en vergeten het leerdoel.

Een voorbeeld van een opdrachtenontwerpschema ziet er als volgt uit:

Hoofdleerdoel	<p>De student kan een interview voorbereiden, houden en er een plaatsbaar artikel van maximaal 2.000 woorden over schrijven</p> <p>Het hoofdleerdoel wordt uitgesplitst naar drie leerdoelen. Of de leerdoelen behaald zijn, wordt getoetst met drie opdrachten.</p>
----------------------	---

Leerdoelen	Opdrachten	Beoordelingscriteria	Niveau	Weging
		De student heeft:		
Kan een interview op een goede wijze voorbereiden	Bereid een interview voor	<ul style="list-style-type: none"> ■ Een duidelijk doel voor het artikel gekozen en vastgelegd ■ Relevante literatuur verzameld en geordend ■ Literatuur bestudeerd en de belangrijkste punten samengevat ■ Een goede vragenlijst opgesteld ■ De vragenlijst logisch geordend in relatie tot het te voeren gesprek ■ --- 	Cognitief: creëren Affectief: - Psychomotorisch: -	20%
Kan een interview op een professionele wijze houden	Houd een interview	<ul style="list-style-type: none"> ■ De kennismaking op een persoonlijke wijze gedaan ■ Het doel van het interview goed toegelicht ■ Het interview op een natuurlijke wijze uitgevoerd ■ Tijdens het interviewen overzicht over het interview gehouden ■ Gecontroleerd of alle vragen uit de voorbereiding aan de orde zijn geweest ■ Tijdens het interview samengevat van wat gezegd is en geconcludeerd ■ Afgesproken dat het artikel eerst ter goedkeuring wordt voorgelegd ■ --- 	Cognitief: creëren Affectief: karakteriseren Psychomotorisch: combineren	40%



Leerdoelen	Opgdrachten	Beoordelingscriteria	Niveau	Weging
		De student heeft:		
Kan een artikel schrijven dat aan de normen voldoet die gesteld worden aan een goed artikel	Schrijf een artikel op basis van het interview in maximaal 2.000 woorden	<ul style="list-style-type: none"> ■ Het artikel van een pakkende, duidelijke en uitnodigende titel voorzien ■ In het artikel een lead opgenomen en de lead is een goede samenvatting van het artikel ■ Er voor gezorgd dat de lead uitnodigt verder te lezen ■ Er voor gezorgd dat de inhoud van het artikel correct is en van een passend niveau ■ Een goede conclusie en/of een goed slot opgenomen ■ Een goede opbouw in het artikel gerealiseerd ■ Gebruik gemaakt van tussenkoppen en de tussenkoppen zijn goed gekozen ■ Een helderde schrijfstijl gebruikt ■ De grammatica is goed toegepast ■ De spelling correct uitgevoerd ■ --- 	Cognitief: creëren Affectief: karakteriseren Psychomotorisch: combineren	40%
				100%

Tabel 1: Voorbeeld opdrachtenontwerpschema

3. Ontwikkelen en beoordelen van opdrachten lopen bij opdrachten in elkaar over

Beoordelen is het toekennen van een waardering aan een geleverde prestatie van een student op een toets aan de hand van geobjectiveerde (zover als mogelijk) beoordelingscriteria (Molkenboer, 2015).

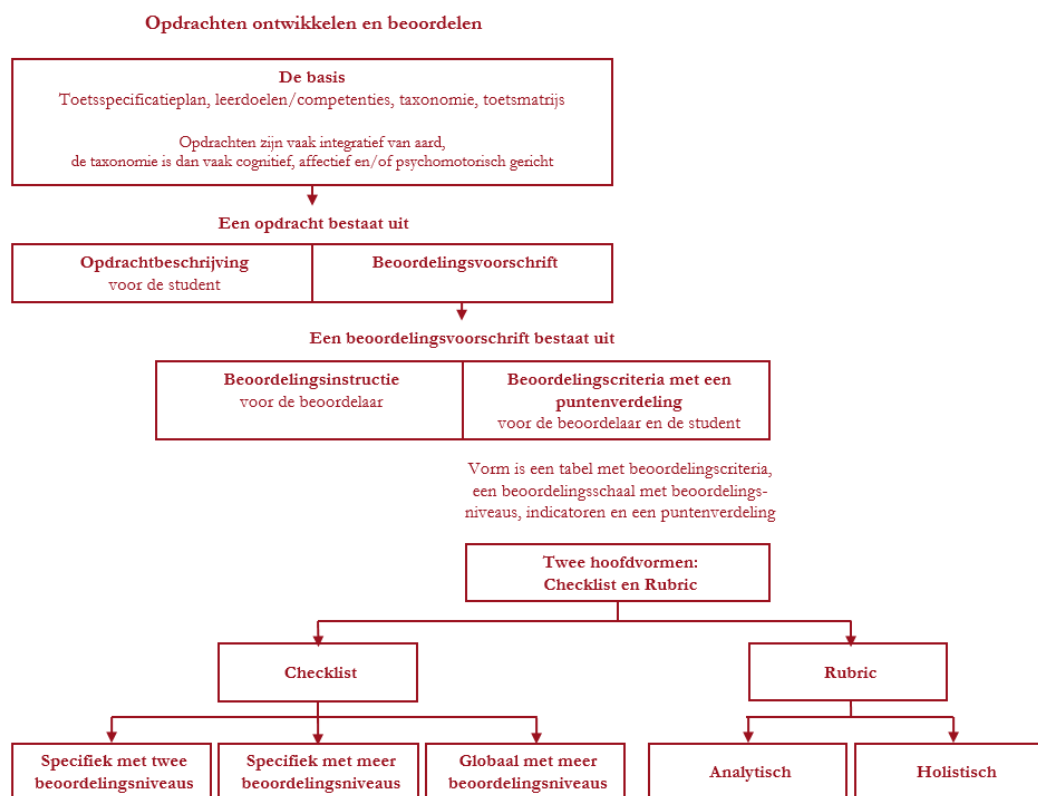
Hoe het beoordelen plaatsvindt is afhankelijk van de gebruikte toetsvorm. Zo wordt een toets in de vorm van bijvoorbeeld een verslag anders beoordeeld dan een toets in de vorm van een gesprek (bijvoorbeeld een adviesgesprek of een gesprek met een patiënt), een gedraging (bijvoorbeeld een presentatie of het bedienen van een hijskraan) of een werkstuk (bijvoorbeeld een technische tekening of een kunstwerk). Voor alle beoordelingen geldt dat een beoordelaar een prestatie moet waarnemen. Aan de waarneming kent de beoordelaar vervolgens een waardering toe.



Om te voorkomen dat beoordelaars teveel subjectief waarnemen en subjectief waarderingen toekennen, wordt vooraf bepaald wat belangrijk is om waar te nemen en hoe een waardering moet worden toegekend. Hierdoor wordt een groot deel van het beoordelen op de tekentafel al gerealiseerd. Als in dit whitepaper hierna wordt gesproken over beoordelen, wordt het beoordelen op de tekentafel bedoeld. Beoordelen in de vorm van waarnemen en waarderen tijdens of aansluitend aan een afname van een opdracht, worden in dit whitepaper niet behandeld.

4. Schema voor het ontwikkelen en beoordelen van opdrachten

Voor het ontwikkelen en beoordelen van opdrachten heeft Bureau voor Toetsen & Beoordelen een schema ontwikkeld. Het schema wordt hierna toegelicht.



Figuur 4: Schema voor het ontwikkelen en beoordelen van opdrachten

Vanuit de basis is duidelijk geworden om wat voor opdracht het gaat. Voor de opdracht wordt een opdrachtomschrijving gemaakt en voor de beoordeling wordt een beoordelingsvoorschrift ontwikkeld.



4.1. Opdrachtbeschrijving

Een goede opdrachtbeschrijving is belangrijk omdat die bepaalt welke prestatie de studenten moeten leveren, hoe het resultaat eruit moet zien, wanneer ze het resultaat moeten leveren en hoe het resultaat wordt beoordeeld (de beoordelingscriteria met de puntenverdeling). Als dit niet volkomen duidelijk is staan de betrouwbaarheid en de validiteit van de opdracht onder druk. Studenten leveren dan mogelijk resultaten op die minder aansluiten bij de gestelde doelen uit de basis.

4.2. Beoordelingsvoorschrift

Een beoordelingsvoorschrift is bedoeld voor de beoordelaars en bestaat uit twee delen: een beoordelingsinstructie en de beoordelingscriteria met een puntenverdeling. De beoordelingscriteria met de puntenverdeling zijn bedoeld voor de beoordelaars en, zoals hiervoor al aangegeven, ook onderdeel van de opdrachtoomschrijving voor de student. De beoordelingsinstructie is vooral bedoeld om ervoor te zorgen dat (verschillende) beoordelaars op een eenduidige en goede wijze beoordelen.

4.3. Beoordelingscriteria en wat daar nog bij komt kijken

Hoewel het bij het beoordelen van een opdracht gaat om de beoordelingscriteria komt er meer bij kijken dan het gebruik van de beoordelingscriteria op zich. Bij het beoordelen van opdrachten spelen vier aspecten een rol die in een tabel worden samengebracht:

1. De beoordelingscriteria

Deze volgen uit het opdrachtenontwerpschema (toetsmatrijs).

2. De beoordelingsschaal met de niveaus waarop wordt beoordeeld

Bijvoorbeeld zeer onvoldoende, onvoldoende, voldoende, goed. Er zijn verschillende beoordelingsschalen mogelijk (zie het whitepaper hierover).

3. De indicatoren

Dit zijn de beschrijvingen van de verschillende niveaus behorend bij de beoordelingscriteria. Hiervoor bestaan twee oplossingen:

- a. Een checklist. De beschrijvingen van de indicatoren zijn per niveau gelijk voor alle beoordelingscriteria. Dit zijn korte en globale beschrijvingen wanneer iets bijvoorbeeld zeer onvoldoende, onvoldoende, voldoende of goed is. De beschrijving van de indicatoren zijn los van de tabel opgenomen (staan dus niet in de tabel).
- b. Een rubric. De beschrijvingen van de indicatoren zijn per beoordelingscriterium en per beoordelingsniveau verschillend en kennen per beoordelingscriterium toenemende concrete inhoudelijke aspecten waaraan moet worden voldaan. De beschrijvingen van de indicatoren staan in de tabel.

4. Een puntenverdeling die leidt tot een score.

Schematisch ziet dit er als volgt uit (met een mogelijke opzet voor een puntenverdeling die leidt tot een score op de opdracht).



Basisschema voor een checklist en een rubric

		Beoordelingsschaal met aanduidingen van de beoordelingsniveaus en een puntentoekening per niveau			
		Niveau 1	Niveau 2	Niveau 3	Niveau n
		0	1	2	m
Beoordelings- criteria	A				
	B	Beoordelingsindicatoren: leeg = checklist beschreven = rubric			
	C				
	D				
	N				
Score					
Totaalscore					

Tabel 2: Basisschema voor een checklist en een rubric

Voor het omzetten van de score op de opdracht naar een uitslag in bijvoorbeeld een cijfer wordt meestal een aparte tabel gebruikt. Daarbij kunnen ook eventuele knock-out-criteria een plaats krijgen. Het omzetten van de scores naar uitslagen wordt in een apart whitepaper behandeld.

5. Vijf voorbeelden van drie hoofdvormen voor het beoordelen van opdrachten

Een checklist kent twee hoofdvormen. Een specifieke checklist die werkt als een afvinklijst al dan niet met aandacht voor een niveau-aanduiding. Of een globale checklist waar beoordelingscriteria worden samengenomen en er dus globaler wordt beoordeeld. Hierdoor zijn drie hoofdvormen voor de beoordeling van opdrachten te onderscheiden:

- 1. Een specifieke checklist**
met specifieke beoordelingscriteria en een beoordelingsschaal met twee of meer niveaus (voorbeeld 1 en 2). Een checklist met meer niveaus wordt ook wel een 'rating scale' genoemd.
- 2. Een globale checklist**
met globale beoordelingscriteria en een beoordelingsschaal met meer niveaus (voorbeeld 3).
- 3. Een rubric**
met analytische beoordelingscriteria (voorbeeld 4) en een vorm met één holistisch beoordelingscriterium namelijk de opdracht als geheel (voorbeeld 5). Het verschil wordt duidelijk uit de voorbeelden.



In de literatuur wordt ook nog onderscheid gemaakt in generieke en taakspecifieke rubrics. Generieke rubrics zijn van toepassing op een verzameling van vergelijkbare opdrachten, terwijl taakspecifieke rubrics alleen van toepassing zijn op die ene specifieke opdracht. Het voordeel van een generieke rubric is dat deze inzetbaar is voor verschillende opdrachten, maar door de generiek toepasbare opzet leveren ze minder betrouwbare scores op dan taakspecifieke rubrics. Taakspecifieke rubrics leveren daarnaast een betere feedback op voor de studenten. Het nadeel van een taakspecifieke rubric is de hoeveelheid werk om voor elke opdracht een specifieke rubric te ontwikkelen.

In dit WhitePaper wordt geen voorbeeld gegeven van een generieke rubric vanwege de betere bruikbaarheid van de taakspecifieke rubric.

Hierna worden de drie hoofdvormen voor de beoordeling van opdrachten met een voorbeeld toegelicht. Alle voorbeelden zijn gebaseerd op de onderstaande vaardigheidsoopdracht, op basis van een rollenspel.

Instructie tentamenopdracht 'Voeren van een klantgesprek'

Inleiding

Het voeren van een goed klantgesprek is een belangrijk onderdeel van je beroep als bouwkundige/architect. Klanten oriënteren zich via internet en voeren enkele kennismakingsgesprekken met een paar bureaus voordat ze de keus maken om verder te gaan met een bureau. Een dergelijk gesprek is een uitgelezen kans om je bureau en jezelf te presenteren. Met een vet verkoopverhaal zal je niet scoren. Met deskundigheid en echte interesse in de klant wordt de kans groter. Aan jou om in deze opdracht te laten wat je hierover geleerd hebt en kunt.

Bestudeer deze opdracht en de bijlagen goed en bereid je op het gesprek voor, voordat je het gesprek ingaat. (Ter informatie: deze tentamenopdracht hebben de studenten aan het begin van de onderwijseenheid al uitgereikt gekregen).

De opdracht

Als student Bouwkunde/Architectuur vervul je in deze opdracht de rol van architect. Je voert een klantgesprek met een mogelijke klant voor een informatief gesprek over het ontwerp van een vrijstaand woonhuis in de nieuwbouwwijk Uylenbosch (zie de bijlage over Uylenbosch).

De klant wordt gespeeld door twee ervaren acteurs (een vrouw en een man). Voor de acteurs is ten behoeve van deze opdracht een instructie opgesteld. Hierin staat onder meer dat de klant representatief is voor de gemiddelde en zich normaal gedragende klant. Je hoeft dus geen bijzondere situaties te verwachten.



Hulpmiddelen en uitvoering

Je mag aantekeningen en andere informatie zoals over Uylenbosch meenemen naar het gesprek. Jouw aantekeningen worden niet gecontroleerd en worden ook niet beoordeeld. Een eigen gespreksleidraad is dus toegestaan. Je mag pen en papier, een rekenmachine en een schetsblok gebruiken. Het gebruik van een tablet of pc is niet toegestaan.

Drinken kun je halen bij de automaat. In de tentamenruimte (een speciaal hiervoor ingericht leslokaal) is een gesprekstafel aanwezig, met een kapstok. De klant zit voor je klaar in de wachtruimte, die ook is ingericht in het lokaal. Jij neemt voor aanvang plaats aan de gesprekstafel en haalt de klant zelf op. Dit doe je op aangeven van de beoordelaars die een apart hoekje hebben in het lokaal, waarna de tentamentijd gaat lopen.

Beoordeling

De beoordeling richt zich op de beoordelingscriteria die je in de bijlagen aantreft. Het maximaal aantal punten dat je kunt behalen is 26. Bij 20 punten of meer ben je geslaagd.

De beoordeling wordt gedaan door twee beoordelaars (waaronder niet je eigen docent) die onafhankelijk van elkaar beoordelen. Eén beoordelaar is docent aan de opleiding en de andere beoordelaar is een bouwkundige of een architect van buiten.

Tijd

Het klantgesprek mag maximaal 30 minuten duren (korter is geen probleem en de tijd is geen beoordelingscriterium). Na 30 minuten wordt de opdracht beëindigd, ongeacht in welke fase het gesprek zich bevindt. Wat dan niet aan de orde is geweest, wordt niet beoordeeld. Houd de tijd dus zelf goed in de gaten.

Time-out

Het kan voorkomen dat je om welke reden dan ook ongewild in een stresssituatie terecht komt. Het is mogelijk om dan een time-out aan te vragen. Het gesprek en de tentamentijd worden dan gestopt. Je krijgt 5 minuten de tijd om te herstellen. Vervolgens hervat je het gesprek. De gestopte tijd loopt vanaf dan weer door. De time-out heeft geen invloed op de beoordeling. De time-out mag niet worden gebruikt om jezelf tussentijds voor te bereiden op een tweede deel in het gesprek. Het moet voor alle aanwezigen duidelijk zijn dat je in een ongewilde stresssituatie terecht bent gekomen. Dit is dus een uitzonderingssituatie. Bij het vermoeden van misbruik zal de time-out niet worden toegekend.

Uitslag

Na afloop verzoeken de beoordelaars je het lokaal te verlaten om na een kwartier terug te komen, waarna je de uitslag krijgt en feedback krijgt op het resultaat. De beschikbare tijd voor de feedback is maximaal 15 minuten.

Herkansing

Gedurende dit studiejaar is één herkansing mogelijk ongeveer een maand na dit tentamen. Herkansers worden hierover nader geïnformeerd. Een tweede herkansing is pas mogelijk in het volgende studiejaar.

Losstaande vervolgo opdracht

Let op! Op basis van het gesprek, werk je in een vervolgo opdracht het gesprek uit tot een brief aan de klant met als bijlage een kostenindicatie en een schetsontwerp. Maak dus aantekeningen. Het maken van aantekeningen en de aantekeningen zelf worden niet beoordeeld.

Veel succes met de uitvoering van de opdracht.

Opmerkingen:

- De beoordeling zou ook gedaan kunnen worden door twee medestudenten of de eigen docent. Dit is afhankelijk van het toetsdoel, formatief of summatief, en/of de zwaarte van het tentamen.
- Of een time-out wenselijk is, is afhankelijk van het toetsbeleid. De hier geschetste tentamensetting met acteurs kan toetsangstigen dermate treffen dat een stressmoment ontstaat. De optie is mede bedoeld als geruststelling dat een stressmoment niet tot het einde van het tentamen hoeft te leiden.

5.1. Voorbeeld 1 van 5: Specifieke checklist met twee beoordelingsniveaus

Beoordelingscriteria		Niet-voldaan	Voldaan
De student:			
1.	Heet de klant welkom		X
2.	Stelt zichzelf voor	X	
3.	Herhaalt de naam van de klant		X
4.	Vraagt of de klant koffie, thee, fris of water wilt		X
5.	Stelt een algemene openingsvraag over de reis, het verkeer, het weer, een actualiteit of iets vergelijkbaars	X	
6.	Laat merken dat hij het adviesgesprek heeft voorbereid door in te gaan op de reden voor de afspraak	X	
7.	Licht de opbouw van het gesprek toe		X
8.	Geeft het woord aan de klant om zijn reden voor de afspraak toe te lichten	X	
9.	Vat wat de klant heeft gezegd kort samen	X	
10.	Stelt inhoudelijke relevante vragen naar de wensen van de klant		X



11.	Vraagt naar eisen (beperkingen) van de gemeente		X
12.	Vraagt door	X	
13.	Komt tot een vorm van een natuurlijke conversatie in een ontspannen sfeer met de klant		X
14.	Laat tijdens het gesprek merken dat hij de klant begrijpt en vooral ook aanvoelt		X
15.	Vraagt tussentijds of de klant nog koffie, thee, fris of water wilt		X
16.	Vat het gesprek op een goede wijze samen	X	
17.	Geeft aan dat hij een vrijblijvende kostenindicatie zal opstellen en een ruwe schets zal maken		X
18.	Wijst de klant op wat een kostenindicatie inhoudt	X	
19.	Geeft aan dat de klant altijd kan bellen of langs kan komen met vragen	X	
20.	Licht toe waarom de schets op kantoor wordt besproken en de klant dan nog nergens aan vast zit		X
21.	Geeft aan wanneer de klant de kostenindicatie kan verwachten en maakt een afspraak voor de schets		X
22.	Vraagt wat de klant vond van het gesprek/advies/voorstel/aanpak	X	
23.	Vraagt of de klant nog vragen heeft		X
24.	Dankt de klant voor zijn bezoek	X	
25.	Begeleidt de klant naar de uitgang		X
26.	Neemt op een goede wijze afscheid van de klant (goede reis of zoiets)	X	

Score		14
--------------	--	-----------

Algehele indruk gesprek (voor feedback)	
Wat ging goed	Je praat makkelijk en duidelijk. Je bent enthousiast. Duidelijk is dat je voldoende inhoudelijke kennis hebt.
Wat kan beter	Teveel zelf aan het woord. Stelt wel goede vragen, maar vraagt niet door. Indruk ontstond dat je zelf wilt scoren in het gesprek. Dat is niet nodig, dat je inhoudelijk goed bent, komt vanzelf wel naar voren. Tip: laat de klant aan het woord. De kostenindicatie is te beperkt aan de orde geweest!



Het nadeel van een checklist met alleen Niet-voldaan/voldaan is dat er geen onderscheid wordt gemaakt tussen minder en meer belangrijke beoordelingsaspecten. Zo is bijvoorbeeld het herhalen van de naam van de klant van een geheel ander orde dan het stellen van voldoende vragen.

Met een checklist met een beoordelingsschaal met meerdere beoordelingsniveaus en een aangepast puntensysteem is dit bezwaar te ondervangen (zie onder Voorbeeld 2 hierna).

5.2. Voorbeeld 2 van 5: Specifieke checklist met meerdere beoordelingsniveaus

Beoordelingscriteria		Gedrag niet laten zien	On-voldoende	Voldoende	Goed
	Beoordelingspunten	0	1	2	3
1.-5.	Zie hiervoor				
6.	Laat merken dat hij het adviesgesprek heeft voorbereid door in te gaan op de reden voor de afspraak	X			
	Feedback	-			
7.	Licht de opbouw van het gesprek toe			X	
	Feedback	Niet afraffelen en ook vragen of de klant zich erin kan vinden.			
8.	Geeft het woord aan de klant om zijn reden voor de afspraak toe te lichten				X
	Feedback	-			
9.	Vat wat de klant heeft gezegd kort samen		X		
	Feedback	Was niet volledig en miste de essentie.			
6.-26.	Zie hiervoor				
Score					
Totaalscore					

Betekenis van de niveau-aanduidingen

	De student heeft het betreffende beoordelingscriterium naar de mening van de inhoudsdeskundige en getrainde beoordelaar:
Gedrag niet laten zien	niet laten zien
Onvoldoende	wel uitgevoerd, maar op een niveau dat niet voldoet aan de standaard
Voldoende	uitgevoerd op het niveau dat voldoet aan de standaard*
Goed	uitgevoerd op een niveau dat hoger is dan de standaard*

* Er valt wat voor te zeggen om bepaalde beoordelingscriteria maximaal de score 'Voldoende' te geven. Bijvoorbeeld '1. Heet de klant welkom', '2. Stelt zichzelf voor', '3. Herhaalt de naam van de klant' en dergelijke. De reden hierachter is dat jezelf voorstellen ook niet overdreven moet worden en voldaan hieraan goed genoeg is. Dit biedt dan ook ruimte om belangrijke beoordelingscriteria juist wel hoger te kunnen scoren zoals '10. Stelt inhoudelijke relevante vragen naar de wensen van de klant' en '12. Vraagt door'.

Voor de student dient duidelijk te zijn voor welke beoordelingscriteria dit dan geldt. Denk er ook aan dat het aantal niveaus om te beoordelen een rol speelt bij deze wijze van beoordelen.

5.3. Voorbeeld 3 van 5: Globale checklist met meerdere beoordelingsniveaus

Beoordelingscriteria		Zeervol- onvol- doende	Onvol- doende	vol- doende	Goed	Uit- stekend
Beoordelingspunten		0	1	2	3	4
1.	Opbouw van het gesprek (structuur vooraf overgebracht naar de klant, logische opbouw gesprek, alle belangrijke onderdelen zijn aan bod gekomen en dergelijke)		1			
	Feedback	Geen structuur vooraf aangegeven aan de klant, opbouw was chaotisch, alle onderdelen zijn aan bod geweest, behalve de kostenindicatie. Tip: werk vooraf de structuur uit en neem die door met de klant en vraag of de klant nog aanvullingen heeft. Vergeet de kostenindicatie niet te bespreken!				



2.	Klantgerichtheid en sociale vaardigheden (ontvangst en afscheid, gastheerschap, omgangsvormen, luisteren, reflecteren en dergelijke)			2		
	Feedback	...				
3.	Technische uitvoering (stellen van relevante vragen, doorvragen, samenvatten, maken van afspraken en dergelijke)				3	
	Feedback	...				
4.	Aandacht voor het zakelijke gedeelte (kostenindicatie, schets, volledigheid, duidelijkheid, belang ervan en dergelijke)		1			
	Feedback	...				
5.	Algehele indruk gesprek			2		
	Feedback	...				

Score		2	4	3	
Totaalscore					9

Betekenis niveau-aanduidingen

Beoordelingsniveau	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
Uitstekend	uitgevoerd op een niveau dat ruim hoger is dan de standaard

In voorbeeld 3 worden vijf beoordelingsniveaus gebruikt. In voorbeeld 2 zijn dat er vier. U ziet dat als meer beoordelingsniveaus worden gebruikt, het lastiger wordt om een goed onderscheid tussen de niveaus weer te geven.

5.4. Voorbeeld 4 van 5: Analytische rubric

Beoordelingscriteria	Zeer onvoldoende	Onvoldoende	Voldoende	Goed
Beoordelingspunten	0	1	2	3
1. Ontvangst	X			
De student ontvangt de klant op een goede wijze en laat merken dat hij weet dat hij de klant kent en het gesprek heeft voorbereid.	De student heet de klant welkom en gaat direct over tot het gesprek. Laat niet merken dat hij weet waarvoor de klant komt.	De student heet de klant welkom, biedt wat te drinken aan en gaat over tot het gesprek. Laat niet merken dat hij weet waarvoor de klant komt.	Heet de klant welkom, biedt wat te drinken aan en laat merken dat hij weet waarvoor de klant komt en gaat over tot het gesprek.	Heet de klant welkom, begint over koetjes en kalfjes, laat duidelijk merken dat hij de naam van klant kent, weet waarvoor die komt en laat merken dat hij het gesprek heeft voorbereid.
Feedback	Op deze wijze ontvang je geen potentiële klanten. Je kon merken dat ze zich niet op hun gemak voelden. Zo is de eerste indruk slecht.			
2. Gesprek			X	
De student tast op een goede wijze af wat de wensen van de klant zijn en laat merken dat hij de klant aanvoelt. Hij laat het gesprek tot een vorm van conversatie komen en vat samen.	De student stelt geen goede of te weinig vragen over wat de klant wil. Van een conversatie is niet echt sprake en de samenvatting ontbreekt.	De student stelt vragen, maar vraagt niet echt door. Het gesprek is er wel, maar voelt niet natuurlijk aan. Er is geen of een slechte samenvatting.	De student stelt goede vragen, voldoende vragen en vraagt door. Het gesprek verloopt natuurlijk. Er is een goede samenvatting.	De student stelt goede vragen, voldoende vragen en vraagt door. Het gesprek verloopt natuurlijk én in een ontspannen sfeer. De student laat merken dat hij de klant berijpt en aanvoelt. Er is een goede samenvatting.
Feedback	Je heb je goed hersteld na een slecht begin. Je merkt de deskundigheid. Je stelt goede vragen, niet te veel en niet te weinig. Het doorvragen moet beter, eigenlijk is dit maar amper voldoende. Zorg voor wat meer contact met de klant en dat je laat merken dat je de klant aanvoelt. Het is nog wat inhoudelijk en technisch.			

3. Kostenindicatie en schets		X		
...				
Feedback				
4. Afscheid			X	
...				
Feedback				
5. Algehele indruk gesprek		X		
Feedback				

Score		2	4	
			Totaalscore	6

Omdat de ruimte in de kolommen beperkt is en de beoordelingscriteria en -indicatoren vaak veel tekst bevatten kan de volgende opmaak van een rubric een handig alternatief zijn.

2. Gesprek			
De student tast op een goede wijze af wat de wensen van de klant zijn en laat merken dat hij de klant aanvoelt. Hij laat het gesprek tot een vorm van conversatie komen en vat samen.			
Zeer onvoldoende	0		De student stelt geen goede of te weinig vragen over wat de klant wil. Van een conversatie is niet echt sprake en de samenvatting ontbreekt.
Onvoldoende	1		De student stelt vragen, maar vraagt niet echt door. Het gesprek is er wel, maar voelt niet natuurlijk aan. Er is geen of een slechte samenvatting.
Voldoende	2	X	De student stelt goede vragen, voldoende vragen en vraagt door. Het gesprek verloopt natuurlijk. Er is een goede samenvatting.
Goed	3		De student stelt goede vragen, voldoende vragen en vraagt door. Het gesprek verloopt natuurlijk én in een ontspannen sfeer. De student laat merken dat hij de klant berijpt en aanvoelt. Er is een goede samenvatting.
Feedback	Je heb je goed hersteld na een slecht begin. Je merkt de deskundigheid. Je stelt goede vragen, niet te veel en niet te weinig. Het doorvragen moet beter, eigenlijk is dit maar amper voldoende. Zorg voor wat meer contact met de klant en dat je laat merken dat je de klant aanvoelt. Het is nog wat inhoudelijk en technisch.		

Voorbeeld 4 betreft een zogenaamde analytische rubric. Daarnaast bestaat een holistische rubric. Bij een analytische rubric wordt elk beoordelingscriterium gescoord. De totaalscore wordt bepaald door de afzonderlijke scores. Bij een holistische rubric geeft de beoordelaar een score op basis van het resultaat als geheel.

5.5. Voorbeeld 5 van 5: Holistische rubric

Beoordelingscriterium	Zeer onvoldoende	Onvoldoende	Voldoende	Goed
De student heeft een adviesgesprek gevoerd.	Het gesprek verliep slecht, meerdere aspecten voor een goed gesprek zijn niet aan de orde geweest en/of de structuur van het gesprek was zeer onvoldoende en/of de wijze waarop met de klant werd omgegaan was zeer onvoldoende.	Het gesprek voldeed als geheel niet aan de standaard, enkele aspecten zijn niet aan de orde geweest en/of de structuur van het gesprek was onvoldoende en/of de wijze waarop met de klant werd omgegaan was onvoldoende.	Het gesprek voldeed aan de standaard, bijna alle aspecten zijn aan de orde geweest, de structuur van het gesprek was voldoende en de wijze waarop met de klant werd omgegaan was ook voldoende.	Het gesprek voldeed ruim aan de standaard, alle aspecten zijn aan de orde geweest, de structuur van het gesprek was goed en de wijze waarop met de klant werd omgegaan was ook goed.

Score		X		
--------------	--	----------	--	--

Het nadeel van een holistische rubric is dat de feedback op basis van de rubric globaal is.

6. Overwegingen bij een checklist en een rubric

6.1. Overwegingen bij een specifieke checklist

Gezien vanuit de beoordelaar

- De beoordelaar heeft de rol van observator en hoeft bij een specifieke checklist niet te interpreteren wat hij ziet (bij twee beoordelingsniveaus) of beperkt te interpreteren wat hij ziet (bij meer beoordelingsniveaus).
- Hierdoor wordt een groot deel van een subjectieve beoordeling weggenomen en neemt de objectiviteit toe.
- Beoordeling vindt 'on the fly' plaats op basis van waar te nemen gedragingen, waardoor de beoordeling relatief eenvoudig is uit te voeren.
- Checklists kennen een vaste volgorde van waar te nemen gedragingen, terwijl in de praktijk de gedragingen niet altijd in die volgorde voor hoeven te komen. De beoordelaar moet de checklist goed eigen hebben gemaakt, om hier soepel mee om te gaan. Voorkomen moet worden dat de beoordelaar gaat zoeken in de checklist en hierdoor vervolgens waarnemingen gaat missen, omdat de student ondertussen doorgaat met het uitvoeren van de opdracht.
- Bij het gebruik van meer beoordelingsniveaus dient de betekenis van de aanduidingen van de niveaus eenduidig te zijn voor de beoordelaars.



- Voorkomen moet worden dat de beoordeling zich richt op de grondigheid van het zichtbare gedrag in plaats van op de beheersing ervan (Wilkinson, Frampton, Thompson-Fawcett, Egan, 2003).

Zelfs voor een specifieke checklist die op het eerste gezicht concreet en goed hanteerbaar lijkt te zijn, is een training in het gebruik ervan gewenst. Met name de laatste drie punten van de opsomming hiervoor vergen, voorafgaand aan de beoordeling, de aandacht van de beoordelaar.

Beoordelaars hebben soms moeite met de inzet van checklists omdat studenten die gebruiken als een afvinklijst bij het leren. Dit is geen probleem als de checklist recht doet aan de (belangrijke) leerdoelen. Daarnaast is het zo dat onderdelen van een checklist niet allemaal gesloten gedragingen betreffen. De gedragingen uit het eerste voorbeeld zoals "Vat wat de klant heeft gezegd kort samen" en "Stelt inhoudelijke relevante vragen naar de wensen van de klant" zijn wel voor te bereiden door de student, maar door de aard van het gesprek en wat de klant zegt, zal de student toch moeten laten zien wat hij werkelijk beheerst.

Gezien vanuit de student

- Vanuit het oogpunt van leren en transparantie heeft de student de checklist vooraf ontvangen en bestudeerd. De student zal zijn voorbereidingen voor de toets (en dus zijn studie) richten op de checklist. Studenten weten waarop ze worden beoordeeld en hoe. Doordat met een checklist meer objectief kan worden beoordeeld ervaren studenten dit als meer rechtvaardig (minder afhankelijk van de beoordelaar) en transparant.
- Feedback kan concreet worden gegeven op de aspecten waaraan de student niet heeft voldaan of die onvoldoende waren. Dit geeft de student duidelijke handvatten waarop hij zijn bekwaamheid moet verbeteren.

6.2. Overwegingen bij een globale checklist

Gezien vanuit de beoordelaar

- Met een globale checklist wordt meer integratief beoordeeld (meer aspecten worden ineen en in relatie tot elkaar beoordeeld). De beoordelaar is niet alleen een observator die waarneemt, maar ook de deskundige die interpreteert wat hij ziet. Er wordt een groter beroep gedaan op zijn professionele deskundigheid om integratief te kunnen beoordelen.
- Globale checklist zijn hierdoor gevoeliger voor de beoordelaar als persoon met zijn persoonlijke eigenschappen, voorkeuren, deskundigheden, ervaringen, humeuren en dergelijke. Beoordelaars moeten zich hiervan bewust zijn en weten dat een beoordeling hierdoor meer subjectief wordt dan een beoordeling op basis van een specifieke checklist. Deze aspecten betreffen met name de intrabeoordelaarsbetrouwbaarheid.



- Een globale checklist dwingt een beoordelaar minder in een keurslijf dan een specifieke checklist, waardoor de beoordelaar meer gebruik kan maken van zijn deskundigheid voor zijn interpretatie van wat hij waarneemt en vindt.
- De interpretatie van de beoordeling (de scoring dus eigenlijk) kan pas plaatsvinden als een onderdeel van de opdracht is afgerond of soms zelfs pas nadat de hele opdracht is afgerond. Hierdoor wordt de beoordeling minder gedetailleerd, minder objectief (de beoordelaar moet veel zelf interpreteren) en meer subjectief.
- Het kan nodig zijn om meer studenten beoordeeld te hebben om tot een eerlijke en vergelijkbare beoordeling te komen tussen de verschillende studenten. De globale checklist kan dan pas achteraf goed ingevuld worden in tegenstelling tot de specifieke checklist. Om studenten uit elkaar te houden moet de beoordelaar tussentijds aantekeningen maken. Het maken van aantekeningen leidt weer af van het waarnemen en interpreteren.

Vanuit het aspect van objectiviteit en het gemak van de beoordeling van een specifieke checklist zouden beoordelaars een voorkeur kunnen hebben voor het gebruik van specifieke checklists, ondanks dat minder een beroep wordt gedaan op hun integratieve beoordelingsvaardigheden op basis van hun professionele deskundigheid. Daar staat tegenover dat met een globale checklist door de beoordelaar meer rekening kan worden gehouden met de totaalindruk (per onderdeel of met betrekking tot de gehele opdracht) van de neergezette prestatie.

Gezien vanuit de student

- De student zal zijn voorbereidingen op de toets nu ook weer, maar anders, richten op de aard van de beoordeling. De globale checklist is meer algemeen, maar nog wel specifiek genoeg, om zich goed voor te kunnen bereiden op de toets. Doordat hij weet dat de beoordeling globaler is, zal hij zich minder specifiek en meer algemeen voorbereiden. Hij heeft minder houvast dan bij de specifieke checklist en zal meer terugvallen op wat tijdens het leerproces is aangereikt aan informatie.
- De feedback die de student terug kan krijgen zal algemener en globaler van aard zijn en daardoor minder concreet (en minder bruikbaar) dan de feedback op basis van de specifieke checklist.

Studenten zouden het meeste hebben aan een combinatie van feedback op basis van een specifieke checklist (concreet) en op basis van een globale checklist met een meer integratieve en overall mening van de professioneel deskundige beoordelaar. De oplossing hiervoor kan worden gevonden in een combinatie van de specifieke en de globale checklist.

6.3. Overwegingen bij een rubric

De overwegingen voor het gebruik van een rubric bij het beoordelen zoals genoemd bij de globale checklist gelden ook voor de rubric.

Zoals hiervoor gesteld bij het gebruik van een brede beoordelingsschaal (dus meer dan twee beoordelingsniveaus) is het kennen van de betekenis van de aanduiding van de beoordelingsniveaus belangrijk. Bij een rubric (zie tabel 2) zijn de indicatoren beschrijvingen van de beoordelingsniveaus en geven weer hoe de meer globaal beschreven beoordelingscriteria beoordeeld moeten worden. Daarbij bevatten de indicatoren zelf ook weer, maar dan meer concreet meetbare of zichtbare, beoordelingscriteria. Al met al is een rubric een nogal tekstueel beoordelingsinstrument. Met een goede voorbereiding kan een rubric worden gebruikt voor het beoordelen van een vaardigheidsoopdracht om gedrag te beoordelen. Het gebruik van een specifieke of een globale checklist is voor het waarnemen van gedrag in veel gevallen praktischer.

7. Kiezen van een beoordelingsvorm

Welke beoordelingsvorm zich het beste leent voor een opdracht is van veel factoren afhankelijk. Een belangrijke factor is het soort opdracht met het soort beoordelingsbekwaamheid dat nodig is.

Oplopende niveaus van beoordelingsbekwaamheid		Hoofdtoetsvorm
1. Geen inhoudelijke deskundigheid en geen beoordelingsbekwaamheid nodig		Toets met gesloten vragen
Inhoudelijk deskundigheid is noodzakelijk plus ...	2. Bekwaamheid om een antwoordmodel en eenvoudige beoordelingscriteria toe te passen	Toets met open vragen
	3. Bekwaamheid om een checklist of een rubric toe te passen op een product of een proces beschrijving	Opdracht met als resultaat een product (bijvoorbeeld een werkstuk of een verslag)
	4. Bekwaamheid om waar te nemen en een checklist of eventueel een rubric toe te passen als waarnemer	Opdracht met als resultaat een te observeren gedragsvaardigheid
	5. Bekwaamheid om een beoordelingsgesprek te voeren en bekwaamheid om waar te nemen en een checklist of eventueel een rubric toe te passen als waarnemer	Opdracht in de vorm van een gesprek zoals bij een mondeling of een criteriumgericht interview

Tabel 3: Beoordelingsbekwaamheid in relatie tot een toetsvorm

Het gebruik van een checklist als afvinklijst, al dan niet met beoordelingsniveaus leent zich goed voor het beoordelen van enkelvoudige gedragsvaardigheden waarbij geobserveerd moet worden, zoals bij medische en verzorgende handelingen zoals het aansluiten van een infuus bij een patiënt in het ziekenhuis of het wassen en aankleden van een cliënt in een verzorgingstehuis. Zijn gedragsvaardigheden meervoudig en dus integratief zoals bij het verzorgen van een presentatie of therapeutische handelingen dan leent een rubric zich weer beter. Vanwege het tekstuele karakter van een rubric, waardoor veel gelezen moet worden, is een goede voorbereiding en een goede kennis van de rubric voorafgaand aan de uitvoering van de opdracht noodzakelijk. Tegelijk goed observeren en lezen gaat namelijk slecht samen. De rubric zal dan meestal pas achteraf en aansluitend aan de uitvoering van de opdracht gebruikt worden.

Omdat met een rubric meer integratief beoordeeld wordt, past een rubric goed bij opdrachten die integratief van aard zijn. Een afvinklijst is dan minder geschikt. Denk bijvoorbeeld aan de kwaliteit van redeneren binnen een verslag of een therapeutische handeling waarbij het niet alleen gaat om de handelingen op zich, maar ook om het gelijktijdige contact met de patiënt.

Opdrachten waarbij meer oplossingen goed zijn, laten zich slecht vangen in afvinklijsten en is een rubric meestal beter geschikt. Bij de beoordeling van een werkstuk zou een combinatie van een afvinklijst met een rubric wel goed kunnen werken. De afvinklijst voor het beoordelen van de karakteristieken van het product (afmetingen, nauwkeurigheid, gewicht, kleurgebruik, materiaalgebruik, afwerking en andere kenmerken), terwijl de rubric zich leent voor het verslag waarin bijvoorbeeld het proces van de ontwikkeling is beschreven en gemaakte keuzes zijn onderbouwd.

Referenties

- Anderson, L. W., Krathwohl, D.R., & Mayer, R.E. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.
- Bloom, B.S. (1956). Taxonomy of Educational Objectives, the classification of educational goals – Handbook I: Cognitive Domain. New York: McKay.
- Dave, R.H., (1970). Psychomotor levels. In R. J. Armstrong et al., Developing and Writing Behavioral Objectives, 33-34. Tucson, AZ: Educational Innovators Press.
- Krathwohl, D.R., Bloom, B.S., & Masia, B.B. (1964). Taxonomy of Educational Objectives, the Classification of Educational Goals, Handbook II: Affective Domain. New York: David McKay.
- Krathwohl, D.R., Bloom, B.S., & Dave, R.H., (1970). Psychomotor levels. In R. J. Armstrong et al., Developing and Writing Behavioral Objectives, 33-34. Tucson, AZ: Educational Innovators Press.
- Masia, B.B. (1964). Taxonomy of Educational Objectives, the Classification of Educational Goals, Handbook II: Affective Domain. New York: David McKay.



Miller, G.E. (1990). The assessment of clinical skills/competence/performance. *Academic medicine*, 65, 63-67.

Molkenboer, H.F.A.M. (2015). *Toetsen volgens de toetscyclus – Deel 1*. Enschede: Bureau voor Toetsen & Beoordelen.

Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. *Acad.Med.* 2003 Feb;78(2):219-22.

