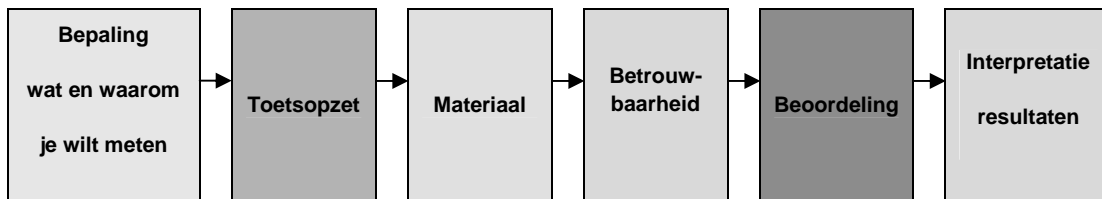


## TOETSTIP 8 – juni 2010



### TIP 8: BEOORDELINGSMODELLEN

Als u als docent toetsen ontwikkelt, zult u wel al gemerkt hebben dat daar heel wat bij komt kijken. U moet goed in de gaten houden *wat* u wilt meten en vooral ook *waarom* u dat wilt. Wat wilt u met de *uitkomsten* van de toets gaan doen? Welke *beslissingen* wilt u (kunnen) nemen? Tot slot zult u ook moeten waarborgen dat de toets *betrouwbaar* en *valide* is, dat wil zeggen dat de resultaten niet afhankelijk zijn van het toeval en dat u meet wat u wilt meten. Voor bovenstaande aspecten van toetsing verwijzen we u naar eerder verschenen ToetsTips. In deze ToetsTip richten we ons op het aspect beoordeling, en meer bepaald op beoordelingsmodellen. We gaan in op de verschillende soorten beoordelingsmodellen, hun voor- en nadelen en mogelijke toepassingen. We beginnen met een definitie van beoordelingsmodellen.

#### 1. HET BEOORDELINGSMODEL: EEN DEFINITIE

Beoordelingsmodellen worden gebruikt om taalgedrag te beoordelen dat een leerder laat zien naar aanleiding van een taak of toets, of het nu gaat om taalgedrag in lesverband of naar aanleiding van een taalexamen. De inhoud van het beoordelingsmodel wordt bepaald door de testdoelen en bijgevolg ook door de inhoud van de taak of opdracht. Zo bestaat het beoordelingsmodel bij een leestoets met meerkeuzevragen uit een overzicht van de juiste antwoorden (1A, 2C, etc.), en zal het beoordelingsmodel bij open vragen in diezelfde leestoets al ingewikkelder zijn. Voor elk model geldt dat het die aspecten bevat waarop de leerder beoordeeld zal worden. In het geval van een schrijftaak zouden dat bijvoorbeeld de volgende aspecten kunnen zijn: spelling, grammatica, woordenschat, samenhang en structuur. Kortom: een beoordelingsmodel is een instrument dat beoordelaars stuurt wanneer ze een score of waardering moeten toekennen aan een product van een leerder en dat hen aanzet om producten telkens op dezelfde aspecten te beoordelen.

#### 2. EEN DIVERSITEIT AAN BEOORDELINGSMODELLEN

Het volledige toetsproces dat u visueel vindt weergegeven bovenaan deze bladzijde, zou de indruk kunnen wekken dat u pas over beoordelingscriteria en -modellen moet nadenken wanneer de eerste vier stappen al doorlopen zijn. Dat klopt niet helemaal omdat u in principe vanaf het moment dat u bepaalt wat u gaat meten (en waarom), ook zult moeten stilstaan bij de beoordeling. Wanneer u de toetsopzet bepaalt en het

benodigde materiaal gaat ontwikkelen kunt u er niet onderuit om tegelijk de beoordelingscriteria te bepalen. Hoe en waarop (welke aspecten) gaat u de kandidaat beoordelen? Als u bijvoorbeeld spreekvaardigheid wilt meten, zult u de aspecten moeten bepalen waarop u gaat oordelen, bijvoorbeeld inhoud, samenhang, vloeiendheid, enz. Wanneer u de toetsopzet bepaalt, moet u er vervolgens voor zorgen dat de kandidaat voldoende gelegenheid krijgt om de aspecten te uiten waarop hij beoordeeld zal worden. Kortom: de ontwikkeling van toetsmateriaal en de ontwikkeling van beoordelingsmodellen gaan hand in hand.

Beoordelingsmodellen zijn nodig om ervoor te zorgen dat de beoordeling zo valide en betrouwbaar mogelijk gebeurt. Betrouwbaarheid kan bekeken worden vanuit twee standpunten: de betrouwbaarheid van één individuele beoordelaar (*intrabeoordelaarsbetrouwbaarheid*) en de mate waarin de waarnemingen van meerdere onafhankelijke beoordelaars overeenstemmen (*interbeoordelaarsbetrouwbaarheid*). Als docent zou u de producten van leerders intuïtief kunnen beoordelen. Uw intuïtie zal echter niet garanderen dat u echt een objectief oordeel geeft. Hoe kunt zeker zijn dat u beoordeelt wat u moet beoordelen (bijvoorbeeld taalvaardigheid) en niet hoe aardig u de student vindt? Of hoe intelligent en ijverig? Een beoordelingsmodel kan de beoordelaar sturen om een solide oordeel te geven, ongeacht zijn humeur, het tijdstip van de dag, de werklast van beoordelen etc. Beoordelingsmodellen minimaliseren ook de variatie tussen beoordelaars. Als elke beoordelaar met hetzelfde model werkt, is het kader waarbinnen beoordeeld moet worden, afgebakend. Het beoordelingsmodel maakt duidelijk welke aspecten en vaardigheden beoordeeld moeten worden en welke eisen je stelt aan het beheersingsniveau van die aspecten en vaardigheden.

Beoordelingsmodellen worden meestal per taak ontwikkeld zodat je de eisen kunt toespitsen op de specifieke eisen van de taak en de context waarin die zich situeert. Dat betekent echter niet dat eenieder die een taak/toets/opdracht maakt het wiel opnieuw moet uitvinden. Er zijn algemeenheden te ontdekken in verschillende soorten modellen. Afhankelijk van welke informatie u over de kandidaat wilt verzamelen op basis van de toets, kunt u kiezen voor een meer of minder gedetailleerd model. Grofweg kunnen we binnen de toetswereld de volgende modellen onderscheiden: *primary trait*-modellen, *holistische* modellen en *analytische* modellen. Onderstaande tabel geeft weer welke kenmerken horen bij de verschillende modellen.

	Taakspecifiek	Generiek
Eén score	Primary Trait	Holistisch
Verschuillende scores		Analytisch

## 2.1 PRIMARY TRAIT

Beoordelingsmodellen van het type *primary trait* bekijken in hoeverre het product van de leerder (bijvoorbeeld de uitwerking van een schrijfpodracht) overeenkomt met de beschrijving van de **ideale uitvoering** die de docent heeft vastgesteld. De docent ontwikkelt een tabel waarin hij per niveau de aspecten beschrijft waarop hij de leerders zal beoordelen. Aan de hand van die niveauomschrijvingen kan de docent daarna het werk van elke individuele leerder beoordelen. Een van die aspecten kan het communicatieve effect zijn; is een argumentatieve tekst bijvoorbeeld overtuigend? Is het onderzoeksverslag relevant en controleerbaar? Zorgt de thriller voor spanning? Om de ideale uitvoering van de (toets)taak te kunnen bepalen moet de docent hypothesen doen over de verwachte prestaties, nog voordat hij de producten van zijn leerders begint te beoordelen. *Primary trait*-modellen zijn voor de docent, ondermeer vanwege het vele voorwerk, erg arbeidsintensief, maar voor de leerder wel erg informatief. Doordat de docent per aspect een niveaubeschrijving geeft, zijn de eisen voor de leerders erg inzichtelijk en transparant. Doordat *primary trait*-beoordelingsmodellen gemotiveerde scores genereren, krijgen leerders bovendien gedetailleerde informatie over hun werkpunten.

Hieronder vindt u een voorbeeld van een (eenvoudig) *primary trait*-scoringsmodel dat gebruikt kan worden om onderzoeksrapporten te beoordelen. De rijen geven de aspecten weer waarop beoordeeld wordt, de kolommen geven de waardering weer. De docent die deze tabel gebruikt, geeft per cel een omschrijving, zodat de eisen per waarderingniveau duidelijk zijn.

	Zeer goed	Goed	Voldoende	Onvoldoende
Methode				
Materialen				
Experimenteel design				
Data-verzameling				
Resultaten en conclusie				

## 2.2 HOLISTISCH

Bij holistische modellen richt men zich doorgaans op de positieve aspecten van leerderproducten door in oordelen uitsluitend te beschrijven wat een kandidaat al kan/kent en niet waaraan hij nog moet werken. De docent onderscheidt geen verschillende aspecten, maar geeft een globaal oordeel, waarbij hij zich vaak lijkt te baseren op oppervlaktekenmerken van het product. Doordat de docent snel en op basis van zijn eigen bevindingen (en niet op basis van een vooraf vastgelegd

scoringsmodel) een oordeel geeft, is de score achteraf vaak moeilijk interpreteerbaar. Voor de leerder zijn scores die het resultaat zijn van een holistische beoordeling weinig informatief; hij krijgt geen enkele informatie over de punten waaraan hij nog moet/kan werken. Holistische beoordelingen sluiten waarschijnlijk het best aan bij de alledaagse onderwijspraktijk. Het is in die zin te bestempelen als het meest authentieke model.

#### Beoordelingsmodel 2

	Beschrijving	Punten
<b>Niet bekwaam<sup>1</sup></b>	De kandidaat beschrijft geen wensen ten aanzien van activiteiten en ruimtes voor de gegeven doelgroepen. OF De kandidaat schrijft een brief op een manier die verwarring zou opleveren voor het architectenbureau (taalgebruik te onduidelijk). OF De kandidaat schrijft een ongepaste brief.	<b>1</b>
	De tekst van de kandidaat beschikt over een paar elementen van de <i>Niet bekwaam</i> beschrijving and een paar elementen van de <i>Bekwaam</i> beschrijving	<b>2</b>
<b>Bekwaam</b>	De kandidaat schrijft een brief waarin hij minimaal een aantal wensen ten aanzien van activiteiten en ruimtes voor de gegeven doelgroepen beschrijft. Het architectenbureau kan deze wensen meenemen bij het ontwerpen van het buurtcentrum. Niet alle activiteiten en ruimtes hoeven beschreven te worden. De kandidaat schrijft een <u>gepaste brief die voldoende begrijpelijk is</u> . Taalgebruiksfouten mogen nog wel voorkomen.	<b>3</b>
	De tekst van de kandidaat beschikt over een paar elementen van de <i>Bekwaam</i> beschrijving en een paar elementen van de <i>Zeer bekwaam</i> beschrijving	<b>4</b>
<b>Zeer bekwaam</b>	De kandidaat schrijft een brief waarin hij minimaal een aantal wensen ten aanzien van activiteiten en ruimtes voor de gegeven doelgroepen beschreven wordt. Het architectenbureau kan deze wensen meenemen bij het ontwerpen van het buurtcentrum. De kandidaat schrijft een gepaste en stilistisch mooie brief die qua taalgebruik voor het grootste deel in orde is en prettig is om te lezen.	<b>5</b>

NB: de beoordelaar ontvangt een beoordelingsvoorschrift waarin uitgebreid beschreven wordt wat er verstaan wordt onder bijv. 'duidelijkheid van taalgebruik'. Deze is nu niet opgenomen (ook omdat de discussie over de vormelijke elementen tijdens het opstellen van dit document nog niet gevoerd was).

Hoewel holistische beoordelingen aantrekkelijk lijken vanwege het snelle karakter en het feit dat ze aanleunen bij wat gangbaar is in de onderwijspraktijk, zijn ze het vanuit het oogpunt van toetsen en beoordelen niet ideaal te noemen. Denkt u maar aan het gebrek aan diagnostische informatie en de moeilijke interpreteerbaarheid van de gegeven scores. *Primary trait*-modellen, zowat de tegenhanger van het holistische

model, zijn dan wel erg informatief vanuit een diagnostisch oogpunt, maar daardoor ook wel erg arbeidsintensief. Het laatste model dat we in deze ToetsTip bespreken, het analytische model, lijkt de ideale middenweg te vormen.

### 2.3 ANALYTISCH

*Analytische beoordelingsmodellen* combineren de diagnostische waarde van het *primary trait*-model met de gebruiksvriendelijkheid van het holistische model. Prestaties van kandidaten worden op diverse aspecten beoordeeld (grammatica, woordenschat, stijl, etc.), waardoor een beheersingspatroon van de kandidaat zichtbaar wordt. Dat patroon kan als diagnostische informatie gebruikt worden om de taalvaardigheid van de leerder gericht aan te pakken. Het analytische model is een combinatie van taakspecifieke en minder taakspecifieke criteria. Het generieke aspect heeft als voordeel dat modellen breder inzetbaar zijn, bijvoorbeeld voor meerdere taken of voor een volledige vaardigheid. Afhankelijk van de deelaspecten waarop beoordeeld wordt, zal de docent echter wel degelijk rekening houden met de taakspecifieke eisen, bijvoorbeeld voor inhoud, woordenschat of register. Hij moet zich immers afvragen of de inhoud van het afgeleverde product aansluit bij de opdracht. Ook een aspect als woordenschat of register kan enkel vanuit het perspectief van de specifieke taak worden beoordeeld: is de gebruikte woordenschat toereikend en gepast, *gezien de opgegeven taak?* Wordt het register passend *in deze taaksituatie* goed gehanteerd?

Het feit dat elk deelaspect (bijvoorbeeld spelling, woordenschat, grammatica) van de beoordeling in dezelfde mate meetelt voor het uiteindelijke oordeel, beschouwt men vaak als nadelig. Dit (doorgaans ongewenste) effect wordt in de toetswereld vaak ondervangen door een bepaald (mathematisch) gewicht toe te kennen aan elk deelaspect, afhankelijk van het belang van het aspect voor de taak en uitvoering in kwestie. Zo krijgen sommige aspecten een luidere stem in het eindoordeel en -cijfer.

Hieronder vindt u een voorbeeld van een analytisch beoordelingsmodel.

<b>Voorbeeld beoordeling (bron: ESL)</b>		
<b>Criterium</b>	<b>Schaal</b>	<b>Toelichting</b>
<b>Inhoud</b>		
25-22	Uitstekend tot zeer goed	Met kennis van zaken, goed geïnformeerd, zelfstandig, grondige uitwerking van het standpunt - relevant voor het toegewezen topic
21-18	Goed tot gemiddeld	Beschikt over een beetje kennis over het onderwerp - adequaat bereik, beperkte ontwikkeling van het standpunt/onderwerp - meestal relevant voor het toegewezen topic, maar weinig gedetailleerd.
17-11	Voldoende tot zwak	Beperkte kennis van het onderwerp - weinig zelfstandig - weinig adequate ontwikkeling van het standpunt
10-5	Zeer zwak	niet relevant OF niet voldoende om te evalueren OF geen link met het toegewezen onderwerp.
<b>Organisatie</b>		
20-18	Uitstekend tot zeer goed	Structuur duidelijk en onderbouwd - goed georganiseerd en zeer grondige ontwikkeling van inleiding, body and conclusie, grondige ontwikkeling van ondersteunende details.
17-14	Goed tot gemiddeld	Ietswat onsamenhangend - hoofdgedachte steken er bovenuit, maar structuur onduidelijk, beperkte ontwikkeling van inleiding, body en/of conclusie - en/of beperkte ontwikkeling van ondersteunende details
13-10	Voldoende tot zwak	Verwarrende ideeën worden naar voren gebracht - weinig of geen logische opbouw of ontwikkeling van introductie, body en/of conclusie - ondermaatse ontwikkeling van ondersteunende details.
9-7	Zeer zwak	Communiqueert niet, geen structuur OF niet voldoende output om te beoordelen
<b>Zinsconstructie</b>		
20-18	Uitstekend tot zeer goed	Effectief gebruik van complexe tot eenvoudige zinsstructuren – correcte interpunctie – goed gebruik van onderschikking en nevenschikking –

De keuze voor een soort beoordelingsmodel hangt af van meerdere factoren, onder meer van wat u wilt gaan doen met de resultaten; gaat het om een tussentijdse toets met diagnostische informatie voor de leerder of om een toets die tot certificering moet leiden? Het is met andere woorden cruciaal dat u op voorhand duidelijk uw toetsdoel(en) in kaart brengt en die ook tijdens het beoordelingsproces bewaakt.

### **Wordt vervolgd...**

In de volgende ToetsTip komt u meer te weten over de ontwikkeling van beoordelingsmodellen; wanneer besluit u (als docent) om met een beoordelingsmodel te werken en/of om er zelf een te maken, hoe begint u eraan, hoe kunt u beoordelaarstrainingen aanpakken, etc.

Voor vragen & reacties één adres:

**het CNaVT-secretariaat**

Certificaat Nederlands als Vreemde Taal

Blijde-Inkomststraat 7 – bus 3319

B-3000 Leuven

België

0032 (0) 16 53 55 16

[cnavt@arts.kuleuven.be](mailto:cnavt@arts.kuleuven.be)