

**Toetsbeleid De Nieuwe Veste**



## Inhoudsopgave

Nr.		Pagina
1.	Inleiding	2
2.	Doel van toetsen	2
3.	Functies van toetsen	2
4.	Toetssystematiek	3
5.	Kwaliteitseisen aan toetsen	4
6.	Toetscontrole en toetsanalyse	6

### 1. Inleiding

In het toetsbeleid worden de visie en het beleid van De Nieuwe Veste op het gebied van toetsen nader uitgewerkt. De eisen waarop het werk van de leerling wordt beoordeeld, zijn nader uitgewerkt in het toetsreglement.

De ervaring leert, dat het ontwikkelen van goede toetsen en het analyseren en beoordelen van de kwaliteit van toetsen lastig is. Er wordt onder tijdsdruk vaak gebruik gemaakt van de standaardtoetsen die bij de methode worden geleverd. Het zelf ontwikkelen van een toets is een belangrijke vaardigheid, die elke docent dient te beheersen.

De wijze waarop getoetst wordt, de weging van de verschillende onderdelen en de planning is voor de bovenbouw vastgelegd in een Programma van Toetsing en Afsluiting (PTA). De Nieuwe Veste vindt, dat de eisen waarop het werk van de leerling wordt beoordeeld duidelijk zichtbaar en helder moet zijn voor alle leerlingen in alle fases van het onderwijs.

### 2. Doel van toetsen

Evalueren is vaak de laatste stap in het onderwijsproces. Een docent kan allerlei elementen van het onderwijsleerproces evalueren. In deze notitie maken we onderscheid tussen product- en procesevaluatie, waarbij toetsing productevaluatie betreft.

Bij toetsing maken we onderscheid tussen diagnostische toetsing en selectieve toetsing.

Diagnostische toetsing vindt tussentijds (schriftelijk of mondeling) plaats gedurende een onderwijsleerproces, met als doel duidelijk te maken waar het kennen en kunnen van leerlingen nog niet aan het beoogde niveau voldoet. Leerlingen krijgen tussentijds feedback en kunnen vervolgens aan hun zwakke punten gaan werken.

Selectieve toetsing is bedoeld om beslissingen te nemen die de onderwijsloopbaan van leerlingen raken: een voldoende of onvoldoende voor een toets, doubleren of overgaan, zakken of slagen. Hierbij is het belangrijk dat een toets valide en betrouwbaar is en dat hij onder bepaalde condities is afgenomen, door de docent wordt gecorrigeerd en van een cijfer wordt voorzien.

### 3. Functies van toetsen

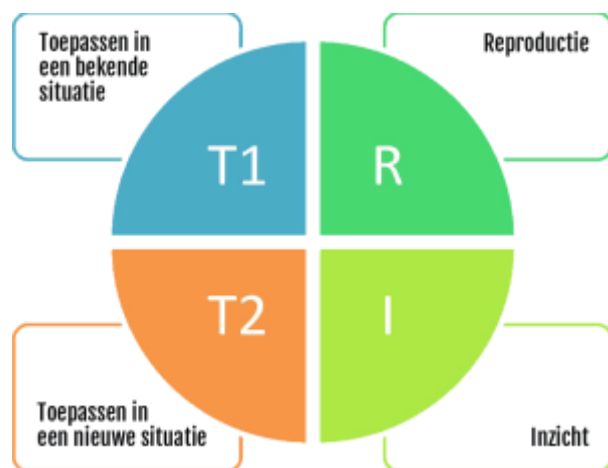
De functies die men aan toetsen toekent, zijn van invloed op de eisen die aan toetsen worden gesteld. Dat men aan toetsen vele (bedoelde en onbedoelde) functies toekent, laat de onderstaande inventarisatie zien:

- Selectiefunctie: afhankelijk van de resultaten op een toets wordt beslist een leerling al of niet toe te laten tot een vervolgtraject.
- Kwaliteitsbewakingfunctie: voldoen aan bepaalde standaarden. Geeft rechten en/of erkenning. Voor de overheid zijn toetsen in het algemeen en met name examens belangrijke instrumenten voor de kwaliteitsbewaking van het onderwijs.
- Prognostische functie: doen van een uitspraak over de kans om een bepaald vervolgtraject met succes te volgen. Als een leerling met goed gevolg een opleiding heeft afgesloten, verwacht men dat hij of zij over kennis en vaardigheden beschikt die rechtvaardigt dat de leerling een bepaalde vervolgstudie aanvangt.

- Operationalisatiefunctie: toetsen zijn de meest concrete uitwerking van nagestreefde onderwijsdoelen. Zij zijn het middel om aan te geven wat als belangrijk wordt gezien. Toetsen zijn richtinggevend voor het onderwijs.
- Didactische functie: voorziet belanghebbenden van informatie over het onderwijsleerproces. Stuur het onderwijsleerproces. Scholen kunnen met toetsen nagaan in hoeverre nagestreefde doelen gehaald zijn en of bijstelling van het onderwijsprogramma noodzakelijk is.
- Kwalificerende functie: geeft inzicht in de kennis en vaardigheden waaraan de leerling moet voldoen.
- Evaluatiefunctie: geeft een bijdrage aan de beoordeling van de kwaliteit van het onderwijs.

#### 4. Toetssystematiek<sup>1</sup>

Er zijn verschillende modellen om de cognitieve niveaus van leren in kaart te brengen waaronder de Taxonomie van Bloom, OBIT<sup>2</sup> en RTTI. Om scherp en transparant de cognitieve niveaus van leren in kaart te brengen, kiest De Nieuwe Veste bij de constructie van toetsen voor RTTI. De reden hiervoor is, dat RTTI duidelijk is, eenvoudig is toe te passen en aansluit bij de kwaliteitseisen van de Inspectie van het Onderwijs.



- R = reproductievragen: vragen die kunnen worden opgelost op basis van herkenning of reproductie. Er hoeft niets wezenlijks te worden toegevoegd aan het geleerde.
- T1 = toepassingsgerichte vragen die gericht zijn op het toepassen van de leerstof in reeds geleerde situaties. Met toepassingsvragen wordt nagegaan of de leerling in staat is om het geleerde zelfstandig toe te passen in een bekende, geoefende context.
- T2 = complexere toepassingsgerichte vragen die gericht zijn op het toepassen van de leerstof in nieuwe situaties. Met deze toepassingsvragen wordt nagegaan of de leerling in staat is om het geleerde te koppelen aan andere situaties.
- I = inzichtvragen zijn vragen die een actievere, eigen inbreng van de leerling verwachten. Inzicht vraagt van een leerling om in een onbekende situatie de oplossing van een probleem te vinden op basis van het geleerde en/of eerder verworven competenties.

Door deze niveaus te beschrijven en markeren kunnen leerprocessen worden gevolgd, gestuurd en verbeterd door zowel docent als leerling. RTTI maakt het ook mogelijk om in de lessen meer aandacht te besteden aan de (hogere) cognitieve niveaus om zo de ontwikkeling van de hersenen bij jongeren te versnellen en/of te stimuleren.

<sup>1</sup> De leerlingen van het PrO werken op een andere wijze en de beoordelingen vinden plaats op wijze van ontwikkeling en niet op prestaties.

<sup>2</sup> OBIT: Onthouden, Begrijpen, Integreren en creatief Toepassen.

 <b>de nieuwe veste</b>  <b>Onderwijs</b>	<b>Beleid</b> Toetsbeleid	<b>206.07.00-BEL</b>
		Versie: 1.0
		Datum: 12-02-14

De ontwikkeling van R naar I in toetsen is afhankelijk van het onderwijsniveau en het leerjaar.

## 5. Kwaliteitseisen aan toetsen

De leerdoelen horen richtinggevend te zijn voor de ontwikkeling van toetsen. Zij geven aan wat een leerling moet kennen of kunnen na afloop van een les of lessenserie.

Toetsen moeten representatief zijn voor het voorafgaande onderwijs, zowel wat betreft inhoud als niveau van stofbeheersing en vaardigheden. Als docent kun je leerlingen niet vragen oorzaak-gevolg verbanden te leggen als het in het boek en de les alleen maar gaat over feitelijke informatie. Een beproefde aanpak om gebrekkige afstemming van leerdoelen en toetsvragen te voorkomen is het samenstellen van een toets voorafgaand aan een lessenreeks. Daardoor wordt meestal wel duidelijk wat als belangrijk wordt ervaren en wordt voorkomen dat dit in de les niet aan bod zou komen. Bij de examenprogramma's in de bovenbouw is het raadzaam vooraf de **eindtermen** erbij te pakken en naar aanleiding daarvan leerdoelen te formuleren. Dat kan, omdat in de **explicitering** van de eindtermen de leerstof vaak tot op begripsniveau is omschreven.

Een toets moet daadwerkelijk meten wat bedoeld is te meten. De functies die aan toetsen worden toegekend, komen alleen dan tot hun recht als de toetsen aan bepaalde kwaliteitseisen voldoen. Vijf belangrijke eisen zijn:

### A. Betrouwbaarheid

Een toets wordt betrouwbaar genoemd als zij bij herhaalde afname onder dezelfde omstandigheden eenzelfde resultaat laat zien. Belangrijke factoren bij betrouwbaarheid zijn:

#### *De kwaliteit van de toets zelf*

- De opgaven moeten helder en duidelijk geformuleerd zijn.
- Er mag geen twijfel bestaan over het soort antwoord dat van de kandidaat wordt verwacht. Als een weegschaal de ene keer 85 kilo aangeeft en de andere keer met dezelfde persoon 90 kilo is de weegschaal niet betrouwbaar.
- Extreem moeilijke of gemakkelijke opgaven moeten vermeden worden.
- De toets in z'n geheel moet een onderscheid maken tussen 'goede' en 'zwakke' kandidaten, hetgeen ook geldt voor de afzonderlijke opgaven (toets moet discrimineren).
- De moeilijke opgaven moeten vooral door de goede kandidaten goed gemaakt worden.
- De betrouwbaarheid wordt ook beïnvloed door het aantal opgaven dat een toets bevat. Bij een toets met weinig opgaven is de invloed van elke afzonderlijke opgave veel groter dan bij een toets met veel opgaven.

#### *De omstandigheden waaronder de toets wordt afgenomen*

Belangrijk in dit kader is standaardisatie en objectiviteit. Omstandigheden kunnen op velerlei zaken betrekking hebben: lokale omstandigheden (het in rust kunnen maken van een toets), het gebruik van hulpmiddelen en de relatie tussen de lengte van de toets en de beschikbare toetstijd. Het is belangrijk dat een toets niet ontaardt in een 'race tegen de klok'.

#### *De frequentie van toetsing*

In principe neemt de betrouwbaarheid van toetsing toe als men vaker toetst.

#### *De wijze waarop de resultaten worden beoordeeld*

Ook bij de beoordeling van de resultaten spelen standaardisatie en objectiviteit een belangrijke rol. Het resultaat van een kandidaat op een toets kan sterk bepaald worden door de beoordelaar.

#### *Oorzaken voor verschillen in beoordeling*

Bij open vragen moet natuurlijk een zo objectief mogelijke beoordeling worden nagestreefd. Het is belangrijk dat beoordelaars de antwoorden van de studenten zo nauwkeurig mogelijk en steeds op

gelijke wijze beoordelen. Dat verhoogt de betrouwbaarheid van de toets. Uit diverse onderzoeken blijkt dat er in de praktijk grote verschillen bestaan tussen de wijze waarop de antwoorden van studenten op toetsvragen worden beoordeeld. De belangrijkste oorzaken zijn:

- Geringe intra- en interbeoordelaarsovereenstemming: de beoordelaar die hetzelfde antwoord op een later tijdstip opnieuw beoordeelt (bijvoorbeeld omdat leerlingen het niet eens zijn met de beoordeling), komt nogal eens tot een ander oordeel. In een dergelijke situatie is er sprake van een geringe intrabeoordelaarsovereenstemming. Van een geringe interbeoordelaarsovereenstemming is sprake indien twee of meer beoordelaars onafhankelijk van elkaar tot aanzienlijk verschillende beoordelingen komen.
- Normverschuiving: het komt vaak voor dat een beoordelaar met open vragen tijdens de beoordeling steeds milder wordt. Antwoorden die aanvankelijk foutief werden gerekend, worden later in de beoordeling goed gerekend.
- Sequentie-effect: voorafgaande beoordelingen zijn van invloed op de beoordeling. Docenten hebben de neiging na een (groot) aantal slechte prestaties een relatief goede prestatie te overwaarderen. Ook andersom komt dit voor. Na een (groot) aantal goede prestaties kan een relatief slechte prestatie worden ondergewaardeerd.
- Contaminatie-effect: sommige docenten kennen aan de antwoorden van de leerlingen nog een andere functie toe. Zij beschouwen de resultaten van 'hun' leerlingen als een maatstaf voor de kwaliteit van het onderwijs dat ze zelf hebben gegeven. Bij het toekennen van scores houden docenten hiermee dan ook rekening.
- Halo-effect: in sommige situaties laten beoordelaars de kennis die zij van hun leerlingen hebben meespelen. Matige prestaties van (in de ogen van beoordelaars) briljante dan wel aardige leerlingen worden dan hoger gewaardeerd dan exact dezelfde prestatie van middelmatige, of minder vriendelijke leerlingen.
- Signifisch effect: ten slotte kan het fenomeen van strenge en soepele beoordelaars voorkomen. De criteria die docenten hanteren bij de beoordeling kunnen verschillen. Een docent die bij de beoordeling van een praktische opdracht let op inhoud, kan dezelfde opdracht veel hoger waarderen als de docent die meer aandacht schenkt aan vormaspecten.

#### Oplossingen voor beoordelaarsverschillen

Hieronder wordt een aantal oplossingen beschreven tegen ongewenste beoordelaarseffecten.

- Om ervoor te zorgen dat de leerlingen van wie de vragen steeds als eerste worden nagekeken niet de dupe worden van de normverschuiving bij beoordelaars, is het raadzaam de volgorde van antwoordvellen van leerlingen regelmatig te wijzigen.
- Het sequentie-effect wordt beperkt door het beoordelen van de antwoorden per vraag (in plaats van per leerling). Ook het wijzigen van de volgorde van antwoordvellen tijdens de beoordeling is hiervoor een oplossing.
- Het contaminatie-effect is moeilijk oplosbaar zolang er geen scheiding is aangebracht tussen doceren en beoordelen. Intercollegiaal uitwisselen van beoordelingsactiviteiten (docent A beoordeelt de toets van leerlingen van docent B, en andersom) biedt een oplossing.
- Het halo-effect wordt beperkt door niet naar de naam te kijken van de leerling die de toets heeft gemaakt. Elkaars toetsen nakijken vermindert het halo-effect eveneens.

Vooraf bij toetsen waarvan de uitslag vergaande gevolgen heeft voor de leerling, is een betrouwbare uitspraak over een voldoende of onvoldoende van groot belang. Het is daarom aan te bevelen de toets van leerlingen met een toetsscore rond de zak/slaaggrens door een tweede beoordelaar te laten nakijken.

#### **B. Validiteit**

De validiteit van een toets is de eigenschap dat de toets meet wat de constructeur bedoeld heeft ermee te meten. Bij inhoudsvaliditeit ga je na of de toets wel werkelijk de kennis, vaardigheden en houdingen meet die je wilt meten. Daartoe dienen doelstellingen helder te zijn. 'Enig inzicht in', 'een

 <b>de nieuwe veste</b>  <b>Onderwijs</b>	<b>Beleid</b> Toetsbeleid	<b>206.07.00-BEL</b>
		Versie: 1.0
		Datum: 12-02-14

voldoende beheersing van' zijn vage omschrijvingen die geen goede basis vormen voor een valide toets. Begripsvaliditeit is de mate waarin de toetsscore beschouwd kan worden als een maat voor het bedoelde begrip. Zo is het niet valide om inzicht te toetsen door een definitie te vragen van een begrip. De toets moet dus op het juiste niveau toetsen. In onderwijssituaties zou een belangrijke rol moeten worden toebedeeld aan de inhoudsvaliditeit van een toets.

### **C. Aanvaardbaarheid**

Het aanvaardbaar maken van beslissingen op grond van een toets is geen eenvoudige zaak en hangt mede af van het belang van de betreffende toetsing. Naast eisen ten aanzien van betrouwbaarheid en validiteit, dient een toets ook voor betrokkenen 'doorzichtig' te zijn ten aanzien van: het kiezen van de te volgen strategie tijdens de voorbereiding, beoordeling en waardering van de toetsresultaten (correctievoorschriften en de procedure voor de bepaling van het cijfer).

### **D. Transparantie**

Transparantie betreft vooral de inhoud van de vragen en de berekening van het cijfer. En wat dat laatste betreft, met als voorbeeld een examen, vooral de vraag hoeveel punten nodig zijn voor een voldoende. Bij de inhoud is het van belang bij de formulering van vragen erop te letten dat deze duidelijk verwijzen naar herkenbare begrippen, situaties of vaardigheden zoals die beschreven staan in examenprogramma's. Het gebruik van standaardformuleringen bevordert de duidelijkheid in dit opzicht.

### **E. Objectiviteit**

Toetsuitkomsten dienen onafhankelijk te zijn van storende invloeden van beoordelaars.

## **6. Toetscontrole en toetsanalyse**

De Nieuwe Veste verwacht van al haar docenten, dat zij toetsen laten controleren door een collega van de vaksectie danwel een collega van één van de scholen binnen het samenwerkingsverband DOVOC. Daarnaast verwacht De Nieuwe Veste van docenten, dat zij zelf toetsen analyseren. Binnen De Nieuwe Veste wordt het centraal examen geanalyseerd met behulp van Wolf.

Tijdens het jaarlijkse formele gesprek tussen leidinggevende en docent legt de docent verantwoording af over de uitgevoerde toetscontroles en toetsanalyses.