

Uit de praktijk

□ Henk van Berkel en Sylvester Draaijer

Gids voor toetsontwikkeling

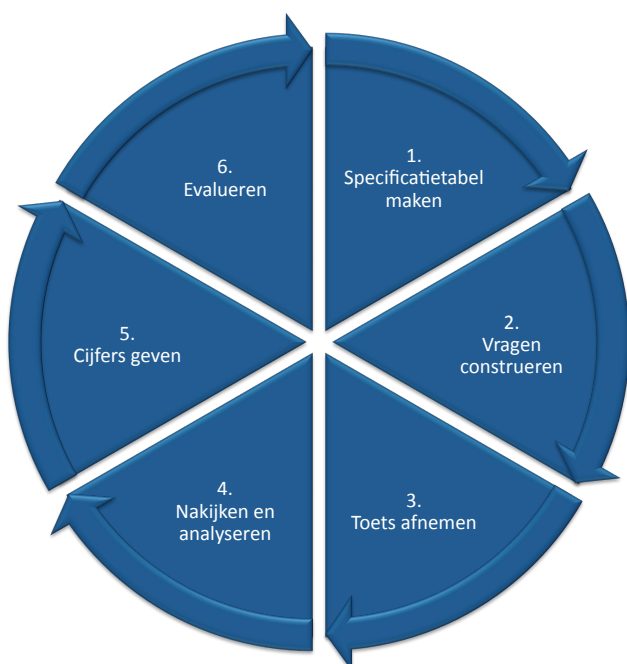
Steeds weer blijkt dat docenten en examinatoren behoefte hebben aan informatie die helpt bij het construeren van toetsen. Niet te veel theoretische achtergrondinformatie, maar concrete informatie 'hoe je het moet doen'. In die behoefte probeert deze bijdrage te voorzien.

Opzet

Deze Gids voor toetsontwikkeling behandelt in de vorm van checklists punten waaraan toetsconstructeurs moeten denken wanneer ze een toets gaan construeren. De theorie hierachter is zoveel mogelijk weggelaten. Die staat in het boek *Zicht op toetsen* waarvan deze bijdrage een zeer beknopte samenvatting is¹.

Met nadruk moet worden gesteld dat de inhoud van deze bijdrage primair betrekking heeft op summatieve schriftelijke toetsen, niet meer en niet minder. Het is niet onwaarschijnlijk dat de inhoud ook relevant kan zijn voor andere toetssituaties, maar die waren niet het uitgangspunt van de auteurs.

De bijdrage is chronologisch opgebouwd. De paragrafen volgen de verschillende stappen in het toetsconstructieproces (zie figuur 1) zoals die zich in werkelijkheid zouden moeten voordoen.



Figuur 1 Het proces van toetsconstructie

¹ Berkel, H.J.M. van. (1999). *Zicht op toetsen. Toetsconstructie in het hoger onderwijs*. Assen: Van Gorcum & Comp.

1. Maak een specificatietabel, kies vraagtype(n) en toetslengte

Een specificatietabel bestaat enerzijds uit een aantal belangrijke onderwerpen waarover u vragen wilt stellen en anderzijds uit het cognitieve niveau waarop u wilt richten met de vragen.

Onderwerpen	Weten – Reproductie Toetsvragen op het niveau van het reproduceren van kennis doen een beroep op het geheugen (feitenkennis). Feiten dienen herkend te worden of feiten dienen in een verband te worden gebracht.	Toepassing Bij toepassingsvragen worden studenten voor een voor hen nieuw probleem gesteld. Het probleem wordt opgelost door verschillende kenniselementen met elkaar in verband te brengen of met een formule, wet of oplossingsprincipe.	Inzicht Inzichtvragen vereisen van studenten een hoge mate van reorganisatie van het probleem en zelfstandig denken. De vragen doen een beroep op intellectuele vaardigheden als interpretatie, analyse, synthese en evaluatie.
1			
2			
3			
Enz.			

1.1. Onderwerp en aantal vragen

- Vermeld de belangrijkste onderwerpen, meestal voldoet een aantal van rond de vijftien.
- Het aantal vragen is van belang voor de representativiteit (inhoudsvaliditeit) van de toets.
Een juiste verdeling van de vragen over de cellen in de specificatietabel zal er voor zorgen dat de toets representatief is.
- Het aantal vragen is van belang voor de betrouwbaarheid van de toets (met meer vragen bereikt u doorgaans een hogere betrouwbaarheid).
- Hoe belangrijker het onderwerp is, des te meer vragen u er over moet stellen.
- Vermeld in de cellen het aantal vragen dat u wilt stellen over het betreffende onderwerp op het betreffende niveau.
- Zorg er voor dat over ieder onderwerp tenminste één vraag wordt gesteld.

Hoeveel vragen er kunnen worden gesteld, is afhankelijk van de vraagtypen en de toetstijd (indien die gelimiteerd is).

1.2. Antwoorttijd open vraag

Soort vraag	Antwoorttijd	Voorbeeldvraag
In- en aanvulvraag	½ minuut	Er wordt volgende week besloten of een van de meest _____ [haten] belastingen verdwijnt. Wat is het onderdeel van het binnenoor waarin zich het orgaan van Corti bevindt? Dat is de/het _____.
Kort antwoordvraag	1 minuut	Wat deed Pilatus volgens het Nieuwe Testament met zijn handen nadat hij Christus ter dood had veroordeeld?
Lang antwoordvraag	10 minuten	Wat waren de langetermijneffecten (noem er 3) van de Kruistochten? Beredeneer je antwoord.
Opstelvraag / Betoogvraag	25 minuten	Beredeneer de ontwikkelingen over de periode van tien jaar van de werkloosheid, als in één klap alle lonen worden gehalveerd. Ga uit van de situatie dat er verder geen andere wijzigingen in de conjunctuur optreden.
Berekeningsvraag	afhankelijk van de complexiteit van het vraagstuk	Stel dat de aandelenprijs momenteel € 50,- is. In de komende zes maanden zal de prijs ofwel dalen tot € 30,- ofwel stijgen tot € 80,-. Wat is de delta van een call optie met een uitoefenprijs van € 50?

1.3. Antwoordtijd gesloten vraag (waarbij geen berekening wordt gevraagd)

Soort	Antwoord-tijd	Aantal normaliter benodigd voor een betrouwbare toets	Voorbeeldvraag
Juist / Onjuist vraag of twee alternatieven	50 sec	80	Een kenmerk van het toepassen van een relatieve cesuur bij tentamineren is dat de kwaliteit van het gegeven voorafgaande onderwijs <i>geen</i> invloed heeft op het aantal geslaagden. (J/O) Wat is binnen de benadering van de Boston Consulting Group de centrale grootheid? A) Cash flow B) Marktgroei
Drie alternatieven	60 sec	60	Aan drie studenten wordt gevraagd een voordracht te houden tegen het gebruik van kernenergie. Alle drie de studenten zijn eigenlijk vóór het gebruik van kernenergie. Benthe krijgt voor de voordracht € 5,-, Maria krijgt € 50,- en Sarah krijgt € 100,-. Welke studente zal conform de cognitieve dissonantietheorie, met de grootste waarschijnlijkheid haar mening veranderen door de voordracht? A) Benthe B) Maria C) Sarah
Vier alternatieven	75 sec	40	De Boston Consulting Group heeft de zogenaamde 'business portfolio' matrix ontwikkeld. Welke van de onderstaande alternatieven hoort bij het kwadrant met een hoge marktgroei en een laag relatief marktaandeel? A) Cash cow B) Dog C) Question mark D) Star

Bedenk

- Het is een hardnekkig misverstand dat met gesloten vragen uitsluitend feitenkennis is te bevragen.
- Het is een hardnekkig misverstand dat vragen over feitenkennis automatisch resulteren in hoge scores en vragen naar toepassing of inzicht in lage scores.
- Lang antwoordvragen en opstelvragen kosten doorgaans veel tijd om na te kijken. Het is bovendien bewezen dat bij het nakijken van open vragen allerlei dingen kunnen misgaan die ten koste gaan van de betrouwbaarheid (en dus de eerlijkheid) van de beoordeling.
- Een vuistregel is: bij meer dan vijftig deelnemers is een toets met gesloten vragen doelmatiger dan een toets met open vragen.
- Lang antwoordvragen kosten veel tijd om te beantwoorden waardoor er minder verschillende onderwerpen in de toetstijd kunnen worden bevestigd. Dit kan snel ten koste gaan van de representativiteit.

2. Vragen construeren en controleren

Inspiratiebronnen

Inspiratiebronnen voor het construeren van vragen zijn, naast uiteraard de eindtermen:

- De problemen/taken in studieboeken, syllabi e.d.
- Hoofdpunten uit de literatuur en colleges
- Het toekomstig beroep
- De vaardigheden en practica tijdens de cursus
- Discussies in de colleges, werkcolleges e.d.
- Vragen van reeds afgenomen toetsen (met name vragen die niet te moeilijk of te makkelijk zijn en die een goed onderscheid maken tussen de studenten die de stof wel en die de stof niet beheersen)
- Grafieken, tabellen, schema's en afbeeldingen van relevante kenmerken of processen kunnen dienen om daarbij meerdere vragen te ontwikkelen

Als u de gangbare indeling 'weten/reproductie – toepassing – inzicht' wilt volgen, gebruik dan bij het ontwikkelen van vragen bepaalde **werkwoorden** zoals die hierna worden gegeven.

Weten -Reproductie	Benoemen, Beschrijven, Citeren, Definiëren, Extrapoleren, Identificeren, Interpretieren, Onderscheiden, Opsommen, Parafraseren, Samenvatten, Schatten, Selecteren, Uitleggen, Vertalen, Vertellen in eigen woorden
Toepassing	Berekenen, Demonstreren, Gebruiken, Maken, Ontwikkelen, Oplossen, Organiseren, Produceren, Relateren, Transfereren, Veranderen, Voorbereiden, Wijzigen
Inzicht	Bekritisieren, Categoriseren, Componeren, Concluderen, Contrasteren, Deduceren, Formuleren, Herschrijven, Illustreren, Interpretieren, Maken, Onderscheiden, Ondersteunen, Ontrafelen, Ontwerpen, Rechtvaardigen, Relateren, Samenvatten, Schetsen, Toelichten, Valideren, Verdedigen, Vergelijken, Verklaren, Waarderen

Onderstaande **startzinnen** geven houvast bij het formuleren van vragen, vooral vragen die een hoog cognitief niveau beogen te meten.

Weten: Nagaan of de student kennis heeft van onderwerpen	Wat is de beste definitie voor ...? Wat is (niet) karakteristiek voor ...? Uit welke onderdelen bestaat het probleem? Wat is de geschiedenis van het probleem? Welke verschillende categorieën zijn er in het probleem?
Kritisch denken (evalueren): Nagaan of de student de eigenschappen van feiten, procedures, principes of theorieën kan gebruiken	Wat is het meest effectief (gepast) voor ...? Wat is beter (slechter) ...? Wat is het meest effectief voor ...? Wat is de meest kritieke stap in de procedure? Als je weet dat X waar is, wat is dan tevens waar over Y? Wat is (niet) nodig in een procedure? Wat is het belang van het probleem?
Kritisch denken (voorspellen): Nagaan of de student op basis van feiten, procedures, principes of theorieën kan aangeven wat gevolgen, consequenties e.d. zijn	Wat zou er gebeuren als ...? Als dit gebeurt, wat zou je doen? Op basis van wat ..., wat zou je doen? Gegeven ... wat is belangrijkste reden dat ...
Probleem oplossen (gegeven een scenario): Nagaan of de student op basis van gegeven problemen oplossingen kan geven of oplossingen kan beoordelen	Wat is de aard van het probleem? Wat heb je nodig om dit probleem op te lossen? Wat is een mogelijke oplossing? Wat is de meeste effectieve (efficiënte) oplossing? Waarom is ... de meest effectieve (efficiënte) oplossing?

Tips

- Construeer minimaal 25% meer vragen dan u uiteindelijk in de toets wilt opnemen. U zult altijd vragen na de eerste constructieronde niet goed achten.
- Denk vooruit; als u tevens al een herkansing moet ontwikkelen of toetsen voor opeenvolgende jaren: ontwikkel direct meerdere vragen.
- Maak gebruik van een vragenbankprogramma om uw vragen in op te slaan.
- Betrek anderen bij het constructieproces.

2.1. Antwoordmodellen

Een antwoordmodel is een hulpmiddel voor de nakijker van open vragen dat dient om de betrouwbaarheid van het nakijken te verhogen. Het bevat de volgende punten:

- Opsomming in trefwoorden van wenselijke antwoorden, en hun puntentoekenning.
- Procedure wanneer het antwoord niet voorkomt in de genoemde trefwoordenlijst.
- Wat te doen met taalfouten? Negeren/puntenaftrek/voorwaardelijke minimumkwaliteit vereist om tot beoordeling over te gaan?
- Bij een lang antwoordvraag, essay of opstel: inhoud, de wenselijke technische vormgeving en vereiste argumentatie en een puntentoekenning op onderdelen.

2.2. Na de constructiefase

Wijd minimaal één bespreking met anderen aan de conceptvragen waarin de inhoud, het antwoordmodel c.q. het correcte alternatief en de vorm ter discussie staan.

2.3. Checklist

Hanteer de onderstaande checklist **nadat** u de vragen heeft geconstrueerd

Inhoud Is de gekozen toetsvorm (open of gesloten) de meest geëigende? Bestaat de toets uit een voldoende aantal vragen? Bevat de vraag slechts één duidelijk probleem? Bevat de vraag geen subjectieve uitspraken? Staat bij een citaat in de vraag de context? Bevat de vraag geen overbodige informatie? Heeft de vraag niet onnodig betrekking op een detail? Is de vraag geen strikvraag? Is bij een stellingvraag de stelling wel 100% juist of 100% onjuist? Bevat de vraag voldoende informatie voor beantwoording?	Is de vraag grammaticaal juist geformuleerd? Bevat de vraag geen ingewikkelde zinsconstructies? Kan de formulering van de vraag geen aanleiding geven tot misverstanden? Zijn ontkenningen als <i>niet</i> onderstreept of gecursiveerd? Bevat de vraag geen dubbele ontkenningen? Bevat de vraag geen woorden als altijd, nooit, meestal, zeker? Zijn de conventies wat betreft spelling, symboolgebruik, interpunctie, en dergelijke in acht genomen? Is de vraag opsplitsbaar in een gegeven en een vraaggedeelte?
Vorm Bevat de vraag geen vage aanduidingen? Is de vraag niet onnodig negatief geformuleerd?	Proces Is er een specificatietabel gemaakt? Heeft er een bespreking met vakgenoten plaatsgevonden?

Open vragen	Gesloten vragen
Geeft de vraag voldoende informatie over de gewenste lengte en vorm van het antwoord? Is duidelijk dat een antwoord verklaard/toegelicht moet worden? Is er een antwoordmodel gemaakt?	Komen er geen overlappingsen in de alternatieven voor (antwoorden die subsets van elkaar zijn veroorzaken veel verwarring)? Zijn alle alternatieven ongeveer even lang? Zijn de alternatieven in oplopende/alfabetische volgorde geplaatst? Zijn alle alternatieven plausibel? Is de vorm 'geen van bovenstaande' of 'alle bovenstaande' echt nodig?

3. De toets samenstellen en afnemen

3.1. Samenstellen

Een toets is meer dan een verzameling toetsvragen.

- Stel uit de goedgekeurde vragen de toets samen met behulp van de specificatietabel.
- Cluster de vragen per onderwerp in de volgorde zoals die in het onderwijs aan de orde zijn gekomen.

Maak – om afkijken tegen te gaan – minimaal twee versies van de toets door de volgorde van vragen te wijzigen.

3.2. De afname

De toetsinstructie voor pen-en-papier toetsen dient informatie te bevatten over de volgende onderwerpen:

- Tijdsduur
- Gesloten vragen: invulinstructie (bijvoorbeeld: het streepje zetten)
- Gewicht van de vragen (bij open vragen kan het gewicht verschillend zijn per vraag; bij gesloten vragen telt iedere vraag in beginsel even zwaar)
- Wijze van de zak/slaaggrensbepaling
- Bekendmaking van de uitslag: termijn en plaats
- Raadplegen van hulpbronnen: welke zijn wel/niet toegestaan?
- Verbod op het gebruik van mobiele telefoons
- Het aantal bladzijden en het aantal vragen in de toets; deelnemers zijn zelf verantwoordelijk voor eventuele misdrukken
- Verwijzing naar het fraudereglement
- Wat te doen bij vragen, bijvoorbeeld hand opsteken
- Regeling toiletbezoek
- Inleverinstructie
- Mogelijkheid tot het maken van opmerkingen

4. De toets nakijken en analyseren

Open vragen	Gesloten vragen
1. Gebruik het vooraf gemaakte antwoordmodel. 2. Kijk de toets 'dwars' na; dat wil zeggen per vraag en niet per deelnemer. 3. Verander af en toe de volgorde van de nakijkstapel. 4. In de praktijk zal het vaak niet mogelijk zijn om alle vragen door twee beoordelaars na te laten kijken, maar kijk in ieder geval de toets van deelnemers die rond de zak/slaaggrens scoren nog een keer na. Middel eventueel beide beoordelingen. 5. Rond de scores pas op het einde af. 6. Ook van een open vragentoets kunt u eenvoudig een toets- en itemanalyse maken in bijvoorbeeld Excel of SPSS. 7. Verzamel het commentaar; leg dit eventueel naast de toetsanalyse en bespreek het met de vragenconstructeur. 8. Neem een beslissing over de antwoordmodellen: handhaven, wijzigen of aanvullen. 9. Kijk, indien daar aanleiding toe is, de vragen opnieuw na met het vernieuwde antwoordmodel.	1. Geef de formulieren aan de instantie die de toets voor u scoort. 2. Meestal kan deze instantie de eerste (psychometrische) analyse uitvoeren. 3. Verzamel het commentaar en leg dit naast de toets- en itemanalyse en bespreek het met de vragenconstructeur. 4. Neem een beslissing over de vragen: handhaven, verwijderen of sleutelwijziging. Besef dat deze verandering over het algemeen slechts kleine wijzigingen in de uitslag oplevert. 5. Laat opnieuw een analyse uitvoeren.

5. Cijfers geven

Er zijn vele methoden van cijfers geven. Ze vallen onder drie categorieën: absoluut, relatief en compromis.

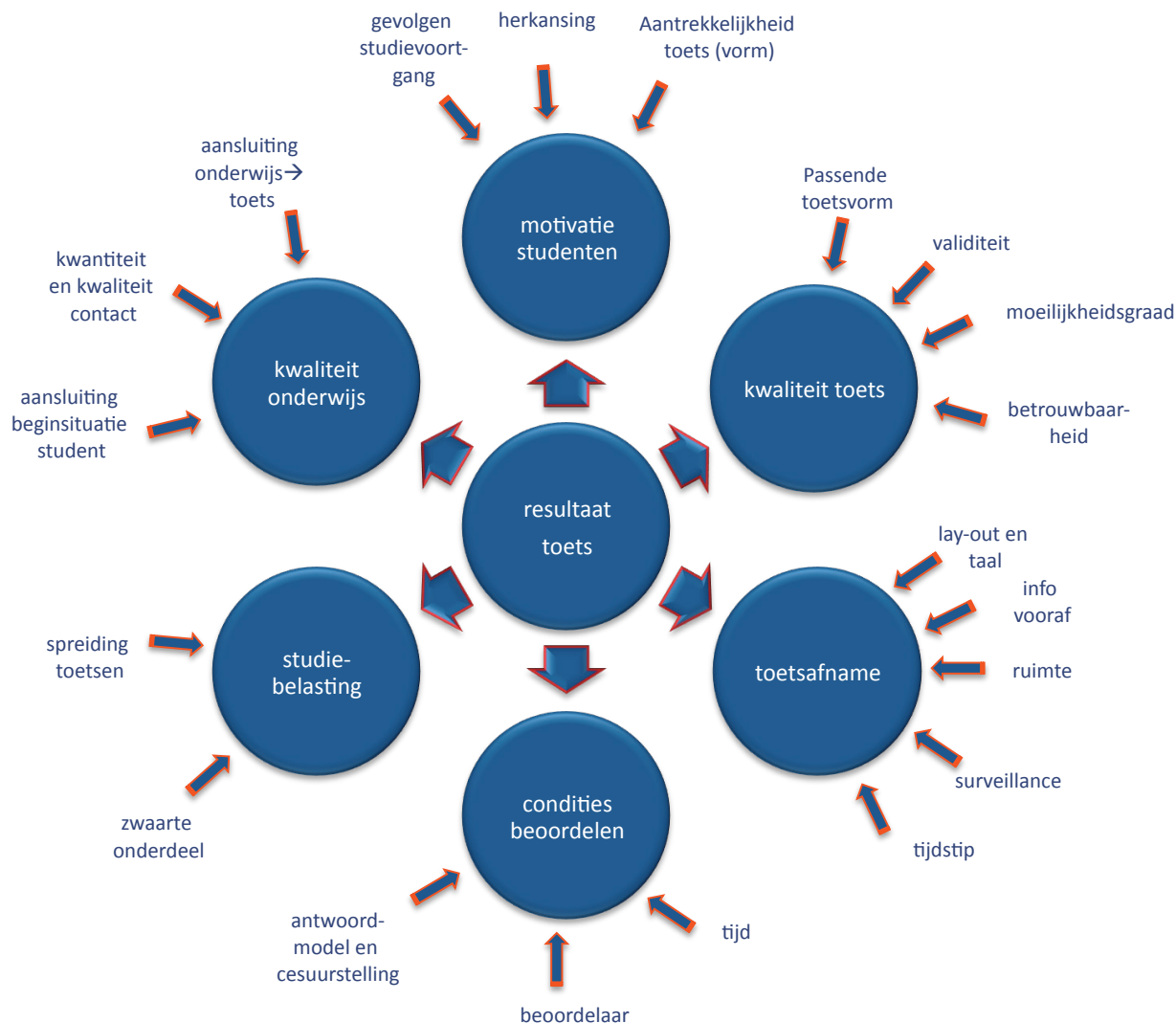
Absoluut	Bij de absolute methode bepaalt u vooraf de voldoende/onvoldoende grens (cesuurscore). Die grens wordt ingegeven door het uitgangspunt dat gecontroleerd moet worden of deelnemers voldoen aan de eisen die kunnen worden afgeleid uit de doelstellingen van het onderwijs.
Relatief	De relatieve methode gaat uit van het idee dat de toets moet zijn afgestemd op de doelgroep waarvan in beginsel de meerderheid in staat moet worden geacht om te slagen voor een toets. Omdat vooraf niet bekend is waartoe studenten bij het gegeven onderwijs en de gegeven toets in staat zijn, kan de norm niet vooraf worden vastgesteld. Het resultaat op de toets moet worden afgewacht om de norm te kunnen vaststellen.
Compromis	De compromismethode probeert de principiële verschillen tussen absoluut en relatief normeren te overbruggen en heeft in praktijk de voorkeur. Bij de compromismethode wordt in het algemeen uitgegaan van een absolute norm en wordt aangegeven onder welke omstandigheden van deze absolute norm zal worden afgeweken.

Open vragen	Gesloten vragen
1. Begin met het bepalen van de grens voldoende/onvoldoende. Het is niet ongebruikelijk die te leggen bij 55% of 60% van de maximaal te behalen score. 2. Indien u vindt dat er teveel deelnemers zakken voor uw toets, zijn de volgende aanpassingen in de procedure mogelijk. a. In plaats van 60% stelt u de grens op 55% van de maximale score en/of b. U neemt niet de maximale score als uitgangspunt, maar het gemiddelde van de vijf hoogste scores. 3. U heeft nu de ruwe score die een 0 oplevert en de ruwe score die een 10 oplevert; de overige cijfers bepaalt u door de ruwe scores op te delen in 10 gelijke eenheden met daaraan gekoppeld een cijfer.	1. Begin met het bepalen van de grens voldoende/onvoldoende. Het is gebruikelijk deze bij 55% of 60% van de hoogst mogelijke score te leggen met inachtneming van de kansscore. 2. De kansscore is het aantal geldige vragen in de toets gedeeld door het aantal alternatieven per vraag. Bijvoorbeeld de kansscore met 105 juist/onjuist vragen is $105/2 = 52,5$. Dus scores in de range tussen 52,5 en 105 zijn betekenisvol. 3. Een score van 52,5 ($105 - 52,5$) geeft een cijfer 0 en een score van 105 een 10. De overige cijfers bepaalt u door de range op te delen in 10 gelijke eenheden met daaraan gekoppeld een cijfer. 4. Indien u vindt dat er teveel deelnemers zakken voor uw toets, kunt u besluiten om niet de maximale score met een 10 te belonen, maar het gemiddelde van de vijf hoogste scores.

6. De toets evalueren

1. Kijk naar de betrouwbaarheid van de toets. Is deze lager dan 0,70, dan is aan te raden een volgende keer meer vragen in de toets op te nemen.
2. Bestudeer de zeer moeilijke en makkelijke vragen. Zoek oorzaken hiervoor zodat u de volgende keer dergelijke vragen kunt vermijden.
3. Bestudeer de psychometrisch zeer goede vragen (dat zijn vragen die een goed onderscheid maken tussen competente en niet competente deelnemers). Dat blijkt uit de zogenaamde Rit-waarde van de vraag. Dergelijke vragen zijn het waard om ook in de volgende toets op te nemen.

Bedenk dat het resultaat op een toets door vele aspecten wordt beïnvloed (zie figuur 2).



Naar idee van de Haagse Hogeschool

Figuur 2 De invloed van het resultaat op de toets

□ De heer dr. H.J.M. van Berkel is hoofdredacteur van *EXAMENS* en werkzaam aan de Universiteit van Maastricht.

E-mail: h.vanberkel@maastrichtuniversity.nl.

De heer ir. S. Draaijer is onderwijskundig adviseur bij de afdeling hoger onderwijs van het Onderwijscentrum van de Vrije Universiteit Amsterdam. E-mail: s.draaijer@vu.nl.