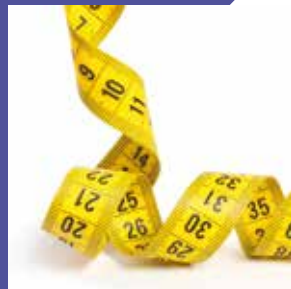


Klopt dit wel?



Met een betrouwbare toets meet je wat je wilt meten

■ Silvester Draaijer

Silvester Draaijer is Onderwijsadviseur & Adviseur ICT en Onderwijs bij de Vrije Universiteit te Amsterdam
Email: s.draaijer@vu.nl.

■ Jackelien ter Burg

Jackelien ter Burg is Toegepast Onderwijskundige en werkzaam als freelance toetsdeskundige en lid van de redactie van EXAMENS.

Deze uitspraak hoor je vaak, maar de twee belangrijkste kwaliteitseisen aan toetsen worden hier verhaspeld. Als de toets inderdaad meet wat bedoeld wordt om te meten, is de toets met name valide. Een toets is betrouwbaar als de toets consistent meet en het resultaat niet op toeval berust, ongeacht de validiteit van de toets.

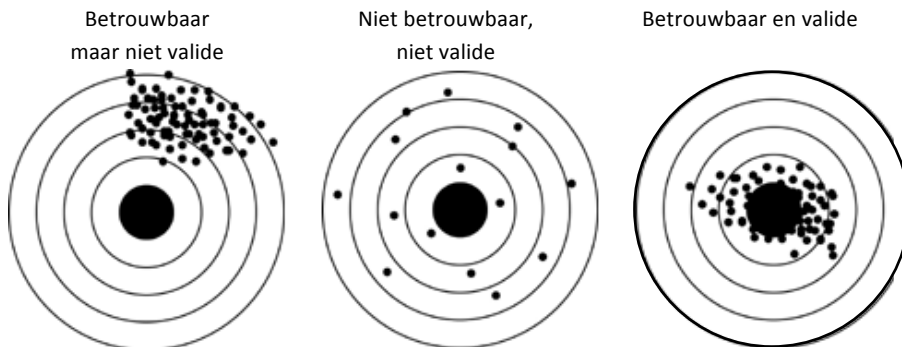
Validiteit versus betrouwbaarheid

Stel je wilt spreekvaardigheid Engels toetsen met meerkeuzevragen. Je hebt hiervoor een grote vragenbank samengesteld en de vragenbank is zo ingericht dat elke toetsafname uniek is. Je geeft Mo een week lang elke dag een toets en Mo krijgt op elke toets dezelfde score. Deze toets is dus

heel betrouwbaar. Maar de toets zal niet valide zijn; je kunt zaken als Engelse uitspraak immers niet valide toetsen met meerkeuzevragen. Validiteit en betrouwbaarheid zijn in onderstaande figuur weergegeven als pijlen op een dartbord. De pijlen staan voor de vragen, de roos is hetgeen we willen toetsen. Op het eerste dartbord komen alle pijlen mooi bij elkaar terecht, maar wel boven de roos. De toets is betrouwbaar, maar niet valide. Bij het derde dartbord komen de pijlen bij elkaar rondom de roos, de toets is betrouwbaar én valide.

Het belang van validiteit en betrouwbaarheid

Men vindt in het algemeen dat een toets als meetinstrument op de eerste plaats zo valide mogelijk moet zijn. Bij een gebrekkige validiteit kun je geen onderbouwde uitspraak doen over



wat het behaalde resultaat eigenlijk voorstelt in het licht van wat je wilde weten; wat stelt dan de hoogte van de score voor?

Een studietoets moet meestal de onderwijsinhoud dekken (inhoudsvaliditeit), maar een studietoets moet ook het gedrag toetsen dat van belang wordt geacht (construct of begripsvaliditeit). Vaak wordt hiervoor een toetsmatrijs opgesteld waarin de verschillende inhoud en het te verwachten in te zetten cognitieve proces, zoals feiten herkennen, begrip hebben of het kunnen toepassen van concepten tegen elkaar worden afgezet.

Naast valide dient een toets echter ook zo betrouwbaar mogelijk te zijn. Dat wil zeggen dat de scores van een kandidaat op de toets vergelijkbaar zijn bij herhaalde afname (en bij gelijkblijvende vaardigheid). Als dat namelijk niet zo is, dan krijgt het toeval te veel de overhand.

Een toets zal helaas nooit 100% betrouwbaar kunnen zijn. Er zijn altijd toevallige factoren die de prestatie van een kandidaat kunnen beïnvloeden en dus een meetfout veroorzaken. Denk aan een buurman die veel herrie maakt of vermoeidheid bij de kandidaat door een slechte nacht. Je kunt als toetsontwikkelaar wel proberen de meetfout zo klein mogelijk te houden door onder andere elk vraag zo eenduidig mogelijk te formuleren, voldoende vragen in de toets op te nemen en/of eventuele beoordelaars te trainen. Ook kun je overwegen vaker te toetsen en het eindoordeel over een kandidaat te baseren op meerdere toetsen.

Psychometrische gegevens

Vanuit de psychometrie wordt aangegeven dat een niet-betrouwbare toets per definitie niet-valide is. Wat stelt de score op een toets immers voor, als de uitslag op een toets te veel van het toeval afhangt? Voor toetsen opgebouwd uit gesloten vragen zijn hier vrij eenvoudig psychometrische gegevens over te verkrijgen, zoals Cronbach's alfa waarden (waarover in latere afleveringen van deze rubriek meer). Voor meer complexe toetsen met bijvoorbeeld rapporten en analyses of performance assessments zijn de psychometrische gegevens moeilijker te verkrijgen. Dan moet de mate van betrouwbaarheid worden opgemaakt uit de dekkendheid van bijvoorbeeld beoordelingsvoorschriften, de nauwgezetheid van de beoordelingsprocedure en de mate waarin beoordelaars het met elkaar eens zijn over de score op een taak.

Conclusie

Een betere stelling is: met een valide toets meet je wat je wilt meten, bij een betrouwbare toets is de toetsscore consistent. ■

Meer lezen?

- Piet Sanders, De betrouwbaarheid van toetsscores http://www.toetswijzer.nl/html/toetsenopschool/hoofdstuk3_herzien.pdf
- H.J.M. van Berkel, Anneke Bax, Desirée Joosten-ten Brinke. (2014) Toetsen in het hoger onderwijs. Houten: Bohn Stafleu van Loghum.