



Wereldwijd bestaan er veel systemen en standaarden waarmee een uitspraak kan worden gedaan over de kwaliteit van toetsinstrumenten.

Waarborgen van kwaliteit van complexe toetsen

■ Barbara Spieringhs

B. Spieringhs MEd. is opleidingscoördinator bij de Master toetsdeskundige en lerarenopleider bij Fontys Educatie. E-mail: b.spieringhs@fontys.nl

■ Peggy de Jong

P.W.G.A. de Jong MSc MEd is hoofddocent bij de HAN. E-mail: peggy.dejong@han.nl

■ Cor Sluijter

Dr. C. Sluijter is docent bij de Master Toetsdeskundige van Fontys, vice-president van de Association for Educational Assessment Europe en oud-voorzitter van de NVE. E-mail: c.sluijter@fontys.nl

■ Erik Roelofs

Dr. E.C. Roelofs is universitair hoofddocent bij de Universiteit Twente, faculteit Behavioral Management and Social sciences (BMS), afdeling Learning, Data analytics and Technology (LDT). E-mail: e.c.roelofs@utwente.nl

Onder invloed van een veranderde kijk op opleiden en leren, is in het beroepsonderwijs steeds meer nadruk komen te liggen op de ontwikkeling van wendbare vaardigheden en competenties, die ingezet worden in authentieke beroepssituaties. De benodigde kennisbasis wordt in samenhang met het te ontwikkelen professioneel repertoire verworven. Dit maakt dat bestaande toetsvormen, gericht op het afzonderlijk meten van kenniselementen, niet meer toereikend zijn. Er is dan ook een groeiende focus te zien op authentieke toetsen die beter aansluiten bij het beoogde beroepsprofiel.

Authentieke toetsen zijn toetsen waarmee de beheersing van meerdere vaardigheden tegelijk of complexe vaardigheden (professionele competenties, zoals het kunnen doen van onderzoek, lesgeven of het stellen van een diagnose) worden gemeten, die een grote variatie aan bewijsmateriaal vereisen. Het waarborgen van de kwaliteit van deze complexe toetsen is echter een uitdaging en een systeem voor het in kaart brengen en verbeteren van die kwaliteit is wenselijk. Het Systeem voor Assessment-evaluatie (SVA) is zo'n systeem.

Waarom een systeem?

Wereldwijd bestaan er veel systemen en standaarden waarmee een uitspraak kan worden gedaan over de kwaliteit van een toetsinstrument. Het doel van dergelijke systemen is tweeledig: enerzijds het waarborgen van de kwaliteit van toetsen en anderzijds het bieden van inzicht in mogelijkheden om deze kwaliteit te verhogen. Het SVA is bedoeld voor gebruikers die de kwaliteit van een zelfontworpen assessment in kaart willen brengen. Door systematisch de validiteit en betrouwbaarheid



Het doel van dit systeem is het evalueren van de kwaliteit van een assessment om op basis daarvan vervolgacties te formuleren waarmee die kwaliteit verhoogd kan worden.

van toetsen te evalueren, kunnen onderwijsinstellingen en examencommissies niet alleen de kwaliteit van hun huidige toetsen garanderen, maar ook continu verbeteringen doorvoeren.

Bestaande systemen

In Nederland zijn de systemen van COTAN en RCEC het meest bekend. De COTAN (Commissie Testaangelegenheden Nederland), verbonden aan het Nederlands Instituut van Psychologen (NIP), doet met behulp van een concreet beoordelingskader uitspraken over de kwaliteit van een breed scala aan tests, toetsen en vragenlijsten (Evers et al., 2010). Het RCEC (Research Center voor Examinering

en Certificering) is een onafhankelijke audit- en adviesorganisatie die met een specifiek daarvoor ontwikkeld systeem de kwaliteit van studietoetsen en examens beoordeelt (Sanders et al., 2018).

Ook buiten Nederland bestaat een scala aan codes, richtlijnen, standaarden, reviewsystemen en criteria om de kwaliteit van toetsen en examenprogramma's te borgen (Sluijter et al., 2018). Zo zijn er Standards for Educational and Psychological Testing, uitgegeven door de American Educational Research Association (AERA), de American Psychological Association (APA) en de National Council on Measurement in Education (NCME). Het European Framework of Standards for Educational Assessment (EFSA) is ontwikkeld door de European Association for Educational Assessment (EAEA) en Cambridge Assessment is een internationale examengroep die eigen kwaliteitseisen (principes) heeft ontwikkeld en vastgelegd in de Cambridge Approach to Assessment (Oates, 2017). Deze systemen zijn voornamelijk gericht op (kennis)toetsen of tests en bieden nog onvoldoende handvatten om de kwaliteit in kaart te brengen van complexe toetsen, hierna assessments genoemd. Het SVA is ontwikkeld ter aanvulling op bestaande systemen.

Methode

Bij het ontwikkelen van het systeem is gebruik gemaakt van een iteratief ontwerpproces, dat is vormgegeven door middel van de Design Thinking methode. Deze methode bestaat uit vijf iteratieve fasen waartussen heen en weer bewogen wordt: Empathize (inleven), Define (definiëren), Ideate (ideeën genereren), Prototyping (ontwikkelen) en Testing (uitproberen). Bij het ontwikkelen van het SVA is het prototype in zes opéénvolgende cycli door beoogde gebruikers en experts getest. Op basis van de bevindingen is het prototype bijgesteld en is uiteindelijk tot een definitieve versie gekomen.

Om het systeem te realiseren, zijn bestaande systemen geanalyseerd op toepasbaarheid op de evaluatie van assessments. Op basis van die

analyse zijn er relevante aspecten aangetroffen in verschillende systemen die aan de basis hebben gelegen voor het eerste prototype van het SVA. Dit prototype is geëvalueerd door de ontwerpers en op basis van deze evaluatie bleek er meer nodig te zijn om een uitspraak te kunnen doen over de validiteit van het te evalueren assessment.

Met behulp van inzichten uit het proefschrift van Wools (2015) en Evidence Centred Design (Mislevy, 2014) zijn er aanpassingen gedaan die hebben geleid tot het tweede prototype. Het tweede prototype is getest door beide ontwerpers aan de hand van een assessment uit de eigen onderwijscontext.

De inzichten uit die testfase hebben geleid tot een derde prototype dat voorgelegd is aan twee experts en zes gebruikers uit mbo en hbo.

Op basis van de inzichten van de experts en gebruikers is er onder andere een voorbeeld toegevoegd in de toelichting van prototype 4. Deze vierde versie is getest door de ontwerpers samen met collega's van de eerste twee auteurs. Deze werkwijze leidde tot het inzicht dat de samenwerking meer opleverde bij het invullen van het systeem, waardoor dit is opgenomen als advies voor gebruikers in prototype 5.

Dit vijfde prototype is voorgelegd aan vijf experts wiens input heeft geleid tot aanpassingen voor prototype 6. Hierna is een laatste testronde onder acht gebruikers uitgezet. De

opmerkingen, ervaringen en feedback van deze testronde zijn meegenomen in de totstandkoming van de laatste versie van het SVA. De input die geleid heeft tot de verschillende prototypen is weergegeven in figuur 1. Het SVA is te vinden op: <https://toetsbekwaamheid.nl/tool/systeem-voor-assessmentevaluatie-sva/>

Wat brengt het SVA?

Het doel van SVA is het evalueren van de kwaliteit van een assessment om op basis daarvan vervolgacties te formuleren waarmee die kwaliteit verhoogd kan worden. Het systeem is bedoeld om te bepalen waar verbetering mogelijk of noodzakelijk is. Het systeem is dus ontwikkelingsgericht en niet evaluatief bedoeld.

Het systeem kan snel inzicht geven, maar kan ook als terugkerend onderdeel van een kwaliteitscyclus gebruikt worden of als ondersteuning ter voorbereiding op een gesprek over de kwaliteit van assessments met stakeholders zoals managers, werkveld of accreditatiepanels. Ook kan het systeem fungeren als startpunt voor een gesprek over toetsing binnen een docententeam of kan het systeem worden ingezet om het gesprek aan te gaan met ontwerpers van een assessment.

Hoe werkt het SVA?

Het SVA bestaat uit zeven onderdelen, zoals weergegeven in figuur 2. De kwaliteit van het



Figuur 1. Iteratief ontwerpproces SVA

aspect waar het onderdeel zich op richt wordt in kaart gebracht door verschillende vragen te beantwoorden. Bij elke vraag is een toelichting opgenomen om het correct beantwoorden te faciliteren. Aangezien het systeem is bedoeld voor gebruikers die de kwaliteit van een zelf-

ontworpen assessment in kaart willen brengen is een hoge mate van toetsbekwaamheid van de gebruiker van belang. Denk hierbij aan gebruikers die regelmatig bezig zijn met het in kaart brengen of verhogen van de kwaliteit van toetsen zoals

Onderdelen van het systeem voor assessmentevaluatie (SVA)	
Doel van het assessment	<i>Is helder waar dit assessment voor is?</i> <i>Dit onderdeel brengt in beeld in hoeverre de plaats van het assessment in het toetsprogramma, de doelgroep(en) en het gebruiksdoel van het assessment duidelijk beschreven zijn. Wat wordt er geëxamineerd en waarom?</i>
Bouwplan	<i>Is het bouwplan van het assessment helder?</i> Validiteit is het belangrijkste kwaliteitsaspect van een toets; een bouwplan kan helpen een valide assessment te ontwerpen. Indien helder beschreven is wat het meetdoel is, is het vervolgens van belang om de te meten doelvaardigheden, deskundigheden, of kennisaspecten te definiëren in termen van taakklassen, uit te voeren taakprocessen en bijbehorende producten. Vervolgens is de vraag hoe deze zichtbaar kunnen worden gemaakt binnen toetstaken.
Constructie en standaardbepaling	<i>Is het assessment tot stand gekomen met inachtneming van de kwaliteitseisen?</i> Nadat er is nagedacht over een passende set van toetstaken is het belangrijk te bekijken welke beoordelingscriteria er gehanteerd worden om het werk of de aanpak van de deelnemer te beoordelen, hoe er voor betrouwbaarheid gezorgd wordt en welke standaardbepaling van toepassing is.
Afname	<i>Zijn de procedures voor de afname op orde?</i> De procedures voor het inleveren van uitgevoerde assessmenttaken of afnemen van het assessment moeten helder zijn zodat zowel deelnemers als beoordelaars weten wat er gebeuren moet. Daaronder vallen ook afspraken en procedures rondom fraude.
Beoordeling, verwerking & analyse	<i>Hoe komt de beoordeling tot stand?</i> Het beoordelen van een assessment is complexer dan bij schriftelijke toetsen; het is van belang dat de instructie voor beoordelaars volledig en duidelijk is en dat beoordelaars naar behoren worden geselecteerd, getraind en geïnstrueerd.
Communicatie & registratie	<i>Zijn de processen rondom communicatie en registratie van het resultaat op orde?</i> Nadat de verrichte werkzaamheden (of resultaten daarvan) zijn beoordeeld, is het nodig het resultaat inzichtelijk te maken voor de deelnemer(s). Dit onderdeel legt de focus op het vastleggen van het resultaat en de communicatie richting de deelnemers.
Evaluatie	<i>Welke onderdelen bieden ruimte voor verbetering?</i> Door consequent te evalueren en op basis daarvan onderbouwde verbetervoorstellen te doen kan de kwaliteitszorg voor het assessment geoptimaliseerd worden. Is er sprake van een ingerichte kwaliteitscyclus voor het assessment? En is duidelijk wie daarbij betrokken is (bijvoorbeeld beoordelaars, deelnemers, ontwerpers en/of collega's in een borgende functie)?

Figuur 2. De onderdelen van het SVA

leden van een toets- of examencommissie. Aangeraden wordt om het systeem door twee gebruikers, onafhankelijk van elkaar, in te laten vullen en deze invulling met elkaar te bespreken om op die manier tot een eendoorddeel te komen. Het is ook mogelijk de kwaliteit van een assessment te evalueren met een onderwijsteam, met name als het gaat om een assessment op grond waarvan zwaarwegende beslissingen worden genomen (zoals een afstudeerwerk).

Na het beantwoorden van alle vragen is er per onderdeel zicht op de huidige kwaliteit van het assessment en is duidelijk welke eventuele verbeteringen noodzakelijk zijn. Hierbij is het bouwplan (onderdeel 2) het belangrijkste; de aspecten die hier aan bod komen, zijn voorwaardelijk voor een valide assessment.

Tot slot

Het systeem biedt een waardevol kader voor het waarborgen en verbeteren van de kwaliteit van assessments. De kracht van het SVA is mede te danken aan de betrokkenheid van gebruikers en experts¹. Hun waardevolle bijdragen waren onmisbaar bij de ontwikkeling van het systeem.

Het SVA is ontwikkeld voor de hbo-context, maar blijkt op basis van gebruikerservaringen in bepaalde mate ook bruikbaar voor het mbo. Tijdens de ontwikkeling is voortdurend gezocht naar een balans tussen volledigheid en gebruiksvriendelijkheid. Tegelijkertijd vereist het systeem een aanzienlijke mate van deskundigheid, wat de toegankelijkheid voor minder ervaren toetsontwikkelaars kan beperken.

In de komende jaren zal het SVA worden geëvalueerd en verder ontwikkeld, met specifieke aandacht voor het ondersteunen van nieuwe toetsvormen en het verder optimalise-

ren van de gebruiksvriendelijkheid. Dit proces sluit aan bij de ambitie om een toetspraktijk te creëren die toekomstbestendig is en aansluit bij de dynamische eisen van het beroepsonderwijs. Het optimaliseren van assessments vraagt om een vergelijkbare aanpak, waarbij systematisch input wordt verzameld om onderbouwde aanpassingen door te voeren. ■

Referenties

- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010, mei). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. NIP. Geraadpleegd op 24 oktober 2023, van <https://psynip.nl/wp-content/uploads/2021/11/COTAN-Beoordelingssysteem-2010.pdf>
- Mislevy, R. (2004, mei). *A Brief Introduction to Evidence-Centered Design*. National Center for Research on Evaluation, Standards, and Student Testing. Geraadpleegd op 17 oktober 2023, van <https://files.eric.ed.gov/fulltext/ED483399.pdf>
- Oates, T. (2017). *The Cambridge Approach to Assessment*. Cambridge Assessment. Geraadpleegd op 20 oktober 2023, van <https://www.cambridgeassessment.org.uk/Images/cambridge-approach-to-assessment.pdf>
- Sanders, P., Brouwer, A., Eggen, T., & Veldkamp, B. (2018). *RCEC beoordelings-systeem voor de kwaliteit van studietoetsen en (praktijk)examens (versie 2.0)*. Research Center voor Examinering en Certificering.
- Sluijter, C., Hemker, B. & Eggen, T. (2018). *Beoordelen van de kwaliteit van toetsen en examens. Deel 1: Systemen en criteria*. Arnhem: Cito.
- Wools, S. (2015). *All About Validity. An evaluation system for the quality of educational assessment*. [Proefschrift]. Enschede: Universiteit Twente.

¹De auteurs bedanken Theo Eggen, Desirée Joosten-ten Brinke, Gerard Straetmans, Bas Hemker, Cor Sluijter en Erik Roelofs voor hun inhoudelijke feedback.