

Interview met Dr. Saskia Wools

We moeten proberen om met zo min mogelijk moeite, zoveel mogelijk betekenis te geven aan de prestatie van een kandidaat

■ Harry Molkenboer

Ing. H.F.A.M. Molkenboer MEd is master Toetsdeskundige, eigenaar van Bureau voor Toetsen & Beoordelen en is redactielid van Examens. E-mail: harry.molkenboer@toetsen-beoordelen.nl

■ Jan Jacobs

Ir. J.W.J. Jacobs is Adviseur Onderwijs en Examinering binnen Yuverta en is redactielid van Examens. E-mail: j.jacobs@yuvarta.nl

In de periode 2007-2012 zijn in Examens, tijdschrift voor de toetspraktijk twaalf portretten gepubliceerd van ‘Kopstukken uit de Nederlandse examenwereld’. Het waren portretten van Nederlanders die een bijzondere plaats innemen in de wereld van toetsing en examinering. Nu ruim tien jaar verder is de redactie gestart met een nieuwe serie portretten van ‘Kopstukken uit de Nederlandse toetswereld’. We hebben veel nieuwe Kopstukken!

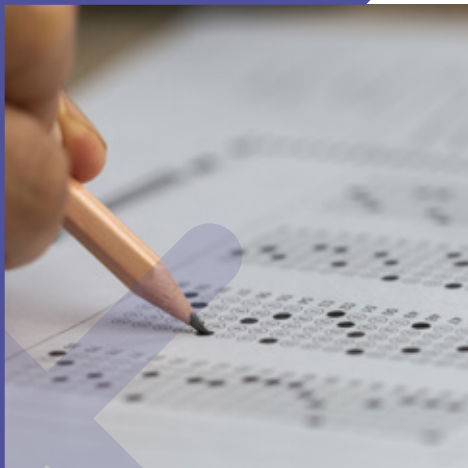
In deze uitgave van Examens het interview met ons zevende kopstuk dr. Saskia (S.) Wools. We spraken Saskia op 23 oktober 2023 online en wel op een avond, want Saskia is als bestuursvoorzitter van de Raad van Bestuur van Cito druk, maar ook flexibel. Dank Saskia, dat het zo kon.

Wie ben je?

“Ik studeerde van 2003 tot en met 2008 aan de universiteit van Leiden (bachelor Pedagogische wetenschappen; onderwijskunde) en aan de universiteit Utrecht (master Onderwijskunde; Educational design en consultancy). In 2006 startte ik bij Cito als stagiaire/onderzoeker bij de afdeling Psychometrisch Onderzoek & Kenniscentrum (POK) met het schrijven van een thesis voor mijn masteropleiding. Vervolgens ben ik niet meer weggegaan bij Cito. De thesis ging over kwaliteit van competentieas-



Saskia Wools



Het besef dat toetsing onderdeel is van een groter geheel dat wordt gevormd door onderwijs, leren en ontwikkelen zie je terug in de visie die ik nu bij Cito uitdraag.

sessments, ik ontwikkelde een variant van het COTAN beoordelingskader gericht op het beoordelen van de kwaliteit van toetsen.

Vervolgens ben ik doorgerold in innovatieprojecten en innovatiemanager geworden. Onderzoek en innovatie trokken mij erg aan. Op 26 juni 2015 promoveerde ik aan de universiteit Twente met het proefschrift 'All About Validity. An evaluation system for the quality of educational assessment'. Ik bleef betrokken bij onderzoeksprojecten over evalueren en valideren van onderwijskundige meetinstrumenten. Daarnaast voerde ik onderzoek uit naar formatieve toetsen en innovatieve toetsvormen. Ik was projectleider Referentiesets en vanaf 2017 was ik manager van het CitoLab, Cito's innovatieafdeling.

In 2019 werd ik samen met mijn team genomineerd voor de NRO-Verbindingsprijs met ons project Groeimeter! Een formatief toetsplatform voor rekenen in het primair onderwijs. De NRO-Verbindingsprijs is in het leven gebracht om projecten die wetenschap en praktijk aan elkaar verbinden, een platform te geven. Het koppelen van wetenschap aan de praktijk en wetenschap nuttig maken voor de praktijk, is wel iets wat echt bij mij past.

Vanaf 2020 was ik directeur Onderzoek, Kennis en Innovatie van CitoLab en verantwoordelijk voor het inrichten en uitbouwen van de publieke kennisorganisatie van Cito. Binnen CitoLab gaf ik leiding aan een multidisciplinair team dat zich richt op onderwijskundig en psychometrisch onderzoek en op de ontwikkeling van prototypes van innovatieve assessments. Daarnaast was ik verantwoordelijk voor de inhoudelijke invulling en uitvoering van Cito's onderzoeks- en innovatieagenda.

Vanaf 1 januari 2023 ben ik Voorzitter van de raad van bestuur van Cito. Samen met bestuurslid mr. Arthur Weijers vormen wij een collegiaal bestuur en sturen wij de organisatie aan. Dat doen we zowel voor Stichting Cito als voor Cito B.V. Van 2006, als scriptie-schrijvende stagiaire/onderzoekende masterstudent van toen 23 jaar, tot heden in mijn nieuwe rol als bestuurder, heb ik veel mogen doen en leren bij Cito. Cito heeft mij dus veel gebracht en ik ken Cito dan ook van binnenuit. Nu mag ik, samen met Arthur en de vele deskundige collega's, mijn inhoudelijke kennis, inzichten en visies, alsmede mijn managerial ervaringen gebruiken binnen mijn nieuwe rol.

Ik geloof erin dat als je een leidende rol in examens wilt spelen, het heel handig is om te weten waarover je het hebt, vooral als je wat in beweging wilt krijgen. Examinering (summatief) en toetsing (breder dan alleen summatief) zijn zo specialistisch, en er zijn zoveel knoppen om aan te draaien, dan moet je wel wat boeken hebben gelezen en

inzicht en overzicht hebben om iets te kunnen betekenen. De vacature van bestuurder van Cito kwam voor mij dan ook eigenlijk op een heel passend moment. Het voelde voor mij een logische stap als vervolg op mijn rol als directeur van het CitoLab. Het tijdsmoment speelde ook mee. De wereld om ons heen veranderd. Binnen examinering en toetsing doen zich zoveel ontwikkelingen voor, daar moet Cito een antwoord op geven. Onze organisatie staat op het punt om zichzelf opnieuw uit te vinden. Dat ik daar leiding aan mag geven vind ik echt een eer.”

Hoe zie je dat uitvinden dan?

“Bij Cito stond het begrip kansengelijkheid altijd al centraal. Ik vind dat zelf ook echt heel belangrijk. Met examens nemen wij kinderen, leerlingen, studenten en kandidaten de maat, en daar hangen soms zwaarwegende beslissingen aan. Beslissingen ook die iemands levensloop bepalen. En dat doen we met een technisch instrument zoals het examen, waar we goed, zorgvuldig en onderbouwd over na hebben gedacht. Maar tevens met de wetenschap dat we niet een 100% zorgvuldige beslissing kunnen nemen.

We scheren feitelijk elke kandidaat met hetzelfde examen over dezelfde kam. Is dat een gelijke kans geven? Technisch wel, want ze krijgen allemaal hetzelfde examen dat we op een zo valide mogelijke wijze hebben ontwikkeld. Het examen doet dus recht aan wat het examen moet meten, en wat aansluit op wat een vervolgopleiding of wat de praktijk aan kennis, inzicht, vaardigheden en houding van de geslaagde mag verwachten. Maar is in de huidige tijd, in klassen met zoveel diversiteit aan leerlingen, een volledig gestandaardiseerd examen wel het best passende meetinstrument? Moeten we kinderen, leerlingen, studenten en kandidaten niet beter accommoderen? Het zou moeten gaan over gelijkwaardige beslissingen en niet over gelijke prestaties aan de hand van een uniform instrument, maar dat is wel wat we nu veelal doen. Over dit soort

zaken nadenken past bij mij. Het laat iets zien over de visie die we bij Cito hebben, maar laat tegelijk ook zien wie ik ben.

Dit sluit aan bij het gedachtegoed van Adriaan de Groot, de grondlegger van Cito in de jaren zestig. Hij dacht juist vanuit gelijke kansen. Niet de schoolmeester die het te goed voor had met de zonen en de dochters van de welgestelden en die dus ook zo werden beoordeeld, maar met toetsen met meerkeuzevragen die objectief te scoren waren. Daardoor konden op basis van gelijkheid beslissingen worden genomen; zie zijn pamflet, want dat was het, Vijven en zessen van De Groot (eerste druk 1966).

Ondertussen is toetsing heel krachtig geworden. We lezen en horen over toetsstress en verschraving van onderwijs. We moeten als Cito nadenken wat onze rol is. We realiseren ons dat toetsen niet allesomvattend zijn. Tegelijkertijd geloven we in verzamelde data om gericht vervolgstappen te kunnen zetten bij het leren en ontwikkelen. Hoe kunnen nieuwe meetinstrumenten eruit zien voor gelijke kansen? We zetten de deur open voor iets nieuws, dat is spannend want het antwoord hebben we nog niet, daar is onderzoek voor nodig.”

Wat is je specialisme?

“Moeilijke dingen makkelijk maken. Dat heb ik altijd al gehad. Daarnaast houd ik ervan om te opereren op het snijvlak van publiek en privaat, en om in samenwerking met anderen het onderwijs echt beter te maken. Mijn specialisme zou je kunnen samenvatten met mijn gerichtheid op validiteit, en daarmee samenhangend de kwaliteit van toetsen en het ontwikkelen en delen van kennis en innovatie op dat gebied. En eigenlijk gaat het over onderwijs, leren en ontwikkelen. Examinering is daar een stukje van. We moeten examinering niet groter maken dan het is.

Wat ik ook wel als mijn specialisme zie is om anderen uit te dagen zich te ontwikkelen en

te groeien. Ik schep daar ook bewust ruimte en tijd voor. Ik stimuleer collega's bijvoorbeeld ook om de masteropleiding tot toetsdeskundige te volgen. Daar heb je zoveel aan, vooral omdat die opleiding zo breed is.

In mijn rol als bestuurder kan ik veel dingen die ik interessant vind samenbrengen: mijn brede toetstechnische achtergrond, mijn kennis over kwaliteit van toetsing, mijn analytisch vermogen, een stuk visie-ontwikkeling en mijn ervaring met innovatieprocessen. Maar bijvoorbeeld ook leidinggeven en coachen. En dit kan ik doen in een sterke organisatie die deel wil en kan nemen aan een anders georganiseerd toetsingsstijfperk, met een aangepaste invulling voor het bieden van gelijke kansen. En dat zonder los te laten waar we sterk in zijn, zoals onderzoek en de ontwikkeling en analyse van kwalitatief hoogwaardige toetsen en examens.

Ik merk, nu ik dit zo samenvat, dat ik hier energie van krijg. Is sta heel positief in het leven, lach veel (en dat hebben we tijdens het interview gemerkt, het was weer een leuk interview (red.)). En als ik niet lach dan is dat niet goed en weet ik dat er een verandering nodig is!”

Hoe zouden we jouw brede expertise op toetsgebied mogen samenvatten?

“Bij jullie vorige vraag is mijn expertise al breed naar voren gekomen. Wat ik nog wil toevoegen is dat veel mensen mij kennen van validiteit. Jullie begonnen daar zelf ook al over. Als ik nu ruim acht jaar na mijn proefschrift en zes jaar nadat ik een hoofdstuk over de validiteit van toetsscores in ‘Toetsen op school; onder redactie van Piet Sanders’ (2017) (zie kader 2) heb geschreven, terugkijk, zie ik validiteit meer als de bruikbaarheid van resultaten van een beoordelingsmoment voor een bepaald doel. Validiteit wordt bepaald in interactie met de praktijk, het accent moet meer komen te liggen op het gebruik. Het gaat dus veel meer over de context en de breedte van het instru-

ment, en juist die breedte van het instrument valt ook onder de noemer van validiteit.

Als je examens of toetsen inzet voor bepaalde beslissingen, dan kijk je heel goed naar de kwaliteit van het instrument in termen van betrouwbaarheid, transparantie maar ook validiteit. Maar wat is de bruikbaarheid, wat zijn gevolgen in de praktijk, hoe worden onze instrumenten ingezet en hoe kunnen we daar rekening mee houden? We moeten wat mij betreft meer op zoek naar de interactie tussen de mens en het instrument. We weten bijvoorbeeld dat studenten en ook ouders voor hun kinderen strategisch gedrag vertonen. Er gebeurt dan iets tussen het toepassen van het beoordelingsinstrument en de praktijk-situatie. Dit beïnvloedt de validiteit van de scores en we meten dus minder goed wat we oorspronkelijk in gedachten hadden. Dit heeft ook weer zijn weerslag op het bieden van gelijke kansen. Kwaliteit is een construct dat ontstaat, kwaliteit heeft dus niet alleen een technische component die in het instrument zit. Daar moeten we over nadenken. De omgeving en de bijbehorende context waarin wij beoordelen moeten we horen en zien en daar aandacht aan schenken. Dit is eigenlijk ook denken vanuit validiteit en dat binnen een breder kader.”

Op wie of waarop richt je je vooral?

“Vanuit Cito richt ik mij heel breed; denk hierbij aan het primair onderwijs (po) en voortgezet onderwijs (vo), het mbo, de doorstroomtoets, volgsystemen po en vo, en denk aan inburgeraars, pabostudenten, beroepsbeoefenaars en certificeringstrajecten. We richten ons eigenlijk op iedereen die leert. Als ik naar mij persoonlijk als toetsdeskundige kijk, dan richt ik mij ook breed en dus ook op iedereen die leert. Ik gebruik daarvoor ook weer dat innovatieve denken. Kinderen en jeugd vormen een belangrijke doelgroep. Zij zitten meer op social media dan dat ze een boek lezen. Als je een examen of toets voor hen ontwerpt moet dat aansluiten bij hun natuurlijke habitat, dus die technologie moet je in onze toetsen en

examens terug gaan zien. Dit vergt innovatief en creatief denken en dat past mij wel. Maar ik ben ook weer niet voor rücksichtslos innoveren. Toetsen betekent dat je gericht uitvraagt om gericht gedrag uit te lokken, zo kom je tot een besluit met alle kennis en inzichten die we hebben opgedaan. Innovatie kunnen we inzetten om beter uit te vragen of bijvoorbeeld om inzichten bij elkaar te brengen, maar wel met beleid en visie.”

Wat versta je onder toetsing en/of examinering?

“Ik zie examinering in brede zin als een uitnodiging om gedrag te laten zien waarover je een beslissing kunt nemen. Dit is de formele definitie. Ik zie daarnaast ook, en maatschappelijk gezien, dat examinering veel meer is dan dit. Deelnemers kunnen tegen een examen op zien, maar het kan hen ook zelfvertrouwen geven als ze een examen behalen. Ik zie het ook als een overgangsritueel, een ‘rite de passage’ (denk aan de vo-examens) met feesten en het enthousiasme om toegang te hebben gekregen tot een vervolgopleiding. Ik zie tevens de stress van een herkansing of van zakken, de onzekerheid bij kandidaten. Examens zijn vaak belangrijke momenten in een leven, waarop je na het behalen trots kunt zijn, over kunt vertellen met iets dat een groot civiel effect heeft. Door corona werd besloten dat examens niet of anders werden afgenomen. Dat zorgt gelijk voor een ander gevoel voor deze scholieren over hun overgang naar het vervolgonderwijs. Ik vraag me af, zullen zij over 25 jaar tegen collega’s zeggen dat zij van het 2020 cohort zijn? Ik denk eigenlijk wel dat het zoveel impact had.”

Hoe en wanneer ben je geraakt door toetsing?

“Binnen onderwijskunde te Utrecht ben ik mij vooral gaan interesseren in onderwijsinnovatie en wilde ik graag tot een grote organisatie behoren waar ik wat zou mogen doen met innovatie. En dat werd dus Cito. Ik houd daarbij van complexe problemen om je in vast te

bijten en tot praktisch toepasbare oplossingen te komen. Ik ben dus eigenlijk in de toetsing gegroeid. Ik ben pas echt geraakt door toetsing bij het promoveren. Toen kwam ook het besef dat toetsing onderdeel is van een groter geheel dat wordt gevormd door onderwijs, leren en ontwikkelen. Dat zie je ook terug in de visie die ik nu bij Cito uitdraag: Toetsen is nooit een doel op zich. We moeten een balans vinden met het onderwijs en de omgeving, lees de maatschappij, dat is wat toetsing goed maakt.”

Wie zijn jouw leermeesters geweest?

“Dat is Piet Sanders (Examens 2009-1, Examens 2023-1). Ik vond het dan ook een eer dat ik op zijn uitvaart mocht spreken afgelopen januari. Naast Piet waren mijn leermeesters Theo Eggen (hij was ook mijn promotor) (Examens 2023-1), Anton Béguin (Examens 2021-1), Erik Roelofs (Examens 2016-3), Gerard Straetmans (Examens 2023-2), en internationaal mensen als Michael Kane (validiteitstheorie) en Stephen Sireci (validiteit) maar natuurlijk ook iemand als Anneke Blok (voormalig Voorzitter raad van Bestuur van Cito, Examens 2022-4).

Eigenlijk ben ik een spons. Ik neem overal wat op, dus ik heb veel leermeesters, maar de acht die ik heb genoemd zijn ze toch wel in het bijzonder.”

Wie beschouw je als vooraanstaand in de toetswereld?

“Dan denk ik direct aan Alina von Davier (psychometrician and researcher in computational psychometrics, machine learning, and education; Chief of Assessment at Duolingo). Veel mensen kennen Duolingo als een taal-leer app. Maar inmiddels brengen zij ook taal-examens uit onder leiding van Alina. Zij doen alles met automatisch gegenereerde content en complexe computational psychometrics. Echt inspirerend! Ook beschouw ik Dominique Sluijsmans, Tamara van Schilt-Mol en Janneke Helsloot als vooraanstaand. Om allerlei ver-

schillende redenen trouwens, maar zij zijn wel vooraanstaand in de toetswereld voor mij. Trouwens is het een idee voor jullie om iets te doen met aanstormend toetskundig talent? (Daar heeft de redactie inderdaad als eens over gesproken, maar eerst zijn nog enkele kopstukken aan de beurt (red..).)”

Wat heeft je gedreven en wat drijft je nog steeds?

“Mij drijft onderwijsverbetering. Dat sluit ook aan bij mijn baan in deze (semi) publieke sector. Werken voor de maatschappij. Dan gaat het onder meer over de impact van de rol van toetsen en examens binnen ons onderwijsstelsel, hoe dat georganiseerd wordt. Er is voor Cito werk te doen voor de plaats van toetsing binnen in het onderwijs, daar zijn we nog niet.”

Wat voor visie heb je op toetsing en hoe ziet die toekomst van toetsing er in jouw ogen uit?

“In aanvulling op wat ik hiervoor heb gezegd noem ik dat we moeten proberen om met zo min mogelijk moeite zoveel mogelijk betekenis te geven aan de prestatie van een kandidaat. Voor de kandidaat betekent dit dat hij met zo weinig mogelijk tijd en ze drempelloos mogelijk een prestatie moet kunnen neerzetten met een zo rijk mogelijke rapportage. Voor de onderwijsorganisatie betekent dit dat de afnames makkelijk te organiseren zijn met weinig beoordelings- en registratiewerk. Hiervoor zijn (digitale) innovaties nodig. Een innovatie die zich nu al aandient voor toetsing is AI (Artificial Intelligence). Karin Gerritsen-van Leeuwenkamp heeft in Examens 2023-2 al laten zien wat mogelijk is (Illustratie van ChatGPT bij toetsing en aandachtspunten voor wel-overwogen gebruik). Ik geloof in AI als een mensgestuurde oplossing. Ik zie het niet als een oplossing voor ‘maak maar een examen over ...’, maar wel ter ondersteuning bij het bedenken van alternatieven bij gesloten vragen, het klonen van vragen, het ontwikkelen van omvangrijke itembanken om adaptief toetsen

makkelijker mogelijk te maken en voor het ondersteunen van docenten bij het nakijken. We zijn nu veel tijd kwijt met nadenken over en ontwikkelen van vragen en opdrachten, AI gaat hierbij echt helpen. We gaan AI omarmen binnen de gehele toetscyclus, maar onderzoek hiernaar is nog wel nodig.

Een ander punt betreft het begrip ‘eerlijk’. In onze toetswereld gebruiken we het begrip ‘eerlijk’ eigenlijk niet. Wij spreken over valide, betrouwbaar en transparant toetsen. Terwijl onze afnemers van examens het wel hebben over ‘eerlijk toetsen’. ‘De toets was niet eerlijk’ (te moeilijk, vroeg niet wat geleerd was, te eenzijdig, we snapten niet wat gevraagd werd, de tijd was te kort en vergelijkbare uitspraken). ‘Eerlijk’ wordt door onze afnemers gebruikt als een intrinsiek kwaliteitsbegrip over examens. Maar kandidaten kunnen overigens ook eerlijk zijn en zeggen dat ze niet geleerd hebben en dus een onvoldoende behaalden. We spreken dus eigenlijk twee talen en dat communiceert lastig en leidt tot onbegrip. Hier zien we terug dat toetsing en onderwijs nog niet voldoende geïntegreerd zijn en twee kanten van dezelfde medaille zijn, terwijl ze samen aan een kant van de medaille horen te staan. Binnen het hbo zie ik dat ze daar volop met die integratie van toetsing, onderwijs en leren bezig zijn. Toetsing heeft dus niet alleen een technische component, maar ook die onderwijskundige component en daar moeten we als Cito wat mee.”

Waar ben je het meest trots op?

“Het meest trots ben ik op mijn zontje van anderhalf lachend (red.!) Maar ook jonge mensen om mij heen die je de gelegenheid kunt geven ons vak te leren. Want wees eerlijk, qua toetsing bevinden wij ons in een niche-markt. Als ik dan zie hoe collega’s van mij dat oppikken, dan ben ik trots, want die mensen hebben we nodig. Mensen moeten mensen om zich heen hebben waarin wordt geloofd. Ook ben ik trots op het validiteitstreintje als metafoor om de breedte van validiteit zichtbaar te maken (zie kader 1).

En dat ik, hoewel begonnen op de afdeling psychometrie van Cito, in mijn proefschrift geen enkele formule heb gebruikt, daar ben ik ook heel trots op (grapje natuurlijk)."

Wat zou je anders hebben gedaan?

"Mijn positie als bestuursvoorzitter bekleed ik nog geen jaar. Om nu al kunnen zeggen wat ik anders had moeten doen, kan ik nog niet. Kom over vijf jaar nog maar eens langs en dan heb ik vast wel wat te melden. Wat ik wel kan zeggen is dat als ik terugkijk over zeventien jaar Cito, ik nergens spijt van heb. En als er spijt zou zijn, dan vind ik dat je iets moet aanpassen of oplossen, maar in ieder geval iets moet doen."

Welke boodschap heb je voor de lezers van Examen?

"Kijk minder naar het instrument en meer naar de impact."



Beslissingen kunnen iemands levensloop bepalen.

https://video.saxion.nl/media/Webinar+Leer+&+deel+met+Eduxion+-+ValiderenA+het+zichtbaar+maken+van+de+kwaliteit+van+een+toets+en+een+toetsprogramma+door+Saskia+Woolts/1_t8coy2tui

Wagon	Claim	Bewijzen
1. Prestatie	Er is een taak, of een set taken, die gedrag uitlokt. Een gelijke prestatie leidt tot gelijke scores.	De toets. Steekproef van beoordelaars.
2. Score	Er is een samenhangend oordeel over de prestatie beschikbaar. De score zou ongeveer gelijk zijn als er andere taken zouden zijn.	(Deel)scores & toelichting. Betrouwbaarheid, of relatie met andere meting/informatie.
3. Test domein	Er is een beschrijving waarom de taken zo geconstrueerd zijn. Taken lokken gedrag uit dat beschreven is.	Rationale van opgaven/taken. Hardop-denk protocol.
4. Comp. domein	Er is een beschrijving van het te meten construct. Het beschreven construct wordt herkend door de praktijk en is relevant.	Omschrijving van het construct. Veldraadpleging, interview.
5. Pt. domein	Er is een analyse van wat nodig is in de beroepspraktijk. De beslissing sluit aan bij de verwachtingen van de praktijk.	Analyse op opleidingsniveau. Beschrijving methode om cesuur te bepalen.
6. Beslissing	Beslissing is voor iedereen goed doorgevoerd.	Omzettingstabel/administratie.

Kader 1. Het validiteitstreintje

Kader 2. Zeven bewijzen voor validiteit

In 'Toetsen op school' (Cito, 2017) heb ik daarover een hoofdstuk geschreven; 'Validiteit van toetsscores'. Daar schrijf ik over validiteit het volgende: 'Validiteit gaat over de interpretatie en het gebruik van toetsscores. Denk bijvoorbeeld aan een luistertoets Frans. Het is de bedoeling dat de scores op deze toets iets zeggen over de luistervaardigheid Frans van de leerlingen die de toets maken. We interpreteren de score op de toets dus als een maat voor luistervaardigheid. Vervolgens gebruiken we de score om te beslissen of een leerling voldoende luistervaardig is om naar de volgende klas over te gaan. Anders gezegd gaat het bij validiteit dus om de vraag of de toetsscores de informatie bieden waarin we geïnteresseerd zijn en of op basis van deze scores de gewenste beslissingen genomen kunnen worden'.

In het spraakgebruik duidt men validiteit aan met 'meet de toets wat die behoort te meten'. Dat is op zich goed, maar wel erg algemeen. We zouden dat moeten verbreden.

In tegenstelling tot betrouwbaarheid kunnen we de validiteit van toetsscores niet uitdrukken in een getal. Toch willen we wel bewijzen hebben of toetsscores valide zijn. De bewijzen kunnen zowel kwalitatief als kwantitatief zijn. We kunnen de validiteit van toetsscores het beste aantonen door bewijzen uit verschillende categorieën te verzamelen. Het is niet noodzakelijk om bewijzen uit alle categorieën te verzamelen, omdat het aantal en het soort bewijzen vooral afhankelijk is van de interpretatie en het gebruik van de toetsscores. We onderscheiden zeven categorieën:

1. inhoudsbewijzen;
2. complexiteitsbewijzen;
3. interne structuur bewijzen;
4. externe structuur bewijzen;
5. betrouwbaarheidsbewijzen;
6. extrapoleerbaarheidsbewijzen;
7. consequentiële bewijzen.

Nog een dingetje om het af te bakenen. Toetsen zelf kunnen niet valide zijn. Het gaat erom of het resultaat van een toets, dus de toetsscore, valide is. Ook bij betrouwbaarheid kunnen we niet zeggen of de toets toetsbetrouwbaar is, maar we kunnen wel zeggen (berekenen) of de scores op de toets betrouwbaar zijn.

Zie voor hoofdstuk 4 uit Toetsen op school: <https://docplayer.nl/13835099-4-de-validiteit-van-toetsscores-4-de-validiteit-van-toetsscores.html>. ■