

Interview met dr. Gerard (G.J.J.M.) Straetmans

Deugdelijke toetsing is onontbeerlijk om onderwijsprocessen efficiënt en effectief te laten verlopen.

■ Harry Molkenboer

Ing. H.F.A.M. Molkenboer MEd is master toetsdeskundige en eigenaar van Bureau voor Toetsen & Beoordelen en is lid van de redactie van EXAMENS. E-mail: harry.molkenboer@toetsen-beoordelen.nl

■ Jan Jacobs

Ir. J.W.J. Jacobs is voorzitter van de Regionale Examen Commissie Heuvelrug Rivierengebied van Yuverta en redactielid van EXAMENS. E-mail: j.jacobs@yuverta.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 vijftien 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 vijfde kopstuk dr. Gerard (G.J.J.M.) Straetmans.

We spraken Gerard online in het kortste interview tot nu toe. Gerard een beetje kennende (vanuit onder meer de masteropleiding Toetsdeskundige) handelt hij gedreven, onderzoekend, nauwgezet, zorgvuldig en graag efficiënt en effectief.

Voorafgaand aan elk interview sturen wij de vragenlijst naar elk kopstuk toe. We maken bij interviews wel vaker mee dat we de antwoorden op de vragen alvast toegestuurd krijgen. Dat zijn meestal korte zakelijke antwoorden die de kern van de antwoorden op de vragen weergeven, maar die het nog geen lezenswaardig portret van de persoon maken. Wel maken we dankbaar gebruik van die antwoorden tijdens het interview en bij het schrijven





Ik kreeg de kans om mijn oude liefde 'itembanking' weer op te pakken en ontwikkelde samen met vakinhoudelijk deskundigen een opgavenbank rekenen/wiskunde.

van het interview-verslag. Van Gerard ontvingen wij eigenlijk al een geheel uitgewerkt interview. Hij schreef hierover het volgende in de begeleidende e-mail: 'Bijgaand treffen jullie mijn schriftelijke reactie aan op de vragen die jullie aan mij wilden stellen. Ik heb hiervoor gekozen omdat het mij efficiënter leek dan (gebrekkig) mondeling reageren op gestelde vragen om daarna jullie teksten wellicht te moeten bijstellen omdat die toch niet precies verwoordden wat ik had willen zeggen. Ik hoop niet dat ik jullie hiermee ontrief.' Getekend Gerard Straetmans; efficiënt en effectief. En nee Gerard, je hebt ons niet ontriefd en het klopt ook wat je schrijft.

Gerard kent als aandachtig lezer van Examens de schrijfstijl van de interviews. Ook die heeft hij toegepast. Zie hier het interview van Gerard, met Gerard over Gerard, met dank aan Gerard.

Wie ben je en wat is je specialisme?

"Ik ben gepensioneerd en heb dus aanzienlijk meer verleden dan toekomst als het gaat om mijn professionele leven. Maar laat ik bij het begin beginnen. Ik kom uit een onderwijsfamilie. Mijn vader stond op zijn 18e al voor de klas en studeerde vanaf die tijd nijver door om zoveel mogelijk bevoegdheden te krijgen voor het geven van onderwijs in hogere onderwijsvormen. Hij was een studiestapelaar avant la lettre. Veel familieleden en de meeste vrienden van mijn ouders waren ook onderwijsmensen. Ik ben daardoor van jongs af aan ondergedompeld in onderwijsproblematiek; dat tekent je uiteraard.

Toen ik in 1965 de overstap moest maken van lagere school naar vervolgonderwijs was er nog geen Cito. Het was de hoofdonderwijzer zelf die bepaalde wie geschikt was voor welk type vervolgonderwijs. Ik was volgens hem voorbestemd om naar de MULO te gaan, wat ik niet leuk vond want mijn beste vriendjes gingen naar het gymnasium. Achteraf denk ik dat dit wel eens de grootste drijfveer kan zijn geweest voor mijn inzet ten dienste van de kwaliteit van beslissingen over onderwijsdeelnemers in mijn professionele leven. Na de MULO ben ik, net als mijn vader, een studiestapelaar geworden. Via HAVO en pedagogische academie belandde ik, na ruim een jaar Hare Majesteits wapenrok te hebben gedragen, op de Katholieke Universiteit Nijmegen waar ik pedagogische en andragogische wetenschappen heb gestudeerd. Ik studeerde af bij de vakgroep Onderwijskunde op het toen nieuwe onderwerp 'Itembanking'.

Anders dan nu was de arbeidsmarkt voor academici begin jaren 80 zeer slecht en ik was daarom de koning te rijk dat ik al na een

half jaar een tijdelijke aanstelling van twee jaar kreeg bij de subfaculteit Tandheelkunde in Nijmegen, om onderzoek te doen naar de effecten van een trainingsprogramma voor beoordelaars van tandheelkundige werkstukken. Wegens succes werd de aanstelling met drie jaar verlengd en omgezet in een promotietraject dat zich naast het genoemde onderwerp richtte op de ontwikkeling en validering van een assessmentmethodiek voor de probleemoplosvaardigheid van tandheelkunde studenten. Bij beide onderwerpen maakten we gebruik van een computer, wat in die tijd nogal opzien baarde want de personal computer was nog niet uitgevonden. Mijn passie voor de inzet van de computer bij toetsing stamt ongetwijfeld uit die tijd.

Na de promotie werd ik aangesteld als universitair docent bij de faculteit Toegepaste Onderwijskunde van wat toen nog de Technische Hogeschool Twente heette. Ofschoon ik daar veel geleerd heb over instructiesystemen en er heel veel interessante collega's waren, beviel me de overgang naar dit vakgebied niet erg en na drie jaar stapte ik over naar het Cito in Arnhem. Al gauw kreeg ik de kans om mijn oude liefde 'itembanking' weer op te pakken en ontwikkelde ik samen met vakinhoudelijk deskundigen van buiten een opgavenbank rekenen/wiskunde voor gebruik in de volwasseneneducatie en het mbo. De opgaven werden grootschalig beproefd bij de diverse doelgroepen en geschaald met behulp van procedures uit de itemresponsstheorie (IRT). Daarna organiseerde ik een samenwerkingsverband tussen Cito, Cinop en Koning Willem I College om voor rekenen/wiskunde een computergestuurd adaptief toetspakket (WISCAT) te ontwikkelen. Gelukkig kon ik collega Theo Eggen overhalen hierin te participeren. Van hem heb ik geweldig veel geleerd over deze elegante en efficiënte toetsprocedure. Enkele jaren later werd de WISCAT-familie uitgebreid met een nieuw pakket, dat speciaal ontwikkeld was voor pabo-studenten. Een tweede onderwerp dat permanent

aandacht heeft gekregen tijdens mijn Cito-loopbaan was performance assessment. De interesse daarvoor was natuurlijk al gewekt door mijn ervaringen bij Tandheelkunde, maar nam verder toe door de groeiende belangstelling in het onderwijs voor hogere-orde leerdoelen. Kennis vergaren is uiteraard belangrijk, maar uiteindelijk gaat het er toch steeds om wat je daarmee kunt doen. In steeds meer curricula werd ruimte gereserveerd voor vaardigheden en bekwaamheden. Verre van duidelijk was hoe het bereiken van deze hogere-orde doelen op deugdelijke en haalbare wijze kon worden vastgesteld. Er werd veel geëxperimenteerd met portfolio's. Het was mij echter al snel duidelijk dat de implementatie van het portfolio als assessmentmethodiek het risico op verkeerde beslissingen over bekwaamheid onverantwoord groot zou maken, aangezien portfolio's vaak uiterst zwakke bewijzen bevatten. Dat laatste was in mijn ogen een gevolg van één van de stilzwijgende uitgangspunten van het portfoliogebruik, dat luidde: 'De student is eigenaar van het portfolio en hij alleen bepaalt wat erin wordt opgenomen.' Dat leek me echt een dwaling. Als de opleiding met het diploma de garantie moet afgeven dat de bezitter ergens bekwaam in is, dan moet die opleiding toch minimaal zeggenschap hebben over wat geldige bewijzen zijn voor die bekwaamheid? Die overweging zette mij ertoe aan een systematiek te ontwikkelen voor het aanleggen van een bewijsdossier voor bekwaamheden. Het is bekend geworden onder de naam Protocol Portfolio Scoring (PPS).

In 2005 zocht Cito naar andere mogelijkheden om haar deskundigheid te promoten in het hoger onderwijs. De oprichting van een bijzonder lectoraat bij Saxion Hogeschool leek daarvoor een geschikte mogelijkheid en het is vooral aan de niet aflatende inzet van Cor Sluijter te danken dat een en ander gerealiseerd werd. Ik was vereerd om die plek te mogen bezetten. Ruim tien jaar lang heb ik samen met mijn kenniskring onderzoek gedaan naar en kennis verspreid over toetsing en examinering in de context van het hoger onderwijs.

Na mijn vervroegde pensionering in 2016 heb ik, op deeltijd-basis, gedurende vijf jaar met heel veel plezier meegewerkt aan de ontwikkeling en uitvoering van een splinternieuwe master op het gebied van toetsbekwaamheid aan de Fontys Lerarenopleiding in Tilburg; de masteropleiding Toetsdeskundige. Een van de hoofdbestanddelen van die functie was het verzorgen van de cursus Systematisch Ontwerpen van Educational Assessments. In die cursus kon ik veel van de in afgelopen jaren verworven inzichten op het gebied van het toetsen van kennis, vaardigheden en bekwaamheden delen met studenten. Die studenten waren veelal ervaren leraren, die na de master een vooraanstaande rol zouden gaan spelen bij de innovatie van toetsing en examinering in hun organisaties. De zeer gemotiveerde studenten van deze master waren gretig om de nieuwe kennis te verwerven en vooral om die te gaan toepassen in hun eigen situatie. Daar doe je het voor!

Ik zou bijna vergeten om te melden dat ik ook allerlei nevenactiviteiten heb vervuld, zoals bestuurswerk, redactiewerkzaamheden bij vakbladen, lidmaatschap van adviescommissies, en dergelijke.”

Hoe zouden we jouw brede expertise op toetsgebied mogen samenvatten?

“Ik weet niet zeker of mijn expertise wel zo breed is; ons vakgebied is verrassend omvangrijk en ik heb beslist niet overal verstand van. Maar uit wat ik tot nu toe gezegd heb over mijn loopbaan, zal duidelijk zijn geworden dat mijn expertise zich toespitst op drie onderwerpen: Itembanking, Computer-based testing, toetsing van vaardigheid en bekwaamheid.

Itembanking

Een oude liefde. Dit onderwerp heeft mij altijd gefascineerd omdat ik al snel begreep dat dit dé manier was om de kwaliteit van toetsing binnen het onderwijs naar een hoger niveau te tillen. Reden waarom ik voor mijn afstudeerproject bij Onderwijskunde stage ging lopen bij

het Cito. Daar werkte ik mee aan de ontwikkeling van een itembank Onderwijskunde voor lerarenopleidingen. Het was 1980 en ik deed er mijn eerste ervaringen op met de item-responstheorie, toen nog latente trektheorie geheten.

[Zie kader op pagina 31 over wat Gerard onder itembanking verstaat]

Computer-based testing

Nog een oude liefde. Mijn eerste ervaring met de inzet van een computer voor toetsing betrof een poging om het proces van het probleemoplossend handelen van de aspirant tandarts vast te leggen en te beoordelen. Dat was een volledig tekst-gebaseerde simulatie. We zijn nu 40 jaar verder en AI maakt het mogelijk om gewoon in gesprek te gaan met een door de software gesimuleerde patiënt. Mijn ervaring met computersimulaties is echter niet erg groot. Wél heb ik redelijk wat ervaring met de inzet van computers voor het samenstellen, presenteren en verwerken van kennistoetsen en met name met de adaptieve variant. Een adaptieve toets wordt tijdens het toetsproces samengesteld uit de itembank en de selectie van de items is zodanig dat de toets als geheel optimaal wordt afgestemd op de gemeten vaardigheid van de kandidaat. De inzet van de computer is daarbij onontbeerlijk omdat na elk antwoord de vaardigheid van de kandidaat geschat moet worden. Adaptieve toetsing is werkelijk een fantastische uitvinding en ik sta dan ook volledig achter de uitspraak dat adaptieve toetsing *le raison d'être* van de item-responstheorie is.

Toetsing van vaardigheid en bekwaamheid

Twee wapenfeiten zijn hier te melden. Het eerste is dat ik heb getracht orde in de chaos van performance assessment methodes te scheppen door het zeer grote aantal methodes dat in de literatuur beschreven wordt, toe te wijzen aan vier hoofdmethodes die wezenlijk van elkaar verschillen ten aanzien van criteria als betrouwbaarheid, generaliseerbaarheid en extrapolatiebaarheid. [Zie kader op pagina 33

voor een toelichting op de genoemde begrip-
pen] Het tweede wapenfeit is de ontwikke-
ling van het al genoemde Protocol Portfolio
Scoring. Deze systematiek zorgt ervoor dat
studenten een bewijsdossier opbouwen op
grond waarvan een deugdelijke beslissing kan
worden genomen over hun beroepsbekwaam-
heid. De volgende eigenschappen moeten dit
bewerkstelligen: Meervoudig bewijs, metho-
demix, meerdere assessoren, standaardset
van prestatiecriteria en een transparante
beslissingsprocedure. Wie graag precies wil
weten hoe dit werkt, kan hoofdstuk 3 lezen in
de reeks Toetsen op School, aflevering ‘Hoger
Onderwijs’: <https://www.cito.nl/-/media/files/kennis-en-innovatie-onderzoek/toetsen-op-school/cito-toetsen-op-school-ho.pdf?la=nl-nl>

Op wie of waarop richt je je vooral?

“Zoals ongetwijfeld voor alle toetsdeskundi-
gen geldt, is mijn werk er steeds op gericht
geweest ervoor te zorgen dat er betere beslis-
singen over onderwijsdeelnemers kunnen
worden genomen. In de afgelopen decennia
is er flink wat bereikt op dit gebied; er is veel
nieuwe kennis ontwikkeld en heel veel toet-
sen, toetsprocedures en andere hulpmiddelen
zijn beschikbaar gekomen om de kwaliteit
van de toetsing te verbeteren. Maar het
merendeel van de in het onderwijs afgeno-
men toetsen wordt nog altijd door docenten
ontwikkeld en schiet, zo hoorde ik geregeld
van mijn studenten in de masteropleiding
Toetsdeskundige, kwalitatief nog vaak tekort.
Dat verbaast me niet. In het onderwijs werken
nogal wat docenten die geen reguliere lera-
renopleiding hebben genoten en als ze die wel
gevolgd hebben, hebben ze daar maar bitter
weinig geleerd over toetsing en examinering.
Ik heb onlangs nog eens wat ‘gesnuffeld’ in de
verplichte literatuur van pedagogische acade-
mies en tweedegraads lerarenopleidingen en
was erg verbaasd daar geen boeken of readers
aan te treffen over dat onderwerp. Doorgaans
blijft de te bestuderen leerstof over toetsing
beperkt tot één of twee hoofdstukken in een
boek over algemene didactiek. Ik leid hieruit af

dat toetsing en examinering binnen leraren-
opleidingen nog steeds een ‘ondergeschoven
kind’ is, totaal niet in lijn met het belang van
dit onderwerp voor de onderwijspraktijk. Ik
juich nascholing op dit gebied (zoals bijvoor-
beeld via de BKE-/SKE-regeling) toe, maar als
er in de initiële lerarenopleiding onvoldoende
aandacht wordt besteed aan zo’n belangrijk
onderwerp, dan is dat toch een beetje dweilen
met de kraan open.”

Wat versta je onder toetsing en/of examinering?

“Ik prefereer de term ‘educational assessment’
en daaronder versta ik al die activiteiten die
iemand ontplooit om informatie te verga-
ren over de vorderingen (*achievement*) of
geschiktheid (*aptitude*) van onderwijsdeeln-
mers, om daarmee beslissingen te nemen over
de in-, door- en uitstroom in een onderwijs-
eenheid of over de kwaliteit van die onderwijs-
eenheid zelf. Als positieve beslissingen leiden
tot een uit te reiken diploma, spreken we van
examinering. En als het doel van educational
assessment is om uitspraken te doen over de
kwaliteit van een onderwijseenheid zelf, dan
duiden we dit aan met de term evaluatie. In de
dagelijkse onderwijspraktijk wordt jammer ge-
noeg niet erg zuiver omgegaan met de begrip-
pen. Dat bemoeilijkt niet alleen de onderlinge
communicatie, maar is misschien ook een
signaal dat men het vakgebied niet helemaal
serieus neemt. Hier is werk aan de winkel voor
lerarenopleiders en nascholers!”

Hoe en wanneer ben je geraakt door het onderwerp toetsing?

“Ik heb al gezegd dat mijn vroege negatieve
ervaringen met selectie waarschijnlijk hebben
geleid tot een bovenmatige interesse voor
het vakgebied educational assessment. Maar
geraakt werd ik pas toen ik kennismaakte met
de itemresponstheorie (IRT); echt enthousiast
werd ik toen ik las hoe een van de grote pro-
blemen van de klassieke testtheorie, namelijk
dat in een toetsscore de vaardigheid van de
kandidaat en de moeilijkheid van de toets on-

ontwaaibaar verenigd zijn, kon worden opgelost. Bij alle toetsen rekenen/wiskunde die we in de periode 1995-2005 hebben vervaardigd, zoals plaatsings- en vorderingentoetsen voor de volwasseneneducatie en voor het mbo, een diagnostische toets voor verpleegkundigen en een instroomtoets voor pabo-studenten, werd gebruik gemaakt van de unieke mogelijkheden die IRT te bieden heeft, zoals het direct met elkaar kunnen vergelijken van scores op verschillende toetsen en, als de toets met een computer wordt afgenomen, adaptieve toetsing.”

Wie zijn jouw leermeesters geweest?

De afstudeerrichting Onderwijskunde kende eind jaren 70 geen mogelijkheden om door middel van keuzevakken een eigen ‘draai’ te geven aan de kwalificatie als onderwijskundige. In het standaardprogramma was wel veel aandacht voor statistiek en voor toetsconstructie (het uitstekende werk van Adriaan de Groot en Bob van Naerssen: ‘Studietoetsen construeren afnemen analyseren’ moesten we zo ongeveer van buiten kennen). Docent Martin Mommers (beter bekend onder zijn kloosternaam frater Caesarius en ontwikkelaar van de methodes ‘Zó leren lezen / Veilig leren lezen’) heeft mij op degelijke wijze ingeleid in de principes van de klassieke testtheorie. Die basis was voldoende om op eigen gelegenheid meer kennis en vaardigheden op het terrein van educational assessment te vergaren, daarbij geholpen door de overvloedig aanwezige expertise van de Cito-collega’s. Eén van hen wil ik hier expliciet vermelden, omdat hij gedurende vele jaren steeds opnieuw de tijd nam om aspecten van IRT en adaptieve toetsing op geduldige en deskundige wijze uit te leggen: Theo Eggen. Behalve dat hij mijn leermeester was, was hij ook mijn *brother-in-arms* bij activiteiten als de WISCAT-pabo, de rekentoetsen VE en MBO, het schrijven van artikelen en, *last but not least*, bij het opzetten en uitvoeren van de masteropleiding Toetsdeskundige.

Wie beschouw je als vooraanstaand in de toetswereld?

“Ik vind dat we in Nederland verwend zijn met heel veel uitstekende toetsdeskundigen, maar wie van hen meer invloedrijk is dan de rest, is moeilijk aan te geven. Ieder heeft vaak zijn eigen specialisme en onderscheidt zich daarmee al snel van de rest. Ik wil ook niemand tekortdoen door hier wat namen te noemen en dus ook een heleboel namen niet te noemen. Wél kan ik wat buitenlandse toetsexperts noemen, maar dan alleen omdat ze mij persoonlijk geïnspireerd hebben, niet omdat ze per se vooraanstaand zouden zijn in de beroepsgroep. **Thomas Haladyna** heeft mij geïnspireerd met zijn boek over het ontwikkelen en valideren van meerkeuzevragen (*Developing and validating test items*). Zijn pleidooi om het format van een toetsitem nauwgezet af te stemmen op zowel de leerstofstructuur als het verlangde cognitieve proces is niet direct opzienbarend te noemen, maar ik denk wel dat docenten heel vaak het antwoord schuldig blijven als hen gevraagd wordt waarom een bepaalde vraag in een bepaald format is gesteld. Haladyna’s boek zet docenten aan om te reflecteren op hun toetsconstructie-activiteiten en zo de validiteit van de items en dus van de toets als geheel van meet af aan te bevorderen. **Michael Kane** noem ik vanwege zijn *‘argument-based approach to validation’*. Sinds de adoptie van ‘validiteit als een eenheidsconcept’ worstelden toetsvalideerders met de vraag of steeds uit alle vijf bewijsbronnen¹ geput diende te worden om toetsscores te kunnen valideren. Kane gebruikte zijn werkervaring bij de opleiding voor de advocatuur door toetsvalidering te beschrijven als een soort juridisch proces van claims pro en contra de argumentatie dat de scores van een toets te interpreteren en te gebruiken zijn voor een nader gespecificeerd doel. Als dit proces maar helder, samenhangend en plausibel was, zou men in principe zelfs met één bewijsbron kunnen volstaan. Dat sprak

¹Evidence based on test content; 2. Evidence based on response processes; 3. Evidence based on internal structure; 4. Evidence based on relations to other variables; 5. Evidence based on the consequences of testing. Bron: AERA, APA, & NCME (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Itembanking

Itembanking verwijst naar al die activiteiten die erop gericht zijn om een georganiseerde verzameling van toetsitems aan te leggen waaruit toetsen kunnen worden samengesteld voor formatieve en summatieve beslissingen over onderwijsdeelnemers. Het belangrijkste argument voor itembanking is dat kwalitatief goede items het verdienen om vaker dan één keer gebruikt te worden.

mij zeer aan, want ik was altijd al van mening dat het voor de doorsnee docent ondoenlijk was om heel veel tijd te besteden aan toetsvalidering. De meesten krijgen amper voldoende tijd voor het construeren van het toetsmateriaal zelf. Verder ben ik zeer gecharmeerd van het omvangrijke handboek over Educational assessment of students van **Susan Brookhart** en **Anthony Nitko**, dat inmiddels zijn 8e herdruk heeft beleefd. Dit boek behandelt alle belangrijke aspecten van het vakgebied voor zowel docenten, studenten als toetsdeskundigen op een didactisch voortreffelijke manier. Vreemd eigenlijk dat er in Nederland, met al zijn voortreffelijke toetsdeskundigen, nog geen handboek is geschreven dat zich inhoudelijk en didactisch hiermee kan meten.”

Wat drijft je?

“Mij drijft het besef dat onderwijs van het grootste belang is voor een samenleving en dat deugdelijke toetsing onontbeerlijk is om de onderwijsprocessen efficiënt en effectief te laten verlopen. Ik vind het daarom onbegrijpelijk dat er in de curricula van lerarenopleidingen nog steeds zo weinig aandacht wordt besteed aan toetsing en examinering. Gedurende mijn hele loopbaan heb ik mij daarom ingespannen om (aanstaande) leraren te laten

ervaren wat de kracht van goede toetsing is en hen kennis laten maken met handzame procedures om kennis- en praktijktoetsen te construeren voor gebruik in de eigen onderwijs situatie. Velen hebben hiervan kennis kunnen nemen tijdens studiedagen, congressen en cursussen en aan de hand van de vele artikelen, hoofdstukken en brochures die ik hierover heb geschreven.”

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

“Ik ben niet zo’n ziener, maar ik wil wel een poging wagen. Zo denk ik dat docenten, meer dan tot nu toe, in bepaalde verbanden (per vakgebied of per beroepspraktijk) zullen gaan samenwerken om itembanken te ontwikkelen. Dat is goed nieuws, want er is denk ik geen directere manier om de kwaliteit van toetsing te bevorderen. Ik hoop dat veel beroepsopleidingen deze samenwerkingsverbanden zullen aangrijpen om de kennisbasis van het betreffende beroep te operationaliseren aan de hand van een itembank, om daaruit voortgangstoetsen samen te stellen waarmee studenten hun studievoortgang kunnen bewaken en waarmee de staf het onderwijs kan evalueren. Als de items in deze banken (op termijn) IRT-geschaald kunnen worden, dan ontstaan er nog betere mogelijkheden voor het beoordelen van de voortgang van (groepen) studenten, voor de evaluatie van didactische modellen en voor geavanceerde vormen van computer-based testing, zoals bijvoorbeeld adaptieve toetsing. Verder verwacht ik dat de snelle opmars van artificiële intelligentie een katalyserend effect zal hebben op de ontwikkeling van computersimulaties voor zowel instructie- als toetsdoeleinden. Dit kan het ontwikkelen en toetsen van beroepsbekwaamheden vergemakkelijken, vooral als er sprake is van dure, gevaarlijke of voor mensen/dieren belastende processen.”

²Wie het betreffende artikel (nog eens) wil lezen: <https://docplayer.nl/13824889-Afrekenen-op-rekenen-over-de-rekenvaardigheid-van-pabo-studenten-en-de-toetsing-daarvan.html>

Waar ben je het meeste trots op?

Ik ga nog steeds glimmen van trots als ik aan de **WISCAT-pabo** denk. Dit was niet de eerste computergestuurde adaptieve toets die we ontwikkelden, maar wel de eerste die in Nederland op verplichte basis werd ingevoerd in een onderwijscontext.

En vooral ben ik trots op de manier waarop wij een positieve besluitvorming hierover hebben ondersteund met een goed getimed artikel in een vakblad, waarin op overtuigende wijze werd aangetoond dat de rekenvaardigheid van eerstejaars pabo-studenten ernstig tekortschoot². De ophef die dit artikel veroorzaakte in de media (zie figuur 1) leidde niet alleen tot invoering van de WISCAT-pabo, maar indirect

ook tot allerlei maatregelen in het voortgezet onderwijs en het middelbaar beroepsonderwijs, zoals de formulering van referentieniveaus voor rekenen en taal en de invoering van een rekentoets als verplicht onderdeel van het eindexamen in het voortgezet onderwijs.

Trots ben ik ook op het al eerder genoemde **Protocol Portfolio Scoring (PPS)**. Het werd ontwikkeld in een tijd dat veel beroepsopleidingen bezig waren hun curricula om te zetten van vakken- naar competentie-georiënteerd. De PPS-methode probeert de bewijskracht van portfolio's te verbeteren door standaardisatie (alle prestaties worden op dezelfde manier beoordeeld), eisen aan de authenticiteit (alleen prestaties waarvan zeker is dat



Figuur 1. Media ophef na publicatie artikel

Betrouwbaarheid, generaliseerbaarheid en extrapoleerbaarheid.

Betrouwbaarheid is een begrip uit de Klassieke Testtheorie dat verwijst naar de verhouding tussen de ware variantie (de variantie van de zuivere, niet door meetfouten vertekende toets-score) en de waargenomen variantie van een reeks toetsscores.

Generaliseerbaarheid is het centrale concept van de Generaliseerbaarheidstheorie en verwijst naar de nauwkeurigheid waarmee vanuit een geobserveerde assessmentscore gegeneraliseerd kan worden naar de gemiddelde score die de geteste persoon gekregen zou hebben onder alle andere assessmentcondities die relevant zijn (bijv. andere taken, andere contexten, andere assessoren, e.d.).

Extrapoleerbaarheid is één van de bewijsbronnen binnen Messick's *unitary concept of validity* en verwijst naar de vraag in hoeverre de assessmentscore ook betekenis heeft voor toekomstige prestaties in vervolgsituaties, zoals bijvoorbeeld de beroepspraktijk.

ze het resultaat zijn van inspanningen van de portfolio-eigenaar mogen worden opgenomen) en transparantie (heldere beslisregels over competentie). De essentie van PPS werd later ook gebruikt in het rapport 'Beoordelen is mensenwerk' van de Expertgroep Protocol, die in opdracht van de Vereniging Hogescholen opleidingen adviseerde over de inrichting van afstudeerprogramma's. PPS is net als programatisch toetsen erop gericht om prestaties van studenten zowel voor formatieve als summatieve doelen te gebruiken en om beslissingen over studenten op zoveel mogelijk prestaties te baseren, maar is bij het beoordelen analytisch

in plaats van holistisch georiënteerd.

Ten slotte moet ik hier mijn bijdrage noemen aan de opzet en uitvoering van de masteropleiding Toetsdeskundige van de Fontys Lerarenopleiding. Ik bewonder Desirée Joosten-Brinke om de systematische wijze waarop ze deze master heeft opgezet, wat heeft geresulteerd in een opleiding die in relatief korte tijd docenten bekwaam maakt voor alle rollen die een toetsdeskundige moet kunnen vervullen. Ik was vereerd om hieraan gedurende vijf jaren een bijdrage te mogen leveren. Ik kon heel blij worden als studenten me vertelden hoeveel ze van mijn cursus geleerd hadden en ik was trots op ze als ik hoorde dat ze het aandurfdën om de nieuwverworven kennis en vaardigheden door te geven aan collega's en te gebruiken voor de verbetering van de toetspraktijk. Wat een mooi vak!"



Toetsconstructie moet niet het sluitstuk zijn van onderwijsontwikkeling, maar juist het startpunt.

Wat zou je anders hebben gedaan?

“Ik houd niet zo van terugkijken met als doel aan te geven wat ik anders/beter had moeten doen. Natuurlijk zijn er dingen verkeerd gelopen, maar mijn ervaring is wel dat je maar zelden iets helemaal voor niets hebt gedaan. Altijd komt er wel weer een moment waarop de gedane inspanning toch weer van pas komt. Loopbanen zijn maar moeilijk te plannen, ze gaan zoals ze gaan. Maar als ik toch iets moet noemen: ik had bij het opzetten van mijn kenniskring (de groep onderzoekers die helpt om de leeropdracht van een lectoraat uit te voeren) de verleiding moeten weerstaan om de beschikbare fte's te verdelen over een relatief groot aantal onderzoekers. Deze versnippering had een negatieve invloed op de efficiëntie van het werk.”

Welke boodschap heb je voor de lezers van Examens?

“Ik heb twee boodschappen. De eerste is dat toetsconstructie niet het sluitstuk moet

zijn van onderwijsontwikkeling, maar juist het startpunt. Op die manier is de kans veel groter dat er congruentie zal zijn tussen leerdoelen, toetsopgaven en instructies en daardoor meer kans dat het betreffende onderwijsprogramma op efficiënte wijze wordt doorlopen door de onderwijsdeelnemers. De tweede boodschap is vooral gericht aan de managers van docenten. Tijdens mijn Saxionjaren, maar ook toen ik werkzaam was bij de masteropleiding Toetsdeskundige, heb ik steeds opnieuw van docenten gehoord dat zij bijzonder weinig tijd kregen voor het construeren van toetsmateriaal. Hun managers waren altijd zeer verbaasd als ik hen vertelde dat er gemiddeld genomen minstens een half uur tijd nodig is om een goede meerkeuzevraag te construeren en dat met de ontwikkeling van een betrouwbare toets dus minimaal 20 uur gemoeid is. Ja, toetsontwikkeling is best duur, maar niet zo duur als de gevolgen van verkeerde beslissingen over onderwijsdeelnemers.” ■