

Kopstukken uit de Nederlandse toetswereld

Prof. dr. ir. Bernard (B.P.) Veldkamp

Het zijn valide en betrouwbare toetsen die de juiste informatie geven over de prestaties van leerlingen.

■ 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 achtste kopstuk prof.dr.ir. Bernard (B.P.) Veldkamp. We spraken Bernard op 21 december 2023 online. Het was lastig om in de drukke periode, die december steeds weer blijkt te zijn, tot een afspraak te komen. Met wat schuiven lukte het.

Wie ben je?

“Ik ben Bernard Veldkamp, hoogleraar onderzoeksmethoden en data-analyse en Vice-Deaan Onderzoek van de faculteit Behavioral Management and Social Sciences van de Universiteit Twente (UT). Net zoals Theo Eggen (zie het interview met Theo Eggen in Examens 2023-01, red.), heb ik wiskunde gestudeerd aan de Universiteit Twente te Enschede (1996) en woon ik nog steeds in Enschede. Aansluitend aan mijn studie ben ik in Twente in 2001 gepromoveerd op het automatisch samen-



Prof. dr. ir. Bernard (B.P.) Veldkamp



Het blijft een uitdaging om de algoritmes zodanig te ontwikkelen dat ze bruikbaar zijn voor het onderwijs.

stellen van toetsen die meerdere vaardigheden meten bij Prof. Wim van der Linden, de founding father van de toenmalige faculteit Onderwijskunde binnen de universiteit Twente en de medeoprichter van de NVE (Nederlandse Vereniging voor Examens) (zie het interview met Wim van de Linden in Examens 2010-2). Bij toetsen die meerdere vaardigheden meten kun je denken aan redactiesommen waar zowel leesvaardigheid als rekenvaardigheid een rol speelt. Na mijn promotie ben ik aan de UT blijven werken.

Internationaal heb ik ervaring op kunnen doen bij verschillende exameninstellingen. In 2007 heb ik een half jaar bij Educational Testing Service in Princeton gewerkt, een grote broer van Cito. In 2011 heb ik bij de Law School Admission Council onderzoek gedaan naar het toelatingsexamen voor alle rechtenfaculteiten in Amerika. Hier heb ik met name gekeken

naar kenmerken van vragen om hun moeilijkheid vast te stellen, onder meer voor toepassing binnen de IRT (itemresponsstheorie; zie Examens 2006-2; gratis te downloaden via www.professioneelbegeleiden.nl).

In 2014 ben ik hoogleraar geworden. Dat werd een voorlopige aanstelling om te kijken of ik wel een echte hoogleraar was (met een knipoog, red.). In 2016 werd ik positief geëvalueerd waarna ik in 2018 mijn oratie heb gehouden met als titel 'Mastering the Data Mass; Er zit goud in big data, maar er zijn te weinig delvers'. In de sociale wetenschappen zijn enorm veel data beschikbaar. De uitdaging is en blijft om de beschikbare data uit die sociale wetenschappen te benutten. Bij die data gaat het dan over vele vormen van (digitale) informatie over de prestaties van leerlingen of studenten gedurende hun opleiding. Belangrijk is om te proberen om relevante informatie uit die data te halen. Eigenlijk ben je dan bezig met het bedenken en ontwikkelen van algoritmes en het optimaliseren daarvan. Dat heb ik mijn hele carrière gedaan. Tegenwoordig werk ik daarbij vooral met machine learning algoritmes. Belangrijk is dat de informatie die we naar boven halen waarde heeft voor de maatschappij.

Toelichting op wat een algoritme is (<https://nl.wikipedia.org/wiki/Algoritme>).

Een algoritme is een stappenplan bestaande uit een set regels in vaste volgorde om tot een oplossing te komen en het einddoel te bereiken. Het wordt ingezet als een handig hulpmiddel bij eenvoudige tot complexe problemen of situaties van het dagelijks leven en de digitale vraagstukken. Eigen aan het algoritme is wanneer de set regels in de verkeerde volgorde worden uitgevoerd ook niet het gewenste resultaat zal opleveren.

Bij machine learning algoritmes gaat het om de ontwikkeling van algoritmes en technieken waarmee computers kunnen leren (https://nl.wikipedia.org/wiki/Machinaal_leren).

Op het gebied van psychometrie en toetsdeskundigheid geef ik onderwijs bij verschillende opleidingen. Ik draag al jaren bij aan de master Educational Science and Technology van de UT. Van 2019-2022 heb ik les mogen geven bij de Master Toetsdeskundige in Tilburg, voor het onderdeel Technologie Enhanced Assessment. Heerlijk om mijn psychometrische kennis over te dragen aan deeltijdstudenten met de mogelijkheid dat ze in hun werk die kennis en inzichten kunnen toepassen. Momenteel geef ik les over psychometrie binnen de Research Master Methodology and Statistics for the Behavioural, Biomedical and Social Sciences van de Universiteit Utrecht in samenwerking met de UT."

Wat is je specialisme?

"Mijn specialisme ligt op het grensvlak van big data, psychometrie en data science. Een voorbeeld op het gebied van data science is het onderzoek of je het beheersingsniveau van leerlingen kunt bepalen met behulp van verschillende soorten onderwijsdata. Dit kan data zijn van (methode)toetsen, aanwezigheidsdata, ziektemeldingen, observaties van de docent, ingeleverd huiswerk, het budget van de school, et cetera. Hierbij loopt de kwaliteit van verschillende soorten onderwijsdata veel uiteen. De uitdaging is om uit de grote hoeveelheid beschikbare data relevante informatie te halen. Informatie over de kwaliteit van de data is niet altijd aanwezig en niet constant. Dat is steeds een uitdaging bij de analyses, maar dat maakt het juist interessant.

Binnen mijn vakgroep zijn we betrokken bij onderzoek om middels algoritmen optimaal gebruik te maken van de digitale data uit

onderwijstoetsing en examinering om te komen tot adaptief toetsen, tot het geven van feedback en om bij summatieve toetsen zo betrouwbaar mogelijk te meten. Ook Lambert Schuwirth (zie Examens 2023-3) is, bijvoorbeeld met dit type onderzoek bezig binnen de opleiding van medisch specialisten, maar dan vanuit een medisch praktische invalshoek.

Voor de laatste jaren is de belangstelling voor het toepassen van algoritmen in de onderwijspraktijk sterk toegenomen. De technische ontwikkeling gaat momenteel heel snel. Een van de voorbeelden is de inzet van ChatGPT. Dit kan ingezet worden als instrument om vraagformulering aan te passen aan het niveau van de student (adaptief toetsen), als instrument om examens te verbeteren maar ook, door leerlingen of studenten, als een instrument om het gebrek aan bekwaamheid te maskeren (fraude). Ik zie het als een uitdaging om technologie constructief in te zetten, grenzen te bewaken bij de inzet van technologie, leerprestaties van studenten te verbeteren, data nog breder en beter te benutten en de mogelijkheden van (digitale) techniek te benutten zonder opleidingen of studenten in gevaar te brengen.

Ik realiseer me dat het voor docenten ingewikkeld is om gebruik te maken van algoritmen. Daar willen wij ze dan ook zo goed mogelijk bij faciliteren zodat zij informatie uit data kunnen gebruiken binnen hun onderwijs zoals het zoeken naar de ware eigen prestatie van studenten (hoe de competentie te meten, wat te toetsen, welke prestaties we vragen of verwachten). Het vraagt om een nieuwe manier van denken in een oud debat. Hoe zorgen we ervoor dat we bij blijven, ook in het onderwijs? Ik zie leerlingen uit de bovenbouw van het voortgezet onderwijs bekwaam ChatGPT inzetten. Dat vereist ook het ontwikkelen van vaardigheden van de docent. Hoe vaardig ben je als docent met alle beschikbare technieken zowel digitaal als vakinhoudelijk? Hoe stel je de juiste vragen aan ChatGPT, prompt-

engineering noemen we dat. Qua specialisme ben ik samen met verschillende van mijn Aio's binnen het primair en voorgezet onderwijs bezig met onderzoek om leerprestaties en leermotivatatie te meten."

Hoe zouden we jouw brede expertise op toetsgebied mogen samenvatten?

"Ik zou zeggen: zie hiervoor; maar jullie vragen naar een samenvatting. In de breedste zin van het woord is mijn expertise; hoe kun je optimaal gebruik maken van de digitalisering van toetsing en examinering zoals bij adaptief toetsen, formatief handelen en bij summatieve toetsen. Optimaal gebruik dus van digitalisering en dan vooral vanuit algoritmes op basis van data.

Het blijft een uitdaging om de algoritmes zodanig te ontwikkelen dat ze bruikbaar zijn voor het onderwijs. Eerst ging mijn onderzoek vooral over het optimaliseren van de technologie, maar meer recentelijk doe ik veel onderzoek naar het implementeren van de algoritmes in de praktijk. Het optimaliseren van de technologie blijft een aandachtspunt door al de nieuwe ontwikkelingen; denk aan ChatGPT. Nu richten wij ons vooral op hoe we ChatGPT positief kunnen gebruiken, gericht op verbetering van de prestaties van studenten."

Op wie of waarop richt je je vooral?

"Ik focus me eigenlijk niet op 'wie', wel op 'waarop'. En dat 'waarop' zijn mijn onderzoeksterreinen zowel qua toegevoegde waarde voor de praktijk als de inhoud. Ik werk vooral met data uit het primair en voorgezet onderwijs; en ik ben gericht op het verkrijgen van inzicht in vorderingen van leerlingen en in differentiëren binnen klassen. Nu ik over 'wie' toch nadenk, richt ik mij op mijn Aio's, niet op goeroes, daar dacht ik in eerste aanleg aan.

Ik weet hoe algoritmes en machine learning algoritmes werken. Ik wil specialisten met verschillende disciplines samenbrengen en daarmee werken om tot betere informatie en

oplossingen te komen. Juist de diversiteit van onderzoekers en onderzoeken levert bruikbare informatie op. Mijn Aio's zie ik als een soort verlengstuk. Ik kan hen ondersteunen en zij komen met de creatieve ideeën en oplossingen. Dat geeft me een trots gevoel, zien wat zij presteren en betekenen voor het onderwijs en de maatschappij."

Wat versta je onder toetsing en/of examinering?

"Hier versta ik eigenlijk alles onder wat te maken heeft met het verkrijgen van betrouwbare vakinhoudelijke/competentiegerichte informatie over leerlingen in de (gesimuleerde) omgeving waarin de leerling functioneert. Het zijn valide en betrouwbare toetsen die de juiste informatie geven over de prestaties van leerlingen."

Hoe en wanneer ben je geraakt door toetsing?

"Ik ben tijdens mijn promotieonderzoek bij Wim van der Linden voor het eerst echt in aanraking gekomen met toetsing. Ik heb mij in de eerste twee jaar ingelezen in de toetswereld want het blijft toch een grote overstap van wiskunde naar toetsing. Mijn promotieonderzoek heeft mij geënthousiasmeerd voor toetsing en Wim heeft daar veel aan bijgedragen. Wim was toen al zoals gezegd founding father van de NVE, en ik ben al vanaf het begin lid van de NVE (1999). Na Hans Vos ben ik twee periodes secretaris van de NVE geweest tussen 2006 en 2012."

Wie zijn jouw leermeesters geweest?

"Ik vind leermeester een groot woord. Maar ik heb veel geleerd van mijn promotor, Wim van der Linden, met wie ik jaren heb samengewerkt, en van Theo Eggen en Piet Sanders, met wie ik het Research Center voor Examinering en Certificering (RCEC) heb opgericht. Daarnaast van Cees Glas. Cees was jarenlang voorzitter van de vakgroep waar ik werk en is echt een kopstuk op het gebied van psychometrie (zie het interview met Cees Glas in

Examens 202-4). Internationaal noem ik Ron Hamilton, waarvan ik veel geleerd heb over item response theory. In Amerika liggen voor mij veel samenwerkingsverbanden waaraan ik heb deelgenomen en waar ik ben gevormd. Binnen de big data/data science heb ik geen grote leermeesters; in ben erin gegroeid. Omgekeerd is het wel zo dat ik mijzelf als een soort leermeester voor mijn Aio's zie. Het is net zoals bij Theo Eggen, we willen onze Aio's echt begeleiden om daarmee kennis en inzichten te verbreden. Impact van onderzoek zit mede in bijvoorbeeld de Aio's die echte experts worden binnen hun vakgebied. Dus minder echte leermeesters, maar meer overal wat leren van de juiste mensen en dat weer delen, dat eigenlijk."

Wie beschouw je als vooraanstaand in de toetswereld?

"Nationaal noem ik Saskia Wools, nu voorzitter van de Raad van bestuur van Cito, Kim Schildkamp, mijn collega bij de UT die in haar werk een mooie koppeling maakt tussen AI en big data (zie het interview met Kim in Examens 2022-4) en Remco Feskens directeur van het CitoLab. Internationaal noem ik Matthias en Alina von Davier (psychometrician and researcher in computational psychometrics, machine learning, and education; Chief of Assessment at Duolingo). Veel vooraanstaande onderzoekers zijn inmiddels met pensioen daarom noem ik ze niet expliciet, zoals Wim van der Linden, Cees Glas, Theo Eggen en Cees van de Vleuten. Aanstormend talent zie ik gelukkig ook, binnen de UT, bij Cito, bij het Ministerie van OCW, bij Bureau ICE en bij Oberon. Jonge mensen die innovatief bezig zijn om toetsing aantrekkelijker, informatiever en minder belastend te maken."

Wat heeft je gedreven en wat drijft je nog steeds?

"Wat mij drijft is de rol van het onderwijs in de samenleving; elk kind gelijke kansen geven zich te ontwikkelen en tot bloei te laten komen. Een hulpmiddel daarbij is het ontwikkelen van



Wat mij drijft is de rol van het onderwijs in de samenleving; elk kind gelijke kansen geven zich te ontwikkelen en tot bloei te laten komen.

betrouwbare toetsen. Dit is een drijfveer voor mijn onderzoeken waarmee ik betrouwbare informatie op wil leveren voor de leerkracht. Op 13 december 2023 heeft in Twente de jaarlijkse IRT-bijeenkomst plaatsgevonden waarin de helft van de inbreng lag op het snijvlak van big data en IRT. Dit soort bijeenkomsten geven daar mede inhoud aan.

Vroeger was mijn drijfveer de combinatie van verschillende vakgebieden wiskunde, digitalisering en de mogelijkheden die dat bood en biedt. Maar van de techniek an sich neig ik steeds maar naar het nut maken daarvan; naar de toepassing voor de samenleving."

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

“De uitdaging voor de nabije toekomst is om met behulp van technologie de toetsdruk in het onderwijs te verminderen. Dat kan door het efficiënt gebruik van data. Bij formatief handelen is sprake van narratieve feedback. Ik zie mogelijkheden om met technologie dat narratief te ondersteunen op een meer objectieve wijze en subjectiviteit in dat narratief te verminderen. Dat betekent wel dat de breedst mogelijke vulling van de data (zoals gesprek, video, responsietijd bij digitale vraagstukken) aan de orde is. Video- of audio-opnames van gesprekken kunnen gedigitaliseerd worden en vervolgens geanalyseerd worden met algoritmes waaruit informatie wordt verkregen. Denk aan praktijkonderwijs met studievoortgang in de vaardigheden van leerlingen in de vorm van camerabeelden van praktisch werk. Of een GoPro opname vanuit de borst van kabelleggers als ze in een gegraven sleuf een glasvezelkabel leggen en vertakkingen aanbrengen (kritieke handelingen binnen een praktijkopdracht vastleggen om gerichte feedback te kunnen geven), zelfs online tijdens het werk. Maar voor een leerling blijft de interactie met de leraar altijd belangrijk, technologie gaat de leerkracht echt niet vervangen maar wel ondersteunen en helpen waardoor het werk van de leerkracht makkelijker en beter kan worden.

Hoewel ik veel met data science en algoritmes te maken heb en informatie wil verkrijgen uit big data en technologie wil inzetten voor het onderwijs en leren, ervaar ik dat ik toch binnen de universiteit uiteindelijk de docent (de mens) ben van mijn studenten en de coach/mentor van mijn Aio's en de onderzoeker die

met papers een bijdrage levert aan de kennisoverdracht naar de studenten.”

Waar ben je het meest trots op?

“Op mijn Aio's!!! En dat is echt zo. Zoveel Aio's hebben onderzoek gedaan met mooie resultaten, dat dus! Ik ben ook wel trots op wat ik zelf ontwikkeld heb en vastgelegd heb in meer dan honderd papers, om daarmee die kennis te kunnen delen. Hiermee doel in niet op de verschillende papers zelf, hoewel er hele mooie tussen zitten, maar het delen ervan als geheel met mijn Aio's, dáár ben ik trots op.”

Wat zou je anders hebben gedaan?

“Wat ik anders zou hebben gedaan weet ik niet. Ik sta zo ook niet in het leven. Met mijn verblijf in het buitenland (in Verenigde Staten en tweemaal in Italië), daar had ik misschien meer mee moeten doen, maar dat is te klein om te noemen.”

Welke boodschap heb je voor de lezers van Examens?

“Mijn boodschap aan jullie als redactie is om met Examens een bijdrage te blijven leveren aan de doorontwikkeling van het vakgebied van toetsing en dat op een laagdrempelige manier; deel kennis en kunde! En zoals jullie dat nu doen met Examens, met een afwisseling tussen theorie en praktijk. Het gaat vooral om validiteit en betrouwbaarheid. De uitdagingen voor examenorganisaties blijven groot en door alle veranderingen zijn ze ook steeds weer nieuw.

Mijn boodschap aan de lezers is dus: lees Examens want we hebben veel deskundigheid nodig om te komen tot kwaliteit in ons onderwijs en om tot valide en betrouwbare toetsen te komen die de juiste informatie geven over de prestatie van de leerlingen.” ■