

IN GESPREK MET BEN WILBRINK

# Onderzoek vanuit de studeerkamer

**Een freelance onderzoeker die zijn activiteiten op internet etaleert en die examinering en toetsing tot zijn specialiteit heeft gemaakt en via zijn website discussieert over de vorderingen van zijn wetenschappelijk onderzoek is geen alledaags verschijnsel. In het restaurant van het Jaarbeursgebouw in Utrecht hebben we een gesprek over de schaduwkanten van de testwetenschap met Ben Wilbrink.**

*We kennen u als auteur van de publicatie Toetsvragen schrijven en sinds kort ook als pleitbezorger voor voortschrijdend inzicht in de psychometrie. Wie is Ben Wilbrink? Vertel ter introductie iets over uw achtergrond.*

Ik studeerde differentiële psychologie met als specialisaties functieleer en daarnaast deed ik als bijvak onderwijsresearch bij Meuwese. In 1968 kwam ik als onderwijs-onderzoeker in dienst bij het Centrum voor Onderzoek van het Wetenschappelijk Onderwijs (COWO), het onderzoeksbureau van de Universiteit van Amsterdam (UvA).

*Was u van meet af aan betrokken bij onderzoek gebaseerd op examenresultaten?*

Dat niet direct. Mijn specialisaties waren in de eerste plaats toetsen als selectiemiddel en de relatie tussen toetsen en studiestrategieën. Daarnaast de aansluiting van onderwijs en arbeidsmarkt, en dan vooral wat daar vanuit de psychologie over valt te zeggen of te onderzoeken. Dit laatste ligt doorgaans dichterbij de werkvloer of de mensen dan wat er sociologisch aan valt te onderzoeken. De overgang van eerste- (COWO) naar derde- (SCO-Kohnstamm Instituut) geldstroomonderzoek bracht een uitbreiding van het aantal aandachtsvelden met zich mee. Ik schreef vele nota's in die jaren, zoals over het overheidsbeleid op het terrein van

technisch-wetenschappelijk onderzoek en onderwijs. Die nota's verdwenen dan weer in de ministeriële bureaula. Van 1968 tot 1999 ben ik in dienst van de UvA gebleven; het COWO is bij de eerste grote bezuinigingsoperatie in het hoger onderwijs gesneeuwd, en in 1985 opgenomen in het SCO, het latere SCO-Kohnstamm Instituut. Daar ben ik in 1999 vervroegd weggegaan. Sinds 1999 ben ik bezig met de ontwikkeling van een theorie over strategische voorbereiding op examens, naast incidenteel freelance opdrachten.

*In uw artikelen zet u zich nogal af tegen de klassieke testtheorie. Op grond waarvan eigenlijk?*

Ik zet mij niet zozeer af tegen het bouwwerk van de psychometrie, maar tegen de verheerlijking van het begrip objectiviteit, waaraan alles lijkt ondergeschikt gemaakt te moeten worden. Eind jaren zeventig hield ik eens een lezing op een onderwijscongres waarbij ik het begrip objectiviteit ter discussie stelde aan de hand van enkele voorbeeldvragen in meerkeuzevorm. Ik liet zien dat de deskundigen het niet eens waren over de alternatieven en van mening verschilden over het goede antwoord. De aanwezige directeur van het Cito verzette zich daartegen, maar misschien dat mijn presentatie het Cito op dat moment heeft geholpen om ook voor open eind-examenvragen diensten te gaan aanbieden.

*De discussie over objectiviteit was in die tijd inderdaad nogal en vogue. Maar het was tevens opvallend dat de bewijsvoering over het ontbreken van objectiviteit bij meerkeuzevragen altijd aangetoond werd aan de hand van slechte items. Het was dus meer een discussie over de kwaliteit van de vragen dan over objectiviteit.*

Mogelijk. Dat argument is ook wel omgekeerd gebruikt, vergelijking met slechte open vragen; je moet wel een eerlijke vergelijking maken. Maar terug naar mijn vermeende bezwaren. De psychometrie gaat voorbij aan de vraag hoe cursisten zich op examens voorbereiden. Kwaliteit van examens heeft daarom alles te maken met hoe doeltreffend die cursisten dat doen. A.D. de Groot had dat heel goed gezien, hij was een pleitbezorger van doorzichtigheid bij examens. Job Cohen, die bij hem promoveerde, noemt dat de kenbaarheid van de exameneisen.

## De nota's verdwenen weer in de ministeriële bureaula

*Maar we hadden het over het begrip objectiviteit en uw mening daarover.*

Wat ik ernstiger vind is dat ter wille van de objectiviteit en de eerlijkheid enorm veel tijd verknoeid wordt. Op de universiteit heb ik gemerkt dat studenten bij tentamens in de juridische faculteit lappen tekst moeten produceren waarbij het uitsluitend gaat om blijk te geven een aantal feitelijkheden begrepen te hebben. Bij de studie rechten, met grote aantallen studenten, worden nogal eens schriftelijke toetsen afgenomen die dan heel zorgvuldig door twee docenten onafhankelijk worden nagekeken. Met een serie korte antwoordvragen zou dezelfde kennis getoetst kunnen worden. De vele uren die doorgebracht worden met het nakijken en beoordelen van al deze tentamens zijn zonde van de tijd. Die had beter besteed kunnen worden aan het doceren, daar hebben studenten allereerst recht op.

*Er is daarvoor een goed alternatief: meerkeuze-toetsen die machinaal gecoörd kunnen worden.* Er is niets tegen een goed gebruik van keuzevragen. Maar ik vraag me wel eens af: waarom die toetsen? Het altijd maar willen toetsen mag wel een tikje minder vanzelfsprekend zijn. Het is mijn stellige overtuiging dat docenten de kwaliteit van hun studenten alleen al door middel van waarnemen tijdens de studie op waarde kunnen schatten. Volgens mij kan een optelsom van subjectieve oordelen heel goed concurreren met objectieve toetsing, zeker wanneer je die oordelen op een goede manier over een hele opleiding heen combineert. Er zou veel meer belangstelling moeten komen voor wat A.D. De Groot het h-gedeelte, het handelingsgedeelte, van het examen noemde. Alles wat niet objectief toetsbaar is, zou moeten plaatsvinden in het h-gedeelte. Wat in werkgroep en practicum al is gepresteerd, hoeft niet nog eens te worden getoetst. Het behoort toch tot de taak van docenten om in hun onderwijssituatie ook zonder al die toetsen tot een oordeel te komen over studenten? Dat kunnen ze prima.

*Wellicht zou dit een oplossing kunnen zijn voor een kleine onderwijssetting waarin contacten tussen studenten en docenten vrij intensief zijn. Maar hoe moet dat dan in onderwijssituaties*



Ben Wilbrink

*waarin de band tussen student en docent nauwelijks aanwezig is? Bijvoorbeeld bij particuliere opleidingen en in het afstandsonderwijs.*

Tja... daar gelden natuurlijk andere regels. Die kunnen waarschijnlijk niet zonder kwalitatief goede toetsen. Wanneer er minimale onderwijsvoorzieningen zijn en kandidaten alleen uit hun examenresultaten gekend zijn, dan komt alles aan op de kwaliteit van toetsen en examens. Maar ook daar moet niet al te veel tussentijds getoetst worden. Iets anders is dat je in die situaties vaak cursusboeken tegenkomt met toetsen aan het eind van elk hoofdstuk, bedoeld als oefenmateriaal, als een soort instrument voor zelfevaluatie. Dat is een goede zaak. Maar wat er op universiteiten gebeurt met al die tentamens, dat is werkelijk uit de hand gelopen.

*Probleem is dat de kwaliteit van toetsen en examens vaak ontbreekt bij dit soort opleidingen.* Daar moeten die opleidingen dan wel in willen investeren, bijvoorbeeld door externe deskundigen in te schakelen als eigen kennis onvoldoende is. Externe verantwoording is ook noodzakelijk. Vooral ten aanzien van de kwaliteit van de examenopgaven zelf. Want ik heb begrepen dat die bureaus van het type KCE alleen maar toezien op het correct toepassen van procedures. En waar mogelijk zou ik bij tussentijdse opleidingen ook naast de examenuitslag het advies van de cursus-

leider over de kwaliteit van de student laten meetellen. Dat systeem werkt al jarenlang goed in het basisonderwijs waar het gaat om de uitslag van de Citotoets én om het advies van de schooldirecteur. Particuliere opleidingen zijn vooral gebaat bij goede afsluitende examens. Dat wil zeggen: valide opgaven, doeltreffend voor te bereiden door de kandidaten, die er dan ook een realistische slaagkans voor moeten hebben.

Naar bestede studietijd vragen en het laten uitvoeren van betrouwbare analyses.

*Dus toch weer de psychometrie?*

Jazeker. Begrijp me goed: ik ben geen tegenstander van de psychometrie op zich. Ik ben alleen tegen het eenzijdig doorvoeren van de theoretische uitgangspunten. Ik ben ook sterk gekant tegen het ondeskundig gebruik van de psychometrie. Een verkeerd begrepen theorie kan vervelende gevolgen hebben wanneer het gaat over examens. Voor *high stakes*-toetsing, met gestandaardiseerde toetsen, is uiteraard de psychometrie een belangrijk component. Maar het hele psychometrische begrippenapparaat is bij tussentijdse toetsen en bij toetsen die door docenten zelf worden gemaakt en afgenomen, overbodig. Bovendien staat die wetenschap niet stil. Na de overwaardering van het begrip objectiviteit is in wetenschappelijke kringen de aandacht voor de validiteitsvraag veel sterker geworden. De UvA werkt aan een ontwikkeling die je als 'oorzakelijke validiteit' zou kunnen benoemen. Dat vind ik alleen al winst. Ook zal er bij examens veel meer een relatie gelegd moeten worden tussen de inzet van de student en de zak-slaag-discussie.

*Wat bedoelt u daarmee? Cesuurbepaling op basis van bestede studietijd?*

## Ik zet mij af tegen de verheerlijking van het begrip objectiviteit

Waar ik een voorstander van ben is het leggen van een relatie tussen de aan de leerstof bestede studietijd van een student of cursist en de toets. Als cursisten gemiddeld minder tijd besteden dan het programma voorschrijft, dan kan en moet de eis, dus de cesuur op toets of examen, omhoog. De reden is dat er in het onderwijs geen absolute normen zijn, alles is relatief ten opzichte van de groep waarvoor de cursus is bedoeld en de tijd die beschikbaar is. Die gemiddelde studietijd is makkelijk door middel van een vragenlijst te achterhalen.

*Zodra studenten weten waar die vragenlijst voor bedoeld is, ontstaat er op dat punt toch een vertekend beeld?*

Dat ligt eraan hoe je die vraag inkleedt en uiteraard moet je daarbij geen relatie leggen met een te verwachten toets of examen. Het zou toch te proberen zijn? Ik heb het jaren lang gevraagd in de propedeuse rechten aan de UvA, dat ging prima.

*Is die methode wel eens op die manier toegepast?*

Niet dat ik weet... Ik vrees ook dat er nauwelijks kennis is van het begrip cesuurbepaling. Niet bij docenten op universiteiten noch bij de particuliere instellingen. Toetsing is een ondergewaardeerd begrip. Er lopen vele 'ondeskundigen' rond in deze wereld. De intuïtieve manier voor het kiezen van de cesuur lijkt er wel op: docenten zullen de lat hoger leggen wanneer ze de indruk hebben dat ze te lage eisen stellen. De zwakte zit in die 'indruk': het kan zomaar gebeuren dat er veel kandidaten zakken en dat daar de verkeerde indruk uit ontstaat dat het examen te moeilijk zou zijn: dat weet je pas wanneer tegelijk bekend is dat zakkers wel degelijk keihard hebben gewerkt. In de praktijk zie je dat toetsen of examens met veel zakkers 'makkelijker' worden gemaakt, met mogelijk als resultaat nog meer zakkers. Hier ontstaat op den duur een negatieve spiraal.

*U hebt nogal een sombere kijk op het hele 'examengebeuren' in Nederland.*

Er is in het algemeen (en zeker in het hoger onderwijs) te weinig aandacht voor de kwaliteit van toetsen en examens. Men heeft niet echt scherp op het netvlies dat gebrek-

kige kwaliteit van examenvragen grote gevolgen kan hebben. Het examen is geen bijzaak. Nog daargelaten dat gebrekkige examens tot enorme kosten leiden, ook bij cursisten die meermalen voor hetzelfde examen opgaan omdat ze gewoon niet kunnen weten wat precies van hen wordt verwacht. Dat is geen kwestie van procedures, maar van inhoud. Ik zie te vaak dat er geen relatie is gelegd tussen de analyse van de leerstof en het examen zelf. Voorbeeld: ik ben betrokken geweest bij het opstellen van examens voor artsenbezoekers. Dat zijn studenten met een betrekkelijk hoge vooropleiding. Zij moeten zich nogal wat kennis eigen maken op het gebied van fysiologie en farmacologie. De opleiders van deze artsenbezoekers vertellen dan aan hun studenten dat ze de stof maar uit hun hoofd moeten leren en dat ze dan een goed examen zullen doen. Ik ben bang dat deze houding ten aanzien van de leerstof, exemplarisch is. Het heeft niets te maken met de cognitief psychologische analyse van die leerstof, maar uitsluitend met het aanleren en toetsen van kunstjes en feitelijkheden. En daar schiet het onderwijs en de maatschappij toch weinig mee op, dunkt me. Een extra probleem is dat de cognitieve wetenschap de inzichten levert die nodig zijn om examenvragen adequaat te ontwerpen, maar trainers en docenten zijn meestal geen cognitieve wetenschappers. Zij kunnen daarom ook niet goed zien wat er in dit opzicht ontbreekt aan de kwaliteit van hun examenvragen. Dat is hen natuurlijk niet aan te rekenen, maar evengoed schiet de

samenleving niet echt op met examens die feitelijk niet veel meer toetsen dan geleerde trucs en feitjes.

*Wat zijn uw plannen voor de nabije toekomst? Wat kunnen we nog van u verwachten?*

Ik kan nu mijn tijd nuttig besteden door de vele losse einden, thema's uit een rijk onderzoeksveld, te verbinden en losse einden uit te werken, en dat doe ik op mijn website. Het belangrijkste losse eind is het werken aan de theorie over strategische voorbereiding op toetsen. Een stevige basis is voor die theorie is al rond 1980 gelegd, waarop tot mijn verbazing niemand anders wenste voort te bouwen, of kritisch te reageren. Het is in 2004 gelukt het project een voor iedereen direct toegankelijke vorm te geven, inclusief instrumenten – als Java-applets – die binnen de browser zijn te gebruiken. Het is een ontwikkeling waar ik erg gelukkig mee ben. De vorderingen in dit project zijn voor de buitenwereld te volgen op [www.benwilbrink.nl/projecten](http://www.benwilbrink.nl/projecten). Zodra enige afronding mogelijk is, zal misschien die promotie er ook nog komen. Ondertussen houd ik contact met de buitenwereld door incidenteel freelanceprojecten te doen. Ik hoop dat het mogelijk is op een paar thema's een kring van geestverwanten te vormen: er moet een herziene versie van *Toetsvragen schrijven en ontwerpen* komen waarvoor ik docenten zoek die mee willen werken en ik zoek mogelijkheden om die theorie van strategische voorbereiding op examens in concrete situaties toe te passen.



**Ook de ATEX examens  
explosieveilgheid...**

**...staan onder  
toezicht van de  
Examenkamer.**

