

Norman Verhelst – teamwerk en elkaars taal begrijpen

□ Annemarie de Knecht-van Eekelen en Ed Kremers

In april 2010 vindt het gesprek met Norman Verhelst plaats op Cito. Natuurlijk gaat het over psychometrie. Maar voordat wij bijgespijkerd worden op het gebied van IRT en OPLM, krijgen wij een historische vraag voorgelegd. Of wij pater Verbiest wel kennen, dat is namelijk een beroemdheid uit de geboorteplaats van Verhelst.¹

Hoe ontwikkelde u zich tot psychometricus?

Verhelst schetst ons eerst een beeld van de situatie aan de universiteit van Leuven in de jaren zeventig waar hij als research-assistent, wetenschappelijk medewerker, zijn proefschrift bewerkte. College geven was de taak van de hoogleraar dus een assistent had niet veel anders te doen dan te werken aan het probleem van het proefschrift en het geven van practica. Het predictie-experiment, het onderwerp van Verhelst, was een moeilijk onderzoek waarbij hij weinig begeleiding kreeg, terwijl er destijds voor de uitkomst nauwelijks belangstelling was. In zijn onderzoek bestudeerde hij het gedrag van mensen die moeten voorspellen of na een groen licht een rood licht zal aangaan of weer een groen licht. Ondanks dat rood of groen licht *ad random* worden getoond,



Prof. dr. Norman D. Verhelst is geboren in 1946 te Pittem, een plaats in de Belgische provincie West-Vlaanderen. Hij studeerde experimentele en mathematische psychologie aan de Katholieke Universiteit te Leuven (1964-1968). Na het afronden van deze studie bleef hij als research-assistent verbonden aan de universiteit van Leuven waar hij zijn proefschrift bewerkte. In 1975 promoveerde hij op dit werk getiteld *Een Analyse van het Predictie-experiment: Steekproefmodellen en een Perseveratiemodel*. In de periode daarna werkte hij achtereenvolgens aan de KU te Nijmegen (1976-1977), KU Leuven (1977-1978) en RU Utrecht (1978-1985). In 1985 trad hij als wetenschappelijk medewerker bij de afdeling psychometrisch onderzoek in dienst van Cito waar hij tot op heden werkzaam is gebleven.

Vanaf 1992 was hij gedurende tien jaar bijzonder hoogleraar Psychometrie aan de Universiteit Twente bij de vakgroep 'Onderzoeksmethodologie, Meetmethoden en Data-Analyse' van de faculteit Toegepaste Onderwijskunde. Zijn leerstoel aanvaardde hij op 9 juni 1994 met het uitspreken van de rede *Zwarte dozen*. Voor zijn publicaties kan worden verwezen naar <http://catalogus.cito.nl>.

Verhelst heeft internationale bekendheid gekregen met het ontwikkelen van het One Parameter Logistic Model (OPLM). Hij is als adviseur betrokken bij een groot aantal internationale projecten op het gebied van toetsing. Hij zal eind 2010 met pensioen gaan.



Eigenlijk ben ik autodidact op het gebied van psychometrie

gaan waarnemers redenen bedenken waarom er zeker groen of zeker rood zal volgen. Het is hetzelfde gedrag dat leerlingen vertonen bij het beantwoorden van meerkeuze vragen waarvan ze verwachten dat na zoveel keer antwoord C toch echt niet nog een antwoord C zal komen.

Eenmaal gepromoveerd vertrok hij voor een jaar naar Nederland. De faculteit psychologie te Leuven had geen plek voor hem en een promovendus van deze katholieke universiteit maakte bij de andere Vlaamse universiteit met een psychologie faculteit, de niet-katholieke universiteit van Gent, geen kans op een baan. In Nijmegen gaf hij statistiek aan pedagogiestudenten. In die tijd was B.Th. Brus hoogleraar onderwijskunde. Verhelst noemt een toets die in die tijd door Brus en Voeten is ontwikkeld, de Een-Minut-Test, een toets voor technische leesvaardigheid, die nog steeds in gebruik is.

Terug naar Leuven op een tijdelijke plaats, maar toen besloot Verhelst toch om met zijn gezin naar Nederland te verhuizen. In Utrecht waren vacatures bij de faculteit psychologie ontstaan door het vertrek van Mellenbergh naar Amsterdam en van Van der Linden naar Twente. In de vakgroep Psychologie, Statistiek, Modelvorming (PSM) gaf hij weer veel college: statistiek, methodologie, psychometrie. "Een echte lesboer was ik." Ondertussen studeerde hij, "eigenlijk ben ik autodidact op het gebied van psychometrie", vooral uit de boeken van Lord & Novick² en van Gerhard Fischer³ over het Raschmodel.

Hoe en wanneer bent u geraakt door het onderwerp Examens?

Verhelst heeft in 1993 al over het modelleren van examens gepubliceerd.⁴ Hij was geïnteresseerd in het positie-effect van een item in een toets, een onderwerp waarnaar weinig onderzoek is gedaan. Mede door de introductie van computergestuurde adaptieve toetsen (CAT) is dit meer in de belangstelling gekomen, want bij CAT verandert de positie van het item voortdurend, en wat is daarvan het gevolg? Het is de vraag of een kandidaat ook tijdens de toetsafname leert, met als mogelijk gevolg dat eenzelfde opgave een andere p-waarde kan hebben, afhankelijk van de plaats in het examen. Hiernaar is meer onderzoek nodig.

Wie zijn uw leermeesters?

"Zoals gezegd Lord en Novick en van Georg Rasch (1901-1980) heb ik veel geleerd. Fischer heb ik goed gekend. En de toonaangevende Nederlandse hoogleraren in de jaren tachtig met wie zulke stimulerende discussies konden worden gevoerd. Ik noem Eddy E.C.I. Roskam (1932-1997), hoogleraar cognitieve en mathematisch psychologie in Nijmegen, Ivo W. Molenaar, hoogleraar mathematische statistiek in Groningen, Jan de Leeuw van de Vakgroep Datatheorie FSW Rijksuniversiteit Leiden, een pionier op het gebied van de data analyse.

Hij mist de gedrevenheid van de mensen van toen. Pittige discussies, waarin men het oneens kon zijn, die prikkelden tot nadenken. "Ook onder Cito-collega's in de afdeling psychometrisch onderzoek bestond die stimulerende sfeer, met sommigen is dat nog wel, maar tegenwoordig is het toch anders."

Wat drijft u?

Verhelst is gefascineerd door het vak en zou willen dat er een nieuwe doorbraak komt. "Misschien schort het moderne psychometrie aan fantasie. De modellen waarmee men werkt zijn al vijftig jaar oud, wordt het niet eens tijd voor een nieuw model? De statistische onderbouwing is beter geworden en er wordt veel tijd besteed aan de technische uitwerking van de modellen, maar niemand komt met een echt innovatieve gedachte. Men kent het IRT-model en wil dat te pas en te onpas gebruiken."

Verhelst laat de gelegenheid niet voorbijgaan om ons een klein college te geven. "Een van de grote problemen was het schatten van de parameters in het Raschmodel. De opzet van de itemrespons theorie (IRT) dateert van eind jaren vijftig. De Amerikaanse statisticus Allan Birnbaum (1923-1976) heeft er de basis voor gelegd." Volgens IRT wordt een expliciet model aangenomen voor de respons op elk afzonderlijk item. De itemkarakteristieken zijn onafhankelijk van de specifieke toets waarin de items zitten en er is de mogelijkheid om het aangenomen model expliciet te toetsen. "Veel tijd wordt besteed aan parameters schatten en valideren. Ook CAT is gebaseerd op de oude modellen."

Verhelst waarschuwt tegen het te pas en te onpas inzetten van IRT. Hij geeft als voorbeeld: "Is Haarlem

Het schort moderne psychometrie aan fantasie



moeilijker dan Amsterdam?”. Hij doelt op een toets over topografie waarvoor IRT niet geschikt is. In zo’n toets kan het beheersingsniveau veel beter worden beoordeeld op basis van een steekproef uit de te kennen plaatsen. Hetzelfde geldt voor het toetsen van de woordenschat van een leerling. Verhelst kijkt met argusogen naar sommige computertoetsen. Waarom zijn zoveel computertoetsen niets anders dan rechtopgezette papieren toetsen? Waar blijft de fantasie? “Ik zie relatief weinig computertoepassingen van toetsen die niet evengoed met pen en papier kunnen worden afgenomen, terwijl het gebruik van de computer itemformaten toelaat die niet te realiseren zijn met pen en papier. Denk alleen maar aan het registreren van de antwoorttijd per item. Daar ligt een bijzonder interessant veld van psychometrisch onderzoek te wachten op ontginning. Bovendien is het niet zo dat een computertoets equivalent is aan de papieren toets. Onderzoek van Jan Janssen (Cito) laat zien dat een rekentoets voor jonge leerlingen op de computer veel moeilijker is dan dezelfde toets op papier.”

Wat voor visie heeft u op examens? Wat is het doel?

“Ik vind toetsen en centrale examens heel belangrijk, in België is het anders georganiseerd. In het katholieke Vlaanderen worden jaarlijks de diocesane examens afgenomen, dat zijn ‘eindexamens’ die informatie over de onderwijsopbrengst geven. Helaas doet in Nederland het systeem van de ‘eigen correctie’, de docent die het werk van de eigen leerlingen nakijkt, in belangrijke mate afbreuk aan de centrale examens; die correctie is heel onprofessioneel. Het toezicht op het centraal schriftelijk en praktisch examen (cspe) laat ook te wensen over, scholen hebben veel te veel ruimte om te sjoemelen. Het is trouwens de vraag of vmbo-leerlingen in de basisberoepsgerichte leerweg (bb-leerlingen) wel zo nodig een certificaat moeten krijgen. Het model van eindexamens is niet geschikt voor die lage niveaus.”

De examens voortgezet onderwijs in Nederland kunnen op dit moment te weinig profiteren van de verworvenheden van IRT. “Eindexamens zouden daarvoor anders ingericht moeten worden. Psychometrisch is er meer te bereiken als vragen van jaar tot

OPLM – One Parameter Logistic Model

OPLM is een itemresponsmodel dat de aantrekkelijke eigenschappen van het Raschmodel combineert met de flexibiliteit van het twee-parameter logistisch model. In het OPLM worden moeilijkheidsparameters geschat en discriminatieparameters geïmputeerd als bekende constanten. Voor dit model zijn verantwoorde methoden ontwikkeld voor het formuleren en toetsen van hypothesen ten aanzien van de grootte van de discriminatie-indices. Door het versoepelen van de assumpties ten aanzien van de gelijkheid van discriminatie-indices binnen het Raschmodel vergroot het OPLM de toepasbaarheid van het Raschmodel. Door het imputeren in plaats van het schatten van de discriminatie-indices blijven binnen het OPLM (in tegenstelling tot de twee-parameter logistische modellen) de krachtige mathematische eigenschappen van de exponentiële familie van modellen behouden.

OPLM is tevens de naam van een (MS-DOS) computerprogramma voor het analyseren van dichotoom en polytoom gescoorde items binnen het raamwerk van het een-parameter logistisch model (OPLM). Omdat het model een generalisatie is van het Raschmodel kan het OPLM pakket ook gebruikt worden voor het analyseren van dichotome en polytome items binnen het kader van het Raschmodel en binnen het Partial Credit model.

Hoewel deze software niet te koop is, stelt Cito de software gratis ter beschikking onder de voorwaarde dat het gebruik beperkt blijft tot niet-commercieel, wetenschappelijk gebruik.

jaar opnieuw gebruikt kunnen worden. In het internationale PISA-onderzoek (Programme for International Student Assessment) wordt een aantal vragen als voorbeeldvragen openbaar gemaakt en de rest blijft geheim. Dat zou in examens ook kunnen. “Examens kunnen dan samengesteld worden uit een opgavenbank, waarbij de psychometrische gegevens van de opgaven bekend zijn.

Wat ziet u als het hoogtepunt in uw loopbaan tot nu toe?

Het hoogtepunt in zijn loopbaan is de ontwikkeling van het OPLM (zie kader). “Een bovenmenselijke inspanning.” Toen hij in 1985 op Cito kwam begon hij met Cees Glas te werken aan het Raschmodel dat door zijn strenge assumpties in de praktijk niet werkte. Vele, inhoudelijk uitstekende, opgaven zouden weggegooid moeten worden, omdat ze niet pasten binnen het model. Rasch gaat er vanuit dat alle items even goed discrimineren, maar in werkelijkheid is dat



Een computertoets is niet equivalent aan een papieren toets met dezelfde inhoud

niet zo. Volgens Verhelst bood het Raschmodel mogelijkheden voor bewerking en dat werd OPLM. "OPLM is mijn kind, daar ben ik het meest trots op. Binnen OPLM hoeft niet ieder item hetzelfde te discrimineren en het model is populatieafhankelijk." Verhelst heeft de theorie voor OPLM ontwikkeld en de software geschreven, toen nog in DOS. In 1988 was de eerste versie klaar. Op de computers uit die tijd draaide het programma zo langzaam dat hij zo nu en dan "iets op het scherm schreef" om de gebruiker te laten zien dat het programma nog rekende. Nog steeds krijgt hij aanvragen van over de hele wereld voor het programma dat zijn naam internationaal heeft gevestigd.

Waar bent u het meeste trots op?

Op het OPLM zoals gezegd. Maar er zijn meer hoogtepunten. Hij werkt momenteel aan de ontwikkeling van psychometrie voor *speed tests*, snelheidstesten voor bijvoorbeeld technisch lezen.⁵ Het is op zich geen nieuwe gedachte, Rasch heeft er al modellen voor gemaakt, maar nu is het door toetsafname op de computer mogelijk om responsietijden te meten. Voor taaltoetsing in het basisonderwijs heeft Cito tempotoetsen gemaakt, toetsen waarin woordjes moeten worden ingevuld. Het gaat erom hoever de leerling in een bepaalde tijd komt. In Groningen heeft dr. G.G.H. (Margo) Jansen hier ook aan gewerkt.

Hebt u leerlingen c.q. opvolgers? Heeft u school gemaakt?

Verhelst heeft de nodige promovendi begeleid, hij telt er een stuk of 14 onder wie vier Cito-collega's. De promovendi zijn altijd zeer lovend over zijn deskundigheid en betrokkenheid. Beate van der Heijden, met wie EXAMENS in 2009 heeft gesproken, is één van zijn promovendi.⁶

De uitspraak van een van zijn Cito-collega's is kenmerkend voor de wijze waarop zij Verhelst waarderen: "Van Norman is nog interessant wat hij in de prullenbak weggooit."

Wie beschouwt u als vooraanstaand in de examenwereld? Waarom?

Als het om examens gaat noemt Verhelst de Nederlandse kopstukken Wim van der Linden en Wim Hof-

stee, maar ook Marc D. Reckase die vroeger bij ACT, nu bij de Michigan State University (USA) werkt. Reckase heeft onder meer belangrijke bijdragen geleverd aan *standard setting*.⁷ Tegenwoordig werkt Reckase veel aan CAT en het scoren van *performance assessments*.

Bent u nu nog actief? En waarmee?

Eigenlijk heeft Verhelst een jaar verlof voordat hij met pensioen gaat. Maar hij blijkt minstens zo druk als anders. Als wij een afspraak voor het interview willen maken blijkt hij in Australië te zijn voor PISA.⁸ Hij is lid van de technische expert groep van PISA en trouwens ook van TIMMS (Trends in International Mathematics and Science Study).

Twee weken later moet hij naar Engeland. Tot eind 2010 is hij Director van het European Survey of Language Competences, een belangrijk initiatief van de Europese Commissie ter ondersteuning van het vreemde-talenonderwijs in Europa. Zijn interesse voor taaltoetsing stamt al van jaren her, met name in relatie tot het Europees Referentiekader Talen (ERK).⁹ Op dat gebied zal hij ook na zijn pensionering actief blijven. Hij heeft acht jaar gewerkt aan een *Manual for relating Language Examinations to the Common European Framework of Reference for Languages* (CEFR) voor de Raad van Europa.¹⁰ "Dat heeft heel veel tijd gekost, alles gratis en uiteindelijk belandt je naam in een voetnoot, zo werken ze daar!"

Een club waarmee hij liever optrekt is EALTA, de European Association for Language Testing and Assessment.¹¹ Hij is een van de *expert members*. In mei 2010 wordt het 7e jaarlijkse EALTA congres gehouden in Den Haag onder auspiciën van Cito. Verhelst zal daar een workshop geven over Classical Test Theory. Hij is een zeer gewaardeerd docent.

Welke missers heeft u gemaakt? Wat is anders uitgekapt dan bedoeld?

Wat betreft de psychometrie vindt hij achteraf dat hij "te langzaam op gang is gekomen." Het duurde een tijd voor hij zijn draai had gevonden. Pas na zijn promotie is hij in de psychometrie gedoken. "Van nature ben ik voorzichtig. Ik wil eerst 100% zeker zijn voordat ik iets opschrijf. Dan duurt het lang voordat ik tevreden ben over een publicatie. Er wordt veel te veel troep gepubliceerd. Dat is nadelig voor de per-

De correctie van de centrale examens in Nederland is heel onprofessioneel



soon zelf, voor de collega's, voor het vak. Bij één idee hoort één publicatie en niet zes, dat geeft alleen maar vervuiling van de bestanden.' Het motto 'to publish or to perish' is daar zeker debet aan.

Verhelst stelt hoge eisen aan zichzelf. In de tien jaar dat hij zowel bij Cito als de UTwente werkte heeft hij gemerkt dat je geen twee banen moet hebben. Bovendien draagt de organisatie van een universiteit niet bepaald bij aan iemands welbevinden. Veel te groot en onpersoonlijk, met veel regels en weinig oog voor individuele verschillen. Ook opereren je collega's vaak als je concurrenten, waardoor je nauwelijks nog over je ideeën durft te praten. Cito was anders, maar ook daar verandert het. "Cito was een huis bevolkt door mensen".

Wilt u een schets geven van de toekomst van examens en examinering?

"Er moet meer in psychometrie worden geïnvesteerd. Examinering is heel belangrijk, dat moet professioneel worden aangepakt. Het probleem van allerlei kleine examenbureautjes is dat zij niet aan psychometrie toekomen.

Psychometrie is geen populair vak op de universiteit. Te weinig studenten kiezen het, dat zal in de toekomst een probleem geven. Voor Cito is het van belang dat het aanbod van psychometrici op peil blijft. Voor studenten liggen er op Cito zoveel interessante onderzoeksvragen die zij in hun wetenschappelijke stage zouden kunnen bewerken. De gegevensbestanden zijn schier oneindig. Cito moet daar veel meer aan doen, zo kweek je medewerkers voor de toekomst."

Welke boodschap heeft u voor de lezers van EXAMENS?

"Je moet in een team werken. De onderwerpen worden zo ingewikkeld dat één persoon die niet meer kan beheersen. In het team moet je elkaars taal begrijpen om zinvol te kunnen communiceren. Als het gaat over taaltoetsing moet de psychometricus iets van taal weten en de talenexpert iets van psychometrie."

□ Mw. dr. A. de Knecht-van Eekelen en de heer drs. E.J.J. Kremers zijn respectievelijk hoofdredacteur en redacteur van EXAMENS. E-mail: a.van.eekelen@gmail.com.

¹Voor diegenen die het – net als wij – niet weten, volgt hier het antwoord: Pittern is de geboorteplaats van Ferdinand Verbiest (1623-1688), missionaris en sterrenkundige, stichter van het observatorium van Peking en hervormer van de Chinese kalender.

²Frederic M. Lord (1912-2000) was psychometricus bij ETS (VS). Op zijn naam staan twee belangrijke boeken *Statistical Theories of Mental Test Scores* (1968), geschreven met Melvin Novick en Allan Birnbaum, en *Applications of Item Response Theory to Practical Testing Problems* (1980). Melvin R. Novick (1932-1986) werkte bij ETS (1963-1970) voordat hij aan de University of Iowa werd benoemd als 'professor of education and statistics' (1970-1986). Van 1970-1974 was hij bovendien werkzaam voor het American College Testing Program (ACT).

³Gerhard H. Fischer (*1938), Faculteit Psychologie van de universiteit Wenen, is in de leer geweest bij G. Rasch in Kopenhagen. Hij heeft een bijzonder belangrijke bijdrage geleverd aan de uitwerking van het Raschmodel en vooral aan de verspreiding van het gebruik van dit model in Europa. In de jaren zeventig en tachtig was zijn moeilijke boek *Einführung in der Theorie psychologischer Tests* (Bern: Huber, 1974) een must voor alle Nederlandse psychometrici-in-de-dop.

⁴Verhelst, N.D. & Glas, C.A.W. (1993). A dynamic generalization of the Rasch model. *Psychometrika*, 58, 395-415.

⁵Verhelst, N.D. & Kamphuis, F.H. (2009). *A poisson-gamma model for speed tests*. Arnhem: Cito.

⁶Knecht-van Eekelen, A. de, & Kemper-Warmerdam, A. (2009). Over het meten van loopbaanpotentieel. Een gesprek met Beatrice van der Heijden. *EXAMENS* 6(2), 24-26.

⁷Mark D. Reckase is sinds 1998 hoogleraar 'Measurement and Quantitative Methods' aan het College of Education, Michigan State University. Voordien was hij Assistant Vice President for Assessment Innovations bij ACT, Inc.

⁸<http://www.pisa.oecd.org/>

⁹<http://www.erk.nl/>

¹⁰http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp#TopOfPage

¹¹<http://www.ealta.eu.org/index.htm>