

Gezien en gelezen

□ Henk van Berkel

4th Annual European Association of Test Publishers (E-ATP) in Berlijn: 12-14 september 2012. <http://www.eatpconference.eu.com/conference/schedule-at-a-glance.aspx> en www.eatp-conference.eu.com/conference/2012-presentations.aspx

Dit is een impressie van de conferentie van de Europese afdeling van de Association of Publishers (ATP). In de E-ATP heeft een aantal commerciële testaanbieders belangen; zij sponsoren ook de conferenties. Deze conferentie telde 240 deelnemers uit 25 landen. Mijn interesse ging vooral uit naar elektronisch toetsen en in het bijzonder hoe je daarbij *cheating*, dat in deze context zowel met 'verduisteren' als met 'frauderen' moet worden vertaald, kunt opsporen wanneer de toets in een thuissituatie wordt afgenomen. Maar er waren uiteraard meer onderwerpen waarvan hier een aantal wordt genoemd. Verder kan worden verwezen naar het programma en de dia's op de website van de E-ATP.

Cheating

Een aantal presentaties ging over *cheating*: preventie, mate van voorkomen, opsporing, strafmogelijkheden. *Cheating* wordt op de eerste plaats gezien als het stelen (*harvesting*) van toetsvragen, bijvoorbeeld via fotograferen, inspreken in een mobieltje, onthouden en doorvertellen. Op de tweede plaats is *cheating* het frauderen tijdens de toets: afkijken, spiekbriefjes, het raadplegen van internet en vriendjes tijdens de toets en dergelijke.

Bij preventie gaat het om voorlichting en om verschillende toetsversies. Maar ook copyright koppelen aan toetsvragen schijnt te helpen. Daarnaast zijn er trainingsprogramma's voor de surveillanten (*proctors*) tijdens de toets.

Cheating komt veelvuldig voor en wat erger is, de helft van de toetsdeelnemers schijnt er niet mee te zitten, zo blijkt uit een recent onderzoek van Caveon. Er zijn opsporingsprogramma's die verdachte antwoordpatronen onderzoeken. Interessant zijn de technische ontwikkelingen om *cheating* op te sporen bij toetsen die online worden afgenomen: het gebruik van camera's, de interactie tussen de surveillanten en de toetsdeelnemer, identificatietechnieken, de antwoordtijd en dergelijke.

Straffen zijn niet mals. De leider van een netwerk – hij vroeg aan mensen om steeds vijf vragen te onthouden en die hem mede te delen en vervolgens verkocht hij de vragen –, moest meer dan twee jaar de cel in. Op de conferentie werd hij met naam genoemd en zijn foto werd groot geprojecteerd, zonder balkje; die Amerikanen doen niet moeilijk over privacyregels wanneer het wetsovertreders betreft.

Aan verschillende testontwikkelaars heb ik gevraagd wat de waterdichte mogelijkheden zijn om fraude op te sporen wanneer er *high stakes* toetsen in een thuissituatie worden afgenomen, bijvoorbeeld via camera's, iris-, stem- of handpalmcontrole. Goede vraag, maar waterdichte garanties zijn niet te geven.

Cheating, of beter het voorkomen hiervan, brengt enorme kosten met zich mee. De kosten zitten in de preventie (formuleren van preventiebeleid, het communiceren hiervan), opsporing (fysiek en via statistische technieken), onderzoek (tijd en kosten voor de opsporing), rechtskosten (college van beroep, advocaten, rechtszaken), herstellkosten (nieuwe toetsvragen construeren). Er was een lezing die stelde dat ieder item in een gerenommeerde toelatingstest \$ 5000 kostte. Iemand die erin slaagde een aantal items van een test naar buiten te brengen en te verkopen, en die vervolgens voor het gerecht werd gedaagd, kreeg een rekening van \$ 650.000.

Itembanking

Een tweede clustering van presentaties betrof itembanken. Het is al een oud onderwerp, maar het blijft actueel omdat bij grootschalige toetsing niet aan itembanken is te ontkomen. Deelonderwerpen waren: itemgeneratie (hoe kun je makkelijk aan een groot aantal gelijkwaardige items komen – overigens werd dit ook gezien als een oplossing voor het *cheating* probleem: als je heel veel items hebt, is het niet erg wanneer er items bekend worden), modellen, bijvoorbeeld Rasch, itemconstructie door itemconstructeurs die ver van elkaar zijn verwijderd, etc.

Het programma LOFT (Linear-on-the Fly Testing) dat werd gepresenteerd, biedt een aantrekkelijke techniek om equivalente toetsen te produceren uit een voorraad gekalibreerde items. LOFT geeft de mogelijkheid toetsen te verkorten zonder verlies aan informatie. Het wezen van een itembank en vooral het vullen ervan blijft een actueel thema. Waaraan moeten de items voldoen? Passen ze in het model? De achterliggende psychometrie hiervan is moeilijk en niet voor iedereen begrijpelijk.

Itemformaten

Velen zijn op zoek naar alternatieven van meerkeuzevragen met behoud van de efficiëntie en inhoudvaliditeit. Een conclusie was dat men beter kan zoeken naar verbeteringen in meerkeuzevragenconstructie zoals screening, toetssamenstelling, cesuurbepaling en feedbackmogelijkheden. Het zal geen verbazing wekken dat deze conclusie werd getrokken door een organisatie die uitsluitend meerkeuzevragen kan verwerken. In een andere presentatie over hetzelfde onderwerp werd gepleit voor verbeteringen in de

presentatie en formuleringen van meerkeuzevragen. Niet het traditionele tekstje met vier of vijf opties en af en toe een grafiekje, maar veel meer opties, meer dan één correct alternatief, vaker grafieken, animaties, geluid, video, simulaties, langere tekstfragmenten waar de deelnemer doorheen kan scrollen, of links naar het internet. Technisch geven dit soort vragen geen problemen meer.

Demonstraties van producten

- Questify, een geavanceerd toetsprogramma (Cito)
- PjP (Predicting Job Performance Selectie), een selectieprogramma dat voldoet aan ISO standaard 10667 (Assesio)
- XamBox, combineert de voordelen van offline en online (Televic Education)
- Transom, visualiseert data (Mountain Measurement Inc.)
- Test security, de laatste ontwikkelingen op het gebied van veiligheid (Pearson)
- Observatietests met behulp van mobile devices (Questionmark)
- Individuele leerpaden (McCann Associates)
- My item writer, het construeren van items zonder internetverbinding (Prometric)
- Authorwise, een itembanksysteem (Certification Management Services)

Andere onderwerpen

De volgende opsomming geeft een indruk van de variatie aan onderwerpen en recente ontwikkelingen:

- Er was een aardige presentatie over hoe data en resultaten te visualiseren zijn, een idee voor het presenteren van bijvoorbeeld onderzoeksresultaten.
- Ontwikkelingen in de apps techniek maken het mogelijk het mobieltje te gebruiken bij het afnemen van toetsen. Over de gevolgen voor de kwaliteit van dit soort toetsen (psychometrie, inhoudsvaliditeit) hield Cito een presentatie.
- CINOP verzorgde een presentatie over de integratietoets die per telefoon wordt afgenomen en die automatisch wordt

gescoord. Wanneer dat laatste om wat voor reden dan ook niet lukt, gaat het gesproken antwoord automatisch naar vier beoordelaars.

- Wanneer men helemaal opnieuw moet beginnen met nadenken over een toetsprogramma, levert dat specifieke problemen op. Wat zijn de opties? Wat zijn de voorwaarden voor de validatie van het programma? De methode die werd voorgesteld was die met behulp van een taakgroep: een overleg met allerlei deskundigen die vanuit hun specifieke invalshoek (inhoud, toetstechnisch) een bijdrage kunnen leveren.
- Wat is eigenlijk een eerlijke test? Met behulp van het aloude onderwerp testbias werd het onderwerp via de psychometrie benaderd. Aangezien er allerlei nieuwe testvormen zijn ontstaan, is het niet uitgesloten dat die vormen onbedoelde invloeden hebben op de toetsdeelnemer.
- Pisa en PIAAC werden in een presentatie gebruikt als voorbeelden van een assessment op grote internationale schaal. De problemen die overwonnen moeten worden, liggen op het gebied van vertalingen, toetsomgevingen, itemformaten en evaluaties.
- ISO 10667 is een nieuwe standaard op het gebied van *workplace assessment*. De standaard formuleert hiervoor de eisen en criteria.
- De National Board for Certification in Occupational Therapy heeft een toetssysteem ontwikkeld waarin deelnemers worden gecertificeerd op basis van hun praktijkervaringen; met *self-assessments* kunnen ze hun professionele ontwikkeling in kaart brengen.
- Er komen hoe langer hoe meer online testen gratis beschikbaar. Dat zal de assessmentindustrie raken (net als het copieren van de CD de muziekindustrie heeft veranderd). Een presentatie gaf aan welke gevolgen dit heeft en hoe daarop kan worden ingespeeld.

De heer dr. H.J.M. van Berkel is hoofdredacteur van EXAMENS en werkzaam aan de Universiteit Maastricht. E-mail. h.vanberkel@maastrichtuniversity.nl.