

Uit de praktijk

□ Rob Rijbroek

Examineren bij het NTI

Al is het NTI MBO-College een particuliere onderwijsinstelling, het moet aan exact dezelfde eisen ten aanzien van examinering voldoen als bekostigde ROC's. Deelnemers worden immers opgeleid voor een wettelijk erkend MBO-diploma.

Een specifieke doelgroep

De doelgroep van het NTI MBO-College kiest niet voor niets voor afstandsonderwijs. De deelnemers zoeken een opleiding waarbij men zelf het tempo bepaalt, zo min mogelijk naar school moet en de stage-uren kan afstemmen op de persoonlijke situatie. Dat vraagt ook om een flexibele invulling en organisatie van de examenmomenten. Voor de examencommissie, docenten en het examenbureau is het een complex geheel en er is heel wat gezocht over KCE, de befaamde negen standaarden, taxonomiecodes en het dekken van eindtermen.

Het NTI MBO-College

Het NTI is in de Tweede Wereldoorlog begonnen als 'Nederlands Talen Instituut'. NTI groeide in de loop der jaren uit tot een toonaangevende instelling voor afstandsonderwijs die ruim 350 opleidingen aanbiedt, van een korte cursus tot een volledige Masteropleiding. In 1997 werd het NTI MBO-College opgericht dat 45 veelgevraagde MBO-opleidingen aanbiedt via afstandsonderwijs.

Eindtermen

Momenteel biedt het NTI MBO-College alle MBO-opleidingen eindtermgericht aan. In de toetsplannen is voor elke deeltkwalificatie nadrukkelijk een verdeling gemaakt tussen theoretische en praktische eindtermen. Aangezien de deelnemers weinig tot geen lesdagen

hebben en ook nog eens zo min mogelijk willen reizen voor het afleggen van examens, worden de theoretische eindtermen van een gehele deeltkwalificatie in één keer afgetoetst. De leerstof is daardoor over het algemeen omvangrijk.

De helft van de theorie-examens wordt zo'n zes keer per jaar aangeboden en kan in Leiden worden gemaakt. Het blijft overigens boeiend om te zien dat er in zo'n examenronde ruim 2000 deelnemers in anderhalve week examen komen doen en in wachtruimtes of lokalen nagelbijtend hun uiterste best doen.

Digitaal examineren

Een nieuw fenomeen is het digitaal examineren. Ongeveer de helft van de theorietoetsen is gedigitaliseerd. Dit vraagt overigens om erg veel kwaliteit en creativiteit van de toetsauteurs. Het is lastig om een deelnemer bijvoorbeeld (ook nog in beperkte tijd) het gehele menselijk lichaam op het scherm te laten tekenen in acht kleuren. Een enorm voordeel is dat de deelnemers deze examens op elke gewenste werkdag op 25 locaties in Nederland kunnen maken.

Praktijk

De praktische eindtermen worden getoetst tijdens de stage of op een van de praktijkdagen die de vakdocenten van het NTI MBO-College verzorgen. KCE – en nu de inspectie – stellen veel en hoge eisen aan de stagebeoordelaars van leerbedrijven. NTI vindt het van belang dat objectiviteit en deskundigheid geborgd worden. De stagebegeleiders zijn echter vrijwel altijd medewerkers van de instellingen waar de deelnemers stage lopen, ze hebben geen hiërarchische relatie met NTI. Van stagebegeleiders wordt verwacht dat ze deskundig zijn in het beoordelen van de examens die in de beroepspraktijk afgenomen worden, maar NTI kan ze niet verplichten om hun deskundigheid te bevorderen. Natuurlijk probeert NTI de beoordelaars te motiveren hun deskundigheid te vergroten. Daartoe zijn de nodige beoordelingsdocumenten ontwikkeld en worden trainingen voor stagebegeleiders aangeboden. Voor NTI is het echter de vraag of de inhoud van de examens en de kwaliteit van de stageplaatsen niet op gespannen voet komt te staan met de vele standaarden waaraan moet worden voldaan.

R. Rijbroek is Business Unit Manager van het NTI MBO-College.

E-mail: rob.rijbroek@nti.nl.

Computer based testing: kansen en risico's

□ Annemarie de Knecht-van Eekelen en Harry Molkenboer

De redactie van EXAMENS was 16 oktober 2008 aanwezig op de vijfde CBT conferentie georganiseerd door Lamark. Hier een impressie van de bijeenkomst en de visie van twee van de sprekers - Kathleen Scalise en Cynthia Parshall – op de mogelijkheden en beperkingen van CBT.

De conferentie

In de stijlvolle ambiance van Kasteel de Hooge Vuursche werd een tachtigtal deelnemers op een onstuimige herfstdag onderhouden over de ontwikkelingen op het gebied van *computer based testing*. Behalve vier sprekers waren er verschillende bedrijven met posterpresentaties. Interessant was de presentatie van professor Dave Bartram (SHL Group) die vooral inging op de beveiliging van persoonsassessment dat via internet wordt afgenomen, maar EXAMENS was speciaal gekomen om te spreken met Kathleen Scalise en Cynthia Parshall, twee Amerikaanse specialisten op CBT gebied. Al snel werd duidelijk dat de ene Amerikaanse de ander niet is. Scalise, oorspronkelijk opgeleid als biochemicus en met een Ph.D. in *Quantitative measurement* van de *University of California* te Berkeley, besprak in hoog tempo de veranderingen die CBT met zich meebrengt in de relatie tussen docent, toetsontwikkelaar en toetsproducenten. Parshall was juist meer geïnteresseerd in de verbetering van toetsproductie die moet leiden tot een verhoging van de kwaliteit van toetsen in het algemeen en CBT in het bijzonder. Zij is de laatste jaren werkzaam als consultant en werkte voordien als onderzoeker bij het *Department of Educational Measurement and Research* van de *University of South Florida*. In die tijd heeft zij samen met anderen het boek *Practical Considerations*

in Computer-Based Testing (Statistics for Social and Behavioral Sciences) geschreven.

Kathleen Scalise

Scalise hield haar voordracht onder de titel *Breaking rules and building success: Next generation for CBT*. De oorspronkelijke titel van haar bijdrage die te vinden is op www.CBTconference.com, is *Using Innovative Items in Computer-Based Testing: Automated Scoring and Measurement Models*. Voor EXAMENS betekende haar voordracht een verbetering, want zij ging meer in op de veranderingen in de praktijk van de toetsafname en minder op de wat onbegrijpelijke modellen die zij in vogelvlucht de revue liet passeren. Voor ons gesprek bood haar presentatie meer dan voldoende aanknopingspunten. Dat waren met name de veranderingen die CBT volgens haar teweeg brengt of zal brengen. Maar eerst gaf ze nader uitleg over een door haar gepresenteerd schema voor een indeling van items (zie figuur 1).



Kathleen Scalise



Intermediate Constraint Taxonomy for Innovative Assessment Questions & Tasks

Most Constrained → Least Constrained

Less Complex ↓ More Complex	1. Multiple Choice	2. Selection/ Identification	3. Reordering/ Rearrangement	4. Substitution/ Correction	5. Completion	6. Construction	7. Presentation/ Portfolio
	1A. True/False	2A. Multiple True/False	3A. Matching	4A. Interlinear	5A. Single Numerical Constructed	6A. Open-Ended Multiple Choice	7A. Project
	1B. Alternate Choice	2B. Yes/No with Explanation	3B. Categorizing	4B. Sore-Finger	5B. Short-Answer & Sentence Completion	6B. Figural Constructed Response	7B. Demonstration, Experiment, Performance
	1C. Conventional or Standard Multiple Choice	2C. Multiple Answer	3C. Ranking & Sequencing	4C. Limited Figural Drawing	5C. Cloze-Procedure	6C. Concept Map	7C. Discussion, Interview
	1D. Multiple Choice with New Media Distractors	2D. Complex Multiple Choice	3D. Assembling Proof	4D. Bug/Fault Correction	5D. Matrix Completion	6D. Essay & Automated Editing	D. Diagnosis, Teaching

Figuur 1. Taxonomie voor itemconstructie

De taxonomie in figuur 1 is bedoeld als een hulpmiddel bij de constructie van items voor CBT. Als itemconstructeurs het item in één van de vakken van de matrix kunnen plaatsen, kan vervolgens de psychometrische kennis die over dat type item bekend is uit *paper&pencil* toetsen, toegepast worden op het CBT item. Het schema is in zijn oorspronkelijke vorm ontworpen door Bennett; Scalise heeft kolom 4 eraan toegevoegd. Het werken met deze taxonomie maakt het karakter en de complexiteit van het item voor toetsconstructeurs en psychometrici meer inzichtelijk.

“De definitie van CBT verandert,” zegt Scalise, “we zouden moeten spreken over *Technology Based Testing*.” Het gaat niet meer om het gebruik van alleen de computer, in de nabije toekomst worden toetsen met allerlei nieuwe technologieën afgenomen. Een leuk voorbeeld geeft ze van het gebruik van de digitale pen. Leerlingen van groep 4 tekenden daarmee hun beeld van het heelal. De tekeningen konden geana-

lyseerd worden en gescoord. De scores konden weer psychometrisch worden bewerkt.

“Als je kunt kiezen uit toetsvormen zou je de eenvoudigste en dus de goedkoopste moeten kiezen,” vindt Scalise, “maar zo makkelijk ligt het niet.” Want toetsen hebben vaak meer dan één doel, toetsen hebben invloed op het leerproces, op de motivatie en de betrokkenheid van leerlingen en docenten. In de USA worden desondanks toch veel MC-toetsen gebruikt, omdat die als het meest betrouwbaar gelden. Hoewel de acceptatie onder leerkrachten laag is, zijn zij niet invloedrijk genoeg geweest om de toetscultuur te veranderen. Al lijkt daar nu toch een omslag te komen. Dat komt door CBT, “dat geeft de leerkrachten meer macht en vermindert de macht van de *test agencies*,” aldus Scalise.

“CBT geeft de mogelijkheid voor het afnemen van meer complexe items,” daar doet Scalise onderzoek



CBT vermindert de macht van de 'test agencies'

naar. Het gaat haar vooral om een psychometrisch model voor het scoren van een groep vragen - een testlet - waarbij voor het totaal aan vragen van het complexe scenario één eindscore wordt gegeven. Scalise lacht tevreden, met het introduceren van zulke complexe items is volgens haar het probleem van het stelen van items en toetsen van de baan. Want dat is de zorg van grote *test agencies*: een aantal Chinezen van wie ieder een stel vragen uit een toets onthoudt, kent binnen de kortst mogelijke tijd alle vragen uit de itembank. De *item pool security* is de grootste zorg van de huidige computer adaptieve testprogramma's. Als een set vragen zo ingewikkeld is dat het niet te onthouden is, betekent dat een enorme kostenbesparing ook al zou de constructie van zulke items veel duurder zijn dan van gewone MC-vragen.

Dat brengt ons weer terug bij de basis: de itemconstructie en daarover kan Parshall ons alles vertellen.

Cynthia Parshall

Cynthia Parshall presenteerde zes regels waaraan toets- en itemconstructie van innovatieve vraagvormen moet voldoen. Het artikel hierover dat zij samen



Cynthia Parshall

met haar collega J. Christine Harmes van de *James Madison University* heeft geschreven onder de titel *How to Design High-Quality Innovative Item Types* is eveneens op www.CBTconference.com te vinden.

Het Six step design process

Het door Parshall en Harmes beschreven proces bestaat uit de volgende stappen om te komen tot kwalitatief hoogwaardige, goed meetbare en logistiek praktische vragen:

1. Analyseer de doelen en de eisen van de toets.
2. Selecteer de juiste vraagvorm en houd rekening met de bijzonder aspecten van de vraagvorm in relatie tot wat je met de vraag wilt meten.
3. Ontwikkel één of twee prototypen van de vraagvorm die je wilt gebruiken, evalueer en bespreek intern alle denkbare aspecten van de vraagvorm, test de vraag en breng eventueel verbeteringen aan.
4. Verfijn op iteratieve wijze (drie tot vijf iteraties) de prototypen, ontwikkel een richtlijn en templates, test de vraagvormen bij kandidaten, stem de vraagvorm af met andere betrokkenen, voorzie en bespreek alle mogelijke problemen zoals bijvoorbeeld scoring en het gebruik van media.
5. Organiseer een pilottest waarbij het hele proces van ontwikkeling tot afname wordt doorlopen. Zorg ook voor een statistische analyse van de toets en de vragen; analyseer de gegevens. Breng zonodig nog verbeteringen aan in het ontwerp van de vraagvormen.
6. Maak van alle ontwerpregels, processen en documenten die tijdens het ontwikkelproces zijn ontstaan een definitieve versie. Het resultaat is een ontwerp en procedurehandboek voor het ontwikkelen, afnemen en analyseren van innovatieve vraagvormen.

Het zes stappen model biedt een raamwerk voor itemconstructie



De redactie van *EXAMENS* had dit artikel al eerder onder ogen gehad en was in eerste instantie niet onder de indruk van de zes regels voor het ontwikkelen van innovatieve vraagvormen: het *Six step design process* zoals Parshall het noemt. Innovatieve vraagvormen zijn bijvoorbeeld de matching vraag, hotspot vraag, rangschikvraag, drag&drop vraag; dus niet de standaard MC-vraag en de juist/onjuist-vraag. Overigens hoort het gebruik van media in vragen ook tot het innovatieve gebruik. De komst van Parshall naar de *CBT-Conference* was een goede gelegenheid om haar aan de tand te voelen.

Waarom heeft Parshall het *Six step design process* ontworpen? Het proces ziet er degelijk, maar ook tijdrovend en niet direct vernieuwend uit. Zijn het geen open deuren die zij beschrijft? En wie neemt de moeite om bijvoorbeeld een meer-uit-meer vraag met een videofragment in zes stappen te introduceren in de eigen organisatie? Ook die vraag stelden wij Parshall. Zij gaf aan dat in de USA voornamelijk standaardvraagvormen worden gebruikt en dat innovatieve vraagvormen vooral worden gebruikt om toetsen meer "flashy" te maken. Vragenmakers wilden zich met "flashy items" onderscheiden van anderen, terwijl een kwalitatieve, toetstechnische reden ontbrak. Het gevolg was dat slechte vragen op de markt kwamen. Bovendien toonden opdrachtgevers vaak weinig begrip voor de hoge kosten die het ontwikkelen van kwalitatieve vragen met zich meebrengt. Het *Six step design process* maakt inzichtelijk hoe complex het is om tot een goed ontwerp en dus tot goede vragen te komen. Parshall benadrukte dat het genoemde artikel het ideale proces beschrijft en dat de iteratieve stappen ook minder vaak doorlopen kunnen worden. Ook kan het model als groeimodel gebruikt worden, waardoor het zich in de loop van de tijd binnen een organisatie kan ontwikkelen. Een organisatie die met Parshall werkt volgens het *Six step design process* is Pearson VUE.

Als het *Six step design process* op de bedoelde wijze wordt gebruikt, zouden er min of meer bewezen standaarden moeten ontstaan waarvan anderen kunnen profiteren. Ze vermoedde dat dit het geval is, maar wist dat niet zeker, omdat in de USA weinig kennis wordt gedeeld, organisaties houden dat voor zich.

Eigenlijk is het vreemd dat innovatieve vraagvormen zo in de aandacht staan, het gaat immers om de keuze van de juiste vraagvorm bij een bepaalde eindterm met een bijbehorende taxonomie en een goede validiteit. En zo ingewikkeld is het toch niet om bijvoorbeeld een goede hotspot vraag te ontwikkelen? Parshall was het daarmee eens, maar juist omdat het zo vaak fout gaat vond ze het nodig om met het procesmodel te komen.

Toen wij de opmerkingen van Parshall nog eens overdachten, kwamen wij tot de conclusie dat er inderdaad genoeg voorbeelden zijn waaruit blijkt dat ook in Nederland soms te lichtvaardig over het constructieproces en de implicaties daarvan wordt gedacht. Denk aan het artikel over de scoringsregel (Vos & De Graaf, 2008). Hoe was die regel ook weer ontstaan? Een examenorganisatie kwam er achter dat de scoringsregel voor de meer-uit-meer vraag in een toetssysteem veel te streng rekende. En waarom worden in Nederland innovatieve vraagvormen nog maar beperkt gebruikt, ondanks de vele nieuwe toetsmogelijkheden van die vragen? En waardoor ontstaan er vaak problemen bij het afspelen van videofragmenten in een vraag? Door die ontbrekende Codecs, de stukjes soft- of hardware die toelaten data te coderen / decoderen of te comprimeren / decomprimeren.

Na het interview met Parshall lijken de deuren toch minder open te staan dan wij dachten. Voor haar is het duidelijk: innovatieve vraagvormen ontwikkel je niet zomaar even, daar is een zes-stappen-proces voor nodig.

Literatuur

- Bennett, R. E., Goodman, M., Hessinger, J., Kahn, H., Ligget, J., Marshall, G., et al. (1999). Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior*, 15(3-4), 283-294.
- Kubiszyn, T. & Borich, G. (2000). *Educational testing and measurement: Classroom application and practice* (6th ed.). New York: Wiley.
- Nielsen, J. (2003). *Usability 101: Introduction to usability*. Retrieved April 4, 2008 from <http://www.useit.com/alertbox/20030825.html>
- Parshall, C. G. & Becker, K. A. (2008, July). *Beyond the*



Snelle beschikbaarheid van resultaten door automatische scoring

technology: Developing innovative items. Presented at the bi-annual meeting of the International Test Commission, Manchester, UK.

Parshall, C.G., & Harnes, J.C. (2008, March). *Stages in designing innovative item types.* Presented at the annual meeting of ATP, Dallas, TX.

Parshall, C.G., Spray, J.A., Kalohn, J.C., & Davey, T. (2002). *Practical considerations in computer-based testing.*

New York: Springer-Verlag.

Scalise, K., & Gifford, B. R. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. *Journal of Teaching, Learning and Assessment*, 4(6).

Vos, H.J. & de Graaf, A-M. (2008). De quotiëntregel. *EXAMENS 4* (1), 18-21.



NATIONAAL Exameninstituut

Examens voor VCA opleidingen?

Het Nationaal Exameninstituut
....een geslaagde keuze!

Erkend Examen centrum VCA

Onder toezicht van:



Data 2, 7741 MG Coevorden - Postbus 340, 7740 AH Coevorden

Tel. 0524-563415 Fax 0524-563416

www.exameninstituut.nl - info@exameninstituut.nl

