

In: Heijnen G, Meeder S. ~~Heijnen~~ (Eds.). Toetsen en ICT
in het hoger onderwijs. Congresbundel seminar ~~in~~ Assen
~~de~~, december 1995. Utrecht: Stichting SURF, 2000.

Lit 080.

INLEIDING

De rol van ICT in studietoetsen: een verkenning

*prof.dr. C. van der Vleuten, dr. L.Schuwirth, dr.ing. A. Muijtjens en E. Driessen
Universiteit Maastricht, capaciteitsgroep Onderwijsontwikkeling en Onderwijsresearch*

Inleiding

Informatie- en communicatietechnologie (ICT) heeft onze maatschappij verstrekkend beïnvloed. Het is dan ook logisch dat ICT niet aan het onderwijs voorbij kan gaan (vgl. Tijdschrift voor Hoger Onderwijs, oktober, 1999). De vraag of ICT moet worden toegepast in het hoger onderwijs is dan ook niet aan de orde. De vraag hoe ICT in het hoger onderwijs moet worden toegepast echter des te meer, zeker gezien de snelle onderwijskundige ontwikkelingen van de laatste jaren. Het is duidelijk dat het hoger onderwijs aan een fundamentele onderwijskundige heroriëntatie bezig is. Het conventionele model van leerstofoverdracht door middel van frontale lesmethoden komt steeds meer onder druk te staan, en wordt aangevuld met vormen van onderwijs waarbij meer de nadruk ligt op het leerproces. Van een meer docentgecentreerde benadering van onderwijs gaan we

in toenemende mate over op studentgericht onderwijs. Hierbij ligt het accent op de zelfwerkzaamheid, zelfsturing en leren in een betekenisvolle authentieke context. Ook wordt nadruk gelegd op het aanpakken van complexe problemen en vaardigheden, en op het samenwerken met anderen (medestudenten). Deze beweging in het hoger onderwijs gaat echter langzaam, veel langzamer dan menigeen zou wensen (Tuninga, 1999). ICT wordt gezien als een middel om dit proces te versnellen. Strategisch gezien kan dit een geschikte aanpak zijn, maar het draagt ook het risico in zich dat ICT niet op de juiste wijze wordt ingezet of dat mogelijkheden ervan niet optimaal benut worden. Meer aandacht voor de specifiekere bijdragen die ICT al dan niet kan leveren bij de geschetste onderwijskundige ontwikkelingen is daarom gewenst.

Deze bijdrage richt zich op de vraag welke rol ICT kan spelen met betrekking tot het toetsen van studiepres-

taties. Daartoe wordt eerst teruggeblikt op wat tot op heden bereikt is, om vervolgens enkele mogelijkheden voor de toekomst te schetsen.

Een terugblik

Voor een bespreking van de rol die ICT tot nog toe bij studietoetsen heeft gespeeld kan een onderverdeling gemaakt worden naar haar rollen bij het beheer van toetsen, de analyse en scoring van toetsen en de afname van toetsen.

Elk zullen we apart bespreken. Voor elk onderdeel zullen we enkele observaties toevoegen dan wel onderzoek bespreken. Van elk toepassingsgebied zullen niet alleen de ervaren voordelen maar ook de ervaren nadelen besproken worden, en waar mogelijk zullen enkele kanttekeningen geplaatst worden. We zullen hieruit proberen enkele lessen voor de toekomst te destilleren.

Beheer van toetsen

Een van de bekendste toepassingen van ICT bij toetsen is de ondersteuning die de computer kan bieden bij het beheer van items en toetsen.

De meeste instellingen kennen specifieke, vaak in eigen beheer gemaakte software voor deze toepassing. Met name valt hierbij te denken aan een itembank. Dit is een geautomiseerd systeem van opslag van toetselementen (items). Soms wordt dit ook gecombineerd met software waarmee scores en itemanalyses kunnen worden berekend. In dat geval wordt veelal gesproken over een toetsservicesysteem.

Itembank en toetsservicesystemen kunnen in principe grote voordelen bij het beheer van toetsen hebben (Van der Linden, 1983). Er zijn een aantal, meest Amerikaanse, toetsservicesystemen commercieel verkrijgbaar (Schade heeft daarover een bijdrage geleverd op de conferentie). Hoewel de verschillende systemen een uitgebreide variëteit bieden voor wat betreft hun opties, is

het ook zo dat de opties die aangeboden worden meestal gestandaardiseerd zijn en relatief weinig mogelijkheden bieden voor aanpassing aan de lokale wensen. Dit kan problemen geven met de aansluiting op de lokale organisatie en de specifieke gebruikerswensen. Het gevolg is dan dat de gebruikerswensen aangepast moeten worden aan de mogelijkheden van de software, in plaats van andersom.

Een belangrijker aspect is echter het verschil tussen de algemene verwachtingen ten aanzien van itembanken en de praktische beperkingen ervan.

Een eerste verschil betreft de misconceptie dat als een itembank eenmaal gevuld is, hieruit met een druk op de knop eenvoudig toetsen samen te stellen zijn. Het blijkt bijvoorbeeld dat mogelijkheden tot hergebruik van items in de praktijk beperkt is. Veelal zijn items verouderd of voldoen ze niet meer aan de itemconstruktie-eisen van de organisatie. Het blijkt dat items die reeds afgenomen zijn en een tijdlang geblokkeerd zijn (om te voorkomen dat ze te snel weer in een toets zouden verschijnen) bij een zorgvuldige review maar voor ongeveer 60% zonder wijziging weer in een toets toegelaten kunnen worden. Ook is het goed te beseffen dat een goede toets meer is dan een willekeurig getrokken verzameling items. In een normaal kwaliteitszorgproces is het van belang dat een concepttoets nog een keer integraal bekeken wordt, bijvoorbeeld om te voorkomen dat gegevens van het ene item het antwoord op een ander item weggeven.

Een verder probleem bij het gebruiken van een itembank is het definiëren van de database structuur. Wanneer een docent items wil trekken die aan bepaalde kenmerken voldoen dan is het namelijk van belang dat ook alleen die items gevonden worden die aan de kenmerken voldoen en dat ook *alle* items gevonden worden die aan de kenmerken voldoen. Wanneer niet of onvoldoende aandacht aan het opzetten van de structuur besteed wordt, leidt dit ertoe dat de itembank al snel onbruikbaar wordt.

Ondanks deze nadelen zijn er grote voordelen verbonden aan een goed opgezette itembank. Op de eerste plaats kan een itembank dienen als medium voor opslag van items, toetsen en afnamegegevens, waardoor de kwaliteitscontrole van toetsmateriaal systematischer en beter verloopt. Dit kan nog verder ondersteund worden als de itembank ook gebruikt wordt als itemvolgsysteem, waardoor getraceerd wordt waar een item zich in de productiecycclus bevindt. Ook is een itembank natuurlijk een uitgelezen instrument om hergebruik van items in de hand te houden en te voorkomen dat items te vroeg hergebruikt worden.

Helaas zijn vele initiatieven die in het verleden genomen zijn om flexibele itembanksystemen te ontwikkelen toch niet succesvol geëindigd. We noemen hier een project uit de eigen instelling als voorbeeld. De Universiteit Maastricht is inmiddels bezig met haar tweede project. Het eerste project is nooit opgeleverd en het tweede project is nu lopende. Het betreft de ontwikkeling van een toetsservice systeem voor voortgangstoetsen en casustoetsen. Hoewel vorderingen worden gemaakt, is de planning inmiddels geruime tijd overschreden en zijn de kosten vele malen hoger uitgevallen dan oorspronkelijk werd ingeschat. Het toont aan dat de problematiek weerbarstiger is dan men op het eerste gezicht zou vermoeden en dat dat blijkbaar voor iedereen geldt. Al snel worden de eisen te ambitieus en is te veel tijd nodig voor realisering. Vaak blijken ook nog eens de eisen inmiddels weer veranderd te zijn als gevolg van een veranderende onderwijsomgeving of is de gebruikte technologie achterhaald door een volgende.

De les die uit het bovenstaande getrokken kan worden is dat ICT weliswaar grote mogelijkheden biedt bij het beheer van toetsen, maar dat de verwachtingen niet te hoog gespannen moeten zijn: hoge ambities worden duur betaald. Verstandiger is de ambities in te perken en waar mogelijk gebruik te maken van bestaande pakketten, of samen te werken met andere instellingen. Ontwikkeling van software in eigen beheer vergt een

toegesneden professionele organisatie, met kwalitatief en kwantitatief voldoende deskundigheid op toets- en informaticagebied met voldoende garanties voor continuïteit en permanente ontwikkeling, die beter in samenwerkingsverband dan alleen te realiseren zijn.

De scoring en analyse van toetsen

Vanuit de testpsychologie zijn statistische theorieën ontwikkeld die een belangrijke ondersteuning kunnen vormen bij de kwaliteitsbewaking van toetsen en waarmee nieuwe technieken van toetssamenstelling en toetsafname kunnen worden bereikt. Deze zogenaamde psychometrische modellen vereisen het nodige rekenwerk, en daar is de computer bij uitstek voor geschikt. Er zijn een drietal klassen van psychometrische theorieën. We zullen ze elk kort bespreken en nagaan in welke mate hun ICT-inzet in de onderwijspraktijk gerealiseerd is.

De klassieke testtheorie is de oudste van de drie. De theorie levert een aantal gegevens op die het mogelijk maken de kwaliteit van toetsitems en de toets als geheel te onderzoeken (zie de bijdrage van Van Berkel in deze bundel over psychometrische indicatoren en de bijdrage van Hofstee daarover op de conferentie). Met relatief eenvoudige statistieken worden schattingen verkregen van de moeilijkheidsgraad en het onderscheidend vermogen van de afzonderlijke items en de toets als geheel. Deze parameters kunnen gebruikt worden om de kwaliteit van de toets te verbeteren.

De klassieke testtheorie wordt in de praktijk zeer veel gebruikt. De meeste toetsverwerkingspakketten maken gebruik van dit psychometrische model. Ook standaard statistische pakketten (bijvoorbeeld SPSS of SAS) hebben toets- en itemanalyse-onderdelen in hun arsenaal. Veel docenten zijn daarmee in de loop der tijd vertrouwd geraakt.

De beschikbaarheid van (eenvoudige) software heeft wellicht in belangrijke mate bijgedragen aan de verbetering van de kwaliteit van toetsen en het veelvuldige gebruik ervan in de praktijk. Theoretisch gezien is Klassieke Testtheorie echter niet zo'n sterk model, wat aanleiding heeft gegeven voor ontwikkeling van andere modellen.

De generaliseerbaarheidstheorie is een uitbreiding op het klassieke model. Het is een krachtiger model, omdat in de schattingen van de kwaliteitsparameters meerdere factoren betrokken kunnen worden die van invloed zijn op de kwaliteit. Daarmee is het een buitengewoon handig model, vooral in de meer complexe toetsvormen waarbij er sprake is van meerdere factoren die de toetsing beïnvloeden.

In de onderwijspraktijk komen we het gebruik van de generaliseerbaarheidstheorie nauwelijks tegen. De reden daarvoor is mede gelegen in het feit dat er geen echt bruikbare software voor bestaat. Begin tachtiger jaren is een mainframe computerprogramma geschreven (GENOVA, (Brennan, 1983)), dat later zonder al te veel wijzigingen beschikbaar is gemaakt voor een PC. Het gebruik vereist behoorlijk specialistische kennis. Het is interessant te constateren dat een praktisch zinvol kwaliteitsinstrument nauwelijks doordringt omdat er geen goed ICT-hulpmiddel bestaat om het te gebruiken.

Beide voorgaande modellen worden 'zwakke' theorieën genoemd, omdat zij enkele statistische nadelen kennen. De moderne psychometrische benadering, de zogenaamde item-respons modellen, kennen deze nadelen niet. Daarmee zijn in potentie belangrijke onderwijskundige voordelen te boeken. Zo kan de moeilijkheidsgraad van een toets kan beter in de hand worden gehouden, en kunnen toetsen afgestemd worden op de vaardigheid van het individu (cf. Van der Linden, 1983).

De verspreiding en het gebruik in de praktijk van deze 'krachtige' psychometrische modellen is echter betrekkelijk gering. Hoewel specialistische software beschikbaar is, is het gebruik daarvan te ingewikkeld voor de doorsnee gebruiker. Het zou handig zijn als deze modellen onderdeel zouden zijn van bestaande toets-service-systemen, maar dat is helaas (nog) niet het geval. Het gebruik blijft dan ook beperkt tot de professionele instanties die zich met toetsing bezig houden, dan wel tot de wetenschapper met interesse op psychometrisch gebied. Ondanks de toch al lang bestaande kennis op dit gebied is er van de mogelijkheden in de praktijk nog slechts weinig gerealiseerd: de belofte is nog niet waargemaakt.

De les uit deze observaties is vergelijkbaar met de conclusie uit de voorgaande paragraaf. Blijkbaar zijn er in principe belangrijke toepassingen voor ICT in toetsing die de kwaliteit van de toetsing in belangrijke mate kunnen bevorderen, maar laat de feitelijke realisatie daarvan nog op zich wachten. Dat wat eenvoudig en (softwarematig handig) beschikbaar is, wordt gebruikt en voor het overige moeten we ambities bijstellen aan de realiteit van de praktijk. Veel beloftes moeten dus nog worden waargemaakt.

Afname van toetsen

Afname van toetsen met gebruikmaking van ICT biedt een aantal grote voordelen boven afname in mondelinge of schriftelijke vorm. Niet alleen zijn interfaces voor computermatige afname van eenvoudige toetsen vrij snel te maken en algemeen commercieel verkrijgbaar, maar door te werken met netwerken is het produceren van grote aantallen papieren kopieën niet nodig. Indien het toetsen betreft die op verschillende locaties afgenomen moeten worden is de verspreiding ook eenvoudiger. Het gebruik van multimedia bij toetsen heeft een verder voordeel; indien plaatjes en geluid geschikt ingezet worden, kan het sterk bijdragen aan de authenticiteit

van de toets. Een extra voordeel betreft tenslotte de mogelijkheid om de kandidaat opties te bieden die met een papieren toets onmogelijk zijn. Hierbij valt te denken aan de mogelijkheid om de lettergrootte te veranderen, of het programma mee te laten tellen hoeveel items ingevuld zijn en hoeveel nog niet, etc.

Natuurlijk zijn er ook keerzijden van de medaille. Open vragen die lange stukken tekst vereisen zijn (nog) niet met voldoende accuratesse computermatig te scoren, het lezen van een computerscherm is over het algemeen minder prettig dan van papier, wat weer leidt tot een snellere vermoeibaarheid van de kandidaat. Ook zijn de eisen die gesteld worden aan de beveiliging en voorkoming van fraude van een andere orde dan bij papieren toetsen.

Een zorgvuldige afweging van deze voor- en nadelen is dus belangrijk. Veel belangrijker echter is de bepaling of de mogelijkheden die bij computertoetsen bestaan ook werkelijk hun onderwijskundige meerwaarde hebben, of dat ze alleen maar beter lijken. Met andere woorden: we moeten afwegen of we willen wat we kunnen of dat we kunnen wat we willen. In een voorbeeld willen we dit aspect verduidelijken.

ICT biedt de gelegenheid tot het ontwikkelen van dynamische simulaties. Met name op dit gebied zijn belangrijke onderwijskundige en psychometrische lessen geleerd, omdat de toetsing van probleemoplossende vaardigheden in het medisch onderwijs reeds een lange geschiedenis heeft en er het nodige onderzoek naar verricht is (Swanson, Norcini, & Grosso, 1987; Swanson, Norman, & Linn, 1995; zie ook de bijdrage van Schuwirth et al. in deze bundel). Een zwaarwegend struikelblok bleek uit het gegeven, dat de gemeten competentie op de ene simulatie nauwelijks voorspellende waarde had voor de gemeten competentie op een andere simulatie. Zelfs relatief beperkte wijzigingen in de aangeboden context hebben andere resultaten tot gevolg. Competentie bleek systematisch inhoudsspecifiek. Om met een

toets bestaande uit simulaties betrouwbare of reproduceerbare conclusies te kunnen trekken, dient tengevolge hiervan het aantal simulaties (erg) groot te zijn. Er is dus een (erg) lange toetstijd noodzakelijk en de kosten van ontwikkeling worden zeer hoog. Een tweede struikelblok was dat uitkomsten op eenmaal betrouwbare metingen met (kostbare) complexe simulaties erg voorspellend bleken te zijn voor eenvoudige (minder kostbare) toetsresultaten (zoals meerkeuzetoetsen). De incrementele informatiewaarde is blijkbaar beperkt. Ook meer complexe scorings-methodieken bleken hierin geen verandering te brengen (integendeel zelfs). De ICT mogelijkheden bleken hier dus onvoldoende onderwijs- en toetskundige basis te hebben. Er moet daarom helaas geconcludeerd worden dat een hoog ambitieniveau kan leiden tot een complexiteit die niet bijdraagt aan de meerwaarde van het toetsinstrumentarium.

Algemeen

Deze laatste gevolgtrekking wijst in de richting van de algemene conclusie van deze terugblik. ICT kan wellicht de nodige voordelen hebben voor toetsing, maar het is duidelijk dat gewaakt moet worden voor te veel ambitie. Dit kan eerder contraproductief werken en tot teleurstellingen leiden. De winst van ICT zit eerder in kleine stapjes, dan in grote sprongen. Tot op heden blijken de voordelen van ICT vooral van logistieke aard te zijn of de verbetering van efficiëntie. Veel onderwijskundige mogelijkheden (beloftes) moeten nog worden waargemaakt.

Een vooruitblik

Gewapend met deze wat ontnuchterende conclusies kunnen we naar de toekomst kijken. Wat kan ICT wel betekenen voor toetsing, wat zijn relevante nieuwe onderwijsontwikkelingen en hoe moeten we daarop met toetsing en ICT op inspelen?

Het hoger onderwijs van de toekomst

Zowel nationaal als internationaal ondergaat het (hoger) onderwijs een fundamentele koerswijziging. Het conventionele frontale onderwijs maakt in toenemende mate plaats voor moderne onderwijsvormen. Daarin is meer aandacht voor het leerproces en vaardigheden. De maatschappelijke ontwikkelingen vereisen andere kwaliteiten van de afgestudeerden uit het (hoger) onderwijs. Kennis is sterk aan verandering onderhevig en een levenslang leren traject is onvermijdelijk. De bekwaamheden om kennis eigen te maken zijn even belangrijk als de kennis zelf. Bovendien worden andersoortige vaardigheden in toenemende mate belangrijk geacht, zoals bijvoorbeeld het vermogen in teams te kunnen werken of de bekwaamheid om zowel in woord als geschrift te kunnen communiceren met anderen. Tegelijkertijd is de onderwijsfilosofische basis aan het veranderen. In plaats van een objectivistische visie op kennis, die kennis beschouwt als een absoluut gegeven (wetenschappelijk vastgesteld, feiten als objectieve gegevens), dat overgedragen dient te worden aan degene die deze kennis niet heeft, wint in toenemende mate een constructivistische visie op kennis terrein (Jan Vermunt in het handboek over de onderwijskunde). Kennis is hierin geen vaststaand gegeven, maar kennis dient door de lerende actief te worden 'geconstrueerd'. Kennis is niet objectief, maar wordt bepaald door het begrip dat de lerende ervoor heeft.

De wijze waarop met kennis wordt omgegaan bepaalt ook de wijze waarop kennis zal worden gebruikt. Tenslotte verandert het onderwijs door de invloed van de toenemende kennis over leerprocessen en de effectiviteit van instructiemethoden (Van der Vleuten, 1997). Door deze ontwikkelingen verandert het onderwijs van een meer docentgerichte benadering naar een meer studentgerichte benadering. Typerend is dat het leren in een zinvolle context plaatsvindt, bijvoorbeeld door studenten herhaald te confronteren met realistische problemen uit de (professionele) praktijk en hieraan (zelfstudie)

opdrachten te verbinden. Expliciete aandacht wordt besteed aan het zelfregulerend vermogen van studenten: zelfverantwoordelijkheid, reflectie en bijsturing zijn belangrijk. Typerend is ook dat groepsgericht onderwijs en samenwerkend leren een centrale rol krijgen. Deze fundamentele verandering in het onderwijs is zich breed aan het verspreiden. De rol van toetsing zal hierin van cruciale betekenis zijn. Toetsing dicteert namelijk in belangrijke mate het leren (Frederiksen, 1984). Voor studenten vormen de toetsen de sleutel voor succes in hun onderwijsloopbaan. De vorm, de inhoud, de programmering en de regelgeving bepalen feitelijk wat en hoe studenten zullen leren. Voor studenten is de toetsing het feitelijke curriculum. Wanneer onderwijsdoelen en de doelen van het toetsingsprogramma strijdig zijn met elkaar, dan 'wint' de toetsing. Met andere woorden, de zich aftekenende verandering in het onderwijs kan alleen slagen wanneer de toetsing zich eveneens aanpast. De vaardigheden die worden benadrukt in modernere vormen van onderwijs (reflectief vermogen, metacognitie, zelfsturend leren, leren leren, presentatievaardigheden, interpersoonlijke vaardigheden, presentatievaardigheden...) zullen in een modern toetsprogramma aan bod moeten komen. Toetsing dient strategisch gebruikt te worden om gewenste effecten te bereiken. We gaan van een 'toetscultuur' naar een 'assessment cultuur' (Dochy & De Rijke, 1995). De leer-taak en de toetsing worden als het ware in elkaar geschoven: taak is toetsing. Daarmee wordt een maximale authenticiteit bereikt. De 'toetsinstrumenten' worden dus ook veel diverser: productevaluaties, presentatiebeoordelingen, werkstukbeoordelingen, dag- of logboekbeschrijvingen, zelfevaluaties, peeroordelen, observaties in authentieke (gesimuleerde) situaties, etc. De cruciale vraag die zich vervolgens opwerpt is in welke mate ICT in deze nieuwe ontwikkeling een rol kan spelen. Afgezien van het feit dat de leertaken in toenemende mate elektronisch ondersteund zullen worden, en er dus feitelijk ook een ondersteuning van de toetsing plaatsvindt (bij-

voorbeeld: werkstukken zullen elektronisch beschikbaar zijn, presentaties zullen met digitale ondersteuning plaatsvinden, onderlinge communicatie zal deels elektronisch geschieden), noemen we hieronder enkele gebieden waar de inzet van ICT in onze ogen zinvol zou zijn.

De elektronische portfolio

Portfolio-assessment is een vorm van toetsing die tegemoet komt aan een aantal van de veranderingen in het onderwijs. Portfolio-assessment komt voort uit kunsten en de architectuur. Een kunstenaar legt een portfolio aan met zijn (beste) werk en biedt dat ter beoordeling aan, aan een potentiële werkgever of koper. Deze analogie wordt doorgevoerd in portfolio-assessment. Kenmerkend voor portfolio-assessment is dat de lerende 'bewijs' moet aanvoeren voor gebleken competentie. Dit kunnen allerlei producten zijn, zoals bijvoorbeeld verslagen, presentaties, werkstukken, maar ook beoordelingen van anderen of beoordelingen van de student zelf. Alle eerdergenoemde instrumenten kunnen onderdeel van de portfolio zijn. Meestal wordt een periodieke zelfreflectie verwacht op de inhoud van de portfolio en vinden begeleidingsgesprekken plaats met een docent of met een mentor.

Enkele nadelen van een papieren portfolio kunnen uitstekend worden opgevangen met een elektronische variant (zie ook de bijdrage van Van Tartwijk e.a. in deze bundel):

- De fysieke omvang: al snel worden schriftelijke portfolio's lijvig en onhanteerbaar voor gebruik.
- Toegankelijkheid: de lezer of de beoordelaar van een elektronische portfolio is makkelijker in staat om door de portfolio te navigeren; verwijzingen ('links') naar andere relevante delen kunnen eenvoudig worden gemaakt, multimedia (bijvoorbeeld video) kunnen makkelijker worden opgenomen en doorgenomen; een structuur is eenvoudiger aan te brengen.
- Meerdere kopieën: van schriftelijke portfolio's bestaat

meestal slechts een exemplaar (meerder exemplaren geeft onmiddellijk administratieve rompslomp) en moeten 'heen en weer' worden gezonden, een (webgebaseerde) elektronische toegang is flexibel. Toegangsrechten en rechten voor aanpassing en toevoeging kunnen goed worden gereguleerd.

- Het beoordelen van de producten in een portfolio is niet eenvoudig. Een portfolio kan grotere hoeveelheden tekst bevatten en het is moeilijk te controleren of het werk dat een student in zijn portfolio heeft gestopt ook echt zijn oorspronkelijk werk is. Hoewel de ontwikkeling van ICT applicaties die hierbij kunnen helpen nog in de kinderschoenen staan zijn ze wel veelbelovend. Een eerste ontwikkeling betreft applicaties die open teksten volgens een voorgedefinieerde antwoordsleutel zover voor kunnen beoordelen dat de workload voor de examinerator drastisch verminderd wordt, omdat hij na toepassing van de applicatie nog slechts gericht enkele studenten te beoordelen heeft om voor alle studenten een beoordeling te kunnen geven (zie bijdrage van De Mulder in deze bundel). Een tweede ontwikkeling betreft applicaties die teksten met elkaar kunnen vergelijken om plagiaat op het spoor te komen.

Een elektronische portfolio biedt dus uitstekende mogelijkheden om de informatie hanteerbaar te houden en de toegang tot de informatie goed te regelen. Zo wordt een longitudinaal en inhoudelijk rijk profiel van een student opgebouwd en wordt de sterke kant van ICT goed benut met een modern middel van toetsing.

Adaptieve vormen van toetsen

ICT biedt ook mogelijkheden ten aanzien van alternatieve vormen van afnameprocedures, zoals sequentiële en adaptieve toetsing. De bedoeling van deze procedures is de toetstijd (en dus het gebruik van items) zodanig te reguleren dat met een minimale toetstijd een maximum aan zekerheid over de beheersing van de stof door de kandidaat bereikt kan worden.

Bij sequentiële toetsing wordt daartoe aan alle kandidaten een relatief korte 'screenende' toets voorgelegd. Na afloop hiervan berekent de computer de score van ieder van de kandidaten, en vergelijkt deze met de zak-slaaggrens of cesuur. Dan wordt een beslissing genomen: die studenten wier score ver genoeg van de cesuur af ligt om met voldoende zekerheid te kunnen stellen dat ze gezakt dan wel geslaagd zijn hoeven eigenlijk niet verder getoetst te worden. Die studenten echter, die te dicht bij de cesuur gescoord hebben om een zekere uitslag te geven krijgen een extra set vragen voorgelegd. De scores hierop worden opgeteld bij die van de eerste set. In feite wordt een toetsverlening gerealiseerd. In principe zou dit net zo vaak herhaald kunnen worden totdat van iedere student met voldoende zekerheid bepaald kan worden of hij gezakt of geslaagd is.

Een gecompliceerdere aanpak is de zogenaamde computer adaptive testing of adaptieve toetsing. Hierbij wordt uitgegaan van een itembank waarvan van alle items de moeilijkheidsgraad bekend is. De items worden voor iedere student individueel getrokken. Het eerste item dat een student gepresenteerd wordt heeft een middelmatige moeilijkheidsgraad. Afhankelijk van het antwoord van de student wordt nu bepaald aan welke eisen het volgende item moet voldoen. Heeft de student het eerste item goed dan wordt een iets moeilijker item getrokken, heeft hij het eerste item fout dan wordt een iets makkelijker item getrokken. Dit wordt herhaald totdat voldoende zekerheid over de competentie van de student verkregen.

Het conceptuele verschil tussen sequentiële en adaptieve toetsing is dus dat bij sequentiële toetsing de toetslengte varieert afhankelijk van het niveau van de student en bij adaptieve toetsing de toetsinhoud varieert afhankelijk van het niveau van de student. Het moge duidelijk zijn dat de statistische modellen die de itembanken bij dergelijke procedures ondersteunen niet eenvoudig zijn: bij sequentiële toetsing zou eventueel nog met klassieke test theorie of generaliseerbaarheidstheo-

rie gewerkt kunnen worden, maar bij adaptieve toetsing zijn IRT-modellen onontbeerlijk.

Feedback

Een laatste punt betreft de mogelijkheden van feedback. Een bekende reclameslogan van enige tijd geleden stelde dat een ontbijt pas voedzaam is als iemand het eet. Parallel daaraan zou als 'onderwijswet' gesteld kunnen worden dat feedback pas nuttig is als de student deze ook daadwerkelijk in zich opneemt. Bij normale toetsen is feedback veelal een standaard formulier. Dit kan variëren van het terugrapporteren van de score, de grens en de kwalificatie (of cijfer) tot zeer uitgebreide feedback over de eigen prestaties op alle onderdelen van de toets vergeleken met de prestaties van jaargroepen, e.d. Het geven van feedback wordt daardoor vaak dansen op het slappe koord: te weinig feedback levert geen effect op, maar te veel kan al gauw leiden tot een informatie-overschot waardoor het doel voorbij gestreefd wordt. ICT biedt hierin twee mogelijkheden. Een eerste en voor de hand liggende mogelijkheid is om de feedback meer in grafische en zelfs in dynamische vorm aan te bieden. Hierdoor kunnen gegevens snel en overzichtelijk gepresenteerd worden, waardoor toch veel informatie gegeven kan worden zonder dat direct het gevaar van een 'informatie-overload' bestaat. Een mogelijk interessantere aanpak is zogenaamde 'query-gebaseerde feedback'. Naar aanleiding van de toetsresultaten zal de behoefte aan informatie van student tot student verschillen. Door de toetsinformatie zo aan te bieden dat de gebruiker zelf op geleide van zijn interesse vragen (query opdrachten) kan formuleren, wordt ingegaan op de specifieke informatiebehoefte van iedere student.

Epiloog

De balans opmakend, zien we dat ICT op toetsgebied nog een hele weg te gaan heeft. Het gebruik van ICT

wordt op dit moment nog sterk gedomineerd door de introductie van toetservicesystemen. Deze ontwikkeling kost de nodige tijd. De meerwaarde van ICT ligt vooralsnog voornamelijk in de bevordering van de efficiëntie en de logistiek. De onderwijskundige meerwaarde moet in veel gevallen nog worden bereikt.

Uit de ontwikkeling van ICT producten blijkt dat ontwikkelingen vooral in kleine stappen worden gezet. Het is verstandig voortdurend waakzaam te zijn voor een te hoog ambitieniveau bij de ontwikkeling en inzet van ICT bij toetsing. Zeker bij het ontwikkelen van eigen ICT-producten is deze waarschuwing op zijn plaats. Zelf maken is duur en gevaarlijk. Een strategie gebaseerd op uitbesteding, samenwerking of deling is aanbevelenswaardig. Zorg er bovendien voor dat ICT-toepassingen eenvoudig blijven en vermijd een verlies in al te grote complexiteit (het z.g. 'KISS principle': Keep It Simple and Stupid).

Voor het bereiken van een goede onderwijskundige meerwaarde van ICT moet eerder de vraag worden gesteld 'waarom wel ICT?' in plaats van de vraag 'waarom niet ICT?'. Het vervangen van het bestaande door ICT-hulpmiddelen levert uiteindelijk niet veel op. De uitdaging voor de komende jaren is de ontwikkeling van ICT als ondersteuning van een toetsontwikkeling die het leren op een moderne manier bevordert. Enkele onderwijskundige voordelen van ICT-inzet in toetsing zijn daarbij genoemd. Maar ook daarvoor geldt dat deze ontwikkelingen niet van vandaag op morgen gerealiseerd zullen zijn. We zullen nog wel in paar valkuilen vallen voor het zo ver is. Het lijkt ons overigens dat de lessen en de eraan verbonden conclusies met betrekking tot ICT niet beperkt zijn tot toetsingsactiviteiten, maar even geldig zijn voor overige ICT-ontwikkelingen op onderwijsgebied.

Literatuur

Brennan, R. L. (1983). *Elements of Generalizability Theory*. Iowa: ACT Publications.

Dochy, F. J. R. C., & De Rijke, T. R. (Eds.) (1995). *Assessment centers: nieuwe toepassingen in opleiding, onderwijs en HRM*. Utrecht: Lemma.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39(3), 193-202.

Linden, W. J. van der (1983). *Van standaardtest naar itembank*. Enschede: Inaugurale rede Universiteit Twente.

Swanson, D. B., Norcini, J. J., & Grosso, L. J. (1987). Assessment of clinical competence: written and computer-based simulations. *Assessment and Evaluation in Higher Education*, 12(3), 220 - 246.

Swanson, D. B., Norman, G. R., & Linn, R. L. (1995). Performance-based assessment: Lessons from the health professions. *Educational Researcher*, 24(5), 5-11,35.

Tuninga, R. S. J. (1999). Internationalisering, netwerken, informatie- en communicatietechnologie en het hoger onderwijs. *Tijdschrift voor Hoger Onderwijs*, 17(3), 173-183.

Vleuten, C. P. M. van der (1997). De intuïtie voorbij. *Tijdschrift voor Hoger Onderwijs*, 15(1), 34-46.