2011
03
jaargang 8

EXAMENS

TIJDSCHRIFT VOOR DE TOETSPRAKTIJK

Een model voor programmatische toetsing

- Meten is weten; vergeet het maar!
- De wet, het rekenen en de rekentoets
- Effectiviteit van een examenstelsel
- Werken met itembanken
- Kopstukken uit de examenwereld: Cees van der Vleuten

reach out to your candidates

Lamark biedt **Flexibele Examinering** aan certificerende instanties en hun kandidaten. **Flexibele Examinering** biedt examenkandidaten de mogelijkheid een digitaal examen te doen op een datum en tijdstip naar keuze in een Test Centre bij hen in de buurt. **Flexibele Examinering** biedt certificerende instanties de mogelijkheid haar dienstverlening optimaal af te stemmen op de behoefte van haar kandidaten.

lamark

P.O. Box 2 - 6800 AA Arnhem
Phone +31 (0)88 51 41 777

beflexible@lamark.com
www.lamark.com

ISO9001: 2008 certified

Inhoud

Een model voor programmatische toetsing 5

Cees van der Vleuten en Lambert Schuwirth

Dit artikel gaat over een onderwijscontext waarin toetsing ondersteunend is aan het onderwijs. In dit onderwijs heeft zelfsturing een belangrijke rol en de onderwijskundige ambities gaan het klassieke beheersingsniveau te boven. Een toetsprogramma in deze context moet het leren maximaal bevorderen en tegelijkertijd informatie opleveren om robuuste beslissingen over studievoortgang te kunnen nemen. Aan de Universiteit Maastricht is een model voor de inrichting van een dergelijk toetsprogramma ontworpen. De auteurs hopen dat ervaringen in het hbo en wo met deze wijze van toetsing worden opgeschreven en gedeeld, zodat ze kunnen leiden tot vernieuwingen in het model.

Meten is weten; vergeet het maar! 10

Henk van Berkel

Je maakt een aantal examenvragen. Je zorgt er voor dat de vragen een goede spreiding hebben over de eindtermen en dan heb je een examen. Dan neem je het af, je telt het aantal punten dat de deelnemers hebben behaald, eventueel na correctie voor raden indien het examen bestond uit meerkeuzevragen. Dat aantal punten zet je om in een cijfer en klaar is Kees.
Ziedaar in een notendop het examenproces. Moet daar nu zoveel heisa over worden gemaakt? Je weet toch hoeveel examenvragen deelnemers correct hebben gemaakt? Dat heb je met het examen gemeten. Dus je weet ook hoeveel kennis ze hebben. Daar is toch geen stok tussen te krijgen?

De wet, het rekenen en de rekentoets 18

Ben Wilbrink en Joost Hulshof

Minister Van Bijsterveldt heeft het voornemen om in 2014 voor havo/vwo een rekentoets een nieuw onderdeel van het eindexamen te laten zijn en onderdeel van de kernvakkenregel (voldoende voor rekentoets, Nederlands, Engels en wiskunde). Er doemen zo ernstige problemen op rond deze rekentoets. Dit artikel betoogt dat de voorbeeld-rekentoetsen die door de commissie-Schmidt zijn gepubliceerd weinig van doen hebben met het rekenen dat vrijwel iedereen voor ogen staat, dat het hoger onderwijs verlangt, en dat bewindslieden en Tweede Kamer voorstaan.

Effectiviteit van een examenstelsel 24

Agnes Dinkelman

De examenbranche heeft te maken met toenemende druk die leidt tot intimidatie door kandidaten en fraude. Voor de sector is het belangrijk om de interne situatie tegen het licht te houden. Waar zitten de mogelijkheden om de tendens te keren?

Werken met itembanken 26

Geert van der Wijk

De toenemende belangstelling voor de kwaliteit van examens en toetsen legitimeren de extra aandacht voor het werken met itembanken. Binnen Hogeschool Windesheim wordt sinds 2009 vanuit een hogeschoolbreed project ingezet op de structurele opzet, uitbouw en het onderhoud van itembanken met behulp van een toetsservicesysteem.

Kopstukken uit de Examenwereld: Cees van der Vleuten 29

Desirée Joosten-ten Brinke en Ad de Jongh

Prof.dr. C.P.M. van der Vleuten (1956) studeerde psychologie in Tilburg, in het bijzonder persoonlijkheidsleer en psychodiagnostiek. Hij promoveerde in 1989 aan de Universiteit Maastricht op het proefschrift *Naar een Rationeel Systeem voor Evaluatie van Studie-prestaties in Probleem-gestuurd Medisch Onderwijs: Studies naar Betrouwbaarheid en Validiteit van Vaardigheidstoetsen*. In datzelfde jaar werd hij bij de vakgroep universitair hoofddocent en in 1996 hoogleraar. In zijn loopbaan heeft hij een omvangrijk oeuvre opgebouwd. Zijn publicatielijst telt honderden titels. Hij heeft vele promovendi begeleid bij het schrijven van een dissertatie. Hij adviseert zowel binnenlandse als buitenlandse organisaties. Vanaf 2005 is hij ook Wetenschappelijk directeur van de UM Graduate School of Health Professions Education.

Rubrieken

Redactioneel 4
Terecht of niet …? 15
Gastcolumn 17
Uit de praktijk 23
Gezien en gelezen 33
Literatuur 34
Agenda 34
NVE 35

EXAMENS
Tijdschrift voor de toetspraktijk

Redactie
Dr. H.J.M. van Berkel, hoofdredacteur
Mw. dr. A. de Knecht-Van Eekelen, hoofdredacteur
Mw. drs. J. ter Burg
Mw. dr. A.H.H.M. Kempers-Warmerdam
Ing. H.F.A.M. Molkenboer
Drs. E.J.J. Kremers
Drs. A.J.C.M. de Jongh
Mw. dr. D. Joosten-ten Brinke
Mw. K.J. Gerritsen-van Leeuwenkamp

Redactieraad
Dr. C. Sluijter
Dhr. H. Gundlach, oud-directeur Raad voor Accreditatie
Prof. dr. C. van der Vleuten, hoogleraar Onderwijskunde Universiteit Maastricht
Dr. O. McDaniel, directeur CBE Consultants
Dhr. H. van der Linden, algemeen secretaris Groep Educatieve Uitgeverijen
Drs. A. Helderman, Inspecteur BVE Onderwijsinspectie
Dhr. S. Jurna, beleidsmedewerker Ministerie OCW
Drs. R. Duvekot, Hogeschool InHolland
Prof. dr. P. Sanders, Cito
Drs. A.J.M. Luijten
Mw. Drs. P. Willemse

Uitgever en advertenties
Kloosterhof acquisitie services
Eric Vullers
Napoleonsweg 128a
6086 AJ Neer
Tel: 0475 597 151 Fax: 0475 597 153
E-mail: info@kloosterhof.nl
Internet: www.kloosterhof.nl

Inzending van de kopij
Alle kopij dient per e-mail gezonden te worden naar: a.van.eekelen@gmail.com

Abonnementen
De abonnementsprijs bedraagt voor particulieren €47,50 per jaar en voor instellingen €58,50 per jaar. Losse nummers zijn alleen voor particulieren te koop voor €14,25. Nieuwe abonnementen kunnen schriftelijk worden opgegeven aan de uitgever. Beëindiging van het abonnementkan uitsluitend schriftelijk voor 1 september van het lopende abonnementsjaar. Bij niet tijdige opzegging wordt het abonnement automatisch met een jaar verlengd.

EXAMENS; Tijdschrift voor de toetspraktijk verschijnt vier keer per jaar.

Vormgeving en realisatie
Kloosterhof vormgeving
Marie-José Verstappen

© Kloosterhof acquisitie services 2008
Door het inzenden van de kopij verleent de auteur toestemming tot publicatie in dit tijdschrift en daarmee tevens tot het in enigerlei vorm elektronisch beschikbaar stellen.Het verlenen van toestemming tot publicatie in dit tijdschrift houdt in dat de auteur aan de uitgever onvoorwaardelijk de aanspraak overdraagt op de door derden verschuldigde vergoeding voor kopiëren, als bedoeld in Artikel 17, lid 2 der Auteurswet 1912 en in het KB van 20-71974 (stb. 351) en artikel 16b der Auteurswet 1912, teneinde deze te doen exploiteren door en overeenkomstig de Reglementen van de Stichting Reprorecht te Hoofddorp, een en ander behoudens uitdrukkelijk voorbehoud van de kant van de auteur.

ISSN 1567-2638

Annemarie de Knecht-van Eekelen

Redactioneel

Cijfers en motivatie?

Bij mijn weten is het nog niet eerder voorgekomen dat toetsing op alle niveaus van het onderwijs zo voortdurend in het nieuws is. Een van de laatste berichten van voor de zomervakantie was de mededeling dat de ministerraad heeft ingestemd met een wetswijziging waardoor een verplichte eindtoets voor taal en rekenen/wiskunde voor leerlingen in groep 8 van de basisschool wordt ingevoerd in 2013. Bovendien regelt het wetsvoorstel dat alle scholen in het primair onderwijs met een leerling- en onderwijsvolgsysteem (lovs) moeten gaan werken. De centrale eindtoets wordt niet meer, zoals de 'Citotoets', afgenomen in begin februari, maar pas eind april. De toets lijkt meer een afrekening voor de school dan voor de leerling, want het schooladvies moet bij de overgang naar het voortgezet onderwijs de doorslag geven en niet de resultaten op deze toets.
Als de leerlingen vervolgens hopelijk op de juiste plaats in het voortgezet onderwijs belanden, krijgen ze ook daar te maken met de referentieniveaus taal en rekenen. Er komen verplichte tussentijdse toetsen aan het einde van de onderbouw voor de vakken Nederlands, Engels en rekenen/wiskunde. De functie van deze toetsen is diagnostisch, formatief. En ook in het vo wordt het gebruik van een leerlingvolgsysteem verplicht.
Inmiddels is het onzalige idee om spelling en grammatica in de eindexamens integraal mee te beoordelen verlaten. Een bijna 80 pagina's dik rapport van Regioplan Beleidsonderzoek vertelt wat wij allang wisten: het is organisatorisch en toetstechnisch 'niet of nauwelijks' uitvoerbaar. Hoe het met de even problematische rekentoets, waarover Wilbrink en Hulshof in dit nummer schrijven, zal aflopen lijkt minder hoopvol. Die komt er vast wel, maar misschien lukt het om de inhoud aan te passen.
Aangenomen dat de leerlingen slagen voor het eindexamen met voldoendes voor de kernvakken dan komen ze een nieuwe barrière tegen. Zij die naar het hbo of wo willen, blijken te worden geselecteerd op hun cijfers. Ook dit bericht verscheen kort voor het zomerreces: eindexamencijfers laten meetellen bij de selectie van nieuwe studenten. Maar cijfers mogen niet het enige criterium zijn; ook een motivatiegesprek moet meewegen.
De discussie over examencijfers is niet nieuw, belangrijk is de vraag op welke cijfers geselecteerd gaat worden: op het gemiddelde cijfer van schoolonderzoek en centraal examen of alleen op cijfers van het centrale examen? De grote verschillen tussen de cijfers van school- en centraal examen die door de Inspectie van het Onderwijs zijn gerapporteerd, geven te denken.
De voorzitter van het Interstedelijk Studenten Overleg (ISO) deed al vast een duit in het selectiezakje: er zouden door deze selectie op cijfers veel studenten naar het hbo in plaats van naar het wo gaan. De universiteit zou elitair worden. Wat is daar eigenlijk op tegen? Maar er is nog het motivatiegesprek dat redding kan bieden. Stel dat een leerling goede cijfers heeft en dan op motivatiegesprek moet komen. Of hoeft hij dat niet vanwege die cijfers. Moeten alleen de zesjes laten zien dat ze toch echt gemotiveerd zijn? Nu hebben luie zesjes vaak een vlotte babbel en een motivatiegesprek is zo subjectief als het maar kan. Ik geef ze een goede kans om binnen te komen. Maar pas op, onze voortvarende staatssecretaris van OCW Zijlstra wil ook minder hertentamens in het hoger onderwijs en de aanwezigheid bij colleges en deelname aan tentamens moeten verplicht worden. Dan moeten die zesjes toch nog gaan studeren.

Mw.dr. A. de Knecht-van Eekelen is hoofdredacteur van *EXAMENS*. E-mail: a.van.eekelen@gmail.com.

EEN TOETSPROGRAMMA VAN DE UNIVERSITEIT MAASTRICHT

Een model voor programmatische toetsing

Cees van der Vleuten en Lambert Schuwirth

Dit artikel gaat over een onderwijscontext waarin toetsing ondersteunend is aan het onderwijs. In dit onderwijs heeft zelfsturing een belangrijke rol en de onderwijskundige ambities gaan het klassieke beheersingsleren te boven. Een toetsprogramma in deze context moet het leren maximaal bevorderen en tegelijkertijd informatie opleveren om robuuste beslissingen over studievoortgang te kunnen nemen. Aan de Universiteit Maastricht is een model voor de inrichting van een dergelijk toetsprogramma ontworpen. De auteurs hopen dat ervaringen in het hbo en wo met deze wijze van toetsing worden opgeschreven en gedeeld, zodat ze kunnen leiden tot vernieuwingen in het model.

Van toets naar toetsprogramma

Een toets moet voldoen aan allerlei kwaliteiten. Betrouwbaarheid en validiteit zijn de bekendste. Invloed op leren, acceptabiliteit en kosten van toetsing zijn daaraan toe te voegen (van der Vleuten, 1996), maar er zijn meer kwaliteiten te onderscheiden. Van belang is dat elke toetssituatie een compromis zal zijn tussen deze kwaliteiten. Er bestaat geen oneindig budget, een perfecte betrouwbaarheid valt niet te bereiken in een gegeven toetsvorm, et cetera. Het concrete compromis zal erg afhangen van de specifieke context. In een certificerende toets waarmee een professionele bevoegdheid wordt verkregen, zal een ander compromis worden gesloten dan in een tussentijdse toets op een bepaald moment in een onderwijstraject. Compromissen worden ook anders gesloten als toetsen een plaats en een functie hebben binnen een geheel van een onderwijsprogramma (van der Vleuten & Schuwirth, 2005). Niet de enkele toets moet maximaal voldoen aan alle toetskwaliteiten, maar het toetsprogramma als geheel. Over een goed toetsprogramma is nagedacht als een arrangement en er wordt regie over gevoerd: de som is meer dan het geheel der delen. Net zoals een goed onderwijsprogramma met verstand in elkaar wordt gestoken, zo is het verstandig om een toetsprogramma te ontwikkelen vanuit een plan. Vergelijkbaar aan een onderwijsprogramma wordt er voortdurend geëvalueerd en bijgesteld in het toetsprogramma. Vergelijkbaar aan een onderwijsprogramma is het na een aantal jaren verouderd of versleten. Een complete revisie zal dan nodig zijn.

Het is interessant om te zien dat in de literatuur bijna uitsluitend aandacht wordt besteed aan de kwaliteitskenmerken van afzonderlijke toetsinstrumenten, terwijl er weinig tot geen aandacht bestaat voor programma's van toetsen. In het proefschrift van Baartman zijn de eerste verkenningen gedaan met betrekking tot de kwaliteiten van toetsprogramma's (Baartman, Bastiaens & Kirschner, 2005). Meer recent is dat gebeurd voor het ontwerpen van toetsprogramma's (Dijkstra, van der Vleuten & Schuwirth, 2009), om uiteindelijk te komen tot een set richtlijnen voor ontwerp. In dit artikel wordt een model beschreven dat op een generieke manier aangeeft hoe een toetsprogramma idealiter in elkaar zou moeten zitten om het leren te bevorderen en tegelijkertijd rijke informatie op te leveren. Cumulatie van die informatie levert vervolgens een solide basis voor het nemen van selectie- of bevorderingsbeslissingen.

Uitgangspunten

Er zijn twee belangrijke uitgangspunten voor een

Het model is bewust generiek gehouden

programmatische benadering van toetsen:
1° naarmate een beslissing belangrijkere gevolgen heeft, moet ze op basis van meer informatie genomen worden;
2° een toets moet maar zeer beperkt een compromis sluiten op informatie- of leerwaarde van de toetsing.
Eén enkele toets is in dit denken slechts één datapunt. Een enkel datapunt zal altijd zijn beperkingen hebben, flinke compromissen moeten worden gesloten ten aanzien van de kwaliteitskenmerken. Taakvariabiliteit vormt een van de grootste bronnen van ruis in vrijwel elke toets. Brede steekproeftrekking uit het taakdomein en lange toetsen zijn nodig voor het bereiken van minimale betrouwbaarheid, veel langer dan wat haalbaar is in de meeste opleidingen.
In veel opleidingen is het volstrekt normaal dat toetsvragen geheim worden gehouden na toetsafname. Dat voorkomt dat telkens nieuwe vragen moeten worden gemaakt. Er wordt blijkbaar een gigantisch compromis gesloten op de leerwaarde van de toets: de enige feedback van zo'n toets is of je geslaagd bent of niet, misschien gekoppeld aan een cijfer (een van de zwakste vormen van feedback).

De discussie over summatief/formatief toetsen wordt bij programmatische toetsing vervangen door te spreken over wat voor gevolg er op het spel staat (*stakes* in het Engels). Wat er op het spel staat is een continuüm, geen dichotoom gegeven zoals in de formatief/summatief discussie. Bij een enkel datapunt staat er weinig op het spel. Dramatische beslissingen nemen op basis van een enkel datapunt is onverstandig. Een enkel datapunt staat volledig in het teken van het stimuleren van goed leergedrag. Dat betekent in essentie dat de toets gericht is op het geven van feedback en gewenste leergedragingen ontlokt. In de praktijk beïnvloeden veel toetsen het leren negatief en ontlokken ze maximalisatie-strategieën die ongewenste effecten hebben (memorisatie, minimale voorbereiding, uitstelgedrag, triviale beheersing van vaardigheden, om er maar een paar te noemen). Staat er veel op het spel, dan is het logisch dat veel datapunten nodig zijn. Voor belangrijke beslissingen gaat men niet over één nacht ijs.

Model voor programmatisch toetsen

In het model worden leeractiviteiten, toetsactiviteiten en begeleidingsactiviteiten onderscheiden. Aan het einde van de onderwijsperiode vindt een tussentijdse beoordeling plaats op grond van alle voorliggende geaggregeerde toetsinformatie tegen een achterliggende beoordelingsstandaard. De onderwijsperiodes en tussentijdse beoordelingen herhalen zich. Het aantal zal afhankelijk zijn van de curriculumindeling en de gemaakte keuzes in bekostiging. Ook hier geldt het adagium: liever minder frequent en goed dan het omgekeerde. Er hoeft niet aan elke onderwijsperiode een tussenbeoordeling verbonden te zijn, maar tenminste één tussenbeoordeling voor de finale beoordeling is nodig om het ontwikkelingsproces zijn gang te kunnen laten gaan. Ook is het systeem van finale beoordeling geloofwaardiger als het tot zo min mogelijk verrassingen leidt voor de lerende. Tussentijdse beoordelingen zijn daarvoor noodzakelijk.

Leeractiviteiten

Een onderwijsperiode bestaat uit allerlei leertaken. Er wordt van uitgegaan dat deze leertaken op een onderwijskundig verantwoorde manier in een arrangement zijn geplaatst, bijvoorbeeld gebaseerd op ontwerpregels voor goed onderwijs (Janssen & van Merrienboer, 2002). Sommige leertaken kunnen een product opleveren (bijvoorbeeld een project, een presentatie, een verkoopproduct, een lijst van gedane taken). In diezelfde periode vinden toetsactiviteiten plaats. Betreft het onderwijs een praktijkleerperiode, dan zijn de praktijktaken gelijk aan de leertaken.

Toetsactiviteiten

Een toetsactiviteit varieert van cognitief (feitgericht, toepassingsgericht) tot gedragsmatig (in vitro, in vivo) (Vleuten, Driessen & Tartwijk, 2006). Het areaal aan toetsmethoden is de laatste jaren sterk uitgebreid. Daarbij geldt dat geen enkele methode perfect is of een wondermiddel is. Het succes van veel methoden (m.n. de observatiegerichte) wordt vaak zelfs meer bepaald door de gebruikers van de methode (m.n. hoe serieus ze invulling geven aan het gebruik van de methode), dan van de methode zelf. Kenmerkend voor de toetsactiviteiten in het model is dat ze maximaal aansluiten op het onderwijs in vorm en

Het model gaat uit van verbondenheid tussen onderwijs en toetsing

frequentie, en maximaal betekenis hebben voor het leren. Dat betekent dat niet wordt volstaan met een beoordeling alleen, maar dat voor het leren nuttige feedback gegeven wordt. Dat kan in kwantitatieve vorm of in narratieve vorm en volgens de spelregels van feedback. Kwalitatieve, beschrijvende informatie is zeker voor complexe en gedragmatige competenties van onschatbare waarde. Cijfers vormen een schrale vorm van feedback die leidt tot een cijfercultuur. Het is soms ontmoedigend te zien hoeveel toetsprogramma's in de praktijk van het onderwijs grotendeels of geheel feedbackloos zijn. Toetsing moet volledig ingebed zijn in het onderwijs en rijke informatie opleveren die gerichte betekenis heeft voor het leren. Uiteraard zullen compromissen moeten worden gesloten. Zo zullen de kosten binnen de marges moeten blijven, maar binnen het haalbare wordt de leerwaarde van de toetsing gemaximeerd volgens het principe 'less is more': liever minder frequent en goed dan vaak en ritualistisch.

Sommige leertaken zullen tegelijkertijd toetstaken kunnen vormen waardoor er de minste kans op incongruentie tussen onderwijs- en toetsdoelen (*constructive alignment*) bestaat. Voor elk individueel datapunt van toetsing staat weinig, maar niet niks op het spel. De informatie uit het datapunt kan gebruikt worden als de informatie wordt geaggregeerd.
Voor sommige toetstaken staat iets meer op het spel. Sommige taken moeten worden beheerst. De tafeltjes van vermenigvuldiging vormen een typische beheersingstaak. Een student geneeskunde moet bijvoorbeeld kunnen reanimeren. Er wordt net zo lang geoefend tot de taak beheerst wordt. In de toetstaak wordt de bekwaamheidsverklaring voor de beheersingstaak afgegeven.

Begeleidingsactiviteiten

De informatie van de leer- en toetstaken wordt door de lerende gebruikt om het leren te sturen. Reflectie en planning en zelfsturing vinden plaats. Nu is bekend dat zelfsturing geen eenvoudig proces is en niet zo maar gebeurt, en dat sociale interactie rondom reflectie van groot belang is. Eén van de paradoxen in het onderwijs is dat veel sturing moet worden geboden om zelfsturend leren plaats te laten vinden, des te meer naarmate de lerende jonger is en het brein onvolgroeid. In het model is daarom een vorm van sociale interactie rondom zelfsturing opgenomen. Deze zou kunnen plaatsvinden door middel van coaching, supervisie of intervisie. De ontwikkeling wordt gevolgd, bijsturing vindt plaats. Wellicht dat in de loop der tijd de begeleidingsactiviteit kan afnemen of, afhankelijk van de lerende, meer of minder gedoseerd kan plaatsvinden.

Tussentijdse beoordeling

Aggregatie van toetsinformatie zal geen eenvoudige sommering of andere kwantitatieve operatie zijn. Het gaat om rijke informatie, deels kwalitatief, waarbij een kwantitatieve aggregatie snel onmogelijk is. De beoordeling zal door een 'expert' moeten gebeuren, liefst een commissie, die onafhankelijk is van de betrokkenen in de begeleidingsactiviteiten. Het besluit dat genomen wordt is diagnosticerend, remediërend en prospectief. Er zal eerst worden vastgesteld of er voldoende informatie is, vervolgens wat er aan de hand is, hoe eventueel dit geremedieerd moet worden, en er wordt voorspeld wat de uitkomst zal zijn indien dat (niet) gebeurt. Dit is een volkomen andere benadering dan in een klassiek zak/slaag besluit gekoppeld aan een herkansing. Herhaling is typisch iets dat past in het klassieke beheersingsgerichte model. Op grond van gelimiteerde informatie, namelijk zakken of slagen, wordt één vorm van remediëring geboden, namelijk herhaling. Dat is tamelijk informatie-arm, weinig efficiënt en niet ontwikkelingsgericht. Het hier beschreven model is gericht op remediëring, op longitudinale ontwikkeling en opvolging. Herhaling wordt in dit model niet uitgesloten, maar het is één van de opties in een geïndividualiseerd remediëringsadvies. De echte beheersingstaken vormen opnieuw een uitzondering. Herhaling en herkansing vormen hier een adequaat remediëringsadvies.
Voor de meeste lerenden zal een uitgebreid tussentijds remediëringsadvies niet noodzakelijk zijn. Hierin zit in een efficiëntieslag van het werk dat gemoeid is met deze vorm van beoordelen en remediëren. Een andere efficiëntieslag zit in de mate waarin het oordeel mede gevormd wordt door degene die het beste bekend is met het werk van de lerende, de begeleider, coach of mentor. Dit typeren wij als een

Veel toetsen beïnvloeden het leren negatief

'firewall-dilemma': de rol van de begeleider komt in een ander daglicht te staan als deze ook een beoordelaar wordt, maar tegelijkertijd is de informatie van de begeleider uniek en kan veel tijd besparen. Tussenliggende oplossingen zijn overigens ook mogelijk, bijvoorbeeld een situatie waarin zowel de lerende als de begeleider een adviserende rol hebben, maar de commissie het uiteindelijke besluit neemt.
Beoordeeld wordt aan de hand van een standaard die beschrijft wat iemand moeten kunnen/kennen op een aantal inhoudelijke domeinen voor een bepaalde fase van de opleiding. De resulterende beoordeling is belangrijk, er staat veel meer op het spel, maar nog steeds staat niet alles op het spel. Dat komt later in de finale beoordeling.

Finale beoordeling

Bij de finale beoordeling staat alles op het spel. Hier wordt een selectie-, doorstroom-, of certificerende beslissing genomen waar grote gevolgen aan vast kunnen zitten. In het model vindt die beslissing plaats op basis van veel informatie en op veel datapunten, de beslissing moet geloofwaardig zijn. Bij toetsing is men gewend te denken in objectivering en standaardisering, maar in dit model kan niet alles worden geobjectiveerd, kan objectivering zelfs tot trivialisering leiden en staat het menselijk oordeel noodzakelijkerwijs centraal. Andere maatregelen zijn nodig om bias in de besluitvorming te voorkomen. De oplossing wordt gevonden in procedurele maatregelen (geïnspireerd op kwalitatieve onderzoeksmethodieken (van der Vleuten, Schuwirth, Scheele, Driessen & Hodges, 2010) die de besluitvorming robuust maken. De term geloofwaardigheid van besluitvorming is bewust gekozen. Verdedigbare antwoorden moeten beschikbaar zijn als een buitenstaander vervelende vragen zou stellen over het besluit (een rechter?).
Enkele voorbeelden die de geloofwaardigheid van de besluitvorming zullen verhogen:

- De grootte en de expertise van de beoordelingscommissie
- De beschikbaarheid en wijze van totstandkoming van de beoordelingsstandaarden
- De mate van training van beoordelaars
- De mate waarin van moeilijke gevallen wordt geleerd (bijvoorbeeld in bijstelling van standaarden)
- De mate van deliberatie in de commissie bij twijfelgevallen
- De mate van beargumentering van het besluit
- De kans tot bijsturing en remediering gedurende de rit
- De mogelijkheid tot beroep tegen het besluit
- De mate waarin één en ander gedocumenteerd is

Al deze maatregelen maken het besluit geloofwaardiger, maar laten het noodzakelijke menselijk oordeel intact èn het besluit is gebaseerd op grond van veel informatie.

Het model is haalbaar

Mocht gedacht worden dat wat hier voorgesteld wordt, in de praktijk niet haalbaar zou zijn, dan wordt gewezen op enkele bestaande praktijken. Los van een van de eigen opleidingen in Maastricht die op dergelijke wijze is ingericht, bestaat er in Cleveland een opleiding geneeskunde die al een aantal jaren op deze wijze werkt (Dannefer & Henson, 2007). Het gehele toetsprogramma is feedbackgericht. Elke week worden open vragen gemaakt door de studenten waarop feedback wordt gegeven (punten worden wel gegeven, maar die gaan naar de docenten!). Periodiek is er een uitgebreide zelf- en peerevaluatie aan de hand van standaarden. Er is een systeem van begeleiding en een commissie van beoordeling met een absolute firewall tussen beide. De studenten zijn zeer tevreden met deze aanpak en zijn nu de grootste ambassadeurs ervan.

Discussie

Sommigen zullen een systeem van portfolioleren en beoordelen herkennen. Een portfolio zou voor het bovenstaande programma van toetsen een nuttig instrument kunnen zijn. Het model is bewust generiek gehouden en spreekt geen enkele voorkeur uit voor welk concreet instrument dan ook. In programma-

tisch denken is de waarde van een instrument erg afhankelijk van de functie in het systeem als geheel en geen enkel instrument wordt daarbij op voorhand in- of uitgesloten (en al beslist niet de klassieke vormen van toetsen). Het model is algemeen bedoeld. Het is bedoeld als een vorm van denken over toetsing, waarbij de vormende en de besluitvormende functie van toetsing worden geoptimaliseerd. Door programmatisch te denken kunnen beide functies hand in hand gaan.

Risico's

Natuurlijk zijn er risico's aan het model verbonden. Geen onbelangrijk risico om een dergelijk model te realiseren zijn de kosten en de haalbaarheid. Het model gaat uit van verbondenheid tussen onderwijs en toetsing, de grens daartussen vervaagt. Dus gaat het om integrale kosten van onderwijs. Compromissen over wat wel en wat niet te doen is dan geen keuze tussen het één (onderwijs) of het ander (toetsing). Conform het 'less is more' principe is 'minder maar goed' beter dan 'veel maar slecht'. Als beoordelingsinstrumenten een doel op zichzelf gaan worden, als lijstjes moeten worden afgewerkt, als puntjes moeten worden verzameld, alles spelletjes gespeeld moeten gaan worden, dan trivialiseert het leerproces (de wet van Goodhart![1]). Om dat te voorkomen moeten de actoren in dit hele proces (lerenden, docenten) goed beseffen wat hun rol is en waarom ze deze vervullen. Dat is niet makkelijk te realiseren. Voorbeeld: de taak van een docent betrokken bij één datapunt van toetsing is vooral het geven van feedback. Misschien moet het daartoe wel beperkt blijven en moet er geen zak/slaag beslissing of cijfer worden gegeven. Docenten vinden dat lastig, zij ervaren zich in hun 'bevoegdheden' beknot. Maar het zou bevrijdend kunnen zijn, de rol van rechter hoeft niet te worden gespeeld en gecombineerd met de rol als hulpgever, als facilitator, als docent.

Tenslotte kan het huidige juridische kader een bedreiging vormen. De wet en het hele stelsel van studiepunten zijn sterk gebaseerd op het klassieke beheersingsleren-gedachtengoed. Toets gehaald, punten binnen.

Literatuur

Baartman, L., Bastiaens, T. & Kirschner, P. (2005). Kwaliteitscriteria voor competentie assessment programma's. *Examens*, 2(2), 13-15.

Dannefer, E.F. & Henson, L.C. (2007). The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Academic Medicine*, 82(5), 493-502.

Dijkstra, J., van der Vleuten, C.P. & Schuwirth, L.W. (2010). A new framework for designing programmes of assessment. *Advances in Health Sciences Education: Theory and Practice*, 15(3), 379-93. Epub 2009 Oct 10.

Janssen, A.M.B. & van Merrienboer, J.J.G. (2002). *Innovatief Onderwijs Ontwerpen: Van leertaken naar complexe vaardigheden*. Groningen/Houten: Wolters-Noordhoff.

van der Vleuten, C.P., Schuwirth, L.W., Scheele, F., Driessen, E.W. & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 24, 703-719.

van der Vleuten, C.P.M. (1996). The Assessment of Professional Competence: Developments, Research and Practical Implications. *Advances in Health Scie nes Education*, 1(1), 41-67.

van der Vleuten, C.P.M. & Schuwirth, L.W.T. (2005). Assessment of professional competence: from methods to programmes. *Medical Education*, 39, 309-317.

van der Vleuten, C., Driessen, E., & Tartwijk, J. van. (2006). Toetsprogramma's. In H. van Hout , G. ten Dam , M. Mirande, C. Terlouw & J. Willems, *Vernieuwing in het hoger onderwijs* (pp. 119-136). Assen: Koninklijke Van Gorcum.

☐ De heer prof. dr. C.P.M. van der Vleuten is als hoogleraar verbonden aan de vakgroep Onderwijsontwikkeling en Onderwijsresearch van de Universiteit Maastricht, en Wetenschappelijk directeur van de UM Graduate School of Health Professions Education; de heer prof.dr. L.W. Schuwirth is als hoogleraar medisch onderwijs aan de bovengenoemde vakgroep verbonden. E-mail: c.vandervleuten@maastrichtuniversity.nl.

[1] "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes" (Goodhart 1975). De wet van Goodhart kan op vele verschillende wijze geformuleerd worden, maar komt er kort gezegd op neer dat zodra een sociale of economische parameter tot doel gemaakt is om sociale of economische politieke beslissingen te nemen hij de informatieve inhoud verliest die noodzakelijk is om deze doelstelling te kunnen realiseren. De wet is genoemd naar Charles Goodhart, een voormalige adviseur van de Bank of England en Emeritus Professor aan de London School of Economics.

OVER HET ZOEKEN NAAR DE WARE SCORE

Meten is weten; vergeet het maar!

Henk van Berkel

Je maakt een aantal examenvragen. Je zorgt er voor dat de vragen een goede spreiding hebben over de eindtermen en dan heb je een examen. Dan neem je het af, je telt het aantal punten dat de deelnemers hebben behaald, eventueel na correctie voor raden indien het examen bestond uit meerkeuzevragen. Dat aantal punten zet je om in een cijfer en klaar is Kees.

Ziedaar in een notendop het examenproces. Moet daar nu zoveel heisa over worden gemaakt? Je weet toch hoeveel examenvragen deelnemers correct hebben gemaakt? Dat heb je met het examen gemeten. Dus je weet ook hoeveel kennis ze hebben. Daar is toch geen stok tussen te krijgen?

Geen enkel examen is perfect betrouwbaar

Zoals met veel zaken, niets is zo eenvoudig als het lijkt. Natuurlijk kun je gemakkelijk tellen wat de score is op een examen. Maar het is naïef om die score klakkeloos te interpreteren als HET kennisniveau van de deelnemers. Examens zijn metingen. En net als alle andere metingen, is een meting nooit 100% betrouwbaar. In de examenliteratuur wordt de betrouwbaarheid gezien als de meest belangrijke beschikbare indicatie voor de kwaliteit van een examen. Als een examen perfect betrouwbaar is, wil dat zeggen dat onwenselijke factoren als vermoeidheid, examenangst, gokgedrag, een verkeerde verdeling van de vragen over de eindtermen, nakijkfouten, onduidelijke vragen, storende toetsomstandigheden et cetera, zich niet hebben voorgedaan of niet aanwezig zijn geweest. Geen enkel examen is echter perfect betrouwbaar. Altijd schort er wel wat aan. Vandaar dat het belangrijk is de betrouwbaarheid van een examen te berekenen. De betrouwbaarheid geeft aan in hoeverre je vertrouwen kunt hebben in het examen als meting. Meet het examen echt iets of kun je net zo goed een muntstukje opwerpen om uitspraken te doen over het kennisniveau van de deelnemers?

Betrouwbaarheid meten

Het is noodzakelijk na ieder examen de betrouwbaarheid te berekenen omdat de hoogte van de betrouwbaarheid grenzen stelt aan de betekenis van de examenuitslag. In *EXAMENS* is daaraan door verschillende auteurs eerder over geschreven (Altemühl-Booltink, (2005); de Gruijter, (2005); Meuffels, (2004a en b); Soeting & Molkenboer, (2006)). Er zijn verschillende methoden om de betrouwbaarheid van een examen te meten. Die meest bekende is coëfficiënt alpha. Overigens, deze bijdrage gaat uit van het paradigma van de zogenaamde klassieke testtheorie waarvan (nog steeds) de bekendste uiteenzetting is gegeven door Lord & Novick (1968).

Coëfficiënt alpha

Coëfficiënt alfa, ook wel aangeduid als Cronbach's alfa of kortweg 'alfa', is bij uitstek geschikt om de betrouwbaarheid te berekenen van een examen. De scores op alle vragen in het examen vormen de basis van de berekening. Via een tussenberekening van onder andere de standaarddeviaties van de vraagscores, wordt coëfficiënt alpha berekend. De mogelijke waarden die alpha kan aannemen, liggen tussen de 0 en 1.

Wanneer een examen niet betrouwbaar is, kun je geen enkele betekenis hechten aan de uitslag. Wan-

Geen enkel examen is perfect betrouwbaar

neer een examen 100% betrouwbaar is, heb je in ieder geval de zekerheid dat de uitslag iets heeft te betekenen. Of de uitslag ook zinvolle betekenis heeft, is uit de hoogte van de betrouwbaarheid niet af te leiden. Dat blijkt uit de bestudering van met name de inhoudsvaliditeit van het examen.
Alle tussenliggende waarden van betrouwbaarheid gaan gepaard met een onzekerheid: hoe dichter de betrouwbaarheid naar het minimum toegaat, des te onzekerder je kunt zijn of de uitslag wat voorstelt. Met andere woorden, er ligt een soort bandbreedte, een interval, rond de score die te maken heeft met de betrouwbaarheid van het examen.

Meetfout

Het interval rondom de eindscore kun je opvatten als een indicatie van de meetfout. Aan iedere meting zit een meetfout. Om het begrip 'meetfout' uit te kunnen leggen, is het noodzakelijk het begrip ware score te introduceren. De ware score is de score die een getrouwe afspiegeling is van de kennis van een deelnemer. Als deelnemers echt 85 van de 100 vragen kennen, moeten in dat geval hun examenscores ook 85 zijn. Hun 'ware' score is dan 85.

Dat wil echter nog niet zeggen dat deelnemers met een ware score van 85 ook een score krijgen van 85. De betreffende deelnemers kunnen tijdens het examen zijn afgeleid en per ongeluk, bij een meerkeuzevraag, het streepje op een verkeerde plaats hebben gezet. Of ze kunnen pech of geluk hebben gehad met de vragen die ze niet kenden en daarom hebben gegokt. Dat gokken kan per saldo slecht of juist goed voor hen uitpakken. Zo zijn er factoren die onbedoeld van invloed zijn op de score. Iedere examenscore heeft daardoor een interval om zich heen waarbinnen de ware score zich moet bevinden. Hoe breed dit interval is, is afhankelijk van de betrouwbaarheid van het examen. Nogmaals, hoe onbetrouwbaarder het examen, des te breder het interval. De breedte is een aanduiding voor de meetfout. De meetfout van een examen is dus afhankelijk van de betrouwbaarheid. Hoe hoger de betrouwbaarheid, des te kleiner de meetfout.

Je zou kunnen zeggen dat een score van bijvoorbeeld 85 bij een 100% betrouwbaar examen inderdaad 85 is, en niet 84 of 86. Bij een examen dat bijna 100% betrouwbaar is, zal er een klein gebiedje rondom de examenscore van 85 zijn waarvan je zeker kan zijn dat de ware score daar ergens tussenin ligt. Een uitspraak die je dan bijvoorbeeld kunt doen, is: wanneer de betrouwbaarheid 0,95 bedraagt en de examenscore is 85, dan weet je met een bepaalde zekerheid dat de ware score dicht in de beurt van 85 ligt.

Met de volgende formule kan men de breedte van het interval berekenen:

$Se = Sx * \sqrt{(1-R)}$
waarin:
Se = meetfout (ook wel aangeduid met standaard meetfout)
Sx = standaarddeviatie van de eindscores
R = betrouwbaarheid van het examen

De standaarddeviatie van de eindscore is een maat voor de spreiding. Als iedere deelnemer dezelfde eindscore behaalt, is de spreiding '0'. Als de betrouwbaarheid van het examen heel hoog is, is het getal onder de wortel heel klein en zal de meetfout ook heel klein zijn.

Een voorbeeld: als de standaarddeviatie van de examenscores 10 is en de betrouwbaarheid is 0,81 dan is de meetfout $4,36 = 10 * \sqrt{(1-0,81)}$.

De interpretatie van de meetfout

Statistisch kun je aantonen dat je 68% zeker kunt zijn dat de ware score ergens ligt binnen het interval dat wordt omgrensd tussen de examenscore plus de meetfout en de examenscore min de meetfout. Indien een deelnemer op bovenstaand voorbeeldexamen een score heeft behaald van 36, kun je er 68% zeker van zijn dat zijn ware score ergens ligt tussen 31,64 (36 - 4,36) en 40,36 (36 + 4,36). Wil je voor 95% zeker weten binnen welk interval de ware score ligt van een deelnemer die 36 punten heeft behaald op de toets, dien je de meetfout met twee te vermenigvuldigen. Het 95%-zekerheidsinterval loopt dan van 27,28 (36 - 2*4,36) tot 44,72 (36 + 2*4,36).
De meetfout is een belangrijk gegeven bij het

Hoe hoger de betrouwbaarheid, des te kleiner de meetfout

bepalen van de zak/slaaggrens van een examen. Het zal duidelijk zijn dat examencommissies slechts die examendeelnemers willen laten slagen die dat ook verdienen. Als je de zak/slaaggrens bij bovenstaand examen (dat dus een 95%-zekerheidsinterval heeft van 17,44) legt bij 35, weet je 95% zeker dat deelnemers die minder scoren dan 26,28 (35-17,44/2) terecht zijn gezakt. Maar je weet ook met 95% zekerheid dat deelnemers met een hogere score dan 43,72 (35+17,44/2) terecht zijn geslaagd. Daartussenin ligt een gebied waarover niet met 95% zekerheid kan worden gezegd dat deelnemers terecht zijn gezakt of geslaagd. Het spreekt voor zich dat dit onzekerheidsgebied zo klein mogelijk moet zijn. Dan pas zal het aantal misclassificaties (onterecht geslaagd of onterecht gezakt) binnen aanvaardbare proporties blijven. De zekerheid over een correcte beslissing is dus afhankelijk van de betrouwbaarheid (hoe lager de betrouwbaarheid, des te onzekerder). Maar het is ook afhankelijk van het percentage gezakte deelnemers (hoe extremer het zakpercentage, hoog of laag, dat maakt niet uit, des te lager de onzekerheid. - Het gaat in deze bijdrage te ver om hier verder op in te gaan; het heeft te maken met de spreiding van de scores -. Met behulp van formules is te berekenen hoe hoog de (on)zekerheid is. In tabel 1 staat de onzekerheid van de zak/slaagbeslissing, uitgedrukt in percentage van het aantal toetsdeelnemers, bij een aantal veel voorkomende waarden van de toetsbetrouwbaarheid en gegeven de zak/slaaggrens.

Op grond van statistische berekeningen is af te lezen dat zelfs bij een redelijke betrouwbaarheid, 0,80, en bij een niet ongewoon slaagpercentage, 70%, de onzekerheid 18% bedraagt. Dat wil zeggen, van 18% van de deelnemers is de beslissing onzeker.

Wenselijke hoogte van de betrouwbaarheid

Om te voorkomen dat er relatief veel misclassificaties optreden, moet de meetfout zo klein mogelijk zijn, en dus de betrouwbaarheid zo hoog mogelijk. Een examen dat 100% betrouwbaar is, is echter onmogelijk. Er zal dus een soort compromis moeten worden

Zak/slaag %	Betrouwbaarheid											
	0.0	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
5	10	8	7	7	7	6	6	5	4	4	3	0
10	18	14	13	12	11	11	10	9	8	6	4	0
15	26	18	18	17	16	14	13	12	10	8	6	0
20	32	23	21	20	19	17	16	14	12	10	7	0
25	38	26	25	23	22	20	18	16	14	11	8	0
30	42	29	27	25	24	22	20	18	15	12	9	0
35	46	31	29	27	25	23	21	19	16	13	9	0
40	48	32	30	29	27	24	22	20	17	14	10	0
45	50	33	31	29	27	25	23	20	18	14	10	0
50	50	33	31	30	29	25	23	20	18	14	10	0

Tabel 1 Percentage deelnemers waarover een niet met zekerheid correcte zak/slaagbeslissing is genomen (Uit: Dousma et al., 1997)

De meetfout is een belangrijk gegeven bij het bepalen van de zak/slaaggrens van een examen

gevonden ten aanzien van wat nog aanvaardbaar is. Het compromis zal rekening moeten houden met het belang van de beslissing. Hoe belangrijker de beslissingen die op basis van de uitslag worden genomen, des te minder misclassificaties je je kunt veroorloven. De vuistregel over de wenselijke hoogte van de betrouwbaarheid in relatie tot de soort beslissing die op basis van de examenuitslag wordt genomen, staat in tabel 2.

Examens die kunnen worden gecompenseerd met een ander examen mogen op zichzelf een relatief lage betrouwbaarheid bezitten (mits hoger dan 0,70), omdat mag worden verondersteld dat beide examens tezamen genomen een acceptabele berouwbaarheid bezitten, hoger dan 0,80. Een uitslag van een examen met een betrouwbaarheid van lager dan 0,60 heeft in het algemeen geen waarde.

Verhogen van de betrouwbaarheid

Bestudering van de formule voor het berekenen van de betrouwbaarheid leert dat het mogelijk is de betrouwbaarheid te verhogen.

Aantal vragen

Een examen bestaande uit veel vragen heeft meestal een hogere betrouwbaarheid dan een met weinig vragen. Dit is te verklaren door de vragen in een examen te zien als een steekproef uit alle mogelijke vragen. Hoe groter de steekproef des te minder de uitslag van het toeval afhankelijk is.
Het effect van het opnemen van extra vragen in het examen is te berekenen met de Spearman-Brown formule voor toetsverlenging die in alle tekstboeken over examineren is te vinden. Een examen kun je echter niet oneindig verlengen. Bij deelnemers gaat dan vermoeidheid een rol spelen. Ook zijn de te examineren eindtermen eindig en eens houdt de inspiratie op om steeds maar weer nieuwe vragen te construeren. Ergens zal een optimum moeten worden gevonden tussen een aanvaardbare betrouwbaarheid en haalbare examenlengte.

Homogene examens

Wanneer de vragen in het examen een spreiding hebben over een groot aantal onderwerpen, verlaagt dit de betrouwbaarheid. Dat heeft er mee te maken dat de overeenkomst tussen de scores op de vragen lager zijn zal in vergelijking tot de situatie waarin slechts een paar onderwerpen worden getoetst. De in praktijk gehanteerde betrouwbaarheidsmaten, bijvoorbeeld alfa, zijn meestal homogeniteitsmaten. Het spreekt dan voor zich dat hoe lager de homogeniteit is, des te lager de betrouwbaarheid. Dus examens die over één onderwerp gaan, zijn doorgaans betrouwbaarder dan examens die meerdere onderwerpen bevatten. Overigens, homogeniteit is een relatief begrip. Examens gaan meestal altijd zowel over één onderwerp, bijvoorbeeld wiskunde, als over meerdere (deel) onderwerpen, bijvoorbeeld analytische meetkunde, algebra, goniometrie et cetera. Het gaat in feite om de veronderstelling of kennis over een deelonderwerp impliceert dat de kandidaat ook kennis heeft over een ander deelonderwerp. Als dat het geval is, is het examen homogeen.

Beslissing	Minimale betrouwbaarheid
High stake examen	0,90
Examen zonder compensatiemogelijkheden	0,80
Examen met compensatiemogelijkheden	0,70
Formatief examen	0,60

Tabel 2 Minimaal wenselijk hoogte van de betrouwbaarheid in relatie tot beslissing

Vragen van extreme moeilijkheid dragen weinig bij aan de betrouwbaarheid

Extreme p-waarden

De moeilijkheid van een vraag is medebepalend voor de betrouwbaarheid. Een vraag die iedere deelnemer correct beantwoordt, of een vraag die iedere deelnemer foutief beantwoordt, draagt niets bij aan de betrouwbaarheid. Dit wordt veroorzaakt door het feit dat de betrouwbaarheid wordt berekend met formules die uitgaan van correlaties tussen de scores op de vragen. Indien iedere kandidaat een vraag (in) correct beantwoordt, is het onmogelijk zo'n correlatie te berekenen. Die vraag wordt daarom bij de berekening van de betrouwbaarheid niet meegenomen. Vragen van extreme moeilijkheid dragen weinig bij aan de betrouwbaarheid. Zeer moeilijke vragen worden vaak gedefinieerd op grond van de (bij gesloten vragen gecorrigeerde) p-waarde die dan minder dan 0,10 is. Een zeer makkelijke vraag heeft een p-waarde van meer dan 0,90.

Verschillen in kennis

Verschillen in het niveau van kennis tussen de examendeelnemers beïnvloeden de betrouwbaarheid: hoe groter de verschillen, des te hoger de betrouwbaarheid. Het omgekeerde is ook het geval: weinig verschillen leiden tot een lage betrouwbaarheid. Deze regel verklaart bijvoorbeeld de in praktijk vaak gevonden uitkomst dat de betrouwbaarheid van herkansingen lager is dan die van de oorspronkelijke toets. Op herkansingen nemen kandidaten deel die eerder zijn gezakt. De verschillen in kennis tussen eerder gezakte kandidaten is kleiner.

Literatuur

Altemühl-Booltink, M.H.M. (2005). Wat is ... cesuur? *Examens, Tijdschrift voor de Toetspraktijk*, 2(3), 18.

Dousma, T., Horsten, A., & Brants, J. (1997). *Tentamineren.* Groningen: Wolters-Noordhoff.

Gruijter, D.N.M. de (2005). Op zoek naar de besten. *Examens, Tijdschrift voor de Toetspraktijk,* 2(4), 17-19.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test score.* Reading: Addison.

Meuffels, H.L.M. (2004a). Is het een zes of een zeven? *Examens, Tijdschrift voor de Toetspraktijk,* 1(1), 6-9.

Meuffels, H.L.M. (2004b). Cijfers en cijferschalen. *Examens, Tijdschrift voor de Toetspraktijk,* 1(3), 9-11.

Soeting, J. & Molkenboer, H. (2006). Wat is ... klassieke testtheorie? *Examens, Tijdschrift voor de Toetspraktijk,* 3(1), 17.

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

Ook de leermeesterexamens van CINSES Examinering & Certificering staan onder toezicht van de Examenkamer kwaliteit door onafhankelijkheid

Terecht of niet?

Henk van Berkel

Plagiaat bij groepsproducten - **Wie** krijgt de **schuld?**

Ze komen hoe langer hoe meer voor, groepsproducten. Drie of vier studenten schrijven dan, bijvoorbeeld, gezamenlijk een paper over een bepaald onderwerp, zetten er hun naam onder en leveren het in. De docent kijkt het vervolgens na en iedere student krijgt hetzelfde punt. Tot zover geen probleem. Maar wat als een groot deel van de paper blijkt overgeschreven van een internetpagina? Wie krijgt dan de schuld?

De zaak

In een opleiding aan de universiteit te S. schrijven studenten in het derde jaar een paper. Zij moeten dat met hun drieën doen. Bij controle blijkt dat een groot deel van de paper is overgeschreven van internetbronnen. Dat is volgens de reglementen ongeoorloofd. De examencommissie legt hen gedrieën één straf op: het ongeldig verklaren van de betreffende paper. Een studente gaat hier tegen in beroep. Zij beweert dat zij niet degene is geweest die haar gedeelte van de paper heeft overgeschreven van internet. Er is volgens haar geen bewijs voorhanden waaruit het tegendeel zou blijken.

Het verweer

Ten eerste stelt de examencommissie dat duidelijk in de handleiding staat dat de hele groep verantwoordelijk is voor uiteindelijke paper, hetgeen ook blijkt uit het voorschrift dat ieder lid van de groep eenzelfde cijfer krijgt. De examencommissie is niet over één nacht ijs is gegaan. De commissie heeft veel moeite gedaan te achterhalen wie van de groepsleden verantwoordelijk is voor het deel van de paper dat geplagieerd is. Geen van de groepsleden evenwel heeft de verantwoordelijkheid genomen.

De uitspraak

Het betreft geen moeilijke zaak voor het College van Beroep. De overtreding is aangetoond en de reglementen zijn duidelijk. De examencommissie wordt daarom in het gelijk gesteld.

Implicaties

Groepsproducten zijn in opmars. Dat is op zichzelf terecht. Opleidingen bereiden studenten voor op beroepen waar hoe langer hoe meer moet worden samengewerkt. Het gezamenlijk schrijven van een paper is een toetsvorm die studenten daarop voorbereidt. Dat er problemen aan kleven, toont de hier beschreven beroepszaak. Zijn die problemen oplosbaar?

Het is mogelijk dat studenten in de gezamenlijke paper aangeven wie voor welk gedeelte verantwoordelijk is. Dat geeft een docent ook de mogelijkheid om niet iedere student eenzelfde cijfer te geven. Maar op die manier is er geen sprake van een gemeenschappelijk werk en daar is het met deze toetsvorm toch om te doen. Bovendien is er nog een ander nadeel. Niet ieder hoofdstuk van een paper is even belangrijk of even zwaar. Het schrijven van een literatuurlijst is weliswaar geen fijne klus, maar meer dan foutloos typen en het toepassen van regels toetst het niet. De rapportage van een literatuurstudie daarentegen is belangrijk en zou zwaar moeten wegen. Kortom, de verdeling van gedeelten over de studenten is niet zonder problemen. Opleidingen nemen daarom in hun reglementen op dat studenten gezamenlijk verantwoordelijk zijn voor de hele paper. De studenten dienen niet alleen zelf de hoofdstukken te verdelen, maar dragen ook verantwoordelijkheid voor het geheel. Studenten moet er op worden gewezen dat ze afhankelijk zijn van elkaar. Dat geldt niet alleen voor de inhoud, maar ook wat betreft het zich houden aan richtlijnen. Dat laatste kan plagiaat betref-

Terecht of niet?

fen, maar ook zaken als vorm van de paper, de presentatie van de inhoud en de inleverdatum.

Als een groepsproduct is opgebouwd uit individuele bijdragen is een tussenvorm mogelijk bij het beoordelen. De docent/beoordelaar kan elk van de individuele bijdragen van een cijfer voorzien én het geheel. Deze twee cijfers worden dan op een van te voren bepaalde wijze gecombineerd tot één eindcijfer.

Daarnaast is het waard de bijdrage van Bax (2004) er nog eens bij te nemen. Zij zet in tabel de verschillende beoordelingssystemen voor groepswerk en onderscheidt daarin de volgende aspecten: Wat wordt beoordeeld? Wie beoordeelt? Wie wordt beoordeeld? Hoe wordt beoordeeld? Wanneer wordt beoordeeld en de gevolgen van beoordelen.

Literatuur

Bax, A.E. (2004). Beoordelingsmethoden voor het toekennen van individuele cijfers aan groepsproducten, *Examens, Tijdschrift voor de Toetspraktijk*, 1(4), 18-21.

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

Gastcolumn

Wynand Wijnen

Mijn leerlingen beoordelen elkaar

Leerling F. ziet het echt nog niet zitten. De docent heeft de klas in drietallen ingedeeld en deze drietallen moesten een presentatie voorbereiden over een milieuonderwerp. Hij werd ingedeeld bij het leukste meisje van de klas en ze hebben met veel plezier gewerkt aan de voorbereiding van de presentatie. Nu een en ander afgerond is vraagt de docent ineens of ze elkaar willen beoordelen. Eigenlijk vindt leerling F. dit maar niks. Hij heeft wel gezien dat zijn klasgenote zich niet bovenmatig inspande, maar omdat hij wat extra tijd heeft ingezet, is de presentatie toch nog heel goed verlopen. Wat moet hij nu? Moet hij zijn klasgenote een lage beoordeling geven, omdat ze eigenlijk niet zo heel veel heeft gedaan of is het beter dat hij uitgaat van de kwaliteit van de presentatie en voor iedereen uit het drietal een goed cijfer voorstelt. Hij weet het niet en daarom vindt hij elkaar beoordelen maar niks.

Er zijn goede redenen om leerlingen in het onderwijs van tijd tot tijd groepsopdrachten te laten uitvoeren. Niet alleen kunnen leerlingen bij de uitvoering daarvan van elkaar leren, maar ook kan op die manier duidelijk worden gemaakt, dat samenwerken in groepsverband ook in de maatschappij van veel belang is. Over het belang van groepsopdrachten is er overigens in de klas geen verschil van mening. Het probleem begint wanneer de docent vraagt of de groepsleden elkaar willen beoordelen. Zo'n verzoek vinden de leerlingen eigenlijk niet fair. 'Je gaat je klasgenoten toch niet afvallen'. 'Hoe kunnen wij nu oordelen over de inbreng van anderen, terwijl we het zelf nog moeten leren?' 'Is het wel redelijk dat de docent de moeilijke beoordelingstaak naar ons doorstuurt?' Deze en andere vragen worden door de leerlingen druk besproken en er zijn verschillende standpunten.
Leerlingen begrijpen dat de docent niet alles kan zien wanneer de klas in kleine groepjes aan het werk is. Maar het beoordelen van elkaar vinden de leerlingen toch geen prettige opgave. Natuurlijk zouden ze wel eens duidelijk willen maken dat sommige leerlingen er de kantjes vanaf lopen, maar wat zijn de gevolgen? Ook willen ze die sympathieke leerling die het moeilijk heeft, graag ondersteunen, maar wat heeft hij daar aan?

Leerlingen realiseren zich natuurlijk dat de docent niet op zoek is naar uitspraken over sympathie en antipathie. Het moet gaan over prestaties en over niets anders. Wanneer men beoordelingen vooral ziet als een afsluiting van het verleden zou men een geringere inzet van medeleerlingen met de mantel der liefde kunnen bedekken, maar het gaat ook om de toekomst. Daarom is het de moeite waard dat leerlingen weten dat hun medeleerlingen niet echt te spreken zijn over hun inzet. Het benoemen van sterke en zwakke punten door medeleerlingen – mits eerlijk en fair bedoeld – kan leerlingen alleen maar helpen bij het verbeteren van hun leerproces. En daar gaat het natuurlijk om.

Wanneer leerlingen in het onderwijs elkaar moeten beoordelen, melden zich achterdochtige gedachten als: 'Leerlingen zullen elkaar wel niet de voet dwars willen zetten.' 'De lievelingetjes in de klas zullen hiervan wel profiteren.' 'Het is toch niet redelijk dat de docent het moeilijke werk aan de leerlingen overlaat.' Deze teksten en variaties daarop worden al snel gehoord. Toch lijkt het leerzaam wanneer leerlingen ervaring opdoen met het beoordelen van elkaar. Gedwongen worden om op feiten te letten kan vooroordelen minder krachtig maken. En meningen kunnen beter uitgesproken worden dan dat ze ongekend en ongecontroleerd hun werk blijven doen. Voorwaarde is dat de docent op een goede manier het 'elkaar beoordelen' begeleidt en zo nodig corrigeert.

Prof.dr. W.H.F.W. Wijnen was hoogleraar Ontwikkeling en Onderzoek van het Hoger Onderwijs aan de Universiteit Maastricht.

INVOERING VAN DE REKENTOETS HAVO/VWO IN 2014

De wet, het rekenen en de rekentoets

Ben Wilbrink en Joost Hulshof

Minister Van Bijsterveldt van OCW heeft het voornemen om in 2014 onder andere voor havo/vwo een rekentoets in het eindexamen in te voeren als onderdeel van de kernvakkenregel (voldoende voor rekentoets, Nederlands, Engels en wiskunde). Er doemen zo ernstige problemen op rond deze rekentoets. Dit artikel betoogt dat de voorbeeld-rekentoetsen die door de commissie-Schmidt zijn gepubliceerd weinig van doen hebben met het rekenen dat vrijwel iedereen voor ogen staat, dat het hoger onderwijs verlangt en dat bewindslieden en Tweede Kamer voorstaan.

Reken- en taalvaardigheden

De achtergrond van de wet 'Referentieniveaus Nederlandse taal en rekenen' is de beeldvorming dat het onderwijs tekortschiet in het bijbrengen van basale reken- en taalvaardigheden. Voor taalverzorging hebben Wilbrink, Borsboom & Couzijn (2010) integraal taalbeleid bepleit. Minister Van Bijsterveldt vindt dat ook: zij heeft op 7 juni 2011 de Tweede Kamer meegedeeld dat zij ervan afziet om spelling en grammatica in alle centrale examens te laten meewegen.

De Periodieke Peiling van het Onderwijsniveau (PPON) 2004 heeft laten zien dat de basale rekenvaardigheden (vermenigvuldigen, delen, breuken) in groep 8 van het basisonderwijs ernstig zijn teruggevallen (van Putten, 2005). Daar zijn tal van verklaringen voor te opperen, maar de belangrijkste is de heersende rekendidactiek waarin juist die basale rekenvaardigheden expliciet zijn verwaarloosd. Deze vaardigheden zouden niet belangrijk zijn, omdat het allereerst om het begrijpen gaat, niet om het beheersen, en omdat deze leerlingen in hun toekomst altijd de rekenmachine onder handbereik zullen hebben.

De wet 'Referentieniveaus Nederlandse taal en rekenen'

Deze wet is op 1 augustus 2010 in werking getreden en strekt zich over vrijwel alle onderwijssectoren uit. De wet beoogt:

- een goede zichtbaarheid van het niveau van beheersing van de Nederlandse taal en het rekenen voor zowel de leerling als de leraar en de school;
- meer eenduidigheid in taal- en rekenonderwijs in de gehele onderwijskolom;
- meer doelgericht taal- en rekenonderwijs door nauwkeurig omschreven doelen;
- een betere overdracht van leerlingen tussen de verschillende onderwijssectoren door de introductie van een eenduidige en gemeenschappelijke taal;
- het ontstaan van beter doorlopende leerlijnen voor taal en rekenen;
- het (opnieuw) doordenken door scholen van de aanpak van taal en rekenen;
- het verleggen van accenten binnen het huidige taal- en rekenonderwijs.

Het referentiekader vormt de basis voor (aanpassing van) lesmethoden, leermiddelen en toetsen/examens. Daardoor zal het ook uitgangspunt zijn bij het ontwerpen van taal- en rekenonderwijs binnen scholen en lerarenopleidingen.

Bron: http://www.taalenrekenen.nl/referentiekader/wet/ (toegevoegd door de redactie)

De basale rekenvaardigheden zijn verwaarloosd

Het probleem is nu onderkend door uitgevers die in de nieuwe edities van hun rekenmethoden en leerlingvolgsystemen weer meer aandacht aan de basale rekenvaardigheden schenken. Ook veel scholen vullen hun rekenmethode aan met materiaal voor oefening van basale vaardigheden. Des te opvallender is het dat in de referentieniveaus voor rekenen dit probleem niet is erkend en opgepakt. Ook de wet maakt er geen melding van.

De rekentoets

De commissie-Schmidt heeft voorbeeld-rekentoetsen opgesteld. Een aantal kenmerken van de rekentoets en voorbeeldvragen zijn in het kader opgenomen. De commissie schrijft dat er geen vragen zullen zijn naar een specifieke rekenprocedure. Dus geen opgaven als: bereken met een staartdeling 74983 : 432. Of: bereken met behulp van vermenigvuldigen onder elkaar 74983 x 432. De impliciete boodschap aan het onderwijsveld is: dit rekenen is niet van belang, er zal niet op worden getoetst. Basale rekenvaardigheden zijn in de voorbeeldtoets inderdaad nauwelijks aan de orde.
Schmidt (2011) heeft in een uitvoerige reactie op de analyse van de eerste auteur in zijn rekenblog op de website van de Stichting Beter Onderwijs Nederland een en ander nader toegelicht. Hij schrijft onder meer: 'De rekentoetswijzercommissie heeft als opdracht gekregen een toetswijzer te ontwikkelen voor havo en vwo op basis van referentieniveau 3F. […] De vrijheid van de commissie was en is zeer beperkt. […] De commissie zou het referentiekader geen recht doen als ze een groot deel van deze rekendoelen niet in de rekentoetsen zou opnemen of zelf rekendoelen zou hebben toegevoegd.'

Kenmerken van de rekentoets havo/vwo en voorbeeldvragen

- 'De rekentoets bevat contextloze rekenopgaven die zonder rekenmachine gemaakt moeten worden en die overeenstemmen met de beschrijving uit referentieniveaus 1F en 1S. Overige opgaven zijn contextopgaven.'
- 'Het aandeel van contextloze opgaven in de score is ongeveer 20%.'
- 'Beantwoorden van een contextloze vraag kan naar keuze van de kandidaat met of zonder toepassing van een standaardcijferprocedure.'
- 'Er zijn geen toetsopgaven die betrekking hebben op een specifieke rekenprocedure.'
- 'Bij contextopgaven is in alle gevallen een digitale rekenmachine met de standaardrekenfuncties plus worteltrekken beschikbaar, ook als gebruik van de rekenmachine niet zinvol is.'
- 'Alle vragen zijn computerscoorbaar.'

Voorbeelden van vragen:
Vraag '6. 48 : 0,12 = ….'
Vraag '9. 286 + 1034 = ….'
Vraag '10. 25 × 128 = ….'
Vraag '29. Mieke heeft met ingang van 1 januari 2010 een loonsverhoging van 4% gekregen. Op 1 januari 2011 kreeg ze opnieuw een loonsverhoging, maar nu van 2%.
Met hoeveel procent is haar loon na deze twee verhogingen in totaal gestegen?
A. Met 6%
B. Met iets meer dan 6%
C. Met iets minder dan 6%
D. Dat kun je niet weten want je weet niet hoeveel Mieke voor de loonsverhoging verdiende.'

Bron: Uitnodiging expertmeeting 12 april 2011. http://staff.science.uva.nl/~craats/#rekentoetsen

Moeten leerlingen zakken op de rekentoets, terwijl ze wel rekenvaardig zijn

Validiteit van deze rekentoets als examenonderdeel

In de nu beschikbare voorbeeld-rekentoets havo/vwo komen geen opgaven voor die basale rekenvaardigheden toetsen. Met andere woorden: deze wettelijke rekentoets havo/vwo dwingt in feite om in het voortgezet onderwijs precies dat rekenonderwijs te gaan verzorgen dat juist in het basisonderwijs voor wegzakkende basale rekenvaardigheden zorgt. Hier liggen ernstige validiteitsproblemen open en bloot. Immers, een rekentoets die niet bijdraagt aan reparatie van de gesignaleerde rekenproblematiek kan geen validiteit hebben.Een rekentoets aan het eind van havo/vwo moet valide zijn naar inhoud. Is het rekenen dat ermee wordt getoetst, inderdaad het rekenen waarover na de PPON 2004 zo'n grote bezorgdheid is ontstaan? Voor havo/vwo ligt het voor de hand dat de criteriumvaliditeit is gelegen in de mate waarin het rekenen in de rekentoets aansluit bij de rekenvaardigheden die het ontvangende hoger onderwijs eist.
Vanuit beide invalshoeken bezien - inhoud en criterium - gaat het om de basale rekenvaardigheden, die daarom de inhoud van de rekentoets havo/vwo moeten vormen. Dat het hoger onderwijs ernstige problemen ondervindt met tekortschietende basale rekenvaardigheden mag voldoende bekend heten. Op tal van plaatsen zijn er bijspijkercursussen en intreetoetsen. Oefenboeken zoals Pfaltzgraff (2009) en Van de Craats & Bosch (2009) gaan over het rekenen dat het hoger onderwijs vraagt. Inhoudelijke overeenkomst tussen de voorbeeld-rekentoets en deze oefenboeken is nauwelijks zichtbaar. Voor de onderbouw vo is Reichard e.a. (2009) passend rekenonderwijs.

Evaluatie van onderwijs of beoordeling van leerlingen?

Wie de beschikbare stukken leest, krijgt de indruk dat de rekentoetsen bedoeld zijn om er leerlingen individueel op af te rekenen. Dat beeld lijkt bevestigd door de uitwerking die de minister op 7 juni heeft gegeven: de rekentoets is onderdeel van de kernvakkenregel (zie figuur 1).

HAVO/VWO examinering afgestemd op referentieniveaus 3F en 4F						
2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016
Prepilot rekentoets	Beperkte pilot rekentoets Oplevering aangepaste syllabi Nederlands, rekentoetswijzer en voorbeeldmateriaal	Eerste pilotjaar rekentoets Aanpassing uitslagregel: CE-onderdelen gemiddeld voldoende	Tweede pilotjaar rekentoets (generale repetitie) Aanpassing uitslagregel: Niet meer dan één onvoldoende (ten minste een 5) voor de eindcijfers Nederlands, Engels en wiskunde	Invoering rekentoets en afgestemd examen Nederlands Aanpassing uitslagregel: Eindcijfer rekentoets ten minste een 5		Aanpassing uitslagregel: De rekentoets wordt onderdeel van de kernvakkenregel: niet meer dan één onvoldoende (ten minste een 5) voor eindcijfers Nederlands, Engels, wiskunde en rekenen.

Figuur 1 Overzicht implementatietraject invoering referentieniveaus in eindtoets/examens (Uit PDF document Kamerstuk | 07-06-2011 | OCW, toegevoegd door de redactie)

De *Handelingen* van 31 maart 2010 geven echter een ander beeld: fracties spraken ernstige zorgen uit over de situatie die kan ontstaan wanneer leerlingen individueel worden beoordeeld op hun rekenprestaties. Dat treft in het bijzonder natuurlijk basisschoolleerlingen. De bewindslieden deelden toen die zorgen. De gedachte ligt dus voor dat rekentoetsen nodig zijn om vast te stellen of de scholen voldoende voortgang maken met het op peil brengen van de (basale) rekenprestaties: het zijn evaluatieve instrumenten. De beslissing daarover werd in het kamerdebat nadrukkelijk opengelaten. Wellicht ten overvloede: hetgeen tussen parlement en bewindslieden is gewisseld en in de *Handelingen* vastgelegd, bepaalt mede de betekenis van de wet, *in casu* de bedoeling van de wetgever. Dat kan pikant zijn wanneer blijkt dat de Tweede Kamer over onvolledige informatie beschikte.
Een heel andere benadering van dezelfde kwestie - evalueren of beoordelen - is een psychometrische. De wetgever kan wel willen dat de rekentoets onderdeel wordt van de kernvakkenregel, maar dan gaan er - op een vak dat niet eens een volstandig eindexamenonderdeel is - zeker slachtoffers vallen onder de leerlingen. Moeten leerlingen het risico lopen om te zakken, terwijl ze wel rekenvaardig zijn? En hun andere examenvakken prima op orde hebben?

Iedere toets is een steekproef uit de leerstof, of een steekproef uit wat de leerling kent. Al naar gelang de steekproef kan de score voor dezelfde leerling hoger of lager uitvallen. Hoeveel hoger of lager? Veronderstel eens dat de stofbeheersing 0,6 is. Dat betekent dat de kans 0,6 is dat de leerling de eerste vraag goed beantwoordt. Idem voor de tweede vraag. De kansverdeling voor zijn score is de binomiaalverdeling, waarmee is te berekenen wat de kans is om onvoldoende te scoren, zeg minder dan 55% van de vragen goed. Zo laat zich prima kwantificeren hoeveel ongelukken er kunnen gebeuren - verkeerde beslissingen - op een rekentoets die meedoet in de kernvakkenregel.

Laat OCW opdracht geven om deze situatie zorgvuldig in beeld te brengen, zoals het eerder door Cito een inventarisatie heeft laten maken van de mogelijke gevolgen van de kernvakkenregel (van Rijn, 2009). Voor de rekentoets zal dat aan de hand van modelberekeningen moeten gebeuren (zie voor een vergelijkbare oefening Wilbrink (1980)).

Integraal rekenbeleid

Bij rekenen is er voorlopig nog een ernstig tekort in basale rekenvaardigheid van veel twaalfjarigen. Dan is er linksom of rechtsom eerst bijspijkeronderwijs nodig. Daarna moet die rekenvaardigheid wel op peil blijven, en daar is integraal rekenbeleid voor nodig: iedere gelegenheid voor oefening is belangrijk, bij ieder vak. Het gebruik van de rekenmachine berooft de leerlingen van die noodzakelijke oefening. Het risico van verkeerd beleid op dit punt is groot, alle mooie en soms gesponsorde praat over de rekenmachine als rekenslaafje ten spijt.
Er is geen goed alternatief voor integraal rekenbeleid. Immers, een afzonderlijk vak rekenen in de bovenbouw van havo/vwo, terwijl elders het gebruik van de rekenmachine blijft toegestaan, komt neer op dweilen met de kraan open. De rekentoets in het eindexamen gebruiken om er de leerlingen op af te rekenen zal zonder integraal rekenbeleid uitlopen op ongelukken en juridische procedures. Draai het dan om: houd scholen verantwoordelijk voor integraal rekenbeleid en gebruik desnoods een rekentoets in het laatste jaar als evaluatief instrument voor het succes van dat beleid.

Afsluiting

Een rekentoets als nieuw onderdeel van eindexamens havo/vwo in 2014 roept tal van dringende vragen op. Een wezenlijke vraag is die over de beperkte aard van de rekenvragen, zoals de commissie-Schmidt die met zijn voorbeeld-rekentoets havo/vwo bekend heeft gemaakt. Het rekenen zoals daarin getoetst, is niet het basale rekenen waarvan PPON 2004 juist laat zien dat het extra aandacht behoeft en het is niet het rekenen dat het hoger onderwijs mag verwachten.
Een rekentoets in de bovenbouw zou prima zijn als evaluatieve toets. Maar een rekentoets waar havo/vwo-leerlingen op kunnen struikelen voor hun eindexamen gaat zorgen voor een hoop gedoe, al was het maar omdat dit als onrechtvaardig zal worden ervaren.
Evenals voor taalverzorging de juiste aanpak ligt in integraal taalbeleid, ligt het in de rede om van scholen een goed integraal rekenbeleid te verlangen. Daar is het instrument van een rekentoets als examenonderdeel -

Er is een integraal rekenbeleid nodig

als stok achter de deur - niet voor nodig. Het gaat hier in principe om een basisvaardigheid die het havo/vwo vooral goed moet onderhouden - zonder rekenmachine - aangenomen dat het rekenvaardige leerlingen van het basisonderwijs mag ontvangen.

Literatuur

Craats, J. van de & Bosch, R. (2009). *Basisboek rekenen*. Pearson Education Benelux.

Kamerbrief voortgang implementatie referentiekader taal en rekenen 7 juni 2011. http://ikregeer.nl/documenten/kst-31332-16

Pfaltzgraff, H. (2009). *Rekenen.* Epsilon.

Putten, C. M. van (2005). Strategiegebruik bij het oplossen van deelsommen. In J. Janssen, F. van der Schoot en B. Hemker: *Balans [32] van het reken-wiskundeonderwijs aan het einde van de basisschool. 4. Uitkomsten van de vierde peiling in 2004.* (125-131). http://www.cito.nl/onderzoek%20en%20wetenschap/onderzoek/ppon/~/media/cito_nl/Files/Onderzoek%20en%20wetenschap/ppon/cito_ppon_balans_32.ashx

Reichard, L. A., et al. (2009). *Getal en ruimte havo/vwo. Rekenboek. Wi 1, wi 2, wi 3*. EPN, Noordhoff Uitgevers.

Rijn, P. van. (2009). Mogelijke effecten van verschillende uitslagregels. *Examens, Tijdschrift voor de Toetspraktijk*, 6(3), 5-8.

Schmidt, V. (22 juni 2011). 'Uw blog van 16 juni j.l.' http://beteronderwijsnederland.net/files/Reactie%20Blog%2016%20Wilbrink.pdf

Wilbrink, B. (1980). *Kansberekening bij Pais' voorontwerp van wet toelating tot numerus fixus studies in het w.o.* UvA: COWO. http://www.benwilbrink.nl/publicaties/80KansberekeningenCOWO

Wilbrink, B., Borsboom, D. & Couzijn, M. (2010). Spelling, grammatica en interpunctie meebeoordelen op eindexamens? *Examens, Tijdschrift voor de Toetspraktijk*, 7(3), 5-9.

Wilbrink, B. Rekenblog 16. http://www.beteronderwijsnederland.nl/node/7791

☐ De heer B. Wilbrink is freelancer onderwijsonderzoek. www.benwilbrink.nl. De heer prof.dr. J. Hulshof is hoogleraar Wiskundige Analyse aan de Vrije Universiteit en bestuurslid van Beter Onderwijs Nederland. E-mail: ben@benwilbrink.nl.

advertentie ten klei

Netwerkbijeenkomsten over toetsing in het wo

De Universiteit van Tilburg en IVA beleidsonderzoek organiseerden gedurende het academisch jaar 2010-2011 drie netwerkbijeenkomsten rond het thema ' Kwaliteit van toetsing in het wetenschappelijk onderwijs'. Na een startbijeenkomst in Tilburg in 2010 zijn in 2011 twee vervolgbijeenkomsten gehouden bij SURF Foundation in Utrecht.

Op initiatief van Claudia Loijens, medewerker toetskwaliteit van de Universiteit van Tilburg, en Paula Willemse, senior onderzoeker van IVA beleidsonderzoek, is een aantal bijeenkomsten over toetsing georganiseerd met deelnemers van 10 verschillende universiteiten. Uit een inventarisatie bleek dat aan de universiteiten behoefte bestaat om kennis te delen en te ontwikkelen, informatie te verstrekken en samen te werken ter verbetering van de kwaliteit van de toetsing.

In de eerste bijeenkomst zijn onder meer de volgende thema's en bijbehorende vragen uitgesproken die leidend zijn geweest voor de invulling de volgende bijeenkomsten:

- Examencommissie: hoe ondersteun je examencommissies in hun nieuwe rol?;
- Docenten: hoe zorg je ervoor dat centraal toetsbeleid de docent bereikt?;
- ICT: wat zijn de mogelijkheden en moeilijkheden van digitaal toetsen?;
- Toetsbeleid: wat is goed toetsbeleid en hoe komt dit tot stand?; wat zijn minimale eisen voor de (nieuwe) accreditatie?;
- Betrouwbaarheid en validiteit van toetsen: kan universiteitsoverschrijdend onderzoek worden opgezet?; hoe voorkom je fraude bij toetsing?

Gerard van de Watering, medewerker kwaliteitszorg en onderwijsinnovatie van de Technische Universiteit Eindhoven, heeft resultaten van een drietal onderzoeken gepresenteerd: 1° ronde vak-evaluaties, 2° een vragenlijstonderzoek over validiteit, betrouwbaarheid, transparantie en efficiëntie (kwaliteitscriteria van toetsing) en 3° naar de opzet van toetsen, het toetsconstructieproces, het beoordelingsproces, de betrouwbaarheid en de cijferbepaling in een steekproef van 35 toetsen uit een totaal van circa 1.400 toetsen.

De examencommissie was het thema van de tweede netwerkbijeenkomst. De plaats van de examencommissie in de faculteit of instelling, de relatie van de toetsdeskundige tot de examencommissie, de MOC-rollen (Managing Organizational Change) en de stappen om tot verandering van de organisatie te komen hebben ruime aandacht gekregen. Een casusbeschrijving over een chaotisch verlopen toets gaf de gewenste praktijksituatie waarin MOC-rollen en positie van de examencommissie werden benoemd en geduid.

In de derde bijeenkomst stond de implementatie van toetsbeleid bij de verschillende universiteiten centraal. Jan Steen gaf een presentatie over de totstandkoming van toetsbeleid bij Wageningen University. Deelnemers maakten daarna een stappenplan van de (mogelijke) implementatie van het toetsbeleid binnen hun eigen instelling. Wie maakt het plan? Hoe landt het plan in de organisatie? Hoe wordt het plan uitgevoerd? Wie heb je nodig om het toetsbeleid te laten werken?

Deelnemers hebben aangegeven het netwerk te willen voortzetten. De aanwezigheid van verschillende gremia wordt als positief en inspirerend ervaren. Het werken met casussen en opdrachten stimuleert om na te denken over de eigen praktijk en de 'theoretische prikkeling' helpt daar zeker bij!

Wie interesse heeft in het bijwonen van de netwerkbijeenkomsten in het academisch jaar 2011-2012 kan contact opnemen met Hans Mariën, onderzoeker bij IVA beleidsonderzoek: tel. 0134668484 of email: j.b.a.marien@uvt.nl.

OVER INTIMIDATIE, OMKOPING EN ONDERMIJNEND GEDRAG

Effectiviteit van een examenstelsel

Agnes Dinkelman

De examenbranche heeft te maken met toenemende druk die leidt tot intimidatie en fraude. Voor de sector is het belangrijk om de interne situatie tegen het licht te houden. Waar zitten de mogelijkheden om de tendens te keren?

Misstanden

Voeger werd de examinator met respect behandeld. Hij was een belangrijke spil in het proces, hij kweet zich van zijn belangrijke taak en was daar trots op. Wangedrag was nog niet doorgedrongen tot in de examenlokalen. Dat is nu anders. De laatste tijd komen meer misstanden bij examinering voor. Dat intimidatie van surveillanten en examinatoren door kandidaten toeneemt was al langer bekend. Nieuw is de ondermijning van zowel gezag als van de examenreglementen. Ook pogingen tot omkoping door zowel kandidaten als docenten/opleiders is nieuw.
Op welke schaal speelt dit zich af? Zijn er hotspots te benoemen? Zijn er terugkerende patronen waar te nemen? En wat zijn de drijfveren van de daders? Welke rol spelen de verschillende actoren in het systeem? Nader onderzoek is nodig om uitsluitsel te geven.
De externe factoren zijn nog onvoldoende in kaart gebracht om verantwoord actie te kunnen ondernemen. De examenbranche kan echter wel de interne situatie nauwgezet onder de loep nemen. Want welke mogelijkheden zijn er om wangedrag te ontmoedigen? In dit artikel worden de twee interne factoren 'Organisatie' en 'Weerbaarheid' verkend.

Organisatie

In een open examenstelsel is het van belang dat er sprake is van optimaal functioneren van alle actoren en wel op een zorgvuldige manier afgemeten aan het gelijkheidsbeginsel. Daarvoor is een eenduidige regelgeving van belang die handhaafbaar is door een toezichthouder met gezag. Men moet in de branche weten wat men van elkaar moet en kan verwachten en op welke wijze men elkaar corrigeert als een van de spelers in gebreke blijft. Afwijkingen van de processen en protocollen zijn precies de punten waar wangedrag de ruimte krijgt. Precies hier ontstaat de *ad hoc* aanpak van situaties, gebrek aan sociale controle en een gebrek aan zelfreinigend vermogen van een organisatie.

Centrale aansturing van de processen kan een goede stap zijn, maar is niet eenvoudig met de vele commercieel zelfstandige exameninstellingen. Als de afstemming van processen en de interpretatie van de invulling van de processtappen uniform en deugdelijk is, is er weinig ruimte voor fraude.
Een samenhangende aanpak van dit thema voor alle betrokken stelselpartners is alleen haalbaar als de focus ligt op stroomlijning van de hoofdprocessen. Een bedrijfsspecifieke inkleuring hoeft geen probleem te zijn zolang het gelijkheidsbeginsel en de integriteit niet worden aangetast. Instellingen moeten vervolgens met elkaar in gesprek blijven over de interpretatie van de processtappen, de regels en de handhaving.

Weerbaarheid

De examenbranche kan zelf de weerbaarheid van individuen verbeteren. In diverse workshops heeft DMI (Dinkelman Maatschappelijke Interventies) in de personen van Agnes Dinkelman en Ineke Dane groepen surveillanten en leidinggevenden van examenbureaus getraind op weerbaarheid. Ondermijning, omkoping en intimidatie zijn besproken en in kleine groepen zijn ervaringen van deelnemers uitgewisseld.

Intimidatie

Veel surveillanten/examinatoren hebben ervaring met intimidatie, meestal is dat spontane intimidatie waarop zij adequaat kunnen reageren. In geval van teleurstelling en boosheid bij een kandidaat kan de

De branche moet zich buigen over de mores inzake wangedrag

surveillant duidelijk maken dat hij alleen in gesprek wil gaan als de kandidaat hem op een niet-agressieve manier bejegent. Als het om een groep gaat, kan hij de leider eruit pikken en zich op deze persoon richten om tot een gesprek te komen, hem oproepen zijn agressieve gedrag te laten varen en te vertellen wat er aan de hand is. In deze interventies is het belangrijk dat er oogcontact is tussen examinator/surveillant en kandidaat. Vaak bepaalt de kwaliteit van het contact of de surveillant de situatie weet te de-escaleren. Het is belangrijk om de kandidaat uit de anonimiteit te halen door zijn/haar naam te vragen en te gebruiken in het gesprek; of om een hand te geven en je eigen naam te noemen en vervolgens te vragen wat er precies speelt. Begrip en respect voor zijn of haar situatie tonen is belangrijk. De kandidaat moet begrijpen dat de surveillant/examinator niet van de opgelegde regels kan afwijken. Dergelijke interventies helpen om de situatie tot aanvaardbare proporties terug te brengen.

Bij vooropgezette intimidatie ligt de situatie natuurlijk fundamenteel anders. De toename van dit soort incidenten vraagt om maatregelen op het gebied van preventie en op het gebied van bescherming van de surveillant. Het is belangrijk dat de branche gaat nadenken over maatregelen die de prikkels voor wangedrag wegnemen. Als de examens niet ter plekke maar op een andere plaats en op een andere tijd worden beoordeeld, wordt de surveillant niet langer gezien als beoordelaar en verdwijnt een van de prikkels om te intimideren. De surveillant/examinator zou in risicogebieden in kleine teams van twee of drie mensen kunnen werken in plaats van alleen. Als de bestrijding van fraude en intimidatie daadwerkelijk een hoge plaats op de agenda inneemt, is dit een aan te bevelen stap.

Omkoping en ondermijnend gedrag

Met enige regelmaat komen situaties voor waarin opleiders aan examinatoren geld bieden of de examenruimte zo inrichten dat er makkelijker gefraudeerd kan worden. Het is van belang dat de branche zich buigt over de mores betreffende wangedrag. Het moet duidelijk worden wat er van personeel van exameninstellingen binnen een stelsel verwacht wordt op het moment dat er wangedrag wordt gepleegd.
Daarnaast zou er een meldpunt ingericht kunnen worden dat zich erop toelegt dat meldingen juist worden opgepakt, dat advies geeft over te nemen stappen, trends en cijfers verzamelt, en extern rapporteert. Het is ook een plek waar richtlijnen en beleid ter bestrijding van wangedrag kunnen worden ontwikkeld.

☐ Mw. A. Dinkelman, Dinkelman Maatschappelijke Interventies (DMI) conflicthantering en contextanalyse. E-mail: agnesdinkelman@hotmail.com. www.dinkelman.eu.

Ook de leermeesterexamens van NBC Kennis- en adviescentrum voor de bakkerij staan onder toezicht van de Examenkamer kwaliteit door onafhankelijkheid

DE PRAKTIJK BIJ HOGESCHOOL WINDESHEIM

Werken met itembanken

Geert van der Wijk

De toenemende belangstelling voor de kwaliteit van examens en toetsen legitimeren de extra aandacht voor het werken met itembanken. Binnen Hogeschool Windesheim wordt sinds 2009 vanuit een hogeschoolbreed project ingezet op de structurele opzet, uitbouw en het onderhoud van itembanken met behulp van een toetsservicesysteem.

Meerwaarde van een toetsservicesysteem

- Eenvoudig samen bouwen en sleutelen aan een itembank
- Samenstellen van toetsen gebeurt minder ad hoc
- Geautomatiseerd nakijken en verwerken van toetsen
- Structurele aandacht voor de kwaliteit van toetsvragen met behulp van toets- en itemanalyses
- Uitwisseling van goede toetsvragen op diverse niveaus (docenten, opleidingen, domeinen, hogescholen)
- Mogelijkheden voor gebruik van multimedia (video, flash, audio, foto)
- Gebruik van innovatieve toetsvragen en verschillende vraagtypes
- Per student een unieke toets
- Resultaat direct beschikbaar

Waarom werken met itembanken?

Sinds 1 september 2010 hebben examencommissies een vergaande verantwoordelijkheid voor de kwaliteit van toetsen en examens. Uit het rapport *Boekhouder of wakend oog?* (2009) van de Onderwijsinspectie kan worden afgeleid dat examencommissies daar alle hulp bij kunnen gebruiken. Verder is toetsing veel nadrukkelijker opgenomen in het nieuwe accreditatiestelsel. Waar beoordeling en toetsing voorheen nog één van meer dan 20 verschillende facetten was, is het nu één van de drie standaarden waarop een opleiding zal worden beoordeeld.

Naast de gestegen aandacht voor kwaliteit vanuit de externe omgeving spelen ook diverse redenen voor de instelling zelf om met een toetsservicesysteem te willen werken. Met een toetsservicesysteem is er meer aandacht voor de kwaliteit van (MC) toetsvragen, terwijl het beheer van itembanken met behulp van toetsautomatisering eenvoudiger is. Bovendien ontstaan bij digitaal toetsen nieuwe mogelijkheden voor vorm, inhoud en afname.

Belangrijke mijlpalen voor implementatie

Sinds 2009 wordt vanuit een hogeschoolbreed project, op basis van een olievlekmodel, het werken met Questionmark Perception (QMP) steeds verder uitgebouwd binnen de domeinen van de hogeschool.[1] In 2009 is gestart met een paar 'early adopters' binnen de hogeschool. Eén van hen, de opleiding journalistiek, was al in 2006 begonnen met het in eigen beheer opbouwen van itembanken. Deze opleiding heeft de afgelopen jaren een belangrijke voorbeeldfunctie gehad voor de andere opleidingen.

Verder zijn vanuit het project de volgende mijlpalen bepalend geweest in de verdere uitbouw:

- Inrichting van het toetsservicesysteem op een centrale server met decentrale gebruikersrechten voor docenten en beheerders in de domeinen.
- Inrichting van een centrale beheersorganisatie (inclusief beschrijving van de processen) binnen de staande organisatie, ter ondersteuning van de gebruikers in de domeinen (o.a. technisch beheer, functioneel beheer en decentraal beheer).
- Diverse instructiebijeenkomsten en trainingen voor docenten en beheerders gericht op het maken van toetsmatrijzen, goede mc-vragen en het werken met het toetsservicesysteem.
- Ontwikkelen van een conversietool waarmee bestaande toetsvragen in Word eenvoudig kunnen worden omgezet naar het toetsservicesysteem.
- Ingebruikname van een maatwerk printtool waarmee

[1] De opleidingen van Windesheim zijn ondergebracht in vier domeinen: Economie, Management, Media en Communicatie; Gezondheid en Welzijn; Educatie en Beweging; Techniek. Daarnaast is sinds 2011 een start gemaakt met de zelfstandige Hogeschool Windesheim Flevoland.

Voor het eerst wordt een centrale screening van itembanken aangeboden

eenvoudig papieren toetsen in pdf-formaat worden gegenereerd, die geautomatiseerd kunnen worden nagekeken door het centrale toetsbureau.

Daarbij is dankbaar gebruikgemaakt van de bestaande expertise die bij veel Nederlandse onderwijsinstellingen aanwezig is. Zo is er een actieve gebruikersgroep Perception Nederland en is veel informatie verzameld door te gaan buurten bij de collega-hogescholen Arnhem-Nijmegen, Zuyd, Hanzehogeschool en Stenden.

Huidige stand van zaken

Alle inspanningen hebben er toe geleid dat binnen Hogeschool Windesheim momenteel zo'n 70 itembanken met een totaal van ongeveer 9.000 vragen in het toetsservicesysteem zijn ondergebracht. Een klein deel van de toetsen uit deze banken wordt digitaal afgenomen (circa 15%). Het grootste deel van de toetsen wordt op papier afgenomen en met behulp van schrapkaarten geautomatiseerd nagekeken. Overigens zal de verhouding tussen digitale en papieren toetsen wijzigen in verband met een groeiende behoefte aan digitale afname.

De stadia waarin alle itembanken verkeren zijn divers. Sommige itembanken zitten nog in de opstartfase en moeten nog flink worden uitgebreid, terwijl andere banken met een dekkingsgraad van 5/6 keer het aantal toetsvragen goed op orde zijn. Naast kwantiteit is de kwaliteit van de toetsvragen een belangrijk aandachtspunt. De docenten zijn beheerder van hun 'eigen' toetsbank en worden geacht de kwaliteit van de eigen bank op orde te houden. Daarbij kunnen ze gebruikmaken van toets- en itemanalyses die worden aangeleverd door het centrale toetsbureau. Als extra stimulans vanuit het hogeschoolbrede project wordt dit jaar voor het eerst een centrale screening van itembanken aangeboden. Op verzoek van bijvoorbeeld een examen- of toetscommissie kan een itembank worden doorgelicht door een toetsdeskundige.

Voorbeelden van itembanken

Itembanken opleiding Journalistiek

De opleiding journalistiek werkt met itembanken voor alle conceptuele toetsen in de Major (circa 25 itembanken). Voor elke toets heeft één docent de verantwoordelijkheid gekregen om samen met collega's de itembank op orde te houden. Docenten hebben rechten gekregen om rechtstreeks in de itembank vragen te ontwikkelen en/of wijzigingen door te voeren. Een afgevaardigde binnen de toetscommissie van de opleiding heeft het toetsservicesysteem in de portefeuille en fungeert als aanjager en vraagbaak voor de docenten. Binnen het bedrijfsbureau is een medewerker aangesteld die de docenten ondersteunt bij het genereren van toetsen, de organisatie van toetsen en die hen voorziet van relevante toets- en itemanalyses.

Wet financieel toezicht - certificering

Voor werknemers die een adviesfunctie binnen de financiële sector vervullen, gelden vanuit de Wet Financieel Toezicht deskundigheidseisen die zijn opgesteld door het College Deskundigheid Financiële Dienstverlening (CDFD). Om te voorkomen dat studenten van de business opleiding na hun afstuderen direct weer de schoolbanken in moeten, is de WFT-certificering van studenten ingebed in het curriculum. In het zogeheten EGAZ-verband (Eindhoven, Groningen, Arnhem, Zwolle) onderhouden vier hogescholen gezamenlijk een WFT-itembank opdat deze certificering voldoet aan de eisen van het CDFD.

Nederlandse taalvaardigheid

Binnen de instelling wordt een hogeschoolbreed talencentrum opgericht dat zich onder andere zal richten op de toetsing van Nederlandse taalvaardigheid bij instromende studenten. Veel opleidingen beschikken al over eigen taaltoetsen. Deze worden momenteel geïnventariseerd, geconverteerd naar QMP en vandaar uit centraal ontsloten voor de belanghebbenden. Ook op het niveau van de gebruikersgroep Perception Nederland is een initiatief gestart om toetsvragen met elkaar uit te wisselen.

Ontwikkelingspsychologie

Samen met de Hanzehogeschool en de hogeschool Arnhem-Nijmegen is een initiatief gestart om tot een gezamenlijke itembank ontwikkelingspsychologie te komen. In de samenwerking zijn afspraken gemaakt over de structuur van de itembank en de samenstelling van de toetsvragen. Dit initiatief bevindt zich nog in de opstartfase en is een voorbeeld van een serieuze poging om door een bundeling van krachten tot een goed gevulde itembank te komen.

Ambities voor de toekomst

Binnen de hogeschool zijn alle domeinen in meer of mindere mate bezig met de (op-)uitbouw van (extra) itembanken. Om te zorgen dat de ontwikkelingen op dit vlak

Het plan is om een extra omgeving voor formatieve toetsen in te richten

niet stagneren en de mogelijkheden van het toetsservicesysteem zo goed mogelijk worden benut, is het plan om de komende tijd extra aandacht te besteden aan de vergroting van de capaciteit voor digitaal toetsen, aan het uitbreiden van de mogelijkheden tot formatief toetsen en aan de koppeling met het cijferregistratiesysteem.

Door digitaal te toetsen komt een uitgebreid arsenaal aan innovatieve toetsvormen binnen handbereik en wordt het eenvoudiger om hogere cognitieve vaardigheden te toetsen. Op dit moment is de capaciteit voor digitale toetsafname beperkt tot 35 studenten op hetzelfde tijdstip. Dit komt door de beperkte aanwezigheid van computerlokalen, die specifiek zijn ingericht voor summatieve toetsdoeleinden (toetsopstelling, tussenwandjes, airco e.d.). Het streven is daarom om de toetscapaciteit verder uit te breiden met een lokaal waarin bij 100 studenten op hun eigen laptop een toets kan worden afgenomen. Deze vorm van afname past bij het instellingsbeleid om eigen laptopgebruik van studenten te ondersteunen. Ook is het een oplossing die een flexibeler lokaalgebruik mogelijk maakt, omdat lokalen met een vaste toetsopstelling een relatief lage bezettingsgraad hebben. Aan de andere kant vergt afname op studentenlaptops extra aandacht voor beveiliging, omdat een student volledig beheer heeft over de eigen PC. Questionmark heeft hier een goede oplossing voor bedacht in de vorm van de Questionmark Secure. Dit is een extra beveiliging in QMP die ervoor zorgt dat een student geen andere applicaties (mail, chat, internet e.d.) kan openen als de toets gemaakt wordt. Echter, ook dit systeem is niet waterdicht. Zo is recent door de Hogeschool Zuyd een lek in de beveiliging van qm-secure geconstateerd. Het gevolg van dit lek is dat een student een toets op afstand zou kunnen laten invullen door een medestudent. Ook kan de medestudent op afstand screenshots van de toets maken, waardoor vragen uitlekken. Dit probleem zal technisch verholpen moeten worden door de leverancier. Daarnaast is het zaak om ook bij digitale toetsen altijd aandacht te hebben voor extra fraudeverminderende maatregelen, zoals:

- Zorg voor een uitgebreide itembank, waarvan de vragen en antwoordkeuzes altijd gerandomiseerd worden aangeboden. Een student die van een uitgebreide itembank alle vragen en antwoorden uit het hoofd heeft geleerd beheerst immers ook de leerdoelen.
- Maak gebruik van een vaste groep surveillanten, met specifieke ervaring voor surveillance bij digitale toetsen. Instrueer deze groep in actuele zaken waar ze op moet letten.
- Zorg voor een scheiding tussen de summatieve en de formatieve toetsafname-omgeving.

Itembanken met een hoge dekkingsgraad en gevuld met vragen inclusief feedback zijn zeer geschikt voor formatieve toetsing. Formatieve toetsing is een waardevolle aanvulling op het leerproces, doordat de student feedback krijgt waarmee hij zijn leerproces kan bijstellen en optimaliseren. Het plan is om naast de huidige summatieve toetsomgeving, die stevig afgeschermd is, een extra omgeving voor formatieve toetsen in te richten. De toegang tot de formatieve omgeving wordt dan laagdrempelig doordat studenten vanuit een willekeurige locatie kunnen inloggen, toetsen ook benaderbaar zijn met tablet-pc's en smartphones e.d. en de omgeving geïntegreerd wordt in de elektronische leeromgeving. Docenten zijn veel tijd kwijt met het invoeren van cijfers in het cijferregistratiesysteem. Door de toetsscores geautomatiseerd vanuit het toetsservicesysteem in te lezen in het cijferregistratiesysteem wordt een administratieve last bij de docenten weggehaald, terwijl de kans op foutieve invoering van cijfers een stuk kleiner wordt.
'Er wordt te weinig met itembanken gewerkt. Met een itembank bouw je iets op, waardoor je niet steeds opnieuw hoeft te beginnen. Je kunt daardoor beter leren van je fouten. Slecht scorende vragen worden verwijderd of aangepast waardoor de kwaliteit van de itembank toeneemt', aldus Dato de Gruijter in dit tijdschrift (Molkenboer & van Berkel, 2010). De ervaringen van de hogescholen sluiten bij deze uitspraak aan: een itembank biedt nieuwe mogelijkheden die voor alle betrokkenen van voordeel zijn.

Literatuur

Inspectie van het onderwijs. (2009). *Boekhouder of wakend oog? Verslag van een onderzoek bij examencommissies in het hoger onderwijs over de garantie van het niveau.* Utrecht. http://www.onderwijsinspectie.nl/actueel/publicaties/Boekhouder+of+wakend+oog.html.

Molkenboer, H., en Berkel, H. van. (2010). Dato de Gruijter – ik wil de psychometrie toegankelijk maken, *Examens: Tijdschrift voor de Toetspraktijk,* 7(4), 24-29.

☐ De heer G.R.S. van der Wijk is als onderwijsadviseur werkzaam bij de afdeling Student- en onderwijsservices van de Hogeschool Windesheim in Zwolle. E-mail: GRS.vander.Wijk@windesheim.nl

KOPSTUKKEN UIT DE EXAMENWERELD

Cees van der Vleuten - voorstander van longitudinaal toetsen

Desirée Joosten-ten Brinke en Ad de Jongh

Op een zonnige februarimiddag hadden wij een ontspannen gesprek met Cees van der Vleuten in zijn werkkamer van de Universiteit Maastricht. Ondanks dat hij nog maar 55 jaar is, is Van der Vleuten een echt kopstuk uit de examenwereld. Hij heeft zijn sporen op dit gebied verdiend en is nog volop actief om zijn boodschap en missie uit te dragen. Belangrijke thema's zijn: toetsen gericht op de ontwikkeling van het leerproces, het belang van voortgangstoetsen en toetsen van competenties, toetsen van werkplekleren en het toetsprogramma.

Hoe en wanneer bent u geraakt door het onderwerp examens?

Van der Vleuten geeft aan dat hij van meet af aan tijdens zijn studie persoonlijkheidsleer en psychodiagnostiek interesse had voor het onderwerp examens. Meten, meetproblemen en oordeelsvorming op basis van het meten hebben altijd al zijn belangstelling gehad. Al vrij snel na zijn opleiding kwam hij in Maastricht en daar werkte hij in de specifieke onderwijssetting van de medische opleidingen met Probleemgestuurd Onderwijs (PGO) waarvoor de universiteit bij de oprichting had gekozen. Maastricht speelde in de vormgeving van PGO een voortrekkersrol. Zijn psychometrische insteek is sterk gekleurd geraakt is door PGO. Hij constateert dat zijn ontwikkeling niet puur vakmatig vanuit de psychometrie heeft plaatsgehad. Het is vooral de combinatie van werken in een onderwijssetting waarin psychometrie een belangrijke rol speelt bij het waarderen van studentenprestaties en het geven van feedback aan studenten over het leerproces dat ze doormaken.
Zijn wetenschappelijk werk wordt sterk gevoed vanuit

Prof.dr. C.P.M. van der Vleuten (1956) studeerde psychologie in Tilburg, in het bijzonder persoonlijkheidsleer en psychodiagnostiek. Na zijn studie werd hij in 1981 docent Persoonlijkheidsleer en Psychodiagnostiek aan de Katholieke Universiteit Brabant. In 1982 verruilde hij Tilburg voor de Universiteit Maastricht (UM) waar hij universitair docent werd bij de vakgroep Onderwijsontwikkeling en Onderwijsresearch. In 1989 promoveerde hij aan de UM op het proefschrift *Naar een Rationeel Systeem voor Evaluatie van Studie-prestaties in Probleem-gestuurd Medisch Onderwijs: Studies naar Betrouwbaarheid en Validiteit van Vaardigheidstoetsen*. In datzelfde jaar werd hij bij de vakgroep universitair hoofddocent en in 1996 hoogleraar. Kortom, hij is al 30 jaar aan de UM verbonden. In zijn loopbaan heeft hij een omvangrijk oeuvre opgebouwd. Zijn publicatielijst telt honderden titels. Hij heeft bijna 50 promovendi begeleid bij het schrijven van een dissertatie. Hij adviseert zowel binnenlandse als buitenlandse organisaties. Vanaf 2005 is hij ook Wetenschappelijk directeur van de UM Graduate School of Health Professions Education.
Zie verder: http://www.fdg.unimaas.nl/educ/cees/cv/CV%20Cees.PDF.

Toetsen hebben een functie bij het innoveren van het onderwijs

de praktijk van het onderwijs en het leren zoals dat in zijn omgeving plaatsvindt. Toetsen beschouwt hij altijd in het licht van wat je met je onderwijs bereiken wil en hoe toetsen daaraan ondersteunend kunnen zijn. Traditionele toetsing en het klassieke systeem van examens passen niet goed in PGO. Immers in PGO wordt onder andere uitgegaan van actief leren, zelfsturend leren, en authentieke taken. En wanneer je in het onderwijs uitgaat van authentieke taken komen die op de één of andere wijze ook in het toetsen terug. Van der Vleuten wil eigenlijk niet over het systeem van examens spreken, maar liever spreekt hij over toetsen en beoordelen waarvan examens een onderdeel zijn. Zijn ontwikkeling en zijn denken worden steeds gevoed door het innoverend onderwijs. En dat koppelen aan empirie, aan onderzoeksbevindingen. "Toetsen hebben een functie bij het innoveren van het onderwijs. Daarin hebben we heel wat slagen gemaakt" zegt hij.

Wie zijn uw leermeesters geweest?

Zonder aarzelen noemt Van der Vleuten Wynand Wijnen als zijn leermeester. Iemand met een hele sterke denklijn, heel erg begaan met innoveren in het onderwijs. In de verschillende periodes van zijn leven heeft hij wel meer leermeesters gehad. Hij noemt David B. Swanson van de National Board of Medical Examiners, zeg maar het Cito van de geneeskunde in de VS. Op het gebied van data-analyse was Swanson zeer sterk, "hij kon datasets laten spreken". Zijn uitgangspunt was dat alle interpretatie datageoriënteerd moet zijn. Op het gebied van theorievorming heeft hij veel geleerd van Geoff Norman (McMaster University Clinical Epidemiology & Biostatistics), een kernfysicus die een hele grote naam is geworden op medisch-onderwijskundig gebied. Hij gebruikte de psychologische kennis als basis voor voortdurende theorievorming. "Data waren bij hem bijna secundair: als je de theorie rond hebt, dan is het slechts een kwestie van de juiste gegevens te verzamelen. Beide personen vormden dus een mooi leerzaam contrast."

Op de vraag of hij nog anderen wil noemen, denkt hij na en zegt in alle bescheidenheid dat hij voortdurend leert van zijn medewerkers, zijn promovendi en zijn studenten. Door de praktijk wordt hij scherp gehouden. En dat typeert Van der Vleuten: toetsen relateren aan innovatie. Met andere woorden: al diegenen die meewerken aan het verduidelijken en het verhelderen van deze relatie zijn leermeesters voor hem.

Wat drijft u?

De drijfveer voor al zijn activiteiten is het optimaliseren van het onderwijs. En dat doe je door middel van onderzoek. Onderzoeksresultaten onderbouwen de onderwijskundige verbeteringen. Duidelijk met een link naar de praktijk. Van der Vleuten vindt dat het optimaliseren van onderwijs door middel van het toepassen van onderzoeksresultaten zeker in de afgelopen decennia vruchten heeft afgeworpen. Het werken met PGO is van grote invloed geweest op onderwijskundige vernieuwingen. "PGO of elementen daaruit zijn terug te vinden in bijvoorbeeld het studiehuis en het competentiegericht onderwijs. Daarin hebben wij een enorme professionalisering gerealiseerd. Die is mede gestoeld op empirisch onderzoek. De combinatie met de praktijk, dat motiveert."
Wat Van der Vleuten wil is het samenbrengen van onderwijs en onderzoek: "De studenten zijn wel de beste ambassadeurs voor wat wij doen. Zij overtuigen."

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

In zijn visie op examens onderscheidt Van der Vleuten vier ontwikkelingslijnen:

1. Van momentaan naar longitudinaal toetsen. Toetsen wordt in de klassieke opvatting gezien als een momentopname. Op het moment dat de student de toets aflegt moet hij zijn leerstof beheersen. Van der Vleuten hecht veel meer belang aan longitudinaal toetsen, waaronder hij verstaat dat er voortgangstoetsen worden afgenomen. "Als je alleen cursusgerelateerd toetst, mis je een stuk dat je met longitudinaal toetsen wel dekt. Ook op basis van voortgangstoetsen kun je oordelen uitspreken, maar nog belangrijker is het geven van feedback op het leerproces."
2. Toetsen is meer dan het beoordelen van kennis alleen, daarvan is Van der Vleuten al vanaf zijn studietijd overtuigd. Zijn proefschrift ging niet voor niets over 'stationstoetsing', het in een laboratorium zetten van authentieke, realistische situaties. Toetsen van kennis is van belang, maar toetsen van vaardigheden en competenties is natuurlijk - zeker

Voortgangstoetsen geven feedback op het leerproces

voor de toekomstige beroepsuitoefening - net zo belangrijk.

3. Toetsen is een manier om het informatiegehalte over het leerproces te verhogen. "In het onderwijs is men vaak te reductionistisch bezig: de student krijgt een stuk onderwijs en dat sluit hij af met een examen waarvoor hij zakt of slaagt. Alle informatie is daarna weg. Als hij slaagt, dan is hij eigenlijk 'immuun' voor het leven, want dan gaat hij door naar de volgende cursus. Slaagt hij niet, dan doet hij de toets of de cursus over, maar ook dat is een tamelijk informatieloze, diagnostisch arme, reductionistische toetspraktijk." Van der Vleuten constateert dat we in het onderwijs zeer slecht zijn in het feedback geven, terwijl toetsen enorm veel informatie geven die voor het geven van feedback van belang is. Hij pleit dan ook voor het "informatierijk maken van toetsen", voor feedbackrijke toetsen.
4. Toetsing als een programma. Van der Vleuten stelt dat de literatuur gedomineerd wordt door onderzoek naar individuele instrumenten. Op elk instrument is wel het één of ander aan te merken. "Je kunt niet altijd met één instrument alles bereiken dat je graag zou willen bereiken. Ook bij toetsen moet je compromissen sluiten en liefst het compromis dat maximaal past bij het doel van je toets." Hoe Van der Vleuten het toetsprogramma vorm wil geven is te lezen in het artikel dat hij samen met Schuwirth voor *EXAMENS* heeft geschreven en dat in dit nummer is opgenomen.

Voor Van der Vleuten blijft het optimaliseren van het leren een zeer belangrijk punt. "Je moet de regie voeren over hoe er geleerd wordt en over de toetsen die je afneemt. Je moet informatie longitudinaal verzamelen van meerdere 'puntmetingen'. Elke puntmeting heeft zijn beperking, maar is maximaal afgestemd op het verzamelen van rijke informatie. De verzameling van puntmetingen levert een rijk longitudinaal beeld op. Je moet dus denken in hele toetsprogramma's en dan niet alleen in onderwijsprogramma's, maar ook in toetsprogramma's. Beide programma's moeten hand in hand gaan. De toetsen moeten dus heel dicht tegen het onderwijs aan liggen. En op beide moet je regie houden. Alle afzonderlijke toetsen zijn belangrijke metingen, maar tegelijkertijd beperkt in waarde. De verzameling van metingen zorgt echter voor de rijke informatie om robuuste beslissingen te kunnen nemen."

Toetsen is gericht op ontwikkeling van het leerproces (longitudinaal toetsen) in tegenstelling tot de puntmeting (een meting op één moment waarop je een oordeel velt). Longitudinale toetsing is mede gericht op het geven van feedback op het leerproces. Het feedbackproces moet volgens Van der Vleuten ondersteund worden. Hij wil zo min mogelijk de toetsen buiten het onderwijs zetten. "Druk van buiten om kennis te toetsen is prima, maar laat dat niet door een extern instituut gebeuren. Bouw je eigen expertise op en zorg dat de toetsing iets van jezelf is."

Desgevraagd gaat Van der Vleuten nader in op de relatie docent-examen. "Ik ben TEGEN de ontkoppeling van examens en de docent. Maar dat betekent niet dat alles subjectief moet blijven. Ik ben ook voor objectieve toetsen die de docent en de klas overstijgen. Maar ook bij dergelijke toetsen ben ik voor een sterke betrokkenheid van de docent. De verantwoordelijkheid voor (het bewijzen) van de kwaliteit zie ik graag dicht bij het eigen instituut en de eigen docenten. Een externe partij kan je mogelijk iets opleggen dat je onderwijs de verkeerde kant opstuurt. Een voorbeeld is hoe we in de geneeskunde voortgangstoetsen hebben geïntroduceerd die volledig door samenwerkende opleidingen tot stand komen. Er is volledige transparantie. We weten precies hoe we het doen. We kunnen verantwoorden. We kunnen onderzoek doen. En we zijn eigenaar van het hele proces."

Waarmee bent u nu bezig?

Tot nu toe hebben Van der Vleuten en zijn collega's zich voornamelijk gericht op het binnenschoolse leren. Op dit moment houdt hij zich ook bezig met het leren en beoordelen op de werkplek. In de medische opleidingen is leren op de werkplek uitermate belangrijk. Werkplekleren vraagt een bepaalde aanpak wat betreft toetsen en beoordelen. "Je ontkomt niet aan subjectieve holistische beoordelingen waarin de feedback op het werkplekleren centraal staat. Je beoordeelt dus het werkplekleren als een geheel en wel vanuit je eigen waarneming. Het gaat daarbij om de rol die je als beoordelaar van werkplekleren moet vervullen. Het is

Zorg dat de toetsing iets van jezelf is

dan niet alleen een kwantitatief maar ook een kwalitatief oordeel. Dat maakt de informatie die je verkrijgt veel rijker, waardoor je besluiten beter kunt funderen." Het hard maken van 'zachte informatie' gaat niet door alles te objectiveren en te standaardiseren, dat leidt nogal eens tot trivialisering. Het hard maken van zachte informatie is mogelijk door allerlei procedurele maatregelen in te bouwen in de beoordelingsprocedure die het holistisch oordeel als zodanig intact laat.
Voor Van der Vleuten is een belangrijke vraag wanneer je kunt ophouden met het verzamelen van informatie waarop je een oordeel gaat vormen. Het is interessant om hier weer formele modellen te gebruiken die kunnen omgaan met complexe informatie en aangeven of nieuwe informatie nog iets toevoegt aan bestaande informatie. En daar zal de psychometrie of de statistiek weer behulpzaam kunnen zijn.

Wat ziet u als het hoogtepunt in uw loopbaan?

Van der Vleuten vindt dat zijn kracht ook zijn zwakte is. Hij zit met heel veel publicaties in het gezondheidszorgdomein. Wat hij niet gedaan heeft, is buiten dit domein treden. In het gezondheidszorgonderwijs heeft hij veel bereikt, er is samenwerking tussen onderzoekers en onderwijsmensen, er is een gemeenschap gecreëerd waarin uitwisseling plaatsvindt. "De kloof tussen onderwijsonderzoek en praktijk bestaat bij de gezondheidswetenschappen niet."
Hij ziet de transfer van al zijn activiteiten in het gezondheidszorgonderwijs naar andere sectoren van onderwijs als een hoogtepunt, dat zou hij graag realiseren. Hij zegt: "Ik blijf leren, ik blijf me ontwikkelen. Ik ben nog niet klaar en wil mijn klus afmaken."

Waar bent u het meest trots op?

Van der Vleuten hoeft over deze vraag niet na te denken. Hij is trots op zijn navolgers. Zijn navolgers die anderen kunnen coachen om ze te bekwamen voor het geven van onderwijs dat is gebaseerd op evidentie. Hij is trots op zijn promovendi en ook op de docenten. Hij raadt zijn navolgers aan om hun eigen onderwijs te problematiseren en dan iets te gaan doen aan die problemen. Daar moeten veranderaars uit voortkomen, de leiders voor de toekomst.
Hij is ook trots op het programmatisch denken over toetsen en beoordelen. "Je krijgt zaken voor elkaar door met elkaar het debat in te gaan, elkaar te overtuigen, voortdurend te zoeken naar nieuwe aanknopingspunten."

Heeft u schoolgemaakt?

De door hem ontwikkelde visie op examineren maakt school. De beoordeling op de werkplek is een nieuw terrein waarop hij naam maakt. Er is grote behoefte aan feedback op het werkplekleren.

Wie beschouwt u als vooraanstaand in de examenwereld?

Van der Vleuten kan of wil geen namen noemen. Uit zijn eigen omgeving kunnen wel een aantal van zijn promovendi als vooraanstaand worden geschouwd.

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

Hij hoopt dat de reductionistische praktijk veranderd wordt. Volgens Van der Vleuten zal longitudinale toetsing in de toekomst erg belangrijk worden. Die vorm van toetsing zal het leren bevorderen zonder daarbij de beslisfunctie uit het oog te verliezen. Beoordelen zal steeds meer gebaseerd worden op gegevens uit die longitudinale toetsing.

Welke boodschap geeft u aan de lezers van *EXAMENS*?

"Verdiep je in de problematiek van toetsen en beoordelen. Werk niet vanuit intuïtie en traditie. Er is meer dan de traditionele methoden waaruit je inspiratie kunt krijgen. Kijk naar de beschikbare technologie. Er is heel veel technologie om hogere cognitieve vaardigheden of authentiek te kunnen toetsen. Daar weten docenten nog veel te weinig van. Gebruik de kansen die zich in het onderwijs voordoen, niet alleen in het wetenschappelijk onderwijs, maar ook bijvoorbeeld in de lerarenopleiding en in het middelbaar beroepsonderwijs. Ontwikkel in gezamenlijkheid. Ga niet uit van opgelegde structuren. Zorg ervoor dat je de kwaliteit van je onderwijs op orde hebt wanneer je verder gaat ontwikkelen."

☐ Mw. dr. D. Joosten-ten Brinke en de heer drs. A.J.C.M. de Jongh zijn redacteur van *EXAMENS*. E-mail: Desiree.Joosten-tenBrinke@ou.nl.

Gezien en gelezen

Desirée Joosten-ten Brinke

Sanders, P. (red.) (2011). Toetsen op school. Arnhem: Cito. 190 pp. Te downloaden op: www.toetsenopschool.nl, gratis

Het boek ***Toetsen op school*** is eigenlijk geen boek, maar een website die Cito ontwikkeld heeft om een bijdrage te leveren aan beter onderwijs, een cadeau aan de lerarenopleidingen. De kennis en kunde van docenten op het gebied van toetsen en beoordelen kan vergroot worden met behulp van dit boek. Dat het boek niet als boek verschijnt, maar als website heeft het voordeel dat bijstelling mogelijk is als daar aanleiding voor is. Er zijn of komen nog aanvullende publicaties met specifieke informatie voor primair, voortgezet en middelbaar beroepsonderwijs.
Toetsen op school is geschreven vanuit de gedachte dat docenten ten aanzien van vijf aspecten van toetsen deskundig moeten zijn:

1. Docenten moeten weten dat toetsen verschillende doelen hebben;
2. Docenten moeten toetsen kunnen ontwikkelen;
3. Docenten moeten weten waar ze informatie over toetsen kunnen vinden;
4. Docenten moeten toetsresultaten kunnen beoordelen en interpreteren, en
5. Docenten moeten de kwaliteit van toetsen en examens kunnen beoordelen.

In het eerste hoofdstuk wordt door Sanders het doel van toetsen beschreven. Hierbij wordt een onderscheid gemaakt naar toetsen voor het beoordelen van leerlingen, toetsen voor het beoordelen van het onderwijsleerproces, toetsen voor het beoordelen van groepen leerlingen en scholen en toetsen voor het beoordelen van de kwaliteit van het onderwijs. Roelofs en Visser richten zich daarna in hoofdstuk 2 op de stappen die gezet moeten worden om te komen tot het samenstellen van een toets. Hoe kom je van een visie op toetsen via einddoelen en tussendoelen naar toetsbare leerdoelen en hoe kan je deze omzetten in betrouwbare en valide toetstaken, de juiste keuze van toetsvorm en de uiteindelijke samenstelling van de toets.

Hoofdstuk 3 en 4 zijn twee hoofdstukken die op een begrijpelijke manier twee kernconcepten van toetsen en beoordelen uitleggen: betrouwbaarheid en validiteit van toetsscores.
Het vijfde hoofdstuk is een vreemde eend in de bijt. Meijer beschrijft de bronnen die docenten kunnen raadplegen om meer te weten te komen over toetsen en beoordelen. Een heel mooi overzicht dat beter aan het einde van het boek geplaatst had kunnen worden. Opvallend is de verwijzing naar Google (p. 95) alsof docenten wezens van een andere planeet zijn. Hoofdstuk 6 en 7 respectievelijk geschreven door Noijons en Erkens gaan in op het construeren van gesloten en open vragen. Zeer praktische hoofdstukken met handige voorbeelden. Het geven van feedback door docenten had bij het corrigeren van open vragen meer aandacht mogen krijgen. Kuhlemeier beschrijft in hoofdstuk 8 wat praktijktoetsen zijn en hoe deze ontwikkeld en beoordeeld moeten worden. Hoofdstuk 9 behandelt het beoordelen van toetsscores. De problematiek van normeren, relatief of absoluut, het vaststellen van de cesuur en de score-cijfer-transformatie worden beschreven.
Hoofdstuk 10 heeft als titel 'De kwaliteit van toetsen en examens'. Een hoofdstuk met deze titel had ik aan het begin verwacht bij een boek dat bedoeld is om de deskundigheid van docenten te verhogen. Bij het lezen blijkt het echter niet te gaan om kwaliteit van toetsen in het algemeen, maar gaat het in op twee beoordelingssystemen, Cotan en het beoordelingskader van het mbo.
Toetsen op school is een boek dat zeker een bijdrage kan leveren aan de deskundigheidsbevordering van docenten. Lezers moeten daarbij wel genegen zijn om over het hoge Cito-gehalte heen te lezen. Daarnaast is toetsen en beoordelen op school ook veel breder dan het toetsen zoals dat in dit boek behandeld wordt. De formatieve functie van toetsen komt in het boek slechts minimaal aan de orde. Vanuit het oogpunt van Cito is dat logisch, maar voor de doelgroep waar het boek voor geschreven is, is dat ontbrekende informatie.

Mw. dr. Desirée Joosten-ten Brinke is redacteur van *EXAMENS*, lector Eigentijds toetsen en beoordelen bij Fontys lerarenopleiding Tilburg en universitair docent bij de Open Universiteit. E-mail: desiree.joosten-tenbrinke@ou.nl; d.tenbrinke@fontys.nl.

Ook de leermeesterexamens van Svh staan onder toezicht van de Examenkamer
kwaliteit door onafhankelijkheid

Literatuur

Assessment reform in education policy and practice
R. Berry & B. Adamson
Dordrecht: Springer, 2011
ISBN: 9789400707283

Centraal ontwikkelde examens Nederlandse taal en rekenen in het mbo, Hoe? Zo!
M. Bijleveld
Zoetermeer: Stichting Kennisnet, 2011

Educational assessment: measuring progress in schools
O.T. Ricoha
New York: Noca Science Publishers, 2011
ISBN: 9781608768189

Examen survival gids
S. Haarhuis
Amersfoort: Thieme-Meulenhoff, 2011
ISBN: 9789006076332

Making assessment matter
G. Butt
London: Continuum International Publishing Group, 2010
ISBN: 9781847063830

Manual for language test development and examining for use with the CEFR
Strasbourg: Language Policy Division Council of Europe, 2011

Meting taal en rekenen 2010; tweede meting: een indicatie van leerprestaties in termen van het referentiekader
Arnhem: Cito, 2011

Normering bij de centrale examens in het voortgezet onderwijs
Arnhem: Cito; Utrecht: CvE (College voor Examens), 2011 (online: www.toetswijzer.nl)

Practical language testing
G. Fulcher
London: Hodder Education, 2010
ISBN: 9780340984482

The effects of educational systems, school-composition, track-level, parental background and immigrants origins on the achievement of 15-years old native and immigrant students: a reanalysis of PISA 2006
A. Dronkers, R. van der Velden & A. Dunne
Maastricht: ROA, 2011

Toetsen op school
P. Sanders (red.)
Arnhem: Cito, 2011

Zakboek toetsen beoordelen in zorg en welzijn: praktische handleiding voor toetsen en beoordelen in competentiegericht onderwijs
N. van Halem, T. Stuur & H. Verbeek
Houten: Bohn Stafleu Van Loghum, 2010
ISBN: 9789031384464

De invloed van het ERK op de toetsing van vreemde talen
E. van Hest
Levende talen tijdschrift 2011 (2): 25-32

De rekenvaardigheid van de Nederlandse verpleegkundige
T. Lampe, G.Straetmans & Th. Eggen
Onderwijs en Gezondheidszorg 2011 (3): 3-9

Meester van mens en wereld: het niveau van eerstejaars pabostudenten voor zaakvakken
H. Notté & R. Baltus
Tijdschrift voor lerarenopleiders 2011 (2): 27-29

Agenda

3nd Annual European-Association of Test Publishers (E-ATP) Conference, 28-30 september 2011, Praag.
Derde Europese conferentie met als thema: Growing talent in Europe: gaining advantage through assessment. (www.eatpconference.eu.com)

2nd Annual International Association for computerized adaptive testing (IACAT), 3-5 oktober 2011, Pacific Grove, California (USA).
(www.iacat.org)

RCEC Workshop on IRT and Educational Measurement, 12-14 oktober 2011, Enschede.
(www.rcec.nl/cursussen)

IAEA (International Association for Educational Assessment), 23-28 oktober 2011, Manila.
Jaarlijkse conferentie met dit jaar als thema: 'Assessment and the challenge of globalization'. (www.iaea2011.com)

AEA-Europe Annual Conference, 10-12 november 2011, Belfast.
Thema: Assessment and opportunity to learn. (www.aea-europe.net)

Toetskwaliteit in het hoger onderwijs, 3 en 17 november, 1 december 2011, Utrecht.
Driedaagse cursus. (www.sbo.nl)

NVE (Nederlandse Vereniging voor Examens), Achtste nationale examencongres, 17 november 2011, Papendal/Arnhem.
Tweejaarlijks nationaal examencongres met dit jaar als thema: 'Competentiegericht toetsen en beoordelen … hobby of noodzaak?'. (www.nvexamens.nl)

17th International Conference on Technology Supported Learning and Training, 30 november-2 december 2011, Berlijn.
Belangrijkste bijeenkomst over ICT en onderwijs in Europa. (www.online-educa.com)

De redactie heeft afscheid genomen van redactielid drs. Ed Kremers die sinds de oprichting van het tijdschrift zitting in de redactie heeft gehad. Het tijdschrift heeft veel te danken aan Ed Kremers. Hij verzorgde de rubrieken Literatuur en Agenda, hij bracht vele artikelen aan en zijn kritische en toetsdeskundige inbreng bij de bespreking van de kopij legde de basis voor de gewenste kwaliteit van ons blad. Kremers wordt per september 2011 opgevolgd door drs. Alex van de Kerkhof.

In het volgende nummer van *EXAMENS* komen onder andere de volgende onderwerpen aan bod:

- Studietoetsen: Zinvolle inhoud of betrouwbare score?
- Digitaal Toetsen op de HBOV
- Examens contactlensonderwijs
- Kopstukken uit de examenwereld: Pieter Drenth

EXAMENS 4, 2011 verschijnt in november

Het NVE-congres 2011

Op 17 november 2011 wordt het achtste Nationaal Examencongres gehouden. Het thema van dit congres is 'competentiegericht toetsen en beoordelen: hobby of noodzaak?'. Het thema wordt benaderd vanuit de invalshoek van het hoger onderwijs, het middelbaar beroepsonderwijs en het bedrijfsleven. Een zeer actueel thema, zeker voor het middelbaar beroepsonderwijs. Al een aantal jaren wordt geëxperimenteerd met competentiegericht onderwijs en competentiegericht leren en dus ook met competentiegericht toetsen. Het competentiegericht onderwijs zou ook vastgelegd worden in de Wet Educatie en Beroepsonderwijs. Maar competentiegericht onderwijs staat onder druk.
De politiek, met name de Tweede Kamer heeft er meer dan gewone belangstelling voor. Vaak wordt competentiegericht onderwijs op één hoop gegooid met het nieuwe leren. En dat laatste heeft zeker in de huidige politieke constellatie een negatieve klank.
Je kunt je afvragen hoelang competentiegericht onderwijs de kans krijgt om volledig tot ontwikkeling te komen nu de minister heeft besloten om het woord competentie niet op te nemen in de Wet Educatie en Beroepsonderwijs. Volgens haar gaat het om kennis en vaardigheden in het onderwijs. Toch een beperkte opvatting: houdingen en motivatie spelen blijkbaar geen rol meer. Waar bestaan competenties dan uit? Alleen kennis en vaardigheden? Of toch ook uit attitudes? Kortom het congres heeft een boeiend thema waarvan waarschijnlijk ook de minister en politici nog het nodige kunnen leren alvorens relevante besluiten, die het onderwijs in hart treffen, genomen zullen worden.

De toetsbaarheid van integriteit
Gaat het in het onderwijs alleen om kennis en vaardigheden? Of toch ook om attitudes. Dik van Velzen, consultant bij NIBE-SVV, stuurde ons tijdschrift een opiniërend artikel over de toetsbaarheid van integriteit. Hij stelt dat de minister van Financiën het plan heeft opgevat om 'integriteit' toe te voegen aan de exameneisen voor financiële adviseurs (zie kader). Hij vraagt zich af of het mogelijk is om integriteit te toetsen. Als het aan de minister ligt, wordt integriteit een voorwaarde voor het diploma dat financiële adviseurs verplicht moeten gaan afleggen.

Integriteit is natuurlijk een mooie doelstelling die overigens voor meerdere beroepen geldt. Als je integer wilt handelen, in welk beroep dan ook, moet je veel weten, maar dat geeft geen garantie dat je integer zult handelen in de beroepssituatie. Gaat het om een vaardigheid of gaat het om een professionele houding? Operationalisatie zal nog het nodige vragen.

Hoe moet je integriteit toetsen? Een vraag waar toets- en examenconstructeurs zich in vast kunnen bijten. Een uitdaging voor de leden van de NVE en natuurlijk ook voor de NVE zelf. Leden worden uitgenodigd hun mening over het kunnen en mogen examineren van integriteit kenbaar te maken. Stuur een e-mail naar redactielid Ad de Jongh, e-mail addejongh@home.nl, wij zijn benieuwd.

Minister De Jager van Financiën heeft op 2 mei 2011 aan de Tweede Kamer meegedeeld dat hij voornemens is om het vakbekwaamheidsstelsel voor financiële dienstverleners aan te scherpen door de volgende maatregelen: 1e invoeren van een diplomaplicht voor iedere adviseur; 2e instellen van een register waarin alle adviseurs zijn opgenomen, 3e aanpassingen van de wettelijke vakbekwaamheidsmodules en examinering. Wat dat laatste betreft schrijft hij:
'Daarnaast ben ik van plan de vakbekwaamheidsmodules aan te vullen met de elementen vaardigheden en professioneel gedrag, waarbij ook aandacht zal zijn voor ethiek en integriteit van de adviseur. Hierbij zal ik per module kijken welk niveau van vaardigheden en professioneel gedrag gewenst is. In een aantal gevallen is ook het kennisniveau van de modules te laag. Waar nodig zal ik het kennisniveau van de modules gericht verhogen.'

Bron: Vakbekwaamheid financiële dienstverleners; Kamerstuk : Kamerbrief | 02-05-2011 | Financiën

NVE website
De NVE website is vernieuwd. Ga naar http://***www.nvexamens.nl*** en lees interessante nieuwtjes. Er is onder meer een forumpagina waarop ALLE NVE-leden worden aangemoedigd om een korte introductie over zichzelf te posten: 'Leer elkaar kennen en deel uw interesses'. En er is een Ideeënbus: 'Heeft u feedback of ideeën voor de NVE die u wilt delen? Aarzel niet en laat het ons weten. We horen het graag van u en streven er naar onze site beter en gebruikersvriendelijker te maken voor onze gasten en leden'.

Het doel van de NVE is
De kwaliteit van de examenpraktijk te bevorderen;
Normen en gedragscodes voor het examineren te ontwikkelen;
Vakbekwaamheid en deskundigheid in de examenpraktijk te stimuleren;
Er te zijn als netwerkorganisatie en informatiebron.

Geïnteresseerd in toetsen en in de toetspraktijk? Dan bent u toch ook lid van de NVE
NVE, Nederlandse Vereniging voor Examens,
Postbus 217, 7500 AE Enschede
Tel.: 053-489 36 16 Fax: 053 - 489 42 39
e-mail: nve@edte.utwente.nl
website: www.nvexamens.nl

Andriessen
toetsen voor de toekomst

Start nu een efficiënt en online toetsproces

ISO gecertificeerd
Processen & Informatiebeveiliging

Neem vandaag nog contact op met Andriessen:

Surf naar www.andriessen.nl
of bel/mail met Piet Philipsen:
06-4616 5401 of
piet.philipsen@andriessen.nl

- Efficiënter toetsproces
- Toetsen onafhankelijk van locatie & tijd
- Netwerk toetslocaties
- Gecertificeerde toezichthouders
- Inzicht in kwaliteit toetsen
- SAAS oplossing