

2010

03

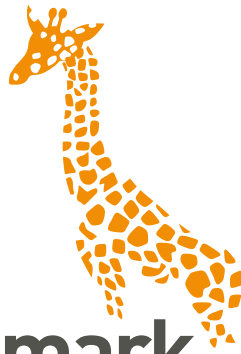
jaargang 7

EXAMENS

TIJDSCHRIFT VOOR DE TOETSPRAKTIJK

Spelling, grammatica en interpunctie meebeoordelen op eindexamens?

- Peer assessment in het basisonderwijs
- Optimale uniforme scoringsregels voor innovatieve vraagvormen
- Kopstukken uit de examenwereld: Norman Verhelst



lamark

secure online examination

www.lamark.com | examination@lamark.com



REACH OUT TO YOUR CANDIDATES

lamark biedt uw kandidaten

- Flexibele examenmogelijkheden
- In beveiligde Test Centers door heel Nederland
- Op het moment van hun keuze

→ www.lamark.com • +31 (0)88 5141 777

en wat levert u dit op?

- **Examens binnen 1 dag beschikbaar**
- **Volledige bescherming van intellectueel eigendom**
- **Online boeking- en betalingsmogelijkheden**
- Continu inzage van examenvoortgang
- Op maat communicatie met kandidaten
- Kandidaat- en certificaatmanagement
- Uitgebreide managementinformatie & rapportage
- Wereldwijd Test Center Netwerk

Inhoud

Spelling, grammatica en interpunctie mee-beoordelen op eind-examens? 5

Ben Wilbrink, Denny Borsboom en Michel Couzijn

Begin april van dit jaar is in de Tweede Kamer een motie ingediend die de regering vraagt te bevorderen dat 'correcte spelling en grammatica op enigerlei wijze meetellen in de beoordeling van elk centraal examen-onderdeel'. De indieners, Jan Jacob van Dijk (CDA) en Jasper van Dijk (SP), zagen dit als de logische sluitsteen op de wens van de commissie-Meijerink dat taal en rekenen niet alleen de zorg van de docenten Nederlands respectievelijk wiskunde zijn, maar van alle docenten.



Peer assessment in het basisonderwijs - Het effect van peer assessmenttraining op het geven van een spreekbeurt 13

Kelly Meusen-Beekman en Desirée Joosten-ten Brinke

Een kritische denkhouding, capaciteit om eigen werk te evalueren en goed ontwikkelde communicatieve vaardigheden: dit zijn eisen die aan mensen in de hedendaagse maatschappij worden gesteld. Het basisonderwijs kan hierop inspelen door kinderen



te leren reflecteren op zichzelf en anderen, en kritisch te zijn ten aanzien van eigen ontwikkelingen en vaardigheden. Peer assessment, een toetsvorm waarbij personen elkaar beoordelen, biedt deze mogelijkheid. Dit artikel laat zien dat peer assessment in het basisonderwijs veelbelovend is.

Optimale uniforme scoringsregels voor innovatieve vraagvormen 21

Hans Vos, Marlies Kloppenburg en Onno Tomson

Gebruikers van toetssystemen rekenen erop dat deze betrouwbaar zijn en uitgaan van de optimale, uniforme scoringsregels voor innovatieve vraagvormen. De discussie over de quotiëntregel en de ontwikkeling van de productregel voor scoring van multiple responsvragen, waarbij het aantal goede antwoorden niet wordt weergegeven, heeft eerder in dit tijdschrift laten zien waar knelpunten liggen. Dit artikel gaat verder in op scoringsregels. Hans Vos houdt zich aanbevolen voor reacties.

Kopstukken uit de examenwereld: Norman Verhelst 25

Annemarie de Knecht-van Eekelen en Ed Kremers

Prof.dr. Norman D. Verhelst heeft internatio-

nale bekendheid gekregen met het ontwikkelen van het One Parameter Logistic Model (OPLM). In 1985 trad hij als wetenschappelijk medewerker bij de afdeling psychometrisch onderzoek in dienst van Cito waar hij tot op heden werkzaam is gebleven. Vanaf 1992 was hij gedurende tien jaar bijzonder hoogleraar Psychometrie aan de Universiteit Twente bij de vakgroep 'Onderzoeksmethodologie, Meetmethoden en Data-Analyse' van de faculteit Toegepaste Onderwijskunde. Hij is als adviseur betrokken bij een groot aantal internationale projecten op het gebied van toetsing.



Rubrieken

Redactioneel	4
Terecht of niet ...?	10
Gastcolumn	17
Uit de praktijk	18
Gezien en gelezen	30
Literatuur	31
Agenda	31
Wie zijn wij?	32
NVE	34

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

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. Helderma, Inspecteur BVE
 Onderwijsinspectie
 Dhr. S. Jurna, beleidsmedewerker
 Ministerie OCW
 Drs. R. Duvekot, Hogeschool InHolland
 Mw. drs. A.J.M. Bakker, directeur Paepo
 Prof. dr. P. Sanders, Cito
 Drs. A.J.M. Luijten
 Mw. Drs. P. Willemse, loopbaanontwikkeling
 en competentiegericht beoordelen, Cinop

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 parti-
 culieren €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. Beeën-
 diging van het abonnement kan uitsluitend
 schriftelijk voor 1 september van het
 lopende abonnementsjaar. Bij niet tijdige
 opzegging wordt het abonnement automa-
 tisch 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 Auteurs-
 wet 1912 en in het KB van 20-7-1974 (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-vanEekelen

Redactioneel

Waardevastheid van diploma's

In Nederland is het vertrouwen in de examinering in het voortgezet onderwijs hoog. Dat betekent dat de waardevastheid van diploma's hoog is, dat leerlingen met een diploma zonder meer kunnen doorstromen naar andere onderwijsvormen of naar de arbeidsmarkt. Maar het vertrouwen in de waarde van het diploma staat onder druk. Niet voor niets heeft dit jaar de demissionair staatssecretaris van Onderwijs, Marja van Bijsterveldt, besloten dat de resultaten van het centraal schriftelijk eindexamen zwaarder gaan tellen. Vanaf het examenjaar 2012 kunnen leerlingen in het hele voortgezet onderwijs alleen slagen als het gemiddelde van de cijfers die zijn behaald bij het centraal schriftelijk voldoende is.

Deze beslissing moet van invloed zijn op de zak/slaagregeling, want – zo zegt de staatssecretaris – 'De lat moet omhoog, dat komt de waarde van het diploma en daarmee de kwaliteit van het onderwijs ten goede'. Zal de lat inderdaad omhoog gaan? Op het eerste oog lijkt dat logisch, maar dan is er geen rekening gehouden met de wijze van correctie van de centrale eindexamens. In Nederland kijkt de leraar van de eigen school de examenwerken van haar/zijn leerlingen na en daarna volgt een tweede correctie door een leraar van een andere school. Het cijfer hangt dus in hoge mate af van de eigen leraar.

Als ik dat aan buitenlandse collega's vertelde, kregen ze mij vol ongeloof aan. De eigen leraar kan toch nooit objectief beoordelen, was de reactie. En dan lieten ze corruptie nog buiten beschouwing. In zoveel landen is het leraarsalaris dermate mager, dat enkele steekpenningen maar al te makkelijk worden geaccepteerd. In de meeste landen worden groepen leraren getraind om als corrector op te treden. Zij kijken grote aantallen werken van hun onbekende leerlingen na. Hun scores worden besproken in vergaderingen van de examencommissies waar cesuren en cijfers worden vastgesteld.

De objectiviteit is gewaarborgd, maar nadelen van een dergelijk systeem zijn de hoge kosten en de lange tijd voordat de resultaten bekend zijn. Er liggen toch snel twee maanden tussen examen en uitslag.

In Nederland gaat het dus anders, maar zijn we daar tevreden mee? Examenprocedures zijn een pijler van het systeem waarop de waardevastheid van diploma's is gebouwd en ook die moeten betrouwbaar en transparant zijn. De Onderwijsraad adviseert het omdraaien van de correctievolgorde. Een leraar van een andere school wordt dan eerste corrector. Bovendien heeft de Tweede Kamer de Onderwijsraad om advies gevraagd over de waarde van de behaalde diploma's. Er zijn drie vragen gesteld: Kan de samenleving rekenen op de betrouwbaarheid van het diploma? Is het diploma voldoende herkenbaar? Houdt het diploma ook zijn waarde door de jaren heen?

Het verzwaren van de eisen van het eindexamen is één stap, er zullen er nog meer moeten volgen om de waardevastheid van diploma's te blijven garanderen.

De redactie van het EXAMENS verwelkomt Jackelien ter Burg als nieuw lid. Zij is freelance onderwijskundige en is onder meer gespecialiseerd in item- en toetsenbankontwikkeling, beoordelen van items en toetsen, en training en advies op het gebied van beoordelingsinstrumenten.

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

Spelling, grammatica en interpunctie **meebeoordelen** op eindexamens?

□ Ben Wilbrink, Denny Borsboom en Michel Couzijn

Begin april van dit jaar is in de Tweede Kamer een motie ingediend die de regering vraagt te bevorderen dat 'correcte spelling en grammatica op enigerlei wijze meetellen in de beoordeling van elk centraal examenonderdeel'. De indieners, Jan Jacob van Dijk (CDA) en Jasper van Dijk (SP), zagen dit als de logische sluitsteen op de wens van de commissie-Meijerink dat taal en rekenen niet alleen de zorg van de docenten Nederlands respectievelijk wiskunde zijn, maar van alle docenten.

De motie

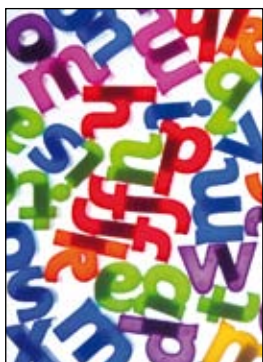
Het NRC Handelsblad berichtte over deze motie op 1 april, wat kon doen vermoeden dat de krant een intelligente aprilgrap presenteerde. Een aangehaalde uitspraak van staatssecretaris Van Bijsterveldt versterkte die indruk: zij zou het niet eens zijn met de motie en er de voorkeur aan geven dat scholen zelf uitmaken of en hoe zij in het schoolonderzoek bij alle vakken taalfouten meerekenen. Sommige ideeën druisen zozeer in tegen wat verstandig examenbeleid is, dat ze moeilijk serieus zijn te nemen. Na publieke discussie over deze motie werd de tekst afgezwakt tot '... verzoekt de regering te onderzoeken in hoeverre correcte spelling en grammatica op enigerlei wijze kan meetellen ...'. De zo gewijzigde motie werd met brede steun aangenomen. De formulering 'correcte spelling en grammatica meetellen' is die van de indieners en moet hetzelfde betekenen als 'spelling en grammatica meebeoordelen'. Interpunctie hoort er overigens ook bij; het is gebruikelijk om spelling, interpunctie en grammatica samen taalverzorging te noemen. Procedures die

32 290 Vaststelling van regels over referentieniveaus voor de taal- en rekenvaardigheden van leerlingen (Wet referentieniveaus Nederlandse taal en rekenen)
nr. 14 MOTIE VAN DE LEDEN JAN JACOB VAN DIJK EN JASPER VAN DIJK
Voorgesteld 31 maart 2010
De Kamer,
gehoord de beraadslaging,
overwegende, dat een verbetering van de kwaliteit van taal en rekenen het best kan worden bereikt indien de verantwoordelijkheid hiervoor integraal wordt gedeeld binnen de onderwijsinstelling;
overwegende, dat de commissie-Meijerink ervoor heeft gepleit om de verantwoordelijkheid voor het behalen van de referentieniveaus niet alleen bij de docenten Nederlands en wiskunde neer te leggen;
constaterende, dat het in theorie mogelijk is, een voldoende te halen voor een centraal eindexamenonderdeel, niet zijnde Nederlands, terwijl de beantwoording qua spelling en grammatica ver beneden peil is;
overwegende, dat dit strijdig is met de eerdergenoemde integraliteit en het doel van het verhogen van het niveau van de taalvaardigheden van de leerlingen;
verzoekt de regering te bevorderen dat correcte spelling en grammatica op enigerlei wijze meetellen in de beoordeling van elk centraal eindexamenonderdeel,
en gaat over tot de orde van de dag.

aan het tot stand komen van de adviezen ten grondslag liggen, zijn eveneens voor iedereen te controleren.

Beeldvorming

De achtergrond van de motie is de beeldvorming dat het onderwijs tekortschiet in het bijbrengen van basale reken- en taalvaardigheden. De commissie-Meijerink heeft nagegaan of empirische gegevens deze beeldvorming ondersteunen (Meijerink, 2008). Voor het mbo: 'Onderzoekers constateren dat er te weinig onderwijsaanbod in



Sommige ideeën druisen zozeer in tegen verstandig examenbeleid dat ze moeilijk serieus zijn te nemen

taalvaardigheid is; er wordt te weinig tijd aan taalvaardigheid besteed, de aanpak is niet functioneel genoeg en de eindtermen Nederlands zijn vaag of niet omschreven. Volgens de commissie steunen de empirische gegevens deze beeldvorming niet voor de hele breedte van het onderwijs; wat niet wil zeggen dat alles oké is; bovendien zijn voor het mbo en hbo maar weinig onderzoeken beschikbaar. En de problemen met taalvaardigheid zijn bepaald breder dan alleen spelling en grammatica. Het rapport van de commissie geeft een nuttig overzicht, met een uitwerking van spellingniveaus. Overigens doet de commissie geen aanbeveling om taalverzorging ook bij andere vakken dan Nederlands mee te beoordelen.

Niet van vandaag of gisteren

Het is duidelijk dat er zorg bestaat over de beheersing van spelling en grammatica. Maar ligt het dan wel voor de hand om een maatregel in de sfeer van de eindexamens te nemen? Gezien het feit dat de klacht van alle tijden is, heeft het zin om de opvattingen van verre voorgangers in deze discussie te betrekken.

Didactische oplossing

Zo beschreef dr. J.G.M. Moormann - taalkundige, leraar Nederlands en vakdidacticus - in het didactiekboek *De moedertaal* (1936) dat leerlingen in proefwerken van andere vakken de ergerlijkste spellingfouten maken. Door de betreffende docenten te vragen de belangrijkste fouten te noteren, kon hij de leerlingen in de eigen lessen Nederlands daarop bijspijkeren, met goede gevolgen voor het Nederlands in het eindexamenwerk. Moormann deed dat in de schaarse uren Nederlands die in klas vier en vijf beschikbaar waren: twee uren! Moormann gaf de juiste benadering van het probleem: niet via de examens, maar door een didactische aanpak, mede gericht op het besef dat taalverzorging ertoe doet, ook bij andere vakken. Daarom is een Algemene Maatregel van Bestuur die sleutelt aan het centraal schriftelijk eindexamen (CSE) ongewenst, op zijn minst totdat empirisch is aangetoond dat een didactische oplossing niet volstaat. De volgende uitspraak van een commissie van neerlandici in de toonaangevende didactische handleiding van de commissie Van Dis (1962, p. 121) raakt het hart van de motie over meebeoordelen die een halve eeuw later zou worden aangenomen. 'Tenslotte geldt voor gymnasium, h.b.s.-A en h.b.s.-B de

bepaling, dat men bij de berekening van het eindcijfer van het schriftelijk-Nederlands rekening moet houden met het Nederlands van de vertalingen uit het Frans, Engels en Hoogduits. Dit is een dood voorschrift. In de eerste plaats heeft de neerlandicus in het drukste deel van de cursus geen gelegenheid om naast zijn eigen twee werkstukken ook nog drie vertalingen per kandidaat door te nemen, zonder daardoor de correctietijd van gecommiteerden of deskundigen aanzienlijk te verkorten. Maar afgezien van dit praktische bezwaar, vertalen is een integrale bezigheid, waaruit men goed of slecht gebruik van de ontvangende taal niet los kan pellen. Daarom moet men de kwaliteiten van een werkstuk waarboven staat 'Vertaal in goed Nederlands' niet in twee rubrieken onderbrengen, maar overeenkomstig de opdracht als een geheel waarderen. Zo gebeurt het over het algemeen ook.'

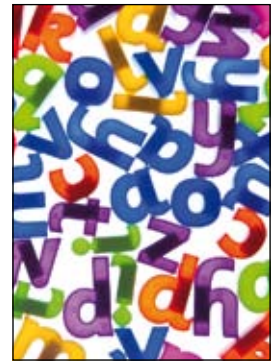
Pro en contra de beoordeling van taalverzorging in eindexamenwerk

Zowel de didactiek van Moormann als de analyse van de commissie Van Dis bieden op zich een argument dat sterk genoeg moet zijn om de huidige poging tot sleutelen aan de eindexamens te stoppen. Wellicht zijn er ook argumenten die pleiten vóór beoordeling van taalverzorging in eindexamenwerk. Wat zijn bijvoorbeeld de argumenten geweest om in het huidige CSE Nederlands havo-vwo taalfouten in de samenvatting mee te tellen, terwijl dit examen toch uitdrukkelijk de leesvaardigheid bedoelt te toetsen? Hierover is twintig jaar geleden een discussie gevoerd waarin de meningen van de docenten Nederlands verdeeld bleken. Een deel van de docenten vat de samenvattingsopdracht niet alleen als een leestoets op, maar ook als een schrijfoopdracht en wil daarbij de conventionele criteria als spelling- en formulevaardigheid niet geheel negeren. Maar twee zo verschillende vaardigheden in dezelfde toets beoordelen, levert dezelfde bezwaren op die het meebeoordelen van taalverzorging bij een toets natuurkunde zou doen.

Empirische gegevens ontbreken

Alvorens in de examens in te grijpen, is het verstandig om te inventariseren wat precies het probleem is dat zich zo zou laten oplossen. Hoe goed is bijvoorbeeld de taalverzorging van de laatste generatie leerlingen? Wat

Taalverzorging bestaat uit spelling, interpunctie en grammatica



zien we daarvan terug in het CSE Nederlands? Daarover zijn helaas geen empirische gegevens beschikbaar, althans niet in de verslagen van Cito. Dan blijft het dus gissen of, en zo ja welke, tekorten in taalverzorging er zijn aan het eind van het voortgezet onderwijs. De toetsing van schrijfvaardigheid, waaronder ook spelling en grammatica, maakt onderdeel uit van het schoolexamen; daarover zijn evenmin empirische gegevens beschikbaar. Hoe dat ook zij, als de taalverzorging in principe goed is, is er geen reden om de beoordeling ervan uit te breiden tot ander examenwerk; als de taalverzorging onder de maat is, dan schiet het onderwijs tekort en zijn dáár maatregelen nodig, niet bij het eindexamen. Als niemand het weet en ook de commissie-Meijerink er geen uitspraken over doet, waarom dan denken aan meebeoordelen van taalverzorging in alle onderdelen van de eindexamens? Moormann merkte al in 1936 op dat de ergerlijke taalfouten in proefwerken voor andere vakken het gevolg zijn van toegestane slordigheid, omdat dezelfde leerlingen bij het vak Nederlands wél goed spellen. Als dit ook geldt voor het huidige onderwijs - zijn daar actuele empirische gegevens over? - dan ligt het niet voor de hand om daar iets aan te willen doen met een ingreep in de examens.

Invloed van taalverzorging op de beoordeling

Het voorstel om taalverzorging bij elk examenonderdeel in de beoordeling te betrekken, roept bij lezers die hun methodologie van De Groot (1961, 7:3:2) kennen de tegenvraag op: zijn er niet juist maatregelen nodig om te voorkomen dat taalverzorging onbedoeld de beoordeling van het gemaakte werk beïnvloedt? Dit is een interessante spanning: een werkstuk moet op A (bijvoorbeeld inzicht in natuurkunde) worden beoordeeld, maar door de aanwezigheid van kenmerk B (bijvoorbeeld de taalverzorging) is de beoordeling van A niet meer zuiver. Het is niet onwaarschijnlijk dat de slechter spellende leerlingen in hun schoolloopbaan zowel als in hun examenwerk onbedoeld zijn meebeoordeeld op hun slechtere spelling; De Groot noemt dit het halo-effect. Van Oudenhoven (1985) rapporteert op dit punt empirische bevindingen in het basisonderwijs. Van den Bergh (1988) wijst op onderzoek hierover in het voortgezet onderwijs door De Glopper (1983). Mogelijke gevolgschade is dat de leerkracht de verwachting voor deze leerling naar beneden bijstelt, in zoiets als een omgekeerd 'Pygmalion-in-the-classroom' effect. Door een negatieve indruk over

het taalgebruik kan de docent een lagere verwachting krijgen van de capaciteiten van deze leerling in zijn schoolvak, zoals geschiedenis; een verwachting waardoor de minder uitgedaagde leerling daadwerkelijk minder kan gaan presteren.

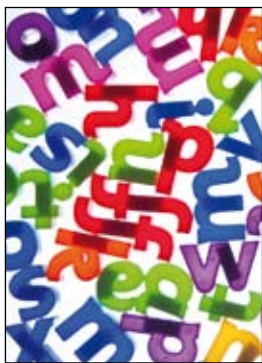
Leerlingen kunnen in hun onderwijsloopbaan door onbedoeld meebeoordelen van hun achterblijvende taalverzorging zijn benadeeld. En dan komt er dit voorstel om taalverzorging van alle examenwerk uitdrukkelijk mee te beoordelen. Het is aan de indieners van zo'n voorstel om aannemelijk te maken dat geen enkele groep door dat meebeoordelen van taalverzorging een partijdige behandeling krijgt.

Taalverzorging beoordelen

De taalverzorging van het werk van de leerling kan dus onbedoeld het inhoudelijke oordeel over dat werk beïnvloeden. Wat is nu een goede aanpak om nadrukkelijk die taalverzorging zelf te beoordelen? In de huidige samenvatting Nederlands tellen de taalfouten letterlijk mee: ze gaan als strafpunten van het puntentotaal af. Deze nogal traditionele aftrekregeling gaat voorbij aan inhoudelijke en toetstechnische eisen voor een deugdelijke beoordeling van taalverzorging. Simpelweg fouten tellen is een eeuwenoude methode in het onderwijs, maar meet niet adequaat het vermogen van de kandidaat om goed Nederlands te schrijven. Voor dat laatste is het bijvoorbeeld nodig om na te gaan of de kandidaat de gangbare maar lastige regels kan toepassen op normale gevallen. Het gaat er hier niet om wie dat uit zichzelf beter of slechter doet dan anderen, maar wie in absolute zin de regels voldoende kent en weet toe te passen. Het is daarom een vanzelfsprekende kwaliteitseis, ook door de commissie Van Dis aangegeven, dat alleen erkend vakdeskundige en dus bevoegde leraren Nederlands de taalverzorging van examenwerk beoordelen.

Een taalstandaard?

Een probleem met het tellen van fouten is bovendien dat beoordelaars verschillen van inzicht over wat in concrete gevallen correct is: Nederlandse taalvormen zijn vaak minder eenduidig te beoordelen dan gemaakte wiskundeopgaven; zie bijvoorbeeld de *Algemene Nederlandse Spraakkunst* (ANS, p. 24). Het College voor Examens en dus ook docenten Nederlands kiezen dan een standaard om deze meningsverschillen uit te sluiten, zoals de



Ergerlijke taalfouten in proefwerken voor andere vakken zijn het gevolg van toegestane slordigheid

ANS, de *Schrijfwijzer* (Renkema, 1979) of de *Woordenlijst Nederlandse Taal*. Een standaard kan in het onderwijs goede diensten bewijzen, maar is voor de beoordeling in eindexamens geen probleemloze oplossing. Vasthouden aan een gekozen standaard levert een evident willekeurige behandeling van de examenkandidaten op omdat veel afwijkingen verdedigbaar zijn, in de publieke media wijd en zijd voorkomen en soms zelfs worden onderwezen. Zo keurt de ene standaard 'lantaren' fout, de andere goed; en ook over 'je zal' wordt verschillend gedacht. Er zijn honderden verschillen tussen standaarden.

Ongestandaardiseerde toetsen

Belangrijker is het probleem dat het meten van taalverzorging in werk dat iets anders beoogt te toetsen, op ongestandaardiseerde wijze gebeurt. Zo zullen sommige kandidaten die weten op spelling te worden beoordeeld, bewoordingen zo proberen te kiezen dat zich zo min mogelijk spellingproblemen voordoen; Van der Velde (1956) signaleerde dit ook al. Omdat de kandidaat zelf de woorden kiest, is het examenonderdeel als meting van taalverzorging dus ongestandaardiseerd. Zo schrijft de ene leerling 'geëvolueerd' waar de ander het veilige 'veranderd' kiest. Het oordeel over taalverzorging weegt in feite de verbale intelligentie mee bij het omzeilen van spellingproblemen. Kortom, als de staatssecretaris staat op adequate toetsing op taalverzorging in het CSE, dan zijn taalproblemen die de leerling voor zichzelf opwerpt of juist weet te vermijden, niet geschikt. Er is een gestandaardiseerde toets nodig op kennis en op het toepassen van taalregels en -vormen. Van kandidaten moet een prestatie worden gevraagd die hun regelkennis en vaardigheid in het toepassen ervan indiceert, niet hun handigheid of zorgvuldigheid.

Validiteit van het eindexamen

Voor de integriteit van de eindexamens is het belangrijkste punt dat meebeoordelen van taalverzorging de validiteit van elk afzonderlijk examenonderdeel aantast en daarmee van het examen in zijn geheel. Als dat meebeoordelen strafpunten voor natuurkunde oplevert, dan is het cijfer voor natuurkunde geen zuivere maat meer voor kennis van natuurkunde. Natuurlijk is het mogelijk om voor taalverzorging van al het examenwerk een afzonderlijk cijfer te geven en dat in het eindcijfer voor Nederlands mee te wegen. Dat lost een deel van dit probleem

op, als de logistiek ervan uitvoerbaar zou blijken - wat niet waarschijnlijk is. Dat neemt niet weg dat kandidaten weten dat bij alles wat zij over natuurkunde opschrijven ook hun taalverzorging ter beoordeling staat. Dus verdelen zij hun aandacht over natuurkunde en het Nederlands; zouden ze dat niet doen, dan benadelen zij zichzelf. Is het met verdeelde aandacht gemaakte examenwerk natuurkunde dan wel de beste prestatie natuurkunde die de kandidaat kan leveren? Wat is met die verdeelde aandacht dan de validiteit van de natuurkundige verklaringen die ze opschrijven? Is dat een marginale kwestie? Wie dat meent, mag dat uitleggen aan kandidaten die op zo'n marginale kwestie gezakt kunnen zijn.

Testtheoretische problemen samengevat

Psychometrisch gezien valt te verwachten dat meebeoordelen van een tweede domein zoals taalverzorging leidt tot een lagere validiteit van het examenonderdeel, tot meer ruis - onsystematische variantie - in de scores en tot het ontstaan van bias. Denk aan jongens versus meisjes of leerlingen met Nederlands als tweede taal: systematische groepsverschillen in scores die niet zijn terug te voeren op het oorspronkelijke kennisdomein, zoals natuurkunde, zijn uitingen van bias ofwel partijdigheid. Het valt moeilijk te voorspellen in welke mate deze problemen zich in de praktijk voordoen, maar het is evident dat een maatregel die de psychometrische kwaliteit van examens eigenlijk alleen maar kan verlagen, hooguit genomen mag worden wanneer het echt niet anders kan. Aan deze voorwaarde is in het voorliggende geval niet voldaan.

Juridische eisen

De rechtsbasis van het examen wordt aangetast door andere zaken mee te beoordelen dan die ieder onderdeel specifiek bedoelt te toetsen. Onzorgvuldig omgaan met het examen zal leiden tot beroepszaken met voorspelbare uitkomst: veroordeling van de inbreuk op wat voor een behoorlijk examen is vereist. Cohen (1981) heeft gedetailleerd de eisen beschreven die in juridische zin voor examens gelden. Zo mogen kandidaten verwachten dat een eindexamen natuurkunde gaat over natuurkunde, niet tevens over Nederlands, en dat een toets taalverzorging gaat over wat daarover is onderwezen.

Kandidaten mogen verwachten dat een eindexamen natuurkunde gaat over natuurkunde en niet ook over Nederlands



Afsluiting

In het CSE is het niet mogelijk om op behoorlijke wijze examenwerk mede op spelling, interpunctie en grammatica te toetsen. De taalverzorging toetsen kan eventueel alleen in een specifiek daarop gerichte gestandaardiseerde toets. En wat in het CSE niet realiseerbaar is, mag in het schoolonderzoek zeker niet gebeuren. Goede taalverzorging - wat iets anders is dan foutloze - is van groot belang, maar moet een zaak van aanhoudende zorg in het onderwijs zelf zijn. Bij dat laatste hoort ook dat alle leerkrachten zelf hun taal goed verzorgen, de taalverzorging van de leerlingen in de gaten houden en bij werk waar geen tijdsdruk op staat - huiswerk, geen proefwerken - structureel een redelijke taalverzorging eisen voor in te leveren werk.

Literatuur

Bergh, H. van den (1988). *Examens geëxamineerd. Een onderzoek naar de validiteit van de examens Nederlands voor het lbo en mavo*. Proefschrift RU Utrecht. Den Haag: SVO Selecta Reeks.

Cohen, M.J. (1981). *Studierechten in het wetenschappelijk onderwijs*. Proefschrift RU Leiden. Zwolle: Tjeenk Willink.

Dijk, J.J. van & Dijk, J. van (31 maart, 14 april 2010). *Motie Tweede Kamer dossier 32290-14 en 19*. ikregeer.nl/.

Dis, L.M. van, en anderen (1962). *Didactische handleiding voor de leraar in de moedertaal*. Amsterdam: Meulenhoff.

Glopper, K. de (1983). *Opstelkenmerken en opstelbeoordelingen: onderzoek naar de validiteit van globale opstelbeoordeling*. Universiteit van Amsterdam: SCO.

Groot, A.D. de (1961). *Methodologie*. Den Haag: Mouton. Ook integraal op dbnl.nl/.

Haeseryn, W., Geerts, G., Rooij, J. de & Toorn, M.C. van den (red.) (1997). *Algemene Nederlandse spraakkunst*. Groningen: Wolters-Noordhoff. let.ru.nl/ans/.

Meijerink, H. (voorz.) (2008). *Over de drempels met taal en rekenen. Eindrapport Expertgroep doorlopende leerlijnen taal en rekenen*. slo.nl/nieuws/dll/Hoofdrapport.pdf/download.

Moormann, J.G.M. (1936). *De moedertaal. Een didactiek voor het middelbaar en lager onderwijs*. Nijmegen/Utrecht: Dekker & Van de Vegt.

Oudenhoven, J.P. van (1985). Spelvaardigheid en de ongelijke kansen van arbeiderskinderen in het onderwijs. In Assink, E. & Verhoeven, G. *Visies op spelling* (155-163). Groningen: Wolters-Noordhoff.

Renkema, J. (1979). *Schrijfwijzer*. s-Gravenhage: Staatsuitgeverij.

Velde, I. van der (1956). *De tragedie der werkwoordsvormen. Een taalhistorische en taaldidactische studie*. Proefschrift RU Groningen. Groningen: Wolters. *Woordenlijst Nederlandse Taal*. woordenlijst.org/ 'Het groene boekje'.

□ De heer B. Wilbrink is freelancer onderwijsonderzoek. www.benwilbrink.nl. De heer dr. D. Borsboom is als universitair hoofddocent verbonden aan de programmagroep Psychologische Methodenleer van de Universiteit van Amsterdam. De heer dr. M. Couzijn is als universitair docent en onderzoeker verbonden aan het Instituut voor de Lerarenopleiding van de Universiteit van Amsterdam en lid van de expertgroep die voor het College voor Examens van OCW de vernieuwing van het CE Nederlands voorbereidt. Zijn bijdrage aan dit artikel is op persoonlijke titel.

E-mail: ben@benwilbrink.nl.

Ook de
examens van



EXAMINERING EN
CERTIFICERING VOOR DE
MOBILITEITSBRANCHE

staan onder
toezicht van de

Examen
kamer
kwaliteit door onafhankelijkheid

Terecht of niet?

□ Henk van Berkel

Een tentamen waaraan veel mis is

Er schortte volgens een klaagster nogal veel aan een tentamen, afgenomen bij een Vlaamse universiteit: te weinig vragen, een oneerlijke raadcorrectie, een zeer laag slaagpercentage, geen voorbeeldvragen, onduidelijke vragen, te korte examenduur en geen individuele feedback.

Wat er aan vooraf ging

Een uitzonderlijk tentamen

Gezegd moet worden dat het tentamen dat deze klaagster had gemaakt, erg uitzonderlijk was en daardoor de zaak op scherp zette. Het tentamen bestond uit tien vragen, waarvan vier driekeuze vragen, de overige zes hadden twee alternatieven. Van deze zes waren vier vragen van het type juist/onjuist. Een correct antwoord leverde twee punten op en voor een fout antwoord werd een punt afgetrokken.

Tentamens hebben vele vormen en eigenschappen. Een tentamen heeft een eindige hoeveelheid vragen, kent een toetstijd, kent (meestal) een norm, de vragen hebben een vorm en ze worden gescoord. Over al deze aspecten van tentamens, en er zijn er veel meer te bedenken, moeten examinatoren een beslissing nemen. Soms schieten de beslissingen bij studenten in het verkeerde keelgat en gaan zij klagen. Zoals de beroepsprocedure voorschrijft, moeten studenten eerst hun klachten indienen bij de examencommissie. Als er dan geen oplossingen wordt gevonden, komt het College van Beroep in beeld. Misschien dat de onderhavige examencommissie al nattigheid rook; ze schakelde twee experts in: een psychometricus (een toetsdeskundige) en een inhoudsdeskundige. Zij bogen zich over de klachten.

De mening van de experts

In het algemeen onderschreven de experts de standpunten van de examencommissie. De psychometricus achtte de scoringswijze fair, oordeelde dat de formulering van de vragen aan de basisregels voldeed en was ook positief over de vooraf gegeven informatie over het tentamen. Hij had echter zijn twijfels over het geringe aantal vragen, vooral als het slagen van studenten louter zou afhangen van dit tentamen (van tien vragen dus). Hij legde daarbij een relatie met de betrouwbaarheid van het tentamen. Het toeval speelt volgens hem bij tien tentamenvragen een te grote rol bij de uitslagbepaling.

De inhoudsdeskundige was van oordeel dat de vragen voldoende duidelijk waren. Maar merkwaardigerwijs plaatste zij toch een aantal kanttekeningen bij haar oordeel. Zij stelde dat vragen bepaalde interpretatieproblemen kunnen hebben die pas tijdens een tentamen 'bij handopsteking' kunnen worden opgelost. (Overigens bestond tijdens dit tentamen niet de mogelijkheid om vragen te stellen wanneer een student problemen ondervond met de duidelijkheid van een vraag.) Over drie vragen uit dit tentamen was de inhoudsdeskundige van oordeel dat de interpretatie van de collegegever over dit onderwerp van belang was om tot een eenduidige beantwoording van de vraag te komen.

Op basis van de oordelen van de experts en het eigen oordeel besloot de examencommissie één vraag achteraf uit het tentamen te verwijderen. Deze tegemoetkoming ging de student niet ver genoeg. Zij stapte naar het College van Beroep.

Het College van Beroep aan het woord

De klacht wordt behandeld

Het College begon de behandeling van de klacht met het bepalen of het tentamen aan de vereisten voldeed: een tentamen moet de bereikte kennis en competenties adequaat meten. Als een tentamen dat niet doet, verklaart het College zich ontvankelijk voor beroepsschriften. Deze stellingname is niet onbelangrijk, omdat alle Colleges van Beroep zich op het standpunt stellen dat het in beginsel aan de examencommissies is om uitspraken te doen over de kennis en competenties van studenten en pas als er sprake is van onredelijkheid en/of onzorgvuldigheid het College op de stoel van de examencommissie kan gaan zitten.

Terecht of niet?

Naar het oordeel van het College was er in dit geval sprake van, anders had ze het beroepsschrift wel naast zich neergelegd. Het College deed vervolgens een niet zo verrassende uitspraak: een meerkeuze tentamen is in beginsel een geaccepteerde tentamen-vorm. In hoeverre het tentamen voldoet aan het redelijkheids- en zorgvuldigheidsbeginsel, moest worden onderzocht. Daarna ging het College over tot de weging van de aangedragen argumenten.

Te klein aantal items

De studente stelde dat tien - en uiteindelijk negen - vragen onvoldoende zijn om een betrouwbaar oordeel te geven. De psychometricus onderschreef dit. De examencommissie stelde daar tegenover dat het een klein vak betrof van slechts drie ECTS met 170 bladzijden studiestof. ¹ De commissie achtte dit in verhouding tot het aantal vragen. Het College koos hierin de kant van de studente: negen vragen zijn in dit geval onvoldoende om een zak/slaagbeslissing op te baseren.

Onduidelijkheid van de vraagstelling

Dan het punt over de onduidelijkheid van sommige vragen. De studente vond een aantal vragen onduidelijk en kon tijdens het tentamen geen vragen stellen om die onduidelijkheid weg te nemen. Ook stelde zij dat de inhoudsdeskundige van mening was dat de interpretatie van de collegegever van belang was bij het vaststellen van het correcte alternatief. Volgens de student was de collegegever hierover niet duidelijk geweest. Bovendien was het niet de collegegever die het tentamen had nagekeken.

De examencommissie stelde daar tegenover dat het correct interpreteren van een vraag onderdeel is van de te testen kennis en omdat volgens de psychometrische expert de vragen correct waren geformuleerd, de student geen punt had. Bovendien zou de collegegever tijdens het college hebben aangegeven welke interpretatie de voorkeur verdiende.

Ongeschiktheid van de vraagvorm

Bovendien vocht de studente de vraagvorm aan: meerkeuze tentamens kunnen volgens haar niet toetsen of studenten zich kritisch kunnen opstellen tegenover bepaalde onderwerpen. En dat was wel een van de doelstellingen van dit studieonderdeel.

De uitspraak

Het College oordeelde dat meerkeuze tentamens in beginsel adequaat zijn om hogere cognitieve vaardigheden als kritisch denken te toetsen. Daarover kreeg de examencommissie dus gelijk. Ten aanzien van de multi-interpreteerbaarheid van de opgaven kwam het College met een verrassend standpunt. De inhoudsdeskundige

was van mening dat bij twee vragen twee alternatieven correct waren. De examencommissie stelde vervolgens dat de collegegever duidelijk de voorkeur had aangegeven voor een van beide en dat de studenten dus hadden kunnen weten wat het enige correcte alternatief was. Dat argument legde het College naast zich neer: 'Indien twee alternatieven correct zijn, moeten die ook als correct worden gescoord'. Kennelijk is de mening van de collegegever niet zo maar doorslaggevend.

Verder stelde het College dat het feit dat er verhoudingsgewijs veel studenten voor dit tentamen waren gezakt, ook een aanwijzing is dat er wat met het tentamen aan de hand was. Hiermee had de examencommissie rekening moeten houden.

De slotconclusie luidde: 'Het tentamen is niet met de vereiste zorgvuldigheid ingericht zodat de meting van de kennis van de student niet voldoende betrouwbaar is geweest'.

De consequentie van de uitspraak

De consequentie van deze uitspraak voor de studente is niet bekend. In formele zin vernietigt het College van Beroep de uitslag van het tentamen en draagt de examencommissie op een andere uitslag te bepalen met inachtneming van het gestelde in de uitspraak. Het is echter niet zo dat de uitspraak op alle studenten die het betreffende tentamen hebben afgelegd, betrekking heeft. Dan hadden die ook in beroep moeten gaan. Het tentamen zal dus niet ongeldig worden verklaard en, bijvoorbeeld, opnieuw worden afgenomen. Meestal zal de examencommissie in dergelijke situaties bij die ene student een mondeling afnemen of anderszins een aangepast tentamen voorleggen.

Wat zijn de lessen?

Het aantal vragen

Het aantal vragen in de toets stond niet in verhouding tot de omvang van de studiestof, althans was onvoldoende om een betrouwbare zak/slaagbeslissing te nemen. Doorgaans zal pas bij dertig vierkeuze vragen de betrouwbaarheid acceptabele waarden kunnen aannemen.

Duidelijkheid

Een te groot aantal vragen was onduidelijk. De examencommissie stelde correct dat de mate van duidelijkheid geen dichotomie is. Het is niet zo dat een vraag duidelijk dan wel onduidelijk is, het is een glijdende schaal. Wat voor de een duidelijk is, is voor de ander onduidelijk. Het stellen van duidelijke vragen, die studenten op het goede spoor zetten in welke richting zij het correcte antwoord moeten vinden, is een recht. Indien studenten meer duidelijkheid nodig hebben, dienen zij de mogelijkheid te krijgen een inhouds-

Terecht of niet?

deskundige een vraag te stellen via handopsteking. Als aanvullende informatie wordt gegeven, hebben alle studenten daar recht op. Dat betekent dat een mededeling wordt gedaan op een aanwezig schoolbord of via een omroepinstallatie. Daarom moet er altijd een inhoudsdeskundige bij de toetsafname aanwezig zijn.

Twee alternatieven correct

In deze beroepszaak komt het interessante aspect van 'twee alternatieven correct' aan de orde. Zoiets komt in de toetspraktijk regelmatig voor omdat het moeilijk is alternatieven te bedenken die helemaal fout zijn. Of omdat 'geleerden' het niet eens zijn met elkaar. Dit probleem is makkelijk oplosbaar. Ten eerste via het aangeven van de context. Dus niet 'De definitie van het Oedipus complex is ...'; maar 'Volgens Freud is de definitie van ...'. Op die manier voorkom je dat een student met een boek van een andere auteur dan Freud in de hand de examencommissie van haar ongelijk kan overtuigen. Ten tweede is het probleem van twee alternatieven correct oplosbaar door in de toetsinstructie te vermelden dat steeds het beste alternatief moet worden aangekruist.

In de praktijk volstaat dit meestal. Als er dan toch nog onduidelijkheid optreedt, resteert niets anders dan twee alternatieven correct

te rekenen. Tentamenverwerkingsprogramma's kunnen dit technisch aan.

¹ECTS = European Credit Transfer System, waarbij een ECTS punt wordt gelijk gesteld aan 28 uur studie, inclusief contacttijd.



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

Gebruikt u wel eens andere vraagvormen dan de mc-vraag?

- ☐ Alleen als het nodig is.
- ☐ Waarom zou ik? De mc-vraag bevat uitstekend.
- ☐ Nee, maar heb je een suggestie?



WWW.TEELN.NL

Peer assessment in het basisonderwijs

□ Kelly Meusen-Beekman en Desirée Joosten-ten Brinke

Een kritische denkhouding, capaciteit om eigen werk te evalueren en goed ontwikkelde communicatieve vaardigheden: dit zijn eisen die aan mensen in de hedendaagse maatschappij worden gesteld. Het basisonderwijs kan hierop inspelen door kinderen te leren reflecteren op zichzelf en anderen, en kritisch te zijn ten aanzien van eigen ontwikkelingen en vaardigheden. Peer assessment, een toetsvorm waarbij personen elkaar beoordelen, biedt deze mogelijkheid. Dit artikel laat zien dat peer assessment in het basisonderwijs veelbelovend is.

Inleiding

Peer assessment in het basisonderwijs wordt nog maar op beperkte schaal ingezet. Men hanteert traditionele toetsvormen om bijvoorbeeld taal- en rekenvaardigheden in kaart te brengen of de ontwikkeling van het kind te volgen. Ontwikkelen, verwerven en toepassen van kennis, vaardigheden, attitudes en competenties worden echter steeds belangrijker. Traditionele toetsmethoden schieten tekort als het gaat om de ontwikkeling van complexe vaardigheden en vormingsdoelen op lange termijn, aangezien ze vooral gericht zijn op kennisdoelen (Dochy, Segers & Sluijsmans, 1999). Onderzoek in het hoger onderwijs laat zien dat prestaties verbeteren ten gevolge van assessments. Studenten worden door het gebruik van assessments opgeleid tot competente beroepsbeoefenaars die in staat zijn zichzelf kritisch te beoordelen, kritisch na te denken en problemen op te lossen binnen een complexe, steeds veranderende sa-

menleving (Sluijsmans et al., 2003). De vraag is of dit voor het basisonderwijs ook geldt.

De rol van assessments in het onderwijs

In een ideale onderwijssituatie wordt een student uitgedaagd om actief deel te nemen aan zijn eigen leerproces, krijgt hij een groot verantwoordelijkheidsbesef, reflecteert hij op eigen kunnen en werkt samen, dit alles in een voortdurende dialoog met de docent (Sluijsmans et al., 2003). De toetsing zal hierop moeten aansluiten. Toetsen is meer dan het meten en verzamelen van gegevens. Assessments geven inzicht in sterke en zwakke punten, brengen vooruitgang in beeld en dragen bij aan de begeleiding van studenten in hun leerproces. Belangrijk is te beseffen dat in de kern een assessment een leeractiviteit is, die op effectieve wijze het leergedrag verandert, de betrokkenheid vergroot en in het algemeen bij lerenden als minder bedreigend wordt ervaren dan traditionele toetsvormen (Dochy & Struyven, 2002).

Assessments geven inzicht in de geboekte vooruitgang van het leerproces. Een bijzondere vorm van assessment is peer assessment. Het doel van een peer assessment is het beoordelen en evalueren van elkaar (Stiggins, 1986). Dit sociale proces leidt tot het ontwikkelen van vaardigheden voor kritische reflectie en analyse en het verbeteren van eigen inzichten en prestaties (Falchikov, 1995; Hargreaves, 1997; Sluijsmans, 2002). Peer assessment wordt daardoor beschouwd als een leerinstrument dat een positief effect heeft op de ontwikkeling van bijvoorbeeld de communicatie van leerlingen.

Assessments in het basisonderwijs

Aan het effect van assessments op het leren in het basisonderwijs is tot voor kort weinig aandacht besteed. Wragg (2001), Briggs, Woodfield, Martin en Swatton (2008) en anderen laten in hun onderzoeken echter zien dat (peer) assessment eveneens goed toepasbaar is in het basisonderwijs. Wragg (2001) geeft aan dat beoordelen geen eenvoudige opgave is, wat zeker van toepassing is op kinderen in het basisonderwijs. Jonge kinderen



Training in peer assessment is noodzakelijk

hebben behoefte aan structuur en ondersteuning, om de vaardigheden te leren waarmee ze elkaar eerlijk, informatief en objectief beoordelen. De context speelt daarbij een belangrijke rol. Een prestatie behoort plaats te vinden in een sociaal veilige, harmonieuze, niet op competitie gerichte omgeving (Wragg, 2001). De gegeven feedback op een prestatie draagt in belangrijke mate bij aan een goede motivatie van de leerlingen. Feedback ontvangen zorgt voor een hoger niveau van zelfkennis en inzicht in de impact van een uitgevoerde prestatie (Berihun, Tesera, & Desta, 2006). Buiten de noodzaak om kinderen te begeleiden in de uitvoering ervan, benadrukken Briggs et al. (2008) dat assessments het daadwerkelijke gedrag en de attitude van kinderen in het basisonderwijs positief beïnvloeden. Leerlingen leren feedback te geven en te ontvangen en leren informatie te beoordelen. Peer assessment in het basisonderwijs draagt bij aan het nemen van verantwoordelijkheid voor eigen leren, aan de betrokkenheid bij het leerproces en het vergroten van aandacht voor de onderwijsinhoud (Lindsay & Clarke, 2001). In de huidige praktijk geven leerlingen elkaar wel beoordelingen bij spreekbeurten, maar doordat ze hierin niet getraind zijn en zich niet bewust zijn van criteria, wordt dit door leerlingen als bedreigend ervaren en zijn de beoordelingen mogelijk gebaseerd op niet-relevante criteria.

Peer assessment implementeren

Ongeacht het niveau van onderwijs waarbinnen peer assessment wordt toegepast, blijkt het niet eenvoudig dit te implementeren en toe te passen. Het is immers geen toetsvorm op zich, er moeten vaardigheden geleerd worden om peer assessment te kunnen uitvoeren. Sluijsmans et al. (2003) geven aan dat leerlingen (peers) het moeilijk vinden elkaar een betrouwbare beoordeling te geven. Kwantitatieve beoordelingen, zoals ontworpen criterialijsten en Likertschalen, worden weliswaar vaak toegepast, maar geven weinig informatie over het leerproces. Een kwalitatieve beoordeling daarentegen, die bijvoorbeeld anoniem geleverd wordt, levert wel waardevolle informatie. Authenticiteit in beoordelingen en toetsen in kwalitatieve zin, zorgen ervoor dat de docent een leerling benadert als een jonge expert of vakgenoot (Sluijsmans, Dochy & Moerkerke, 1999). Men kan echter niet zomaar overgaan tot het afnemen van kwalitatieve toetsen, het vergt een zorgvuldige voorbereiding. Kinderen zijn niet

gewend om op basis van elkaars prestaties feedback te geven en te ontvangen. Deze nieuwe rol, als actief feedbackgevende en beoordelaar, zal op goede wijze getraind moeten worden. Daarnaast zijn duidelijke beoordelingscriteria en een transparante procedure essentieel voor het welslagen van het assessment. Sluijsmans (2002) biedt richtlijnen voor de implementatie van peer assessments in het onderwijs zodat leerlingen de prestaties van medeleerlingen kunnen beoordelen en voorzien van constructieve feedback.

Peer assessmenttraining

Om peer assessment uit te voeren en toe te passen is een training van zes lessen ontwikkeld voor leerlingen van groep 7 in het basisonderwijs. Gedurende de training is aandacht besteed aan het kunnen opstellen van criteria, het kunnen deelnemen aan een peer assessment, het geven en ontvangen van feedback en het verbeteren van de uitvoering van een leerlingprestatie.

De lessenserie heeft de volgende opzet:

1. Selfassessment: Hoe en waarom?
De leerlingen leren dat het belangrijk is om te reflecteren op eigen kennis en vaardigheden en dat zij zichzelf kunnen beoordelen. Deze aspecten worden aangeboden met behulp van een drama, waarin de kinderen in rollenspellen gevraagd wordt om zichzelf te beoordelen.
2. Criteria formuleren.
De leerlingen bespreken onder leiding van de docent wat verstaan wordt onder een 'goede spreekbeurt'. Bijbehorende beoordelingscriteria worden opgeschreven en een voorbeeldspreekbeurt wordt getoond, waarna de leerlingen oefenen in het objectief, kritisch en analytisch observeren ervan.
3. Omgaan met beoordelingscriteria.
De leerlingen gaan in op het kritisch observeren, analyseren en beoordelen van een product of proces van medeleerlingen. Daarnaast passen de leerlingen de beoordelingscriteria toe (zoals taalgebruik, materiaalgebruik en presentatie) op de invulling en uitvoering van een eigen spreekbeurt.
4. Geven van feedback.
In deze les staat het leren geven van constructieve peer feedback centraal. Om dit te oefenen wordt een muziekles aangeboden. De leerlingen bereiden in kleine groepjes een 'rap' voor, die aan het einde

Kinderen zijn niet gewend om elkaar te beoordelen



van de les gepresenteerd wordt. Op basis van vooraf vastgestelde criteria, wordt de rap door de medeleerlingen op grond van argumenten en goede onderbouwing becommentarieerd.

5. Feedback ontvangen.

Het ontvangen van feedback is lastig voor de kinderen. In een verwerkingsopdracht door middel van een opstel wordt hier aandacht aan besteed. Kinderen beschrijven hierbij hoe zij het ervaren om feedback te ontvangen en hoe zij de feedback vertalen naar verbeteringen in hun presentatie.

6. Peer assessment herhalingsles.

In de herhalingsles komen het objectief, kritisch en analytisch observeren en beoordelen van een product van medeleerlingen en het geven van constructieve feedback wederom aan bod. Dit door videobeelden te tonen van onbekende leerlingen die spreekbeurten uitvoeren. Deze spreekbeurten worden door de leerlingen beoordeeld en voorzien van mondelinge feedback en verbetertips.

Effecten van een peer assessmenttraining op een spreekbeurt: het onderzoek

De effecten van het aanleren van de peer assessmentvaardigheid op de uitvoering van een spreekbeurt zijn onderzocht in groep 7 van het basisonderwijs. In het onderzoek gaven de leerlingen van twee vergelijkbare groepen 7 spreekbeurten. Eén van de twee groepen kreeg de peer assessmenttraining. Achtereenvolgens is onderzocht (1) of de peer assessmenttraining effect heeft op de kwaliteit van de uitgevoerde peer assessment, (2) of er een relatie is tussen de peer assessmentvaardigheid en het uitvoeren van een spreekbeurt en (3) of de aangeleerde peer assessmentvaardigheden na verloop van tijd nog aanwezig waren. De gevonden resultaten geven we hieronder weer.

Positief effect van training

Peer assessmenttraining heeft positief effect op de kwaliteit van de uitgevoerde peer assessments bij leerlingen uit de bovenbouw van het basisonderwijs. Voordat de leerlingen de training gevolgd hadden, gaven alle leerlingen van beide groepen vrij oppervlakkige oordelen over elkaars prestaties. De leerlingen die de training volgden, gaven nadien een kwalitatief betere beoordeling van

de spreekbeurten dan de leerlingen die niet getraind waren. De criteria voor de spreekbeurt werden expliciet benoemd en onderbouwd en leerlingen gaven objectief een uitspraak over het niveau van de prestatie. Tevens werden er verbeterpunten gegeven. De niet-getrainde leerlingen bleven steken in beoordelingen als 'de spreekbeurt was leuk', of 'super, 10+'.

Uitvoering van de spreekbeurt

Er is geen relatie tussen het beschikken over de peer assessmentvaardigheden en de uitvoering van een spreekbeurt. Onderzocht is of de leerlingen, nadat zij vaardigheden voor peer assessment hebben aangeleerd, zelf een betere spreekbeurt geven. Deze verwachting was aanwezig omdat de leerlingen bewuster zijn geworden van de criteria die gebruikt worden voor het beoordelen van een spreekbeurt en zoals Lindsay & Clarke (2001) benadrukken dat leerlingen door assessments een grotere invloed hebben op het leerproces en de resultaten ervan. Uit de resultaten komt geen significante relatie naar voren tussen het beschikken over peer assessmentvaardigheden en het cijfer dat de leerlingen krijgen voor de uitvoering van de spreekbeurt. Wel geeft de docent in zijn kwalitatieve beoordeling van de spreekbeurten een duidelijke verbetering aan.

Herhaling van training

De aangeleerde peer assessmentvaardigheden nemen na verloop van tijd af. Wanneer de vaardigheden aangeleerd zijn, worden deze vaardigheden gestimuleerd en meerdere malen aangehaald gedurende de training. Onderzocht is of deze vaardigheden gehanteerd worden wanneer de leerlingen niet meer door een training ondersteund worden. Het resultaat laat zien dat de leerlingen dan weer terugvallen op hun oude gewoonten van oppervlakkige beoordelingen. Dit resultaat sluit aan bij de conclusie van Sluijsmans (2002) dat leerlingen na een enkele training in peer assessment niet als expert beschouwd mogen worden, maar dat herhaling van training noodzakelijk is.

Conclusie

Het geven van een peer assessmenttraining aan leerlingen van groep 7 in het basisonderwijs leidt tot leerlingen met een kritische denkhouding en de capaciteit om het werk van medeleerlingen te beoordelen. Vaardigheden



Getrainde leerlingen geven kwalitatief goede feedback

die belangrijk zijn in de hedendaagse maatschappij. Belangrijk is echter dat zo'n training niet eenmalig wordt aangeboden in het basisonderwijs, aangezien zonder herhaling de leerlingen terugvallen in hun oude gedrag. De peer assessmenttraining zou ook gecombineerd kunnen worden met het aanleren van andere vaardigheden, waardoor peer assessment niet een eenmalig trucje is, maar een structurele bijdrage levert aan het onderwijs.

Peer assessment is als leeractiviteit beschouwd, waarbij de leerlingen zelf een groot aandeel gehad hebben in het vaststellen van de beoordelingscriteria. De interactieve reflecties, het samenwerkend leren, de sociale interactie gedurende de training, en het uitvoeren van peer assessment hebben geleid tot het beschikken over peer assessmentvaardigheden.

Uit het onderzoek blijkt niet dat de spreekbeurten aantoonbaar zijn verbeterd door het beschikken over peer assessmentvaardigheden. Wellicht heeft de beperkte variatie in de docentcijfers daarmee te maken, want de kwalitatieve beoordelingen van de docent laten wel degelijk een verbetering in de kwaliteit van de spreekbeurten zien. De leerlingen die later in de rij aan bod kwamen om hun spreekbeurt uit te voeren, profiteerden mogelijk van de voorbeelden en presentaties van hun voorgangers. Of het beschikken over de peer assessmentvaardigheden ook invloed heeft op andere aspecten van het leerproces dan de prestatie die beoordeeld is, moet verder onderzoek uitwijzen. Wellicht blijken verbeteringen in eigen beoordelingsvermogen en prestaties (Hargreaves, 1997) aantoonbaar ten aanzien van bijvoorbeeld het beheersen van huiswerk of de voorbereiding van toetsen. Dat zou betekenen dat peer assessment een grote rol kan spelen in bijvoorbeeld groep 8 en bij de voorbereiding op het voortgezet onderwijs.

Literatuur

- Berihun, A., Tesera, A. & Desta, D. (2006). Enhancing active learning through teachers' peer and self reflections in selected primary schools. *Review of Educational Research*, 29, 172-187.
- Briggs, M., Woodfield, A., Martin, C. & Swatton, P. (2008). *Assessment for Learning and Teaching in Primary Schools*. Exeter: Learning Matters.
- Dochy, F.J.R.C., Segers, M.S.R. & Sluijsmans, D. (1999).

The use of self, peer and co-assessment in higher education: a literature review. *Studies in Higher Education*, 24(3), 331-350.

Dochy, F. & Struyven, K. (2002). Assessment: betekenis en assessmentvormen. Assessment in onderwijs. In: F. Dochy, L. Heylen & H. Van de Mosselaer (eds.), *Nieuwe toetsvormen en examinering in studentgericht onderwijs en competentiegericht onderwijs*. Utrecht: Lemma, 2002, 33-60.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International*, 32(2), 175-187.

Hargreaves, D.J. (1997). Student learning and assessment are inextricably linked. *European Journal of Engineering Education*, 22(4), 401-409.

Lindsay, C. & Clarke, S. (2001). Enhancing primary science through self and paired assessment. *Primary Science Review*, 68, 15-18.

Sluijsmans, D. (2002). *Student involvement in assessment: The training of peer assessment skills*. Proefschrift. Heerlen: Open Universiteit Nederland.

Sluijsmans, D., Dochy, F. & Moerkerke, G. (1999). Creating a learning environment by using self, peer and co-assessment. *Learning Environments Research*, 1, 293-319.

Sluijsmans, D., Brand-Gruwel, S., Van Merriënboer, J. & Bastiaens, T. (2003). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation*, 29, 23-42.

Stiggins, R.J. (1986). Teacher evaluation: Accountability and growth system-different purposes. *NASSP Bulletin*, 70, 51-58.

Wragg, E. (2001). *Assessment and Learning in the Primary School*. New York: Routledge.

□ Mw. drs. K. Meusen-Beekman is docent en kwaliteitscoördinator op R.K. Basisschool J.F. Kennedy. Zij is afgestudeerd in onderwijswetenschappen aan de Open Universiteit; mw. dr. D. Joosten-ten Brinke is universitair docent aan de Open Universiteit Nederland en lector Eigentijds toetsen en beoordelen aan de Fontys Lerarenopleiding Tilburg.

E-mail: kd.meusen-beekman@studie.ou.nl.

Gastcolumn



□ Joost Dijkstra

Eerlijk duurt het langst

Toetsing moet primair betrouwbaar en valide zijn, maar dat is niet voldoende. Andere criteria als transparantie en eerlijkheid mogen niet genegeerd worden. Over de eerste twee hebben we wel een idee en bij transparantie kunnen we ons ook wel wat voorstellen, maar wat is eerlijkheid? *Eerlijk gezegd* blijkt dat wat moeilijker dan ik op het eerste gezicht dacht. Toetsing moet eerlijk zijn, maar voor wie? Voor leerlingen en studenten? Of voor docenten en staf? Of eerlijk ten opzichte van de maatschappij?

Zeker, we moeten er voor zorgen dat de toets voor alle leerlingen en studenten even moeilijk is en er geen culturele of andere aspecten zijn die sommige leerlingen en studenten bevoordelen ten opzichte van anderen. Maar hoe ver moeten we daarin gaan? Moeten we te allen tijden ervoor zorgen dat verschillende groepen van studenten gelijk scoren? Mag je leerlingen en studenten wel indelen in groepen? Zodra je een indeling maakt krijg je meteen te maken met grensgevallen, want behalve een indeling in jongens en meisjes zijn er weinig (kei)harde criteria te bedenken.

Moeten de resultaten van meisjes en jongens gemiddeld gelijk zijn en is dat dan het eerlijkst voor allebei? Of mag je stellen dat op sommige punten er een verschil is tussen meisjes en jongens (of tussen anders ingedeelde groepen) en dat het dus alleen maar vanzelfsprekend is dat er verschil in prestatie is? Moeten we er dan voor zorgen dat leerlingen en studenten gelijke kansen hebben op een goed resultaat op de toets? Is het eerlijk om alle leerlingen en studenten over één kam te scheren en een gelijke toets voor iedereen te maken? Of moeten we juist zorgen voor een toets op maat per groep studenten?

Of betekent eerlijkheid dat we oprecht en eerlijk moeten zijn tegen studenten? Als beoordelaar of toetsmaker moet je even streng zijn voor alle leerlingen en studenten. We moeten ze niet verrassen met onverwachte inhoud. En als we het aan de leerlingen en studenten zelf vragen? "Dat stond niet in het boek." "Mijn docent is veel strenger dan andere docenten." Tja, als ze de toets gehaald hebben zal het wel een eerlijke toets zijn, maar als niet...

Wat is eerlijk ten opzichte van de leraren en docenten? Is het eerlijk als hun jarenlange toetsing (assessment) van een leerling of student teniet wordt gedaan met een centrale allesbepalende eindtoets? Met als gevolg dat het oordeel van de individuele docent er niet meer toe doet? Het is ook niet erg eerlijk als je steeds meer werk krijgt (of de stapel met na te kijken werken niet kleiner wordt) door het laten zakken van een student.

Als ik *goudeerlijk* ben dan denk ik dat de eerlijkheid van de toets (net als betrouwbaarheid, validiteit en de meeste andere kwaliteitscriteria) nooit voor 100% gehaald kan worden. Hoeveel werk we er ook in steken er zal altijd wel iemand zijn die zich benadeeld voelt. Dus moeten we verschillen maar accepteren. Het zal allemaal wel weer in het midden liggen en de context en het doel zullen wel weer moeten bepalen of het eerlijk is of niet. Maar toch, wat is nou eerlijk toetsen? Okay, eerlijk is eerlijk, niemand kan er tegen zijn, want zeg nou zelf: *eerlijkheid duurt het langst*.

De heer J. Dijkstra is werkzaam aan de universiteit van Maastricht.

E-mail: joost.dijkstra@educ.unimaas.nl.

Ook de vaarbewijs
examens van



staan onder
toezicht van de



□ Harry van Dijk

Aantoonbaar maken vak- bekwaamheid binnen **Achmea** **- De Wft** als katalysator

De Wet op het financieel toezicht (Wft) is ingevoerd om transparantie van financiële dienstverleners en financiële producten af te dwingen. Een van de aspecten betreft het aantoonbaar maken van de deskundigheid/vakbekwaamheid van medewerkers met klantcontacten. Voor Achmea was de wet een kans een lang gekoesterde wens op een gestructureerde en gedegen wijze van de grond te krijgen: toetsing als instrument voor professionalisering.

Wettelijk kader

Sinds 1 januari 2007 eist de wetgever via de Wft dat een financieel dienstverlener er voor zorgt dat haar werknemers die zich rechtstreeks bezighouden met financiële dienstverlening vakbekwaam zijn. Dit houdt in dat werknemers moeten beschikken over een geldig diploma of dat de bedrijfsvoering van de dienstverlener zodanig is ingericht dat naar het oordeel van de toezichthouder Autoriteit Financiële Markten (AFM), de kwaliteit van de financiële dienstverle-

ning aan consumenten en cliënten voldoende is gewaarborgd. De Wft onderscheidt de volgende vakbekwaamheidsmodulen: basis, consumptief krediet, hypothecair krediet, schadeverzekeringen, levensverzekeringen, volmacht schade en volmacht leven. Medewerkers met klantcontacten moeten examen doen in een of meer vakbekwaamheidsmodulen bij een erkende exameninstelling. Exameninstellingen worden erkend door het College Deskundigheid Financiële Dienstverlening (CDFD) dat is ingesteld door de minister van Financiën. De erkenning is gebaseerd op een audit die om de vijf jaar wordt herhaald tenzij er aanleiding is eerder een audit uit te voeren.

Voor kleinere financiële dienstverleners is het aantonen van vakbekwaamheid via examinering een praktische oplossing. Voor grotere financiële instellingen zou het opleiden en laten examineren van de vele, soms tienduizenden, medewerkers met klantcontacten tot hoge out-of-pocket-kosten leiden. Daarnaast zijn grote financiële dienstverleners in staat om een bedrijfsvoeringsmodel in te richten waarmee de vakbekwaamheid van medewerkers zodanig wordt georganiseerd en aangetoond dat het naar verwachting door de AFM zal worden goedgekeurd.

Achmea

Achmea heeft als grote financiële dienstverlener gekozen voor een eigen certificeringstraject om deskundigheid aantoonbaar te maken. Hierbij zijn de vakbekwaamheidseisen uit de Wft door Achmea tegelijkertijd gebruikt als katalysator om het ontwikkelen en meten van kennis en cognitieve vaardigheden naar een hoger plan te tillen.

Analyse van de situatie

De corporate opleidingsafdeling heeft samen met een vertegenwoordiging van de opleidingsafdelingen van de divisies de Wft-eindtermen bestudeerd. Om tot een goed fundament te komen zijn

Op basis van de Wft-eindtermen is een structuur ontwikkeld



de beschikbare functieprofielen, interne eindtermen, opleidingen, cursussen, trainingen, workshops, toetsen en cursusmaterialen van alle divisies verzameld en tegen de Wft-eindtermen gehouden. Op basis van de Wft-eindtermen is vervolgens een structuur ontwikkeld waarin zowel de door de Wft benoemde vakbekwaamheden, de Wft-eindtermen, de divisiestructuur van Achmea alsmede de bestaande eindtermen, opleidingen, cursussen en cetera met elkaar in verband werden gebracht en geordend. Deze structuur is weer-gegeven in figuur 1.

Tevens is vastgesteld waar voor Achmea de Wft-eindtermen verbreed en/of verdiept moesten worden om aan de eigen (hogere) kwaliteitseisen ten aanzien van kennis en vaardigheden te kunnen voldoen. Dit leidde tot een blauwdruk die voorziet in een onderscheid tussen de Wft-eindtermen en de divisiegerichte eindtermen, waardoor het mogelijk is om zowel eigen Wft-certificerende toetsen als niet voor de Wft-certificerende toetsen op divisieniveau af te nemen. Voor de Wft-certificerende toetsen is een CDFD-gecertificeerde exameninstelling geselecteerd, de Stichting Examens Financiële Dienstverlening (SEFD), om de toetsen te certificeren.

De uitvoering

Toetsontwikkeling

Op basis van de blauwdruk zijn cursusmaterialen en cursussen en cetera (her)ontwikkeld. Tevens zijn toetsmatrijzen ontwikkeld op basis van de Wft-specifieke en de eigen eindtermen. Binnen de opleidingsafdelingen (corporate en op divisieniveau) zijn medewerkers

De grootste verzekeraar van Nederland

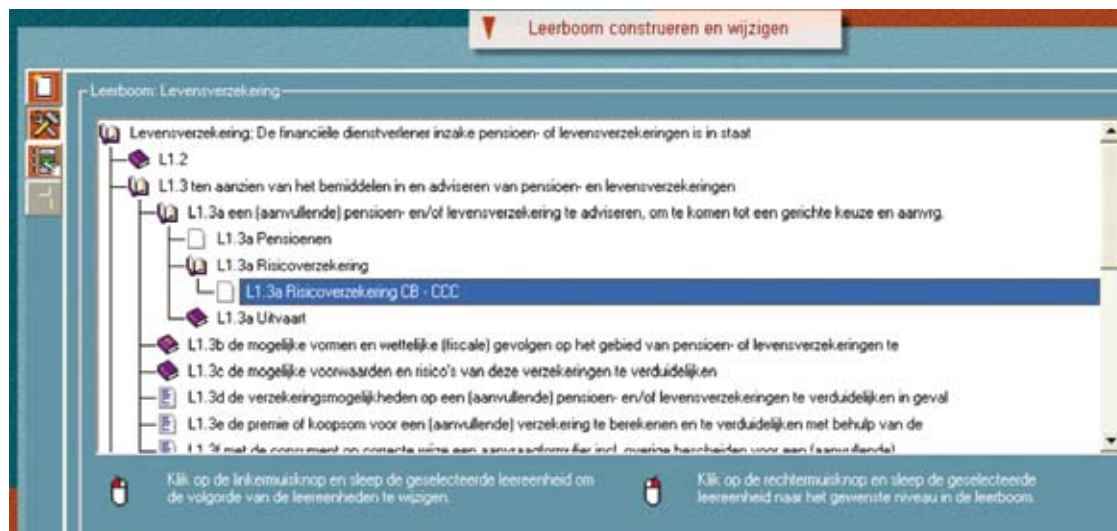
Achmea - onderdeel van het moederbedrijf Eureko - heeft bijna 22.000 medewerkers en enkele miljoenen klanten. Het huidige Achmea is in 2005 voortgekomen uit de fusie tussen Interpolis en het 'oude' Achmea. In 2007 is Achmea gefuseerd met Agis Zorgverzekeringen. Kenmerkend voor Achmea zijn de coöperatieve wortels en de betrokkenheid bij maatschappelijke vraagstukken. De Achmea merken Agis, Avéro Achmea, Centraal Beheer Achmea, FBTO, Interpolis en Zilveren Kruis zijn bij de klanten het meest bekend. Op groepsniveau is er een afdeling die voor geheel Achmea opleidingen en trainingen verzorgt. Daarnaast zijn er in de diverse bedrijfsonderdelen opleidingsadviseurs en -coördinatoren.

http://www.achmea.nl/index.html#ID=/d/over_achmea

opgeleid in het ontwikkelen van goede vragen en het interpreteren van statistische gegevens door middel van een tweedaagse cursus Toetsconstructie en een halfdaagse cursus Statistische analyse. Vanuit de divisies ontwikkelden inhoudsdeskundigen op basis van de toetsmatrijzen toetsen die door externe deskundigen van de exameninstelling getoetst zijn of ze voldeden aan de Wft-eindtermen. Na eventuele aanpassing werden de vragen toegevoegd aan de Wft-itembank.

Elektronisch toetssysteem

De toetsvragen en de toetsen worden ontwikkeld en afgenomen met behulp van een elektronisch toetssysteem. Hiervoor was al een



Figuur 1
De structuur van de
eindtermen



De itembank voor de Wft-examens kent nu ongeveer 1.600 vragen

selectie- en implementatieproject afgerond. Medewerkers binnen de opleidingsafdelingen (corporate en divisies) zijn opgeleid door middel van een eendaagse cursus in het gebruik van het toetssysteem. In totaal zijn er ongeveer 60 gebruikers (itemauteurs en toetsontwikkelaars) van het toetssysteem. Dit grote aantal gebruikers is ontstaan doordat Achmea het ontwikkelen en afnemen van vragen en toetsen bewust laag en breed in de organisatie wilde leggen, dicht bij de business (binnen elk van de zes divisies). Achmea gaat er namelijk vanuit dat het regelmatig afnemen van toetsen medewerkers aanzet tot leren. Hierbij is het formatieve karakter belangrijker dan het summatieve karakter. De Wft-toetsen zijn wel summatieve toetsen.

Wft-itembank

De Wft-itembank is volledig gescheiden van de divisies en wordt, evenals de Wft-toetsresultaten, beheerd door de centrale opleidingsafdeling. Deze afdeling stelt ook de Wft-toetsen samen op basis van Wft-toetsmatrijzen en de Wft-itembank. De itembank bestaat uit vragen die door veel verschillende inhoudsdeskundigen zijn ontwikkeld en door verschillende toetsdeskundigen zijn geredigeerd en gecorrigeerd. De itembank was in eerste instantie beperkt van omvang en voldoende voor de eerste examens. In de loop van de tijd is de itembank steeds verder uitgebreid. (Op)nieuw af te nemen examens zijn hiervoor meestal de aanleiding. De itembank voor de Wft-examens kent nu ongeveer 80 examens met ongeveer 1.600 vragen.

Toetsanalyses

De Wft-toetsen worden op divisieniveau toegewezen aan medewerkers die een Wft-examen willen doen. Nadat een Wft-examen is gemaakt vindt een statistische analyse plaats op de items en het examen als geheel: zoals moeilijkheid, onderscheidend vermogen, a-waarde (aantrekkelijkheid van de afleiders), m-waarde (gemiddelde toetsscore van de kandidaten op het gekozen alternatief) en betrouwbaarheid. Op basis van de analyse worden vragen zonodig aangepast of verwijderd. Soms heeft dit consequenties voor de resultaten. Dit wordt dan voorgelegd aan de examencommissie die volgens een vaste procedure een beslissing neemt.

Organisatie

Er is een toetsbeleid en een examenreglement ontwikkeld en een examenorganisatie ingericht.

Ondersteuning

Een en ander klinkt nu van enige afstand eenvoudig, maar dat was het project zeker niet. Achmea heeft ondersteuning gezocht bij Teelen Kennismanagement die ook het toetssysteem (TestVision)

had geleverd. Samen met die organisatie zijn antwoorden gevonden op vragen als: welke eisen worden gesteld aan goede, concrete en transparante eindtermen, hoe voldoe je aan de Wft-eisen en aan de eigen eisen, hoe structureer je dat, hoe maak je een toetsmatrijs, aan welke kwaliteitseisen en constructievoorschriften moeten vragen voldoen, welke taxonomiecode past het best, wat toetsen we met gesloten vragen, hoe bepalen we de cesuur, wat doen we met herkansingen, welke eisen stellen we aan een toetssysteem, hoe implementeren we dat systeem, hoe implementeren we het totale concept in een grote, geografisch sterk verspreide organisatie en hoe organiseren we het onderhoud en het beheer?

Tot slot

De Wft heeft binnen Achmea voor een totaal vernieuwde opleidings- en toetsstructuur gezorgd waarbinnen zowel de Wft- als de eigen eindtermen een solide basis vormen. Op basis van de eindtermen zijn toetsmatrijzen, cursusmaterialen, cursussen en toetsen ontwikkeld die een consistent geheel vormen. Medewerkers van opleidingsafdelingen hebben leren werken met eindtermen en toetsmatrijzen en kunnen kwalitatief goede toetsvragen ontwikkelen. De opleidingsafdelingen van de verschillende divisies werken autonoom en los van de corporate opleidingsafdeling. Doordat echter vanuit een gemeenschappelijke visie een gemeenschappelijk fundament is ontwikkeld, werken die visie en dat fundament door binnen de divisies zelf. Dit wordt ondersteund door het toetsbeleid en het breed gebruikte toetssysteem. Ook de eigen divisietoetsen worden serieus genomen en doen er toe. Toetsen is geen afrekeninstrument, maar een kwalitatief ontwikkelinstrument geworden, dat laag in de organisatie is geworteld, want dan gaat het leven. In feite is voor Achmea de Wft een katalysator geweest voor een grote professionaliseringsslag die al veel langer gewenst was, maar die in een geografisch sterk gedecentraliseerde organisatie met vele divisies en tienduizenden medewerkers niet zo snel en wellicht nooit met zoveel succes gerealiseerd had kunnen worden.

De heer H.J. van Dijk is programmamanager voor een aantal specifieke Achmea managementleergangen, toetsbeleid en e-learning. Daarnaast is hij (strategisch) opleidingsadviseur en duaal voorzitter van de Opleidingskring (landelijke verzekeraars). E-mail: harry.j.van.dijk@achmea.nl.

Optimale **uniforme scoringsregels** voor innovatieve vraagvormen

□ Hans Vos, Marlies Kloppenburg en Onno Tomson

Gebruikers van toetssystemen rekenen erop dat deze betrouwbaar zijn en uitgaan van de optimale, uniforme scoringsregels voor innovatieve vraagvormen. De discussie over de quotiëntregel en de ontwikkeling van de productregel voor scoring van multiple responsvragen, waarbij het aantal goede antwoorden niet wordt weergegeven, heeft eerder in dit tijdschrift laten zien waar knelpunten liggen. Dit artikel gaat verder in op scoringsregels. Hans Vos houdt zich aanbevolen voor reacties.

Een terugblik

Lampe en Eggen (2005) hebben laten zien dat op item-niveau bij een evenwichtige verdeling over de verschillende kennisniveaus de dichotome scoring doorgaans de hoogste betrouwbaarheid geeft. Op toetsniveau geeft dichotome scoring bij gemengde toetsen geen optimale score. In november 2006 vond via de TOETS-HO-Lijst een e-mail-discussie plaats over het gebruik van innovatieve vraagvormen die leidde tot een studiemiddag over dit onderwerp. Hiervan is een verslag verschenen in EXAMENS (2008) en Hans Vos publiceerde een uitwerking van de quotiëntregel (Vos & de Graaf, 2008). In een reactie op het ontwerp van de quotiëntregel formuleerde Terwisscha van Scheltinga (2009) de productregel. Bovendien heeft Vos op verzoek van Tomson namens ECABO verder onderzoek uitgevoerd om ook voor andere innovatieve vraagvormen wetenschappelijk onderbouwde scoringsregels te ontwikkelen. Dit artikel geeft het onderzoeksresultaat op hoofdlijnen weer. Het onderzoeksrapport is op te vragen bij Vos.

Innovatieve vraagvormen

Onder innovatieve vraagvormen worden gesloten vraagvormen verstaan zoals de multiple responsvraag, ordeningsvraag, matchvraag, matrixvraag en de multiple true/false vraag. Het gebruik van innovatieve vraagvormen komt door de opkomst van elektronische toetssystemen steeds vaker voor. Elektronische toetssystemen scoren deze vraagvormen automatisch. Het komt voor dat een toets met verschillende innovatieve vraagvormen gemaakt met toetssysteem A een andere totaalscore oplevert dan gemaakt met toetssysteem B. Een situatie die niet gewenst is.

Scoring

Vragen kunnen zowel dichotoom als polytoom gescoord worden. Bij dichotoom scoren krijgt het antwoord een score '1' als het item volledig juist is beantwoord en anders een score '0'. Bij een polytome scoring zijn ook tussenliggende waarden mogelijk, afhankelijk van het gegeven antwoord. True/false vragen en de kort antwoord- of invulvraag zijn per definitie dichotoom. Deze vraagvormen kunnen alleen maar óf goed óf fout beantwoord worden. Innovatieve vraagvormen kunnen dichotoom of polytoom worden gescoord. De hier behandelde scoringsregels hebben betrekking op de polytome scoring.

Bij polytome vraagvormen kan een antwoord gedeeltelijk goed zijn, maar ook gedeeltelijk fout. Als een kandidaat de kleuren van de Nederlandse vlag moet selecteren uit de vijf kleuren rood, wit, blauw, oranje en geel en hij geeft als antwoord rood en wit, dan is dat antwoord gedeeltelijk goed. Kiest hij wit, blauw en oranje dan is het antwoord gedeeltelijk goed én gedeeltelijk fout. Het is reëel om voor het foute antwoord aftrek van punten te geven. In feite wordt hiermee dus de traditionele gokkans (bij dichotome vragen) in de scoringsregel van polytome vragen al min of meer verwerkt.

Multiple responsvraag

De multiple responsvraag is geïndiceerd voor het meten van (partiële) kennis over een aantal bij elkaar behorende



Bij polytome vraagvormen kan een antwoord gedeeltelijk goed zijn, maar ook gedeeltelijk fout

onderwerpen (bijvoorbeeld de kleuren van de Nederlandse vlag selecteren uit vijf kleuren). In de stam bij de vraag kan al dan niet worden aangegeven hoeveel juiste alternatieven moeten worden gekozen.

Het aantal juiste alternatieven wordt in de stam aangegeven

Vos & de Graaf (2008) stellen voor dit type vraag de quotiëntregel voor. De quotiëntregel luidt:

$$S_{\text{quotiëntregel}} = \frac{\text{proportie goed}}{1 + \text{proportie fout}}$$

Het aantal juiste alternatieven wordt niet in de stam aangegeven

Als het aantal juiste alternatieven niet wordt aangegeven, kan de kandidaat 'blind' alle alternatieven selecteren en kent de quotiëntregel altijd 0.5 punt toe. Immers, in deze situatie geldt: proportie goed = proportie fout = 1. Terwisscha van Scheltinga (2009) heeft de volgende productregel voorgesteld om aan dit bezwaar tegemoet te komen.

$$S_{\text{productregel}} = \text{proportie goed} * \left\{ 1 - \text{proportie fout} * \left[\frac{1}{a} + \left(\frac{n-k}{n} \right) \left(1 - \frac{1}{a} \right) \right] \right\}$$

waarbij:

n = het totale aantal alternatieven;

k = het aantal juiste alternatieven;

a = weegfactor voor het meetellen van foute antwoorden ($a \geq 1$) (experimenteel is vastgesteld dat $a=2$ het beste resultaat oplevert).

Aanbeveling

De betrouwbaarheden en de discriminerende vermogens van de productregel en de quotiëntregel leveren vergelijkbare waarden op. Als het aantal juiste alternatieven is aangegeven in de stam wordt de quotiëntregel geprefereerd boven de productregel vanwege transparantie-overwegingen.

Orderingsvraag

Bij orderingsvragen moet een aantal 'objecten' in een juiste volgorde worden geplaatst (bijvoorbeeld een aantal Amerikaanse presidenten in een chronologische volgorde). Eggen & Lampe (2005) hebben vier verschillende scoringsmethoden voor orderingsvragen met K vrij te ordenen objecten (minstens 3 objecten) met elkaar vergeleken.

Hieronder wordt een scoringsvoorbeeld gegeven met vier te ordenen objecten met als juiste volgorde ABCD. Stel dat de kandidaat als antwoord 'ABDC' geeft. Hoe kan dit worden gescoord?

1. Tel de juiste objecten op absolute positie (maximaal K punten te behalen, dus hier 4 punten). Het antwoord 'ABDC' levert 2 punten op, omdat de objecten 'A' en 'B' op de juiste positie staan.
2. Tel de juiste ordening van direct naaste paren (maximaal $K-1$ punt te behalen, dus hier 3 punten). Het antwoord 'ABDC' levert 1 punt op, omdat de onderlinge volgorde van 'AB' juist is en die van 'BD' en 'DC' onjuist zijn.
3. Tel de juiste ordening van naaste paren (maximaal $K-1$ punt te behalen, dus hier 3 punten). Het antwoord 'ABDC' levert 2 punten op, omdat de onderlinge volgorde van 'AB' en 'BD' juist zijn en die van 'DC' onjuist is.
4. Tel alle juiste ordeningen (maximaal $K*(K-1)/2$ punt te behalen, dus hier 6 punten).
5. Het antwoord 'ABDC' levert 5 punten op, omdat de onderlinge volgorde van 'AB', 'AD', 'AC', 'BD' en 'BC' juist zijn en alleen de onderlinge volgorde van 'DC' onjuist is.

De door Eggen & Lampe vergeleken scoringsmethoden zijn doorgerekend met de quotiëntregel. Hierbij gaf de bovengenoemde vierde scoringsmethode de hoogste betrouwbaarheden en discriminerende vermogens.

De quotiëntregel voor de vierde methode luidt:

$$S_{\text{quotiëntregel alle juiste ordeningen}} = \frac{\text{proportie goed}}{1 + \text{proportie goed bij raden}} = \frac{\frac{S}{K(K-1)/2}}{1 + \left(\frac{S}{K(K-1)/2} \right)} = \frac{S}{K(K-1) - S}$$

Waarbij:

K = vrij te ordenen objecten;

S = aantal juiste volgordes van alle mogelijke paren van objecten.

Aanbeveling

Hanteer als scoringsmethode alle juiste ordeningen in combinatie met de quotiëntregel.

Als het aantal juiste alternatieven is aangegeven in de stam wordt de quotiëntregel geprefereerd



Matchvraag

Bij een gelijke en ongelijke matchvraag moeten steeds twee zaken met elkaar worden gecombineerd. Dit vraagtype is geïndiceerd bij het toetsen van verbanden leggen (bijvoorbeeld vier steden koppelen aan vier provincies). Lampe & Eggen (2006) scoren matchvragen door voor iedere juist gelegde relatie 1 punt toe te kennen (dus zonder aftrek voor verkeerd gelegde relaties). Teelen Kennismanagement (2006) houdt in haar scoringsregel wel rekening met aftrek voor verkeerd gelegde relaties. De quotiëntregel blijkt na onderzoek ook bij de matchvraag de hoogste betrouwbaarheid en het hoogste discriminerend vermogen te hebben.

$S_{\text{quotiëntregel ongelijke matching}} =$

$$\frac{\text{proportie goed}}{1 + \text{proportie goed bij raden}} = \frac{\frac{S}{K}}{1 + \frac{K-S}{K(N-1)}} = \frac{\frac{S}{K}}{1 + \left(1 - \frac{S}{K}\right)\left(\frac{1}{N-1}\right)}$$

Waarbij:

K = aantal startalternatieven;

N = aantal koppelalternatieven;

S = aantal goed gelegde koppelingen.

Aanbeveling

Hanteer als scoringsmethode de quotiëntregel bij ongelijke matching omdat die ook voor gelijke matching toepasbaar is.

Matrixvraag

Bij de matrixvraag moet van een aantal gegevens worden aangegeven welke kenmerken van toepassing zijn. Dit type vraag is geïndiceerd als men kennis van diverse aspecten van een onderwerp in één vraag wil toetsen. De matrixvraag is ook op te vatten als een variant op de matchvraag, waarbij startalternatieven (gegevens) moeten worden gekoppeld aan alle juiste koppelalternatieven (kenmerken). Een voorbeeld van een matrixvraag is dat van de steden Groningen, Arnhem en Zutphen moet worden aangegeven of zij een provinciehoofdstad zijn, meer dan 100.000 inwoners hebben en aan een grote rivier liggen. De matrixvraag kan beschouwd worden als een verzameling onafhankelijke multiple responsvragen (Lampe & Eggen, 2006; Teelen, 2006). Eerder is aangegeven dat de productregel de optimale scoringsregel is voor iedere

afzonderlijke multiple responsvraag als het aantal juiste koppelalternatieven niet is aangegeven voor ieder startalternatief en de kandidaat dus ieder startalternatief 'blind' aan alle koppelalternatieven kan koppelen (overal 'ja' aankruisen in de matrix). De toegekende score is dan de gemiddelde score over alle afzonderlijke multiple responsvragen.

Stel dat de kandidaat in het voorbeeld voor Groningen, Arnhem en Zutphen de volgende scores realiseert voor ieder van de drie afzonderlijke multiple responsvragen (volgens de productregel (met weegfactor $a = 2$)):

Groningen: 1/6;

Arnhem: 2/3;

Zutphen: 7/12.

Als score voor de totale matrixvraag wordt dan de gemiddelde productscore toegekend:

$$(1/6 + 2/3 + 7/12)/3 = 17/36.$$

Aanbeveling

Hanteer de gemiddelde score op basis van de productregel over de afzonderlijke multiple responsvragen.

Multiple true/false

Bij een multiple true/false vraag moet van meerdere beweringen (proposities) bij een stam worden aangegeven of deze juist of onjuist zijn. Dit type vraag is geïndiceerd bij onder meer het herkennen van begrippen. Er is een ondubbelzinnig juist antwoord en er zijn weinig plausibele alternatieven. De multiple true/false vraag is op te vatten als een speciaal geval van de (gelijke of ongelijke) matchvraag, waarbij er K startalternatieven zijn in de vorm van K true/false beweringen en er $N=2$ koppelalternatieven (true of false) zijn.

$$S_{\text{MTF}} = S_{\text{quotiëntregel (ongelijke) matching met } N=2} = \frac{\frac{S}{K}}{1 + \left(1 - \frac{S}{K}\right)\left(\frac{1}{N-1}\right)} = \frac{\frac{S}{K}}{1 + \left(1 - \frac{S}{K}\right)} = \frac{\text{proportie goed}}{1 + \text{proportie fout}}$$

Waarbij:

K = totaal aantal beweringen;

S = aantal goed aangekruiste beweringen.



De quotiëntregel heeft ook bij de matchvraag de hoogste betrouwbaarheid en discriminerend vermogen

Aanbeveling

Hanteer als scoringsmethode de quotiëntregel bij ongelijke matching met $N=2$.

Verzoek aan de lezers

Het verzoek aan de lezers is om de scoringsregels kritisch te bestuderen en een reactie te sturen naar de auteurs. Het doel is om te komen tot optimale en uniform te gebruiken scoringsregels voor innovatieve gesloten vraagvormen. Ontwikkelaars van de volgende in Nederland veel gebruikte elektronische toetssystemen worden met name uitgenodigd te reageren: Egel, Etude, Question Mark Perception, TestVision, Wintoets. Maar ook reacties van ontwikkelaars van Learning Managementsystemen met toetsfunctionaliteiten zijn welkom. Het resultaat kan zijn dat uiteindelijk de optimale scoringsregels in alle systemen worden opgenomen.

Literatuur

- Eggen, T. & Lampe, T. (2005). Een kwestie van volgorde: de scoring van ordeningsvragen. *Paper ORD 2005*. Gent, België.
- Lampe, T. & Eggen, T. (2005). Hoe scoor je de antwoor-

den? Meervoudige keuzevragen. *EXAMENS*, 2(2), 18-21.

Lampe, T. & Eggen, T. (2006). *Scoring van matchingvragen: Onderzoek naar de betrouwbaarheid van verschillende scoringsmethoden*. Arnhem: Cito.

Teelen Kennismanagement (26 oktober 2006). *Scoring matchingvraag en rangschikvraag*. Notitie uitgereikt op studiemiddag (14-07-07). Gebruikersverenigingen QMP en TestVision.

Terwisscha van Scheltinga, E. (2009). Productregel als scoringsmethode voor meervoudige keuzevragen. *EXAMENS*, 6(1), 26-27.

Vos, H.J. & de Graaf, A.M. (2008). Een nieuwe scoringsmethode voor meervoudige keuzevragen. *EXAMENS*, 5(1), 18-21.

□ De heer dr. H.J. Vos is werkzaam aan de Universiteit Twente, Faculteit Gedragwetenschappen, afdeling Onderzoeksmethodologie, Meetmethoden en Data-analyse. Mw drs. E.F.M. Kloppenburg en de heer O. Tomson zijn werkzaam bij ECABO als Adviseur examinering respectievelijk Beheerder informatiesystemen afdeling Examinering. E-mail: hj.vos@gw.utwente.nl.

Ook de examens van



staan onder
toezicht van de

Examen
kamer
kwaliteit door onafhankelijkheid

Norman Verhelst – teamwerk en elkaars taal begrijpen

□ Annemarie de Knecht-van Eekelen en Ed Kremers

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

Hoe ontwikkelde u zich tot psychometricus?

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



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

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

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



Eigenlijk ben ik autodidact op het gebied van psychometrie

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

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

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

Hoe en wanneer bent u geraakt door het onderwerp Examens?

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

Wie zijn uw leermeesters?

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

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

Wat drijft u?

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

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

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

Het schort moderne psychometrie aan fantasie



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

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

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

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

OPLM – One Parameter Logistic Model

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

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

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

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

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

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



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

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

Waar bent u het meeste trots op?

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

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

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

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

Wie beschouwt u als vooraanstaand in de examenwereld? Waarom?

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

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

Bent u nu nog actief? En waarmee?

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

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

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

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

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

De correctie van de centrale examens in Nederland is heel onprofessioneel



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

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

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

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

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

Welke boodschap heeft u voor de lezers van EXAMENS?

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

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

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

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

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

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

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

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

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

⁸<http://www.pisa.oecd.org/>

⁹<http://www.erk.nl/>

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

¹¹<http://www.ealta.eu.org/index.htm>

Gezien en gelezen

□ Karin Gerritsen-van Leeuwenkamp

Scheuermann, F., & Björnsson, J. (Eds.) (2009). The transition to computer-based assessment. New approaches to skills assessment and implications for large scale testing. Luxembourg: Office for official publications of the European Communities: European Communities. <http://publications.jrc.ec.europa.eu/repository/handle/111111111/8713>.

In september 2008 heeft het Joint Research Centre (JRC, IPSC) van de Europese Commissie samen met het Educational Testing Institute van IJsland, een expert workshop gehouden over 'The Transition to Computer-Based Assessment - Lessons learned from the PISA 2006 Computer Based Assessment of Science (CBAS) and implications for large scale testing'. De bundel die hier wordt besproken, bevat de bijdragen aan de workshop aangevuld met een aantal artikelen die verschillende aspecten van CBAS behandelen. In het bijzonder wordt aandacht gegeven aan de vergelijking tussen 'paper-pencil tests' (PPT) en 'computer based assessment' (CBT). De bundel beoogt een overzicht te geven van de onderwerpen die spelen bij de overgang naar computer-based assessments zowel op nationaal als internationaal niveau.

De 27 bijdragen in deze bundel zijn gerangschikt in de volgende vijf delen: I Assessment Needs and European Approaches; II General Issues of Computer-Based Testing; III Transition from Paper-and-Pencil to Computer-Based Testing; IV Methodologies of Computer-Based Testing; V The PISA 2006 Computer-Based Assessment of Science (CBAS).

In de veelheid van bijdragen kan elke lezer wel iets naar zijn of haar gading vinden. Deze bespreking richt zich op een aantal bijdragen in het derde deel. Hierin wordt duidelijk hoe complex de overgang van PPT naar CBT is. Om de voor- en nadelen van de overgang van PPT naar CBT in kaart te brengen zijn een risico en een financiële analyse gewenst. In de risico analyse worden de risico's, de frequentie en de impact van risico's in kaart gebracht¹. Een voorbeeld uit zo'n risico analyse is het aspect toegang tot toetsen. CBT heeft als voordeel dat studenten makkelijke toegang hebben tot de toets, maar de veiligheidsrisico's zijn een nadeel. De uitdaging is om maximale voordelen met minimale risico's te behalen. Vanuit een risico analyse moet worden besloten of en op welke wijze de geplande verandering van PPT naar CBT wordt gerealiseerd. Naast een risico

analyse is een financiële analyse nodig. Gezien de diversiteit van CBT kan er geen eenduidig antwoord gegeven worden op de vraag of CBT in vergelijking met PPT goedkoper is. Dit dient per situatie geanalyseerd te worden aan de hand van het framework dat gepresenteerd wordt door de schrijvers van het artikel².

CBT kan op verschillende wijzen worden gerealiseerd. Volgens de 'migratory approach' wordt een PPT rechtstreeks zonder aanpassingen overgezet naar CBT. Dit is efficiënt en goedkoop, maar een nadeel is dat het geen innovatief effect heeft op leren, terwijl bovendien de validiteit van de toets vermindert. Daarom is de 'transformational approach' veel wenselijker. Daarbij wordt een geheel nieuw computergestuurd assessment ontworpen, met innovatieve toetsvormen die de mogelijkheden van CBT benutten. CBT beïnvloedt de validiteit van de toetsen doordat naast kennis van de leerstof ook kennis van ICT wordt gemeten. Studenten die ervaren zijn met ICT scoren beter dan studenten zonder die ervaring.³ Meisjes scoren significant beter op een PPT dan op een CBT, bij jongens is geen significant verschil gevonden.⁴

In de bundel worden verschillende onderwerpen over PPT en CBT behandeld. Daardoor is de bundel bruikbaar voor lezers van *EXAMENS* die zich breed willen oriënteren op de verschillende aspecten van de overgang van PPT naar CBT. Lezers die een overgang van PPT naar CBT daadwerkelijk binnen een organisatie willen realiseren kunnen met behulp van deze publicatie aanvullende literatuur zoeken.

Mw. K.J. Gerritsen-van Leeuwenkamp is docent en voorzitter van de onderwijs-toetscommissie bij Hogeschool Saxion. E-mail: k.j.vanleeuwenkamp@saxion.nl.

¹Van Lent, G. Risks and benefits of CBT versus PBT in High-Stakes Testing: Introducing key concerns and decision making aspects for educational authorities. pp. 83-91.

²Farcot, M., & Latour, T. Transitioning to computer-based assessments: A question of costs. pp. 108-116.

³Kiris-Papadakis, K., & Kollias, A. Reflections on paper-and-pencil tests to eAssessments: narrow and broadband paths to 21st century challenges. pp. 99-103; Meijer, R. Transition to computer-based assessment motivations and considerations. pp. 104-107.

⁴Csapo, B., Molnar, G., & Toth, K. R. Comparing paper-and-pencil and online assessment of reasoning skills: a pilot study for introducing electronic testing in large-scale assessment in Hungary. pp. 120-125.

Literatuur

A re-definition of belonging? Language and integration tests in Europe

R.van Oers, E. Ersboll en D. Kostakopoulou
Leiden: Martinus Nijhoff, 2010
ISBN: 9789004175068

Cognitive diagnostic assessment for education: theory and applications

J.P. Leighton en M.J. Gierl
Cambridge: Cambridge University Press, 2007
ISBN: 9780521684217

Educational assessment in the 21st century: connecting theory and practice

C. Wyatt-Smith en J.J. Cumming
Dordrecht: Springer, 2009
ISBN: 9871402099632

Elements of adaptive testing

W.J. van der Linden en C.A.W. Glas
New York: Springer, 2010
ISBN: 9780387854595

Item response theory

C. DeMars
Oxford: Oxford University Press, 2010
ISBN: 9780195377033

The handbook of language teaching

M.H. Long en C.J. Doughty
Chichester: Wiley-Blackwell, 2009
ISBN: 9781405154895

Unlocking assessment: understanding for reflection and application

S. Swaffield
London: Routledge, 2008
ISBN: 9780415453134

De koppeling van centrale examens Nederlands aan referentiekaders (2010)

T. Meestringa en H. de Vries
Levende talen Tijdschrift 2010;11(2):13-21

Agenda

2nd Annual European-Association of Test Publishers (E-ATP) Conference, 29 september tot 1 oktober 2010, Barcelona.

Tweede Europese conferentie met als doel: inform industry professionals about the latest developments and research projects in the field of educational and performance assessment. (www.eatpconference.org)

RCEC Workshop on IRT and Educational Measurement, 12 - 14 oktober 2010, Enschede. (www.rcec.nl/irtworkshop)

RCEC-Cursus: Item response theorie in de toetspraktijk, 21 oktober 2010, Utrecht. (www.rcec.nl)

AEA-Europe Annual Conference, 4 - 6 november 2010, Oslo.

Thema: Assessment as an integral part of learning. (www.aea-europe.net)

Najaarsconferentie voor de tweede fase: thema's, trends en toetsing, 11 november 2010, Utrecht. (www.slo.nl)

RCEC-Cursus: Analyseren van toetsen en examens met behulp van klassieke testtheorie, 15 november 2010, Utrecht. (www.rcec.nl)

Online Educa Berlin (OEB), 1 - 3 december 2010, Berlijn.

Belangrijkste event over ICT en onderwijs in Europa. (www.online-educa.com)

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

- Eindexamenresultaten van allochtone leerlingen
- Rekentuin.nl: combinatie van oefenen en toetsen
- Eerste ervaringen met een interactief beeldschermexamen voor het vmbo
- Kopstukken uit de examenwereld: Dato de Gruijter

EXAMENS 4, 2010 verschijnt in november

Wie zijn wij?



Het tijdschrift *EXAMENS* kent een rubriek *Wie zijn wij*. Een rubriek met als doelstelling de lezer te informeren over ins en outs van relevante organisaties en bedrijven voor de examensector. Gerelateerd aan segmenten uit de markt nodigen we steeds drie organisaties uit om invulling aan deze rubriek te geven.

We hanteren de volgende segmenten:

- exameninstellingen;
- facilitaire bedrijven;
- examenproducten;
- overigen.

De organisaties presenteren zich door middel van een beschrijving van een vijftal vaste kernpunten:

- de organisatie;
- de missie;
- taken/producten/diensten;
- het werkveld;
- algemene gegevens.

Ieder kernpunt heeft maximaal vijftig woorden.

Heeft u interesse om in deze rubriek opgenomen te worden stuur dan een mail volgens bovenstaande specificatie en een voor print geschikt digitaal logo naar: info@kloosterhof.nl.

De uitgever en de redactie behouden zich het recht voor om bijdragen te weigeren of te wijzigen indien mocht blijken dat de inhoud niet in overeenstemming is met de doelstellingen van het vakblad of kennelijke onjuiste informatie bevat die geen recht doet aan een juiste voorlichting van de lezers van het tijdschrift. De uitgever bepaalt de volgorde waarin publicaties zullen worden opgenomen.

Organisatie

Wij zijn een jong, onafhankelijk examenbureau dat zich richt op het aanbieden van examens op markten waar een eenzijdig of onvolgende aanbod is. Wij doen dit met name in de Horeca branche en aanverwante sectoren. Maar zijn bijvoorbeeld ook actief als aanbieder van het examen Examinator.

Missie

Cinses is van mening dat een kandidaat het met een examen vaak al moeilijk genoeg heeft. We maken het examen niet makkelijker - uiteraard -, maar we hebben klantvriendelijkheid hoog in het vaandel staan en streven naar een zo hoog mogelijke service aan kandidaten. Dit uit zich onder andere in examen doen waar je wilt, wanneer je wilt en het snel verstrekken van uitslagen via de website.

Producten en Diensten

Het leermeesterexamen is ons belangrijkste product, officieel gebruiken we de benaming *Cinses LMP- diploma*®. Het Leermeester, Mentor, Praktijkbegeleider diploma, moet je als opleider in een leerbedrijf in de horeca sector behalen als je leerlingen uit het MBO-onderwijs in je leerbedrijf wilt opleiden. Om dit diploma te behalen dient een kandidaat drie onderdelen van het examen te behalen. Het examen bestaat uit, een schriftelijke toets, praktijk: het geven van een instructie aan een leerling een beoordeling van een leerling uitvoeren en een portfolio-beoordeling. In het examen komt het hele traject van het aannemen van een leerling tot het examineren van zijn competenties aan bod.

Werkveld

Ons werkveld ligt zoals duidelijk mag zijn vooral in de horeca en daarbij is het logisch dat dit vooral het beroepsonderwijs betreft.

Gegevens

Cinses Examinering en Certificering
Nassaulaan 19
2514 JT Den Haag
T 070-3462095
E info@cinses.nl
www.cinses.nl



Organisatie

SVH (Stichting Vakbekwaamheid Horeca) stimuleert de professionaliteit en kwaliteit van iedereen die leert, werkt of onderneemt in de horeca.

In opdracht van sociale partners voeren wij opdrachten uit. Daarnaast ontwikkelen we op aangeven van de horecabranche producten en diensten en verbinden we partijen met elkaar. We werken hierbij nauw samen met sociale partners, andere brancheorganisaties, opleiders en beroepsverenigingen.

Missie

Ons oogmerk is dat de branche beschikt over voldoende en goed gekwalificeerde medewerkers.

We promoten het horecavak en bieden producten en diensten aan die vakbekwaamheid stimuleren. Daarnaast fungeren we als vraagbaak, dienstverlener, initiator en kwaliteitsbewaker.

Producten/diensten

Naast de licentie-examens Sociale Hygiëne en Horecaportier, bieden wij beroepsexamens waaronder Kok en Gezel en ondernemers- en managementexamens zoals Leermeester en Basisvakbekwaamheid Horeca. Ter voorbereiding op deze examens leveren we verschillende cursusboeken, examentrainingen en vakboeken. Ook bieden we digitale producten waaronder de online kennisbank Passie voor Horeca met leermateriaal in de vorm van filmpjes, beelden en teksten. Een volledig overzicht van onze producten en diensten kunt u vinden op www.svh.nl.

Werkveld

SVH is hét aanspreekpunt voor iedereen die leert, werkt of onderneemt in de horeca. Heeft u vragen, suggesties of ideeën over leren, werken of ondernemen in de horeca? Neem dan contact op met onze Klantenservice.

Gegevens

SVH
Postbus 303
2700 AH Zoetermeer
T 0900 - 1402 (€ 0,15 p/min)
E info@svh.nl
www.svh.nl

Aequor

Organisatie

Aequor ontwikkelt de kwalificatiedossiers voor het groene onderwijs. Daarnaast verzamelen, analyseren en ontsluiten we arbeidsmarkt- en onderwijsinformatie. Hiermee hebben we zicht op de trends, kansen en knelpunten op de arbeidsmarkt. Kennis die we inzetten voor de afstemming van het onderwijsaanbod op de behoefte aan gekwalificeerd personeel in het bedrijfsleven.

Missie

Aequor is het kenniscentrum bedrijfsleven en beroepsonderwijs voor natuur, voeding en leefomgeving. Wij stimuleren het ontwikkelen en opleiden van mensen in organisaties. Hiermee willen we bijdragen aan het vergroten van de slagkracht van bedrijven.

Producten/diensten

Kennisverspreider, adviseur, maatwerkdienstverlener, analist, verzamelaar en gesprekspartner: Aequor is veel en doet veel voor de bedrijfstakken die actief zijn met voedselproductie en het optimaliseren van de leefomgeving. Zo adviseren we bedrijven over werkgelegenheid en personeelsontwikkeling. We verspreiden kennis die nodig is om als ondernemer bij te blijven en de eigen bedrijfsvoering te verbeteren. En we onderhouden contacten met bedrijven, onderwijsinstellingen en de overheid.

Werkveld

Alle bedrijven die actief zijn in de land- en tuinbouw, recreatie en leefomgeving, en de voedingsindustrie. Daarnaast zijn we de natuurlijke gesprekspartner van aoc's, het ministerie van Landbouw, Natuur en Voedselkwaliteit, groene branche- en werkgeversverenigingen en vakbonden. In de regio werken we samen met gemeenten, UWV WERKbedrijven, sociale diensten, kamer van koophandel en andere partijen.

Gegevens

Aequor
Horaplantsoen 18
6717 LT EDE
Postbus 601
6710 BP EDE
T (0318) 64 95 00
F (0318) 64 19 18
E info@aequor.nl

NVE congres november 2011: Competentiegericht toetsen en beoordelen

Na de implementatie van het competentiegerichte onderwijs is alle aandacht nu gericht op een goede uitvoering van het competentiegericht toetsen. De kwaliteit van dergelijke toetsen is afhankelijk van de professionele wijze van samenstellen, afnemen en beoordelen. Welke protocollen zijn van belang? Welke eisen zijn te stellen aan de constructeur, de assessor en de vaststeller? Wat is het verschil tussen een assessor en een examiner? Het zijn deze thema's die de NVE tijdens het nationale examencongres in 2011 samen met u wil bespreken en uitdiepen.

Professionalisering van de examenfunctionarissen

Door de professionaliseringswerkgroep is het bestuur gevraagd als NVE de lead te nemen in het ontwikkelen van profielen voor verschillende examenfunctionarissen waarmee zowel het bekostigde onderwijs als ook de private sector kan werken ten einde de toetsing dan wel de examine-

ring verder te professionaliseren. Dit verzoek van de werkgroep is door het RCEC uitgewerkt. Er is inmiddels een conceptrapportage aangereikt met verschillende aanpakken voor een verdere uitwerking. Het bestuur heeft er in eerste instantie voor gekozen om bij de functieprofielen uit te gaan van vakbekwaamheidsmatrices aan de hand waarvan duidelijk is wat iemand moet kennen en kunnen, de plaats die de functionaris heeft binnen het proces, de minimale vereisten voor examens en de voorwaarden voor een professionele certificatie.

Met de professionaliseringswerkgroep zal binnenkort over een verdere uitwerking worden gesproken.

Ook in het kader van EVC wordt nadrukkelijk gesproken over te stellen eisen en de professionalisering van in te zetten assessoren. Daarbij is een verwijzing naar de NVE gemaakt om te zien of in gezamenlijkheid tot gelijksoortige profielen kan worden gekomen.

Onderwijsraad: waarde(vastheid) van diploma's

Naar aanleiding van vragen uit de Tweede Kamer voert de Onderwijsraad momenteel een onderzoek uit naar de waarde(vastheid) van diploma's waarbij ook aan de NVE als een van de relevante partijen in het veld met expertise op dit gebied gevraagd is informatie te verstrekken en mee te denken. In het najaar wordt het advies van de Onderwijsraad over dit onderwerp verwacht.

Cursus

Ook dit jaar (2010/2011) zal er in het kader van de verdere professionalisering weer een viertal cursussen door

de NVE worden opgezet. Conform de doelstelling van de vereniging zijn de cursussen bedoeld om de leden te ondersteunen bij hun verdere professionaliseringsproces.

Het is van belang hiervoor de website van de NVE en berichtgevingen via de mail in de gaten te houden.

Bedrijfsbezoeken

28 september 2010:

Bedrijfsbezoek bij PTCplus te Barneveld. PTCplus is ontstaan vanuit verschillende agrarische onderwijsinstellingen en is vandaag de dag een moderne praktijktrainer op het gebied van plant, dier en techniek. Aandachtsgebieden zijn het verzorgen van (praktijk)onderwijs en examinering op het brede vlak van recreatiedieren, dierenspecialzaak, pluimvee- en varkenshouderij en dierenartsenpraktijk.

18 november 2010:

Bedrijfsbezoek bij TNO te Soesterberg. Bezoek aan o.a. één of meer TNO-simulatoren voorafgegaan door een presentatie van TNO/Cito. Tijdens dit bedrijfsbezoek vindt ook de Algemene Ledenvergadering plaats waar de begrotingsplannen met u worden besproken.

Noteert u dus alvast deze data.

Bedrijfsbezoeken beginnen meestal rond 13:30 uur en zijn alleen toegankelijk voor leden van de NVE.

**Bent u nog geen lid? Meld u dan aan via onze website:
www.nvexamens.nl.**

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



Inspectie van het Onderwijs
Ministerie van Onderwijs, Cultuur en
Wetenschap

INSPECTEUR BVE Examens

Wij zoeken voor de Inspectie van het Onderwijs een inspecteur die strategisch inzicht heeft, communicatief vaardig en analytisch sterk is, een visie heeft op en kennis heeft van examinering in het middelbaar beroepsonderwijs en uitstekend kan rapporteren.

Functie-inhoud

Als inspecteur lever je een verantwoorde bijdrage aan het uitvoeren van het toezicht, op basis van het toezichtkader dat voor de sector is vastgesteld. Daarnaast draag je in ruime mate bij aan de ontwikkeling van nieuwe werkwijzen, aan waarderings- en toezichtkaders en de evaluatie hiervan. Je werkt mee aan het verkrijgen van draagvlak voor nieuwe ontwikkelingen, zowel intern als extern.

Als inspecteur voer je onderzoek uit, waarbij je je oordeel uiteindelijk baseert op onder meer documenten die de scholen aanleveren, gesprekken die je hebt met colleges van bestuur, de directeur van een exameneenheid en leden van de vaststellingscommissie. Ook betrek je adviezen van derden in je besluitvorming. Uiteindelijk bespreek je je bevindingen met betrokken partijen en leg je het oordeel vast in een openbaar rapport. Tijdens het onderzoek ben je van begin tot eind verantwoordelijk voor alle stappen en alle contacten die gelegd moeten worden én je bewaakt de voortgang.

Het examentoezicht en het onderwijstoezicht worden stapsgewijs geïntegreerd. Je hebt de interesse en de capaciteiten om over ruim een jaar ook het toezicht op de onderwijskwaliteit uit te voeren.

De Inspectie van het Onderwijs is gevestigd in Utrecht. Dit is ook je standplaats. Het team BVE Examens is verantwoordelijk voor het examentoezicht op het middelbaar beroepsonderwijs (ROC's, AOC's, vakscholen en niet bekostigde instellingen) en bestaat uit ruim 20 inspecteurs en 8 medewerkers toezicht.

Functie-eisen

Je hebt een academisch werk- en denkniveau en beschikt over uitstekende communicatieve vaardigheden. De instellingen verwachten immers een stevige en deskundige gesprekspartner. Je kunt met gezag complexe gesprekken voeren. Ook kun je een diversiteit van beoordelingen van brede onderwerpen met een grote mate van complexiteit uitvoeren, waarbij het afbreukrisico groot is. Je schriftelijke vaardigheden zijn uitstekend. Je bent aantoonbaar goed in het doelgericht sturen en leiding geven aan onderzoeken in het kader van regulier toezicht, aan (grote) onderzoeksprojecten uit het jaarwerkplan én aan door de directeur ingestelde ontwikkelwerkgroepen. Omdat je te maken kunt krijgen met tegenstrijdige belangen en of weerstanden, beschik je ook over strategisch inzicht, geduld, tact en diplomatie om hier effectief doorheen te kunnen breken. Bestuurs- en organisatiesensitiviteit zijn essentieel voor een goede uitvoering van deze functie. Uiteraard ben je analytisch zeer vaardig en heb je een scherpe oordeelsvorming. Je bent professioneel, integer, werkt zelfstandig maar ook in teamverband en je kunt je oordeel to the point en gestructureerd schriftelijk weergeven en vastleggen. Tevens ben je resultaatgericht, flexibel en sta je stevig in je schoenen. Je bent op de hoogte van de meest recente wet- en regelgeving en je hebt ruime kennis van examinering. Je bent goed in staat beschikbare informatie diepgaand te analyseren en ontwikkelingen in de examinering te vertalen naar de wijze waarop het toezicht moet worden uitgevoerd. Je vindt het geen probleem om veel op pad te zijn.

Aanstelling en honorering

Het betreft voorsnog een tijdelijke aanstelling voor de duur van maximaal drie jaar. Het salaris bedraagt maximaal € 6.138,92 bruto per maand bij een volledige werkweek van 36 uur (schaal 14 BBRA 1984). De inschaling is afhankelijk van opleiding en werkervaring.

Informatie

Informatie over de Inspectie van het Onderwijs kun je vinden op onze website: <http://www.onderwijsinspectie.nl>.

Voor informatie kun je contact opnemen met Gerard Bukkems, directeur Toezicht BVE Examens (inhoudelijke informatie) of Wafaa Hdaïda, P&O adviseur (informatie over sollicitatieprocedure), via telefoonnummer 088-6696000.

Sollicitatieprocedure

Stuur je sollicitatie per e-mail, onder vermelding van de functienaam, naar wervingenselectie@owinsp.nl. Je kunt reageren tot en met 24 september.

Kent u Cito al?

De meeste mensen kennen Cito. Het is een naam die synoniem staat voor goed en eerlijk toetsen en beoordelen. De Eindtoets Basisonderwijs, ofwel de Citotoets, is onze bekendste toets. Daar zijn we trots op. Maar Cito doet veel meer. We meten persoonlijke ontwikkeling in het hele onderwijs, en in het bedrijfsleven. We geven advies, verzorgen trainingen en onderzoeken de kwaliteit van het onderwijs. Zowel in binnen- als buitenland.

www.cito.nl



Cito maakt wereldwijd werk van goed en eerlijk toetsen en beoordelen. Met de meet- en volgmethode van Cito krijgen mensen een objectief beeld van kennis, vaardigheden en competenties. Hierdoor zijn verantwoorde keuzes op het gebied van persoonlijke en professionele ontwikkeling mogelijk. Onze expertise zetten we niet alleen in voor ons eigen werk maar ook om advies, ondersteuning en onderzoek te bieden aan anderen.



zeker weten