

Nieuwe examens havo/vwo, nieuwe vraagvormen

Onderzoek naar de objectiviteit van de beoordeling van open vragen moderne vreemde talen

Met de invoering van de Mammoetwet werd de vertaling in het centraal examen mvt vervangen door meerkeuzevragen over een aantal teksten. De beoordeling van de prestaties van de leerlingen werd hierdoor objectief, maar er was ook kritiek op de 'eenzijdige' vraagvorm. Vanaf 2000 voor het havo en vanaf 2001 voor het vwo zullen de centrale examens naast gesloten ook open vragen bevatten. Het examen wordt hierdoor meer gevarieerd. Dat is een voordeel. Daar staat echter tegenover dat objectiviteit bij open vragen niet volledig gewaarborgd kan worden. Om een indruk te krijgen van de invloed van verschillen in beoordeling werd een onderzoek uitgevoerd aan de hand van de havo-voorbeeldexamens Frans en Duits. De verschillen tussen beoordelaars die in dit onderzoek werden gevonden, waren groter dan verwacht.

Sinds het begin van de jaren zeventig bestaan de centrale examens voor de moderne vreemde talen uit een aantal teksten met vijftig meerkeuze-opgaven. De invoering van dit examen betekende een radicale verandering, niet alleen van het examen zelf, maar ook van het voorafgaande onderwijs. Oefening in het vertalen van teksten uit een vreemde taal werd oefening in het begrijpen van teksten, gevolgd door een examen waarin de leerlingen niet meer zelf hoefden te formuleren. Tegen de vertaling als onderdeel van het onderwijsprogramma bestonden verschillende bezwaren. De twee voornaamste waren dat vertalen als taaloefening gezwog op de centimeter is, waarbij het onderkennen van hoofdzaken en bijzaken in teksten onvoldoende aandacht krijgt, en dat de vereiste weergave in goed Nederlands een formidabel struikelblok is dat niets met vreemde-taalvaardigheid te maken heeft. Wat de examinering betreft was bovendien duidelijk geworden dat het cijfer voor een vertaling niet alleen uitdrukking gaf aan de prestatie van een leerling, maar tevens beïnvloed werd door de wijze van beoordeling (mild, streng, slordig...). Na onderzoek werd besloten de vertaling te vervangen

door tekstbegrip-toetsen met vragen in een objectieve vorm (Gras, 1969).

Nu, bijna drie decennia later, zien de tekstbegripexamens er uiterlijk nog bijna net zo uit als bij de invoering ervan, want het gaat nog steeds om teksten met vijftig meerkeuze-opgaven. Toch er is er wel wat veranderd. Het meest in het oog springende verschil is het cloze-onderdeel (de invultekst). Verder valt het grotere aantal teksten en de grotere variatie in lengte van teksten op. Wat de opgaven betreft is het vaste vierkeuze-item gevarieerd met items met drie tot zes alternatieven. De belangrijkste verandering heeft zich geleidelijk voltrokken: de vraagstelling is globaler geworden dan in de beginjaren van de tekstbegripexamens, dat wil zeggen meer gericht op begrip van relatief grote tekstdelen (bijv. op logische verbanden in teksten en op de inhoudelijke functie van tekstdelen).

De betrokken docenten hebben niet altijd onverdeeld gunstig over deze examens geoordeeld. Vooral het backwash-effect (het effect op het onderwijs) is vaak tamelijk negatief beoordeeld. Docenten die nog de vertaling als examen meegemaakt hadden, waren gewend aan het idee dat je moest trainen zoals er getoetst werd. In de jaren zeventig leefde dan ook vrij algemeen de opvatting dat een goede training van tekstbegrip bestond uit woorden leren en het laten oplossen van meerkeuze-opgaven bij teksten. Er was zelfs nauwelijks ander oefenmateriaal verkrijgbaar. Het gevolg was vaak een wat monotone trainingspraktijk en frustraties over bepaalde leerlingen die weinig uitvoerden omdat ze in de gaten hadden dat ze op het examen wel een voldoende zouden halen (ten tijde van de vertaling bleven leerlingen veel langer in het ongewisse over hun kansen op het examen). Na verloop van tijd won bij sommigen het inzicht veld dat een geschikte toetsvorm niet automatisch gebruikt kan worden als didactisch materiaal, dat oefening en toetsing verschillende doelen dienen en dat oefening gebaat is bij gevarieerde, waaronder open, opdrachtvormen. Toch maakten Van Els en Buis nog in 1987 (p. 20) melding van frequent gebruik van examens als oefenmateriaal, ondanks de matige waardering van docenten voor het gebruik van meerkeuzetoetsen voor dit doel.

Nieuwe examens

Bij de reserves van sommige docenten met betrekking tot de bestaande tekstbegripexamens speelt

waarschijnlijk onduidelijkheid of ontevredenheid aangaande de doelstellingen van deze examens een rol. Uit commentaren blijkt, kort gezegd, dat docenten vooral hechten aan toetsing van kennis die specifiek is voor de vreemde taal. De regelmatig naar voren gebrachte kritiek dat het bij de tekstbegripexamens teveel gaat om 'algemene intelligentie' duidt daarop. Het nog geldende examenprogramma geeft slechts aan dat het begrip van teksten getoetst moet worden, zonder nadere explicatie. Wellicht was en is er om deze reden onvoldoende besef dat tekstbegrip, ook in een vreemde taal, niet alleen afhangt van kennis van woorden en de structuur van een taal, maar mede van 'algemene' kennis (en vaardigheid in het gebruik daarvan), zoals kennis van de opbouw en argumentatiestructuur van teksten, kennis van stijl-middelen en schrijfdelen en kennis van de wereld. Het nieuwe examenprogramma geeft wél een invulling aan de leesvaardigheidsdoelstelling. In vijf eindtermen en enige nadere specificaties is een omschrijving gegeven. Van leerlingen wordt, volgens deze eindtermen, verwacht dat zij:

- specifieke informatie in teksten kunnen opzoeken;
- de hoofdgedachte van teksten en tekstdelen kunnen bepalen;
- belangrijke aspecten van de inhoud begrijpen;
- begrip hebben van de structuur van een tekst;
- begrip hebben van bepaalde retorische aspecten.

Weliswaar zijn deze eindtermen en ook de specificaties tamelijk globaal geformuleerd – zij zeggen bijvoorbeeld weinig over de moeilijkheidsgraad van teksten – maar zij geven in elk geval een richting aan en zij kunnen als basis dienen voor communicatie over de examens. Zo kan van de opgaven aangegeven worden welke eindterm zij toetsen. Zeer belangrijk lijkt ons verder dat tegelijk met de invoering van de nieuwe examenprogramma's aan alle scholen een syllabus moderne vreemde talen (Syllabus, 1998) is toegestuurd met ruime uitleg over deze programma's, voorzien van voorbeeldopgaven en (per vak en examenniveau) van complete voorbeeldexamens. Bij beschouwing van de nieuwe examenprogramma's en de nieuwe examens blijkt dat het, anders dan bij de vernieuwing in het begin van de jaren zeventig, niet gaat om een radicale verandering. Alleen de eerste eindterm (het opzoeken van specifieke informatie) is nieuw, althans in de uitwerking die ervan gegeven wordt: simulatie van een 'echte' leessituatie. Bij een tekst wordt een bepaalde informatiebehoefte aangegeven en de leerling moet, gelet op deze behoefte, op een efficiënte manier (zonder de hele tekst te lezen) naar een antwoord zoeken. Nieuw is verder dat leerlingen bij het maken van het examen een woordenboek mogen gebruiken en tenslotte zijn er veranderingen wat betreft de vorm van de opgaven: ongeveer eenderde deel van de opgaven zal een open vraagvorm hebben. In dit artikel gaan wij nader in op het laatste aspect van de nieuwe examens: de open vragen. Voor nadere informatie over de inhoud van deze examens verwijzen we naar een recent artikel in *Levende Talen* (Samson, 1998) en de reeds genoemde syllabus voor de moderne vreemde talen.

Open en gesloten vragen: soorten, voor- en nadelen

Docenten uiten met de regelmaat van de klok bezwaren tegen de meerkeuze-opgaven. Het belangrijkste punt daarbij is waarschijnlijk dat men een discrepantie ervaart tussen de hoogte van scores en het feitelijke begrip van teksten. De indruk bestaat soms dat scores door informatie uit de vragen zelf een geflatteerd beeld geven van wat leerlingen van een tekst begrijpen. Informatie in vragen met betrekking tot het gewenste antwoord is ongewenst, althans onbedoelde informatie. Bij meerkeuze-opgaven valt zulke informatie echter niet altijd te vermijden. Dat kan een nadeel zijn. Maar meerkeuze-opgaven hebben ook belangrijke voordelen.

In de nieuwe examens zijn zowel open als gesloten vragen toegestaan. De opname van open vragen lijkt ons een belangrijke, zo niet de belangrijkste, verandering in de nieuwe examens. Door de diversiteit van vraagvormen wordt het, beter dan in het verleden, mogelijk de optimale vraagvorm te kiezen, de voordelen van een vorm te benutten en de nadelen zoveel mogelijk te vermijden. Aansluitend laten we kort de mogelijke vraagvormen de revue passeren.

Wat betreft de vorm zijn toetsopgaven te onderscheiden in twee typen: gesloten en open opgaven. Bij gesloten opgaven is het antwoord altijd in enigerlei vorm gegeven en het gaat altijd om een keuze uit twee of meer mogelijkheden. Bij open opgaven moeten de leerlingen zelf het antwoord formuleren. Op grond van de lengte vallen de open vragen uiteen in drie soorten:

- kort-antwoordvragen (het antwoord bestaat uit één woord, enkele woorden of een zin);
- lang-antwoordvragen (het antwoord bestaat uit enkele zinnen);
- opstelvragen of essay-vragen (het antwoord bestaat uit een samenhangend stuk tekst).

Ook van de gesloten vraagvorm zijn er diverse soorten en varianten:

- meerkeuzevragen (vragen met twee of meer alternatieve antwoorden);
- combinatievragen (uit reeksen gegevens moeten de bij elkaar passende aangegeven worden);
- classificatievragen (elementen van een tekst moeten geclassificeerd worden in gegeven categorieën);
- herordeningsvragen (gegevens moeten volgens een bepaald criterium worden geordend), enz.

In de reeds genoemde syllabus zijn voorbeelden van deze vragen te vinden.

Welke soorten vragen zijn in de nieuwe examens te verwachten? De examenprogramma's bevatten hierover geen aanwijzingen. Beslissingen op dit gebied horen daarom tot de beleidsruimte van de CEVO (Centrale Examencommissie Vaststelling Opgaven). Bij de open vragen zal het voorsnog vrijwel alleen gaan om kort-antwoordvragen. Lang-antwoordvragen zijn niet uitgesloten, opstelvragen wel. Wegens de nadelen die lang-antwoordvragen kunnen hebben (zie hieronder), zal daarvan slechts spaarzaam gebruik gemaakt worden. Wat betreft de gesloten opgaven zijn alle soorten toegestaan met uitzondering van tweekeuze-opgaven. Dit betekent dat de waar/onwaarvraag (een uitspraak waarvan aangege-

Grandes Randonnées

Lecteur paisible et bucolique,

Sobre et naïf homme de bien,

Jette ce livre saturnien.

Orgiaque et mélancolique.

Si tu n'as fait ta rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Chef d'œuvre de la rhétorique

Een compleet succes



Studiewijzers in de werkboeken

Geïntegreerde courseware op cd-rom

Ondersteuning bij de organisatie van het leerproces

LEZEN WVO

Wij werken samen met u aan beter onderwijs



ven moet worden of deze waar of onwaar is) niet als zelfstandige vraag mogelijk is. Dit vraagtype mag echter wel gebruikt worden in de vorm van de – reeds voorkomende – stellingvraag (van twee of drie uitspraken moet in één item worden aangegeven of zij goed of fout zijn).

In de nieuwe examens blijft bij de meerkeuze-opgaven het uitgangspunt gehandhaafd dat er altijd één alternatief goed is. Varianten als geen of meerdere goede alternatieven zijn dus uitgesloten. Dat geldt evenzeer voor de variant die zekerheidsscoreing wordt genoemd, waarbij een leerling moet aangeven al dan niet zeker te zijn van de juistheid van een gekozen alternatief. Met deze beperkingen – geen zekerheidsscoreing en nooit meer dan één alternatief goed – is naar onze mening de duidelijkheid van de vraagstelling voor de leerling gediend. De vorm van vragen moet zo eenvoudig mogelijk zijn; dat is een eerste uitgangspunt voor het formuleren van vragen. Het tweede uitgangspunt is dat een vraag moet toetsen wat beoogd wordt en niet meer of minder. Dit klinkt buitengewoon vanzelfsprekend, maar de realisering is bepaald niet altijd simpel. Om te beginnen is het nodig dat de vraagsteller precies weet wat (welke inhoud) hij of zij wil toetsen. Maar een tekst is een complex geheel, elementen van de inhoud zijn met elkaar verweven, en daardoor is het risico groot

dat bij een vraag onbedoelde aspecten van de tekst meespelen, of zelfs buitentekstuele informatie die niet zonder meer als bekend verondersteld kan worden. Een voordeel van gesloten vragen is dat hierbij ongewenste elementen makkelijker zijn buiten te sluiten dan bij open vragen. Er is immers meer sturing mogelijk. Mede daarom bieden gesloten vragen ruimere mogelijkheden tot genuanceerde vraagstelling dan open vragen. Maar aan het voordeel kan ook een nadeel kleven: de sturing door middel van vraag en alternatieven kan verder gaan dan eigenlijk bedoeld wordt, zodat uit de vraag en de antwoorden informatie te halen is met betrekking tot de interpretatie van de tekst. Kennis van de wereld en logica – geen vakspecifieke vaardigheden – kunnen daarbij een belangrijke rol spelen. Kortom, met open vragen wordt over het algemeen met de vraag minder informatie 'weggegeven' dan met gesloten vragen. Conceptueel zijn open vragen in principe te prefereren, zij geven een zuiverder beeld van wat een leerling van een tekst begrijpt, omdat zij om zo te zeggen minder toevoegen aan een tekst (er zijn geen alternatieven die kunnen helpen bij het begrip), maar praktisch gezien zijn zij beperkter bruikbaar. Eén reden is al aangeduid: veel relevante vragen zijn vaak (te) moeilijk als open vraag. De tweede reden is dat de leerlingen het antwoord moeten formuleren,

nieuw voor CKV 1

MET PALET



COMPLEET MET CD-ROM

IS KUNST DICHTERBIJ DAN JE DENKT

- Complete leergang Culturele en Kunstzinnige Vorming 1
- Handzaam, rijk geïllustreerd informatieboek havo/vwo
- Boeiende en veelzijdige cd-rom *Hotel Palet* om zelf kunst te beleven
- Vijf themakaternen: 'Helden', 'De Stad', 'Liefde', 'De droom', 'West-Oost'
- Persoonlijke kunstbeleving en reflectie staan centraal
- Compleet pakket voor het studiehuis: handleidingen bij de methode en per katern, studiewijzers, copymasters en uitwerkingenboeken
- Nog dit schooljaar speciale 'voorschooling' voor nieuwe gebruikers van *Palet*
- Nu al een begrip in het studiehuis

U kunt een gratis beoordelingspakket van *Palet* aanvragen.

Bel met onze Docentenlijn: (0575) 59 48 80 of stuur een e-mail: info@thieme.nl

Onze educatief adviseurs geven u tevens alle informatie en advies.

Actuele informatie over de uitgaven van Thieme vindt u ook op internet:

<http://www.thieme.nl>



Thieme

wat kan betekenen dat een (te) sterk beroep gedaan moet worden op de uitdrukingsvaardigheid. Daarmee zal meteen duidelijk zijn waarom de vraagstelling in open vorm vooral beperkt moet blijven tot kort-antwoordvragen. En in deze vragen zal het gaan om conceptueel eenvoudige operaties zoals het noemen van een gegeven, het vermelden van een citaat en dergelijke en niet gauw om een bewerking als het aangeven van de kern van een alinea of het trekken van een conclusie met betrekking tot de bedoeling van de auteur.

Bij de keuze voor open of gesloten vragen speelt nog een andere belangrijke overweging mee: de mate van objectiviteit van de beoordeling. Parallel aan de problemen die leerlingen kunnen hebben om hun gedachten nauwkeurig weer te geven, zijn de problemen die zich voordoen bij het interpreteren van de antwoorden. Dit kan leiden tot (aanzienlijke) verschillen in beoordeling (zie het verslag van het onderzoek hieronder).

Onderzoek naar overeenstemming beoordeling

Rond de nieuwe examens voor de moderne vreemde talen is door het Cito onderzoek gedaan naar de betrouwbaarheid van de beoordeling van open vragen. In het voorgaande is sprake geweest van voorbeeldexamens. Antwoorden van leerlingen op open vragen uit de voorbeeldexamens voor Frans en Duits van havo zijn bij wijze van proef beoordeeld door een aantal docenten ten einde na te kunnen gaan in hoeverre hun beoordelingen onderling overeenstemden (voor een uitgebreid verslag, zie Bügel, Melse & Goldebeld, 1998).

Aan het onderzoek namen 13 docenten en twee medewerkers van het Cito deel. De docenten hebben zich aangemeld in reactie op een wervingsbrief die is gestuurd aan een a-selecte steekproef van ca. 30 scholen. Voorwaarden voor deelname waren ervaring met examenklassen en bereidheid zich te houden aan de aanwijzingen voor de beoordeling. In totaal waren er acht correctoren voor Frans en zeven voor Duits. Dat lijken kleine steekproeven, maar uit eerdere beoordelingsonderzoeken blijkt dat deze aantallen een voldoende representatief beeld geven van het beoordelingsgedrag van docenten in het algemeen.

De voorbeeldexamens zijn op een aantal scholen gemaakt door leerlingen uit examenklassen. Zowel bij Frans als Duits is uit de leerlingwerken een a-selecte steekproef van 30 werkstukken getrokken. Deze zijn gekopieerd en aan de correctoren verzonden om thuis te beoordelen. Zij ontvingen daartoe het voorbeeldexamen, het correctievoorschrift met algemene regels voor de beoordeling (bijv. hoe gehandeld moet worden als een leerling meer informatie geeft dan gevraagd is) en het antwoordmodel (zie voorbeelden). De open vragen van de examens voor de moderne vreemde talen worden in het Nederlands gesteld en zij moeten in het Nederlands worden beantwoord. Schrijfvaardigheid in de vreemde taal speelt bij de toetsing dus geen rol, wel schrijfvaardigheid in het Nederlands. Wanneer een citaat gevraagd wordt, is een letterlijke aanhaling bedoeld en gaat het dus wel om een antwoord in de vreemde taal,

omdat anders vertaald zou moeten worden. In andere gevallen is beantwoording in de vreemde taal niet toegestaan. Bij Frans ging het per werkstuk om 15 vragen waarvoor 32 punten behaald konden worden en bij Duits om 13 vragen en 15 punten. Tot zover de opzet van het onderzoek.

Wat waren de verwachtingen? Op één uitzondering na ging het om kort-antwoordvragen, soms zeer gesloten van aard (bijvoorbeeld citeervragen), vooral bij Frans. Andere goede antwoorden dan in het antwoordmodel waren nauwelijks te verwachten (wel verschillende formuleringen daarvan), zeker niet bij Duits omdat met de betreffende opgaven al een try-out gehouden was. In principe was er dus een goede basis voor een hoge overeenstemming tussen beoordelaars. De onderlinge overeenstemming van beoordelingen (van dezelfde antwoorden/werkstukken) is statistisch uit te drukken in een overeenstemmingscoëfficiënt die kan variëren van 0 tot 1. Nul betekent 'geen enkele overeenstemming' en 1, 'perfecte overeenstemming'. Vooraf werd – onder meer op basis van ervaringen met beoordeling van schrijfvaardigheid – verondersteld dat een overeenstemming van 0,80 haalbaar moest zijn. De uitkomst was evenwel minder gunstig: gemiddeld werd een overstemming van 0,65 bereikt.

De vraag waar het uiteindelijk om gaat, is wat deze uitkomst betekent in termen van verschil op de cijferschaal van 10 punten. Omgerekend naar een examen als geheel (tweederde gesloten, eenderde open vragen) komen de gesignaleerde verschillen tussen de individuele beoordelaars gemiddeld neer op 0,3 cijferpunt voor het hele examen. Tussen de mildeste en de strengste beoordelaar bedroeg het verschil 1 cijferpunt.

Op grond van de antwoorden van de leerlingen en de beoordelingen door de docenten geven wij een aantal mogelijke verklaringen voor de verschillen in beoordeling:

- Leerlingen blijken erg onnauwkeurig te formuleren, wat het beoordelen moeilijk maakt: de corrector moet interpreteren wat de leerling bedoeld heeft (zie voorbeeld 1).
- Bij een aantal vragen bleken de aanwijzingen voor de leerlingen niet voldoende. Het blijkt bijvoorbeeld bij een citeervraag nodig te zijn om precies aan te geven hoe er geciteerd moet worden (zie voorbeeld 2).
- Correctoren hielden zich soms niet aan het correctievoorschrift, mogelijk omdat dit nieuw voor hen was (zie voorbeeld 3).
- Er zijn soms aanzienlijke systematische verschillen tussen beoordelaars. Daarmee worden de verschillen in het gemiddelde van de totaalscores voor werkstukken bedoeld. Er zijn dus duidelijk strenge en milde beoordelaars (zie de figuren in bijlage 1).

Wanneer beoordelaars aantoonbaar afwijken van het correctievoorschrift, ging het om één van de volgende gevallen:

- Zes van de acht beoordelaars Frans gaven soms deelscores bij vragen waar dit niet was toegestaan (zie voorbeeld 2 en 4).
- Het kwam voor dat, in strijd met de scoringsregels (zie bijlage 2), fouten met betrekking tot het Nederlands in de beoordeling werden betrokken.
- Aan antwoorden of delen van antwoorden in de vreemde taal werden soms ten onrechte scorepunten toegekend (zie voorbeeld 1).
- Wanneer er naast goede ook foute en/of overtollige elementen in de antwoorden voorkwamen, werden de scoringsregels niet altijd correct toegepast (zie voorbeeld 5).

Discussie uitkomsten onderzoek

Bij de beoordeling van open vragen in tekstbegrip-toetsen treden, blijkens de uitkomsten van het onderzoek, verschillen op; in het onderzoek gemiddeld 0,3 cijferpunt. Dat lijkt wellicht niet zo hoog, maar bedacht moet worden dat deze verschillen voortvloeien uit ca. eenderde deel van de desbetreffende toetsen. Gelet daarop is het geconstateerde gemiddelde verschil heel serieus te nemen.

De vraag waar het in de examenpraktijk op aankomt, is of deze verschillen dan minder, gelijk of juist hoger zullen zijn. Dat valt moeilijk te voorspellen. Het onderzoek is uitgevoerd met experimentele opgaven. Wellicht functioneren (definitieve) examenopgaven beter, zowel wat betreft de vorm van de antwoorden als de kwaliteit van de beoordeling, bijvoorbeeld doordat na verloop van tijd de ervaring met het hanteren van de scoringsregels en antwoordmodellen toeneemt. Anderzijds zullen correctoren misschien anders beoordelen wanneer het gaat om de beoordeling van eigen kandidaten in de examensituatie. Op grond van de literatuur (bijv. Meuffels, 1994) moet men verwachten dat de docenten prestaties van eigen leerlingen gemiddeld wat hoger waarderen dan prestaties van leerlingen die zij niet zelf onderwijzen (zoals in het hier beschreven onderzoek het geval was). Uit de literatuur is nog een aantal andere ongewenste beoordelareffecten bekend. Sommige daarvan komen voort uit het feit dat de beoordelaar de te beoordelen leerlingen kent (De Groot, 1968; Meuffels, 1994). Het gaat om de volgende effecten:

- Signifisch effect: beoordelaars hebben een verschillende opvatting over de beoordelingstaak. Wanneer correctoren bijvoorbeeld de opbouw van een opstel moeten beoordelen, is het niet zeker dat allen hetzelfde verstaan onder een goede opbouw.
- Halo-effect: andere kenmerken dan het kenmerk waarop het antwoord beoordeeld moet worden, hebben invloed op de beoordeling, bijvoorbeeld eerdere prestaties, een goede stijl of een mooi handschrift.
- Sequentie-effect: invloed van een voorafgaande beoordeling. Een middelmatig antwoord kan na een slecht antwoord leiden tot een hogere beoordeling dan na een goed antwoord.
- Normverschuiving of persoonlijke vergelijking: vrijheid in de beoordelingschaal leidt tot scoreverdelingen die algemeen menselijke of persoonlijke eigenschappen uitdrukken. Zo kan de hoogte van een score beïnvloed worden door de strengheid van de corrector of door het niveau van de hele groep.
- Contaminatie-effect: bij de beoordeling spelen andere doelen een rol dan een objectieve beoordeling. Bijvoorbeeld het besef bij een docent(e) dat veel lage cijfers zijn of haar reputatie nadelig beïnvloeden.
- Competentie-effect (alleen door Meuffels genoemd): de beoordelaar is onvoldoende bekwaam om de beoordelingstaak uit te voeren.

Het halo-effect en het contaminatie-effect doen zich vooral voor bij de beoordeling van eigen leerlingen. De mening die men al heeft over het kunnen van een leerling, de sympathie of afkeer die men koestert, de (pedagogisch wellicht begrijpelijke) wens die men heeft om iemand te belonen of te straffen, al zulke menselijke overwegingen en gevoelens kunnen een objectieve beoordeling van prestaties beïnvloeden. Bij examens gaat het om het verkrijgen van een betrouwbaar beeld van het kunnen van leerlingen en zou het uitgesloten moeten zijn dat ongewenste effecten als gevolg van bekendheid de beoordeling

beïnvloeden. Is dat mogelijk? De oplossing ligt in principe voor de hand: beoordeling van de prestaties van onbekende leerlingen (anonieme beoordeling). Wat betreft de overige genoemde effecten kunnen het signifisch effect en normverschuiving bestreden worden door te zorgen voor een optimale duidelijkheid van de opgaven en het correctievoorschrift. Een mogelijke invloed van het sequentie-effect kan voorkomen worden door per vraag te beoordelen en de volgorde van de te beoordelen werkstukken telkens te variëren. Het competentie-effect is van heel andere aard. Het kan voortvloeien uit onvoldoende vakken-nis of uit onvoldoende aanwezigheid van vereiste eigenschappen als nauwkeurigheid en logisch denkvermogen.

Welke maatregelen men ook neemt, verschillen bij de beoordeling zijn vaak moeilijk in te perken tot een aanvaardbaar niveau. Daarom is bij examens met open vragen in Nederland een tweede correctie verplicht. Een tweede correctie kan op twee manieren worden uitgevoerd:

- De tweede corrector ontvangt de beoordeelde werkstukken en controleert de beoordelingen van de eerste corrector (afhankelijke beoordeling).
- De tweede corrector ontvangt een schoon exemplaar van de te beoordelen werkstukken en beoordeelt deze zonder kennis te kunnen nemen van de scores van de eerste corrector. De eindscore voor een werkstuk is het rekenkundig gemiddelde van de scores van de beide correctoren (onafhankelijke beoordeling).

Uit onderzoek van Meuffels blijkt dat afhankelijke correctoren niet of onvoldoende tot een zelfstandige beoordeling komen. Zij worden te zeer gestuurd door de beoordelingen van de eerste corrector. Het resultaat is dat een afhankelijke tweede correctie minder bijdraagt aan objectivering van de beoordeling dan een onafhankelijke tweede correctie.

Examenpraktijk aan verandering toe?

Uit het voorafgaande blijkt dat anonieme beoordeling, dus beoordeling van leerlingen van een andere school, uit het oogpunt van objectiviteit te prefereren is boven beoordeling van bekende leerlingen. In sommige landen is dat de normale gang van zaken, maar in Nederland is dat geen traditie. In de tweede plaats blijkt dat de tweede correctie, die wel tot de Nederlandse examenpraktijk hoort, tot betrouwbaarder resultaten leidt bij een onafhankelijke tweede beoordeling dan bij de bestaande afhankelijke tweede correctie.

Nu met de invoering van nieuwe programma's voor de tweede fase en binnenkort voor mavo en vbo alle examens (voor een deel) open vragen zullen bevatten, is het ons inziens een goed moment om de bestaande examenpraktijk te evalueren. De recente invoering van kwaliteitskaarten voor scholen levert een extra argument voor een zo objectief mogelijke beoordeling.

Het zou raadzaam zijn de pro's en contra's van beide aanbevelingen te onderzoeken en te beginnen met enkele vergelijkende onderzoeken naar de uitkomsten van afhankelijke en onafhankelijke tweede correctie, alsmede van anonieme en niet-anonieme beoordeling bij echte examens.

Literatuur

- Bügel, K., L. Melse, P. Goldebeld, *Beoordelaarsovereenstemming bij open vragen leesvaardigheid Frans en Duits*, Arnhem: Cito, 1998.
- Els, T.J.M. van, T.J.M.M. Buis, *De praktijk van het onderwijs moderne vreemde talen in de bovenbouw havo/vwo*, Nijmegen: ITI, ITS, 1987.
- Gras, R.R., *Studietoetsen voor de moderne talen*, Groningen: Wolters-Noordhoff, 1969.
- Groot, A.D. de, *Methodologie: grondslagen van onderzoek en denken in de gedragswetenschappen*, 's-Gravenhage: Mouton, vierde druk, 1986.
- Meuffels, B., *De verguisde beoordelaar. Opstellen over opstelbeoordeling*, Amsterdam: Thesis publishers, 1994.
- Samson, D. (1998). Nieuwe examens leesvaardigheid, in: *Levende Talen* 526, p. 2-5.
- *Syllabus moderne vreemde talen voor de centrale examens in de tweede fase havo/vwo*, Arnhem: Cevo, Cito, 1998 (er zijn vakspecifieke uitgaven i.v.m. de voorbeeldexamens).

Karin Bügel



Studeerde Frans en Algemene Literatuur Wetenschap en werkte als lerares Frans in het voortgezet onderwijs. Sinds 1983 verbonden aan het Cito, waar zij o.a. de constructie van

tekstbegripexamens Frans begeleidde. In 1993 promoveerde zij op een proefschrift over sekseverschillen bij tekstbegrip moderne vreemde talen. Op dit moment projectleider van 'Voortgangstoetsing Tweede Fase vo'.
Adres: Cito, Postbus 1034, 6801 MG Arnhem.

Leijn Melse



Germanist; was tien jaar leraar in het voortgezet onderwijs en werkt sinds 1980 bij het Cito. Is betrokken bij de productie van toetsen voor alle vier de vaardigheden en doet

onderzoek met betrekking tot problemen van toetsing van vreemdetaalvaardigheid. Hij promoveerde op een proefschrift over de toetsing van schrijfvaardigheid.
Adres: Cito, Postbus 1034, 6801 MG Arnhem.

Kiliaan

Taalvaardigheid
Nederlands

spreekt de taal van de leerlingen

7PEL
- 2^e -
fase
1

Docenten over Kiliaan:

'Spannend en toch overzichtelijk'

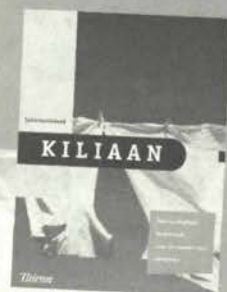
Kiliaan in tien punten:

- Complete methode taalvaardigheid Nederlands voor de tweede fase havo/vwo
- Boeiende, gevarieerde en actuele teksten
- Vaardigheden contextgericht en in samenhang
- Systematische aandacht voor mondelinge taalvaardigheid
- Informatieboek is een uitgebreid naslagwerk
- Video met authentieke fragmenten voor de mondelinge spreekvaardigheid
- Geschikt voor verschillende leer- en doceerstijlen
- Sluit aan op *Op niveau plus* en alle andere methoden voor de basisvorming
- Compleet pakket voor het studiehuis: leerlingendiskettes met extra oefenmateriaal, studiewijzers, toetsdiskettes, video en uitwerkingenboeken
- Nog dit schooljaar speciale 'voorscholing' voor nieuwe gebruikers van *Kiliaan*

U kunt een gratis beoordelingspakket van *Kiliaan* aanvragen. Bel met onze Docentenlijn:

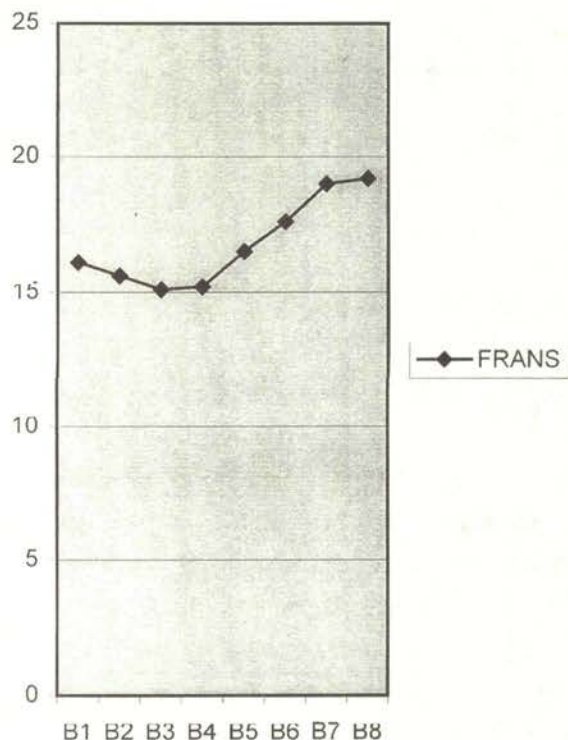
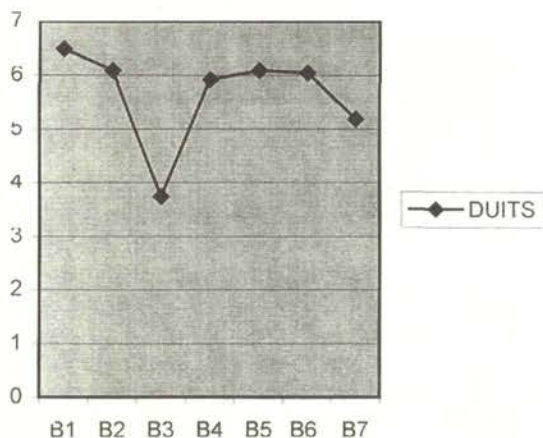
(0575) 59 48 80 of stuur een e-mail: info@thieme.nl

Onze educatief adviseurs geven u ook alle informatie en advies. Actuele informatie over de uitgaven van Thieme kunt u bovendien vinden op internet: <http://www.thieme.nl>



Thieme

Voorbeeld 1 Vraag	Antwoordmodel	Antwoorden van leerlingen met grote verschillen bij de beoordeling (toegekende scorepunten) en <i>Commentaar</i>
<p>Voorbeeld 1 « Douze villes ... lépreux » (regels 5-8). De schrijver vindt dit om twee redenen schokkend. Noem deze redenen door de zin op het antwoordblad aan te vullen.</p> <p>De schrijver vindt dit schokkend omdat..... en omdat</p>	<p>maximumscore 2 Kern van het eerste antwoord-element: (1) - hieruit (groeiende) onverdraagzaamheid (Franse samenleving) blijkt. of: - dit onze / de collectieve / intolerantie aantoont (of woorden van gelijke strekking). Kern van het tweede antwoord-element: (2) - het getuigt van onvermogen van (plaatselijke) autoriteiten om flexibel / vindingrijk / creatief om te gaan met mensen in (financiële) problemen. of: - het laat zien dat (locale) autoriteiten geen goed beleid hebben met betrekking tot het armoede-probleem. (of woorden van gelijke strekking). . Bij beide elementen goed: 2 pt. . Bij één element goed 1 pt.</p>	<p>(1) - de maatregel afdoet aan onze grote collectieve tolerantie (2) - het de incapaciteit duidelijk toont die men zo heeft t.o.v. volken in crises (1x2, 2x1, 6x0).</p> <p>(1) - deze maatregel "témoigne van onze collectieve intolerantie" grandissante. (2) - l'incapacité de nombreuses collectives locales à réinventer leurs dispositifs la crise (2x2, 2x1, 4x0). <i>Commentaar:</i> Weinig leerlingen zijn in staat de inhoud van de tekst correct weer te geven. Er worden in de antwoorden Franse woorden en veel romanismen gebruikt, waardoor de onbegrijpelijkheid van de antwoorden toeneemt. Leerlingen die stukken uit de tekst overschrijven krijgen van sommige beoordelaars toch het maximale aantal punten.</p>
<p>Voorbeeld 2 Citeer een zin uit de tekst waaruit blijkt dat vliegen veiliger is geworden door het werk van Bernard Prévost</p>	<p>maximumscore 2 «Le nombre ... atteindre zéro.» (lignes 117 - 120).</p>	<p>Le nombre d'incidents à Istres a diminué progressivement depuis.....enz. (6x2, 2x1). - Le nombre d'incidents à Istres a diminué progressivement... (6x2, 2x1). <i>Commentaar:</i> Hier ontstaan verschillen bij leerlingen die niet correct citeren: slechts een gedeelte van de goede zin bijvoorbeeld.</p>
<p>Voorbeeld 3 Wat verklaart het succes van de »Lobbytätigkeit« (Zeile 39) van de regeringen van Griekenland en Italië in Brussel? (ongeveer 10 woorden)</p>	<p>Maximumscore 1 Een antwoord met de volgende strekking: Te pleiten voor het recht om te roken. Passabel: Om vrijheid voor rokers te eisen/bevorderen. Fout: Om het roken te bevorderen</p>	<p>- Dat roken geen taboe meer zal zijn in de toekomst (2x 1, 5x 0). <i>Commentaar:</i> dit antwoord beoordelen de meesten als onjuist, maar het lijkt niettemin dezelfde strekking te hebben als het modelantwoord, ook al wijkt het qua formulering geheel af.</p>
<p>Voorbeeld 4 Welke twee dingen deed Bernard Madeleine om het leven in de gevangenis te kunnen verdragen, volgens de 5e alinéa?</p>	<p>Maximumscore 2 De twee volgende • slapen (1) • sporten (1)</p>	<p>- sporten, schrijven, slapen (2x2, 5x1, 1x0). - Hij slaapte veel en hij deed veel aan sport (tafeltennis) (6x2, 2x1). <i>Commentaar:</i> Bij het eerste antwoord had volgens de voorschriften slechts 1 punt toegekend mogen worden..</p>
<p>Voorbeeld 5 Welke gevolgen heeft, volgens Goldkamps zienswijze, de politiek met betrekking tot drugs voor druggebruikers? Noem er twee.</p>	<p>Maximumscore 2 Hoge prijzen (voor drugs) en criminaliteit (onder verslaafden). Indien één van beide elementen ontbreekt: 1 punt</p>	<p>- Doordat het illegaal is, worden gebruikers als criminelen gezien. De prijs wordt veel hoger en de kwaliteit laat ook eerder te wensen over (2x 2, 3x 1, 1x 0). <i>Commentaar:</i> hier worden drie 'gevolgen' genoemd i.p.v. twee. Bij de beoordeling moet naar de eerste twee gekeken worden. Wat in de eerste zin als gevolg genoemd wordt, is fout. Het eerste deel van de tweede zin is goed. → score 1. - Door het verbod betalen druggebruikers veel geld voor de drugs. De gebruikers gaan daardoor door de straten om aan geld te komen (1x 2, 4x 1, 2x 0). <i>Commentaar:</i> De eerste zin bevat een gevolg dat de tekst noemt, de tweede niet. → score 1</p>

**Bijlage 1 en 2**

In de beide figuren zijn de (over alle leerlingen) gemiddelde scores van de beoordelaars weergegeven. Op de verticale as staan de gemiddelde scores. Op de horizontale as zijn de beoordelaars weergegeven. B1 staat voor beoordelaar 1, B2 voor beoordelaar 2, enzovoorts. Het verschil tussen de mildste en de strengste beoordelaar is bij Duits 2,75 punt. De mildste beoordelaar (B1) gaf gemiddeld 6,49 punten en de strengste (B3) gemiddeld 3,74 punten bij een maximumscore van 15 punten. Bij Frans was de maximumscore 32 punten.

De in het onderzoek gebruikte scoringsregels

- Indien een vraag volledig juist is beantwoord, wordt het maximaal te behalen aantal scorepunten toegekend.
- Indien een vraag gedeeltelijk juist is beantwoord, wordt een deel van de te behalen scorepunten toegekend, in overeenstemming met het antwoordmodel.
- Indien slechts één voorbeeld, reden, uitwerking, citaat of andersoortig antwoord gevraagd wordt, wordt uitsluitend het eerstgegeven antwoord beoordeeld.
- Indien meer dan één voorbeeld, reden, uitwerking, citaat of andersoortig antwoord gevraagd wordt, worden uitsluitend de eerstgegeven antwoorden beoordeeld, tot maximaal het gevraagde aantal.
- Indien in het antwoordmodel verschillende mogelijkheden zijn opgenomen, gescheiden door het teken /, gelden deze mogelijkheden als verschillende formuleringen van hetzelfde antwoord.
- Indien in het antwoordmodel een gedeelte van het antwoord tussen haakjes staat, hoeft dit gedeelte niet in het antwoord van de kandidaat voor te komen.
- Indien bij een vraag een maximaal aantal woorden is vermeld, worden uitsluitend de woorden tot dat maximum beoordeeld (deze regel geldt niet indien het aantal woorden met 'ongeveer' is aangegeven).
- Indien Frans/Duits is gebruikt in gevallen waarin geen citaat is gevraagd, wordt het antwoord als onjuist beschouwd.
- Indien taalfouten zijn gemaakt, worden deze bij de beoordeling buiten beschouwing gelaten.