

Talenexamens: waar ligt de grens?

Het lijkt niet meer dan billijk dat de eisen die aan examenkandidaten worden gesteld, over een reeks van jaren gelijk zijn. Examenkandidaten moeten ervan kunnen uitgaan dat het examen waarvoor ze opgaan, even moeilijk is als eerder afgenomen examens van hetzelfde vak en hetzelfde niveau. Dat is gewoon een kwestie van eerlijkheid.

Vanaf 1994 voert het Cito normvergelijkingprocedure uit voor de havo- en de D-examens Frans, Duits en Engels. In 1997 zijn aan deze reeks ook de vwo-examens Frans, Duits en Engels toegevoegd. Door het uitvoeren van deze procedures kan de grens tussen voldoende en onvoldoende voor deze vakken op eenzelfde prestatieniveau worden vastgesteld.

Zijn de exameneisen elk jaar hetzelfde? Deze vraag vormde het startpunt van een in 1991 door de inspectie geïnitieerd onderzoek. In het verslag van dit onderzoek, 'Examens op punten getoetst', wijst de inspectie op het probleem dat ontstaat wanneer het vaardigheidsniveau van de kandidaten wordt vastgesteld met een meetinstrument dat niet geijkt is. Met zo'n instrument is het onmogelijk de moeilijkheidsgraad van examens en het prestatieniveau van kandidaten onafhankelijk van elkaar vast te stellen. Je weet nooit of het nu ligt aan de moeilijkheidsgraad van het examen of aan het vaardigheidsniveau van de kandidaten dat er in een bepaald jaar voor een bepaald vak meer of minder voldoende dan 'normaal' worden gehaald.

Naar aanleiding van dit inspectierapport heeft de toenmalige staatssecretaris van Onderwijs en Wetenschappen in april 1993 de Cevo en het Cito verzocht een plan op te stellen voor een gefaseerde invoering van normhandhaving door middel van normvergelijking.

Normvergelijking maakt het mogelijk uitspraken te doen over het vaardigheidsniveau van de examenkandidaten, zodat we een gefundeerd antwoord kunnen geven op een vraag als: gaat het niveau van de kandidaten voor bijvoorbeeld Frans of Duits in de loop van de jaren achteruit? Ook moet normvergelijking het mogelijk maken examenkandidaten over een reeks van jaren equivalente examens voor te leggen. Dat wil zeggen dat wanneer een examen achter-

af, na het uitvoeren van de normvergelijkingprocedures, moeilijker is gebleken dan gewenst, er een zodanige cesuur (grens tussen voldoende en onvoldoende) kan worden bepaald, dat de eisen aan de kandidaten van een dergelijk examen niet in dat jaar hoger zijn dan in andere jaren.

In dit artikel beschrijf ik eerst hoe het vaardigheidsniveau van examenpopulaties en de moeilijkheidsgraad van nieuwe examens wordt bepaald en op welke wijze aan de hand van de verkregen gegevens voor het nieuwe examen een gelijkwaardige cesuur kan worden vastgesteld.

Vervolgens wordt een overzicht gepresenteerd van

- de verschillen in prestatieniveau van de examenpopulaties 1994 t/m 1997;
- de verschillen in moeilijkheidsgraad tussen de examens van 1994 t/m 1997;
- de verschillen in moeilijkheidsgraad tussen de examens voor het C- en D-programma.

Tot slot geef ik enige opmerkingen en conclusies naar aanleiding van het voorafgaande.

Procedure

De eerste stap in de normvergelijkingprocedure is dat er voor elk vak een *referentie-examen* wordt gekozen. Dit is een eerder afgenomen examen dat wordt gezien als een evenwichtige afspiegeling van het examenprogramma en dat een geschikte moeilijkheidsgraad bleek te hebben.

Elk nieuw examen wordt vervolgens vergeleken met dit referentie-examen. In een extra afname direct na de afname van het nieuwe examen, worden de prestaties van een bepaalde groep leerlingen uit bijvoorbeeld 5-vwo of 3-vwo op het nieuwe examen en op het referentie-examen met elkaar vergeleken.

Uit deze extra afname blijkt hoeveel moeilijker of makkelijker het nieuwe examen voor de afnamegroep is, vergeleken met het referentie-examen. Deze informatie maakt het mogelijk de cesuur van het referentie-examen over te brengen op het nieuwe examen, zodat het nieuwe examen een gelijkwaardige cesuur aan het referentie-examen heeft.

• Opzet van de afname

Het is in de praktijk niet mogelijk een groep leerlingen twee volledige examens achter elkaar te laten maken. In de normvergelijkingprocedure worden daarom kleinere toetsen samengesteld uit vragen die

telkens voor een deel uit het referentie-examen en voor een deel uit het nieuwe examen afkomstig zijn. Dit worden de ankertoetsen genoemd. In alle ankertoetsen samen komen alle vragen uit het referentie-examen en alle vragen uit het nieuwe examen aan bod. Elke ankertoets wordt vervolgens bij een groep leerlingen afgenomen en de gegevens van deze afnames worden gecombineerd.

Voor havo en vwo worden de ankertoetsen afgenomen bij 5-vwo-leerlingen, voor mavo-D en vbo-D bij leerlingen van 3-vwo.

• Afnamegegevens en normvergelijkingsprocedure

Na afname van de ankertoetsen beschikken we over drie soorten afnamegegevens:

- a de afnamegegevens van het referentie-examen uit het jaar waarin het als examen is afgenomen;
- b de afnamegegevens van het nieuwe examen bij de nieuwe examenkandidaten;
- c de afnamegegevens van de ankertoetsen bij de afnamegroepen.

Voor de normvergelijkingsprocedure gebruiken we deze gegevens als volgt.

Door de afnamegegevens van de ankertoetsen weten we of het nieuwe examen voor de afnamegroep moeilijker of makkelijker is dan het referentie-examen.

Door de afnamegegevens van het referentie-examen (a) en de afnamegegevens van de ankertoetsen (c) aan elkaar te koppelen kunnen we een uitspraak doen over de moeilijkheidsgraad van het nieuwe examen, onafhankelijk van het vaardigheidsniveau van de nieuwe populatie examenkandidaten. We kunnen zo een uitspraak doen over de moeilijkheidsgraad van het nieuwe examen ten opzichte van het referentie-examen. De zo gevonden moeilijkheidsgraad wordt vervolgens gerelateerd aan de prestaties van de nieuwe examenpopulatie op het nieuwe examen (b). Hierdoor kunnen we ook uitspraken doen over het vaardigheidsniveau van de nieuwe examenkandidaten.

Resultaten

De normvergelijkingsprocedure maakt het mogelijk uitspraken te doen over het vaardigheidsniveau van de examenpopulaties voor een bepaald vak over de jaren heen. Ook kunnen we de moeilijkheidsgraad van elk nieuw examen bepalen ten opzichte van het referentie-examen en zo de moeilijkheidsgraad van de examens die bij de normvergelijkingsprocedure zijn betrokken, van opeenvolgende jaren vergelijken. Hieronder volgen de uitkomsten die de normvergelijkingsprocedures van 1994 t/m 1997 voor het vaardigheidsniveau van de examenpopulaties en de moeilijkheidsgraad van examens hebben opgeleverd.

• Vaardigheidsniveau van de kandidaten

Zoals we hierboven gezien hebben, geven de normvergelijkingsprocedures ons inzicht in het vaardigheidsniveau van de examenpopulaties. Het vaardigheidsniveau van elke examenpopulatie kan worden uitgedrukt in de gemiddelde score die die examenpopulatie op het referentie-examen zou hebben behaald. In tabel 1 worden deze gegevens gepresenteerd.

vak	referentie-populatie	1994 pop.	1995 pop.	1996 pop.	1997 pop.
Duits-D	35	35	34	34	32
Engels-D	35	36	35	36	35
Frans-D	33	29	31	27	28
Duits-havo	35	33	33	31	32
Engels-havo	32	34	34	33	33
Frans-havo	36	36	35	35	35

tabel 1: Overzicht van het vaardigheidsniveau van de opeenvolgende populaties uitgedrukt in de gemiddelde score op de referentietoets

Tabel 1 laat zien dat de vaardigheid van de examenpopulaties over de jaren heen niet constant is. Opvallend is dat de examenkandidaten bij Engels-D en Engels-havo van de jaren 1994 t/m 1997 gelijk of hoger zouden scoren op het referentie-examen. Bij Duits en Frans zien we een andere beweging. De examenkandidaten zouden hier gelijk of lager scoren op het referentie-examen. Het grote verschil in vaardigheid bij Frans-D tussen de referentiepopulatie en de populaties van 1994, 1996 en 1997 is opvallend. Overbrenging van de cesuur van het referentie-examen naar het examen van 1994 en volgende jaren zou tot onaanvaardbaar hoge percentages onvoldoendes hebben geleid. De norm werd daarom door de Cevo verlaagd, dat wil zeggen versoepeld ten opzichte van de cesuur van het referentie-examen.

We kunnen uit de tabel afleiden dat de vaardigheid van de examenpopulaties vanaf 1994 voor Duits-havo en Frans-D nu vier jaar achtereen geringer is dan die van de referentiepopulatie. Daar kunnen tal van oorzaken voor worden aangevoerd. Als factoren die de prestaties op de examens kunnen beïnvloeden, worden wel genoemd: de soorten vragen in de examens, de tekstkeuze, de samenstelling van het vakkenpakket, gebruikte leergangen, gevolgde didactiek, mavo-instroom bij de havo en bij Frans-D de vermindering van het aantal uren Frans in de urentabel in de laatste jaren.

De eerstgenoemde factoren – soorten vragen in de examens en tekstkeuze – worden door de opzet van de normvergelijkingsprocedure als verklaring voor een eventuele niveaudaling uitgeschakeld. Naar de invloed van de samenstelling van het vakkenpakket op de prestaties van de kandidaten is op het Cito onderzoek gedaan. Bij vergelijking van de examenresultaten van de verschillende groepen kandidaten (groep kandidaten met Frans, Duits en Engels, groep met Frans en Engels, groep met Duits en Engels, groep Duits en Frans) bleek dat bijvoorbeeld de groep kandidaten die Engels, Frans en Duits in hun pakket hadden, niet significant hoger of lager scoorden dan kandidaten met bijvoorbeeld alleen Duits en Frans. Geen van de groepen presteerde significant beter dan de andere groepen.

Naar de andere factoren is nog geen onderzoek gedaan.

• Moeilijkheidsgraad van de examens

De uitkomsten van de normhandavingsprocedures over de moeilijkheidsgraad van de examens staan

vak	In vergelijking met het referentie-examen is het			
	examen 1994	examen 1995	examen 1996	examen 1997
Duits-D	3 pnt moeilijker	even moeilijk	4 pnt moeilijker	3 pnt moeilijker
Engels-D	2 pnt moeilijker	even moeilijk	3 pnt moeilijker	1 pnt moeilijker
Frans-D	even moeilijk	1,2 pnt makkelijker	3 pnt makkelijker	1 pnt makkelijker
Duits-havo	even moeilijk	0,9 pnt moeilijker	0,5 pnt makkelijker	0,5 pnt makkelijker
Engels-havo	even moeilijk	1 pnt makkelijker	2 pnt moeilijker	2,5 pnt makkelijker
Frans-havo	0,5 pnt makkelijker	1,3 pnt moeilijker	1 pnt moeilijker	3,5 pnt moeilijker

tabel 2: Vergelijking tussen de referentie-examens en de examens 1994, 1995, 1996 en 1997 (de puntenaantallen zijn steeds punten op een 50-puntsschaal)

vermeld in tabel 2.

Uit tabel 2 blijkt dat de moeilijkheidsgraad van examens over de jaren heen niet constant is. Het examen Duits-D was in 1996 bijvoorbeeld 4 punten en in 1997 3 punten moeilijker dan het referentie-examen op een maximumscore van 50 punten.

Bij het bepalen van de cesuur voor de examens die in de normvergelijkingprocedure zijn opgenomen, spelen deze gegevens een rol. De Cevo kan de grens tussen voldoende en onvoldoende zodanig leggen dat zij gelijkwaardig is aan de cesuur van het referentie-examen. Bij een dergelijke benadering stelt men aan de examenkandidaten dezelfde eisen als aan de kandidaten die het referentie-examen hebben gemaakt. Zoals wij hiervoor gezien hebben kunnen er redenen zijn om van de gewenste gelijkwaardige cesuur enigszins af te wijken, omdat het percentage onvoldoendes bij een gelijkwaardige cesuur aan het referentie-examen vanuit de traditie van het vak onacceptabel hoog wordt. Ook is het mogelijk dat er in het examen een fout heeft gezeten of dat het examen te lang is gebleken. Bij de bepaling van de cesuur wordt met deze factoren rekening gehouden.

• De examens voor C-niveau

Bij de examens voor het C-niveau vindt ook een vorm van normvergelijking plaats. De cesuur van het C-examen wordt gekoppeld aan die van het D-examen. De hier gebruikte procedure maakt gebruik van de resultaten van de normvergelijkingprocedure bij het D-niveau en van de prestaties van de C- en D-leerlingen op het overlappende deel in het examen.

De C- en D-examens leesvaardigheid bevatten een overlappende deel, 1 à 2 teksten met 10 à 12 opgaven. Door de resultaten die de C- en de D-kandidaten op dit overlappende deel behalen, met elkaar te vergelijken is het mogelijk het verschil in moeilijkheidsgraad tussen het C- en het D-examen, de afstand C-D, in cijferpunten uit te drukken. De afstand C-D geeft aan hoeveel cijferpunten méér een kandidaat met het cijfer 5,45 op het D-examen behaald zou hebben als hij het C-examen zou hebben afgelegd.

Op het moment dat de Cevo de cesuur voor de C-examens moet vaststellen, zijn zowel de normvergelij-

kingsgegevens van de D-examens beschikbaar als de gegevens over de afstand tussen de C- en D-examens. Wanneer voor het D-examen de cesuur zo gekozen is dat deze gelijkwaardig is aan de cesuur van het referentie-examen, kan voor het C-examen ook een gelijkwaardige cesuur worden gekozen door de afstand tussen de C- en de D-examens in de verschillende jaren binnen de gewenste marges te houden. Tabel 3 vermeldt de afstanden in cijferpunten tussen de C- en D-examens van 1994, 1995, 1996 en 1997.

vak	afstand C-D			
	1994	1995	1996	1997
Duits	1,3	1,5	1,2	1,2
Engels	1,3	1,4	1,4	1,1
Frans	1,2	1,0	1,0	1,0

tabel 3: Afstand in cijferpunten tussen de cesuren van het C- en D-examen¹

De cesuur bij de C-examens beweegt mee met de D-examens. Binnen de meeste vakken blijken de onderlinge verschillen tussen de jaren 1994 t/m 1997 gering te zijn. Dit betekent dat ook daar de norm over de jaren heen binnen kleine marges wordt gehandhaafd.

Opmerkingen en conclusies

De gegevens van de normvergelijkingprocedures geven een duidelijk beeld van het verschil in vaardigheid tussen de examenpopulaties van de onderzochte jaren en van het verschil in moeilijkheidsgraad dat er tussen examens over een reeks van jaren kan bestaan.

De normvergelijking levert voor de talen gegevens op die de eerdere conclusies uit het inspectieonderzoek bevestigen: de moeilijkheidsgraad van examens en de vaardigheid van de populaties wisselen over de jaren heen. De kritiek van de inspectie dat gelijke prestaties in onvoldoende mate gelijk beloond worden, geldt door het uitvoeren van normvergelijkingprocedures niet meer voor de examens moderne vreemde talen.

Schat

In *Heer Bommel en de wisselschat* wendt Kwe- tal zich tot een zekere Monkel Oor omdat hij wil weten wat een schat is. Hij blijkt aan het goede adres te zijn, want Monkel Oor is 'een zeer belezen iemand, die al een halve eeuw doende was een woordenboek uit het hoofd te leren'. Gelukkig was hij al gevorderd tot de U, zodat een woord beginnend met een S niet veel geheimen meer voor hem had.

Nu zijn figuren die woordenboeken uit hun hoofd leren toonbeelden van domheid of een pseudo-geleerde levensinstelling. Het gaat daarbij niet eens zozeer om het idee dat er iets uit het hoofd geleerd wordt als wel om het feit dat het een alfabetische lijst betreft. Ook de schrijver van de Bommelstrip geeft toe dat een en ander het denkraam van de dwerg 'niet verruimd' had, maar hij erkent toch dat de opgedane kennis goed van pas kwam.

Opmerkelijk is dat het uit het hoofd leren van een grammatica op heel wat meer

begrip, ja zelfs op enige bewondering kan rekenen. Misschien heeft die kans op bescheiden roem zovelen ertoe gebracht zich aan de studie van de grammatica te zetten. Maar wellicht kan dat ook worden toegeschreven aan de overtuiging dat zoiets het denkraam wél verruimt.

Hoewel de hier aangesneden problematiek geen rechtstreeks verband houdt met het vreemdetalenonderwijs, wil ik toch van de gelegenheid gebruik maken om daar eens iets over te zeggen. Want, anders dan vaak gedacht wordt, is taal niet een kwestie van structuren waarin naar eigen inzicht woorden kunnen worden ingevuld. Uit de psycholinguïstische literatuur komt steeds duidelijker naar voren dat woorden en hun betekenissen centraal staan bij het spreken en verstaan van taal. Woorden bepalen de vorm die een boodschap aanneemt, en dankzij woorden kunnen we chocola maken van al dan niet (grammaticaal) correcte zinnen.

Al jaren lijkt het onderzoek naar de psychologische realiteit van grammaticale

modellen doodgelopen te zijn. Daarentegen sluiten de modellen van het mentale lexicon steeds dichter aan bij de werkelijkheid van de interne woordenschat. En nieuwe, gedigitaliseerde woordenboeken simuleren steeds beter waar mensen goed in zijn: met het grootste gemak en in een fractie van een seconde leggen ze verbanden tussen elementen met vergelijkbare betekenissen of met woorden uit een bepaald domein. En dat is waar het bij het leren van een andere taal voor een zeer groot deel om gaat.

Ook zulke woordenboeken kun je niet uit je hoofd leren. Maar dat is nog geen reden om dan maar door te gaan met grammatica's in je hoofd te stampen. Of om leerlingen daarmee te blijven lastig vallen. Langzamerhand wordt het tijd dat het talenonderwijs zich gaat bezinnen op een nieuwe verdeling tussen lexicon en grammatica. Tijd voor een wisselschat.

Paul Bogaards

Met de gegevens van de normvergelijking en de gegevens over de afstand tussen de C/D-examens beschikt de Cevo namelijk over informatie die het mogelijk maakt de examens te equivaleren. Dat gebeurt door een gelijkwaardige cesuur vast te leggen. Een te moeilijk of te makkelijk examen kan op deze manier worden gecorrigeerd, zodat de eisen die aan examenkandidaten worden gesteld over een reeks van jaren gelijkwaardig kunnen zijn. Normvergelijking biedt de mogelijkheid kandidaten van jaar tot jaar rechtvaardig te behandelen met

betrekking tot een norm die geworteld is in de traditie van de onderwijspraktijk van de verschillende talen. De vraag hoe functioneel de norm die we stellen en handhaven is, is daarmee niet beantwoord. Een norm functioneert altijd in een maatschappelijke context. De functionaliteit van de eens gekozen en vervolgens gehandhaafde norm komt daarmee in het centrum van de belangstelling te staan.

Noot

De afstand C-D geeft aan hoeveel cijferpunten méér een kandidaat met een 5,45 op het D-examen behaald zou hebben als hij het C-examen zou hebben afgelegd.

Literatuur

- Alberts, R.V.J., *Interimverslag normhandhaving examens vbo/avo/vwo 1995*, Arnhem: Cito, 1996.
- *Examens op punten getoetst*, Inspectierapport, november 1992.
- Kreeft, H.P., *Examens vbo/mavo/havo/vwo 1996*, Arnhem: Cito, 1996.
- Kreeft, H.P., *Examens vbo/mavo/havo/vwo 1997*, Arnhem: Cito, 1997.

Harjet Robben



Is werkzaam op het Cito en coördineert daar de talenexamens vbo/mavo/havo/vwo. Werkte daarvoor in het vo en hbo.
Adres: Cito, Postbus 1034, 6801 MG Arnhem.