

Hoe fair is een N-term van 0,0?

Toelichting van het CvE

Vraag van een docent, nét van de opleiding: "Tijdens mijn opleiding heb ik geleerd dat je altijd de normering vooraf moet vaststellen. Anders schuif je mee met een dalend niveau van je leerlingen. Waarom stellen jullie bij de centrale examens de N-term achteraf vast?"

Ameling Algra is sectormanager havo en vwo van het College voor Examens (CvE).

Goede les van de opleiding. Goede vraag ook. Het antwoord: wat in de school past, past niet altijd op de centrale examens, en omgekeerd. De docent heeft zicht op de eisen die je op een bepaald moment aan een bepaalde groep moet en mag stellen. Op grond daarvan maakt hij een toets, stelt vooraf vast wat zijn leerlingen moeten laten zien voor een voldoende. Makkelijker gezegd dan gedaan, maar verantwoord. Nu zal blijken dat de ene toets beter uitpakt dan de andere. Hoe goed het *Fingerspitzengefühl* van de leraar ook is, het zal hem nooit lukken om een serie van precies even moeilijke toetsen te maken. Dat is niet erg. "Dit was een moeilijke toets," zeggen zijn leerlingen. Ze snappen dat er die ene keer meer onvoldoendes vallen. Een volgende toets pakt weer gemakkelijker uit: bijna geen onvoldoendes. Dan volgt iets moois. De cijfers van alle toetsen worden gemiddeld. Er ontstaat een eindcijfer, het gemiddelde van een redelijk aantal toetsen en proeven. De enkele wat 'te' moeilijke toets geeft niet de doorslag. Net zo min heeft de leerling geluk door die ene wat te gemakkelijke toets.

Verskil

En daarmee zien we precies het verschil met de centrale examens. Ook bij centrale examens is het ondoenlijk om exemplaren te produceren die steeds even moeilijk zijn. Net zo min is vooraf nauwkeurig de moeilijkheidsgraad vast te stellen. Dat zou kunnen als het bestond uit een groot aantal vooraf onder examencondities geteste opgaven. En dat past niet in de aard en opzet van ons centraal examen. Maar dan openbaart zich een probleem.

'Er is geen sprake van zwakkere en sterkere cohorten'

Leerlingen maken maar één centraal examen. Ze kunnen dus geluk hebben of pech. Zelfs een paar tienden kunnen de doorslag geven. Daarom moeten bij het centraal examen wél steeds even hoge eisen aan de kandidaten worden gesteld. En daar zorgt de variabele N-term voor. De N-term compenseert als het ware de verschillen in moeilijkheidsgraad. Maar dan komt de cruciale vraag: hoe komen we erachter óf we met een gemakkelijk examen te maken hebben?

De lat op dezelfde hoogte

Met ondersteuning van Cito heeft het College voor Examens (CvE) daarvoor verschillende methoden. Een aantal examens (op havo en vwo een tiental vakken) wordt nauwkeurig geijkt door middel van pretests en posttests. Zo heeft Frans TL een posttest. Het wordt geijkt aan, bijvoorbeeld, Frans TL 2005. Daartoe krijgt een aantal scholen een toets, samengesteld uit gedeeltes van 2005 en 2011. Die wordt afgenomen in leerjaar 3 van de havo. We weten dat havo-3 en TL-eindexamens niet vergelijkbaar zijn. Maar waar het om gaat is, dat *dezelfde* leerlingen een stukje 2005 én een





stukje 2011 maken. Zo zien we of die leerlingen met de stukjes 2005 meer moeite hebben. Dan is dus 2011 eenvoudiger en moet het een lagere N-term krijgen dan 2005. Uit het rekenwerk, met alle scores, volgt de N-term. Die wordt vastgesteld ongeacht het niveau van de kandidaten. Als zij in 2011 bij Frans zwak zijn, zijn er meer onvoldoendes dan in andere jaren.

Deze vorm van equivalering¹ hanteert het CvE niet voor alle vakken. Posttest en pretest kosten veel geld en tijd. Bovendien leent het systeem zich niet voor vakken met wisselende thema's. Daar komt bij: tien vakken per schooltype geven een stevige basis om de vinger aan de pols te houden. Bij deze vakken zien we van jaar tot jaar geen noemenswaardig verschil in vaardigheid. Er is geen sprake van zwakkere en sterkere cohorten. De examenklas 2011 is niet ineens veel vaardiger of minder vaardig dan die van 2010 bij wiskunde A, natuurkunde of Engels. In de loop van decennia kan de vaardigheid in een vak toe- of afnemen. Maar korte-termijnschommelingen zijn er landelijk niet. Op schoolniveau kan dat natuurlijk anders liggen. Naar redelijkheid mogen we dus aannemen dat de populatie ook bij M&O of geschiedenis in 2011 even vaardig was als in 2010. We weten dat het examen van jaar tot jaar niet precies even moeilijk is, we mogen aannemen dat de leerlingen even vaardig zijn, en kunnen dus uitgaan van een gelijk percentage onvoldoendes.

Om deze reden en op deze manier geldt dus: een van jaar tot jaar wisselende N-term die van

jaar tot jaar de lat op dezelfde hoogte houdt. Een fair en billijk systeem. Dat werd in 2011 onderstreept door het LAKS, dat aangaf dat de normeringstermen overeenkomen met de door hun achterban ervaren moeilijkheidsgraad. *So far, so good.* Het systeem van wisselende N-termen, deels gebaseerd op een constant percentage onvoldoendes, is fair omdat de lat op dezelfde hoogte blijft. Dat wordt ook door leerlingen en docenten zo ervaren. Als bij natuurkunde vwo in 2011 leraren en leerlingen het examen pittig vinden, dan weten ze



dat de N-term wat hoger dan 1 zal zijn. Is het examen een keer relatief gemakkelijk, dan weten beiden dat ze rekening moeten houden met een lagere N-term. Dat hoort nu eenmaal bij het principe.

'Een systeem van wisselende N-termen, deels gebaseerd op een constant percentage onvoldoendes, is fair omdat de lat op dezelfde hoogte blijft'

Het ligt anders als de N-term dicht tegen de (overigens wat willekeurig gekozen) grenzen aan komt. Een N-term van 2 (of incidenteel hoger) volgt op een examen dat heel moeilijk was en veel stress gaf bij kandidaten en corrigerende docenten. En weliswaar worden ze met een hoge N-term gecompenseerd, maar dat neemt die stress niet weg. Liever dus een volgend jaar een iets eenvoudiger examen, met (dus) ook een lagere N-term. Vergelijkbare resultaten, minder stress en minder ophef.

De N-term voor geschiedenis

Lastig is wat bij geschiedenis vwo nu jaren achtereen gebeurt. Een N-term van 0 past op het examen. Maar na een 'lekker' examen ►

► Met equivalering wordt er voor gezorgd dat in opeenvolgende examens bij een gelijke prestatie hetzelfde cijfer hoort.



voelt het wel als een domper. We zijn opgevoed met de gedachte 'de helft goed is ongeveer een 5,5'. En terwijl het reëel is dat je bij een relatief eenvoudig examen méér dan de helft goed moet doen, voelt de N-term die immers achteraf komt, toch als een afstraffing. Het CvE realiseert zich dat terdege. Liever dus een wat pittiger examen met een N-term die meer in de buurt van het midden ligt. Maar – wat te doen? De stof voor het centraal examen is nauw omschreven en in de syllabus vastgelegd. Moeilijker maken door uitbreiding van de stof of door een complexer thema, dat gaat niet. Moeilijker vragen? Dat kán. Maar de ervaring leert dat we daar voorzichtig mee moeten zijn. Een moeilijke vraag is soms niet zozeer moeilijk omdat hij veel denkwerk vergt, maar omdat niet helemaal duidelijk is welk antwoord wordt verwacht. En dat is het soort moeilijkheid dat wel de gemiddelde score doet dalen maar niet de betrouwbaarheid van het examenresultaat. Cito en CvE-vaksectie zijn niet bang voor moeilijker vragen (graag zelfs), maar willen de valkuil van strikvragen of te open (lees: vage) vragen vermijden.

'Het kan zijn dat het probleem bij geschiedenis een probleem van corrigerende docenten is'

Alternatief: accepteren dat geschiedenis vwo gemakkelijker is dan andere vakken. De N-term zonder aanpassing in het examen voortaan op 1,0 zetten. Kán. We passen dan wel de hoogte van de lat aan. En of het vak daar wel zoveel mee opschiet? Willen docenten geschiedenis dat hun vak het imago van vluchtvak krijgt, dat je moet kiezen als je wilt



inloten of dat je zo gemakkelijk aan je diploma komt? Ik denk het niet. Op dit moment is de 'succeskans' bij de verschillende vakken in grote lijnen even groot. Voor de ene leerling is natuurkunde gemakkelijker dan geschiedenis, voor de andere andersom, afhankelijk van de aanleg. Maar grosso modo: de decaan kan niet roepen dat je geschiedenis moet kiezen omdat dat een makkie is. Zou hij dat wel kunnen roepen, dan raakt het vak beschadigd.

Intussen: CvE en Cito blijven zoeken naar pittiger vragen, naar goede doordenkers. Maar dat gaat zorgvuldig, met kleine stappen. Geen kandidaat wil proefkonijn worden voor het testen van totaal nieuwe vraagvormen. Dus vooralsnog is de les uit de historie: reken je niet rijk na een overzichtelijk en hanteerbaar examen, ga er maar van uit dat de N-term van 0,0 de passende is. En gebruik je CE-opgaven in het SE, ga dan van dezelfde norm uit!

De rol van de docent

Op één aspect ben ik niet ingegaan: de rol van de docent. Inderdaad, het is *mogelijk* dat het examen vwo geschiedenis niet eenvoudig was. Dat CvE en Cito dat alleen maar denken, omdat de scores goed waren – omdat docenten hun leerlingen matsten. Het *kan* zijn dat het probleem bij geschiedenis een probleem van corrigerende docenten is. Maar CvE gaat daar niet van uit. Het lijkt ons in strijd met de professionaliteit van de docent. Bovendien: waarom dan alleen bij geschiedenis? Elk op hun eigen manier zijn andere examens ook tamelijk open. En daar horen we dit geluid niet. Maar goed, stel dat het zich voordoet (en het zal vast wel eens gebeuren), dan is het aan de beroepsgroep om elkaar daarop aan te spreken. Dat los je niet op met de N-term. ■