

De N van 'niet weer hè'?

Een persoonlijke noot over de N-termen vwo geschiedenis

Wanneer u dit leest, bent u vast en zeker een gemotiveerde en betrokken docent. Vermoedelijk geeft u les aan een of meer examenklassen havo en/of vwo. U maakt uw eigen schoolexamens en probeert daarbij de aansluiting op de exameneisen goed in de gaten te houden.

Joop van den Berg
is historicus en
docent geschiedenis
Tweede Fase

In uw vragen combineert u de bekende historische vaardigheden met kennis op niveau, en u lardeert alles met prachtige beeld- en tekstbronnen. Kortom: u maakt een uitdagende toets, die (omdat u uw lessen ook zo heeft ingericht) vanzelfsprekend zowel aansluit bij het examen als bij het niveau van uw leerlingen. Daarom is uw normering helder: de N-term is gewoon 1,0. Een gemiddeld cijfer tussen de 6,0 en 6,6 mag verwacht worden. Keurig.

Dan volgt het centraal examen, en wanneer u dit even inkijkt terwijl uw klassen zich in het zweet werken in de omgebouwde gymzaal, stelt u tevreden vast: het is te doen, niet te makkelijk of te moeilijk, het sluit aan bij wat ze moeten kunnen.

Eenmaal thuisgekomen gaat u corrigeren. Hoewel het antwoordmodel een *crime* is, er veel te veel alternatieve antwoorden en interpretaties mogelijk zijn en consensus tussen collega's bij slechts zeer weinig vragen bereikt wordt, mag u na dagen ploeteren tevreden zijn: een score van gemiddeld 50 van de 77 punten moet toch wel minstens hetzelfde resultaat opleveren als dat u bij uw schoolexamens zo gewetensvol heeft bereikt. Misschien is een kleine 7 zelfs mogelijk...

...met wat barstjes

Hier eindigt al jaren het sprookje van (met name) de vwo-docent. Midden juni komen de N-termen binnen en wordt al snel duidelijk dat uw resultaten bij lange na niet aan uw verwachtingen voldoen. Een leerling die ongeveer tweederde van alle scorepunten heeft behaald, wordt opgescheept met een waardering die nog maar net of net niet de 6,0 haalt. En zo is het al jaren. Sinds 2002 (verder terug heb ik niet gekeken) is er bij het vwo geen normering van 1,0 of hoger toegepast, zoals uit de tabel blijkt. In zes van de tien jaren is de norm zelfs niet hoger dan 0,5 geweest.

Aan de tabel valt ook op dat het gemiddelde cijfer sinds 2005 (!) een 6,1 is. Vorig jaar werd bij de examenevaluatie in Arnhem expliciet gesteld dat de N-norm moest zorgen voor ofwel een 6,3 gemiddeld, ofwel 30% onvoldoende. Hieraan wordt dus al zeven jaar niet meer voldaan. Dit kan toeval zijn, maar dat wil er bij ons niet zomaar in. Daarom hierbij wat kanttekeningen betreffende de N-termen bij het vwo examen geschiedenis.

De grote boze Wolf

Wanneer je bij het Cito of CvE uitleg vraagt over de constant lage N-termen, word je eigenlijk terugverwezen naar jezelf: wij, docenten, voeren in Wolf de scorepunten in van (minstens vijf, alfabetisch geselecteerde) leerlingen. Zijn wij te genereus, dan trapt men bij het bepalen van de N-term op de rem, met als gevolg een lage norm en dus relatief lagere cijfers. De invloed die dit heeft is aan te tonen door te kijken naar echt goede leerlingen. Iemand die in het afgelopen examen 69 van de 77 punten had gescoord, had slechts 156 leerlingen in heel Nederland vóór zich (die het dus nog beter deden) hetgeen overeenkomt met 1,27%. Toch haalt zo'n leerling door de N-norm 'slechts' een 8,2. Dat is gezien zijn plaats ten

jaar	lengte van de schaal	n-term	% onvoldoende	gem. cijfer
2002	90	0,8	23	6,3
2003	86	0,9	22	6,2
2004	84	0,5	13	6,5
2005	80	0,4	23	6,1
2006	86	0,3	26	6,1
2007	85	0,9	27	6,1
2008	77	0,7	26	6,1
2009	77	0,3	28	6,1
2010	76	0,1	25	6,1
2011	77	0,1	26	6,1



opzichte van de mediaan natuurlijk meer dan vreemd.

Gelet op het consequente gemiddelde sinds 2005, moet het wel zo zijn dat de scorepunten zoals ze worden ingevoerd in Wolf worden gemiddeld en dan 'vertaald' naar de blijkbaar gewenste 6,1 als gemiddeld examencijfer. Hier is wel wat op aan te merken.

Een andere kijk op presteren

Het uitgaan van en toerekenen naar een vaststaand of gewenst gemiddelde, heeft erg veel haken en ogen. Een paar daarvan zet ik graag op een rij.

- sinds de invoering van de nieuwe studielasten krijgen leerlingen meer geschiedenis-onderwijs. De examenomvang, gebundeld in de twee thema's en examenkaternen dus, is ongewijzigd gebleven. Is het gewaagd om te veronderstellen dat bij meer onderwijs de prestaties verbeteren? En als dat zo is, moeten we die prestaties dan gaan nivelleren ('bestrafen' zou hier wat ons betreft ook het woord kunnen zijn) door de N-term te verzwaren? Als een voetbalploeg harder traint en daardoor beter wordt dan de concurrentie, zal de KNVB dan besluiten dat deze ploeg elke wedstrijd reglementair met een 1-0 achterstand moet beginnen? Moet Usain Bolt binnenkort een zandzak meedragen op zijn 100 meter omdat hij anders de rest veel te snel af is? Dat is de omgekeerde wereld.

- nu een lage N-term een jaarlijks fenomeen is geworden, moet ook de vraag worden gesteld of de examens wel het juiste niveau hebben. Blijkbaar slagen examenmakers er niet in om een examen te maken dat uit zichzelf goed determineert en waardeert en dus een in onze ogen wenselijke en logische N-term van 1,0

oplevert. Dat mogelijke onvermogen wordt vervolgens op het bord van de leerling (en de docent, maar daarover straks meer) gelegd. Wij stellen dat als een professionele omgeving een examen produceert, de leerling en docent ervan uit moeten kunnen gaan dat het examen deugt en dus een neutrale N-term oplevert.

- uitgaan van een vast gewenst gemiddelde doet geen recht aan verschillen tussen cohorten: er zullen vast betere en slechtere jaren tussen zitten. Maar bovenal: ik denk dat dit ook samenhangt met de onderwerpen die worden behandeld. Mij staan nog de diepe verzuchtingen voor ogen die in 2004 het thema 'Nederland en zijn gezagsdragers' aan de leerlingen ontlokte.

De docent als factor?

Corrigeren van examens is mensenwerk, en dat zal het altijd blijven. Maar gezien de bovenstaande problematiek mag er langzamerhand wel een discussie ontstaan over de wijze van corrigeren, het correctiemodel en de eerste en tweede correctie. Deze blijken namelijk wel de gelijk invloed te hebben of te kunnen hebben op de N-term en dus het examencijfer.

Een (door werkdruk of te veel meegaandheid ten opzichte van de groep leerlingen die hij of zij zo goed kent) overenthousiast of te gemakzuchtig deel van de docenten, geeft te hoge ►



scorepunten, als je Cito en CvE mag geloven. Of de docenten in den lande ontdekken de beperktheid van het antwoordmodel en zien veel meer juiste antwoorden, waar scorepunten voor kunnen worden toegekend. Elke weldenkende docent zal zijn leerlingen die punten niet onthouden.

Die ruimhartige toekenning zorgt voor een lage N-term, zoals hierboven al uitgelegd. Maar

'In de praktijk is het haast onmogelijk om de tweede correctie even accuraat te doen als de eerste'

de ervaren of zeer consciëntieuze docent, die niet ruimhartig scorepunten uitdeelt, heeft met diezelfde N-term te maken. Zijn leerlingen worden dus benadeeld.

Om deze uitwassen tegen te gaan is de tweede correctie ooit in het leven geroepen. Op zich een prima idee, want twee zien meer dan een en kunnen elkaar bijsturen. In de praktijk is het (tenzij je maximaal ongeveer 20 leerlingen hebt en dus niet meer dan 20 examens na moet kijken) haast onmogelijk om een tweede correctie even accuraat en uitgebreid te doen als de eerste. Een misschien extreem maar al jaren optredend rekenvoorbeeld uit de lespraktijk van diverse collega's:

Eerste correctie:

100 examenleerlingen x 60 minuten per leerling (VGN norm) = 100 uur nakijkwerk, ofwel 2,5 lesweken van 8 uur. Naast de overige lessen, het surveilleren en andere schooltaken eigenlijk al niet te doen. Ook al vallen de examenklassen weg vanaf mei, dit aantal vrijgekomen uren staat in geen verhouding tot de uren correctietijd.



Tweede correctie:

Omdat je meestal een zelfde aantal examens ter tweede correctie krijgt, betekent dit dat je ergens in de meestal twee weken voorafgaande aan de bekendmaking van de norm en de cijfers, nogmaals een slordige 100 uur vrij moet maken, dus zo'n 10 uur per dag. Onmogelijk.

Het bovenstaande toont aan dat de tweede correctie in praktijk de uitwassen van mogelijk te soepel nakijken door collega's niet zal wegnemen. Er zal dus geen zuiverende werking uitgaan van de collegiale check op de toekenning van scorepunten en dus op de N-term. Reden om dit aspect ooit eens nader te bekijken.

Help, de inspectie!

Steeds meer worden collega's door hun schoolleiding op de vingers gekeken en getikt wanneer het gaat om hun examencijfers, en vooral wanneer het gaat om de verschillen tussen het schoolexamencijfer en het centraal examencijfer. Is dit verschil 0,5 of hoger (en naar verluidt wordt dat binnenkort 0,3), dan heeft de sectie of de docent een probleem. Op zich begrijpelijk: controle moet er zijn, want scholen moeten goed presteren om leerlingen uiteindelijk op de juiste plaats in de maatschappij te laten belanden. Maar in al het voorgaande is wel duidelijk geworden dat er op dit moment bij het becijferen van het examencijfer geschiedenis veel meer aan de hand is dat van invloed is op het examengemiddelde. De N-term, die een dwingend stempel drukt op goed of slecht presteren als leerling en docent, wordt door bijna alle docenten geschiedenis als een loterij ervaren. Maar dan wel een loterij volgens het aloude louché casinoprincipe: onder de deelnemers zijn er alleen verliezers.

En verder?

Ik ben van mening dat alle hierboven aangedragen argumenten een rol moeten spelen in de gesprekken die onder andere de VGN voert met Cito en CvE. Ook zullen we als docenten, verenigd in de VGN, het gesprek met de inspectie aan moeten gaan.

Daarnaast, en als laatste, zullen we ons moeten voorbereiden op de impact dat al het bovenstaande zal hebben op de resultaten van het nieuwe examen. Als het inderdaad 'tijdvakken vullen' wordt, zoals twee collega's laatst al in de krant en in Kleio lieten weten, moet er binnenkort misschien een N-term onder de nul vastgesteld worden... ■