

“Vroeger waren de examens Engels veel moeilijker, toch?”

Staatssecretaris Dekker toonde zich in 2013 een tevreden beleidsman. Hij was verantwoordelijk voor de invoering van de aanscherping van de zak/slaagregeling in 2012 en de kernvakkenregel in 2013 en in het examenjaar 2013 waren de gevolgen al te merken. Het stellen van hogere eisen had volgens hem als gevolg dat de leerlingen beter waren gaan presteren¹. Maar veel docenten onderschrijven deze juichende conclusie niet. Met name docenten die Engels op HAVO niveau onderwijzen hebben twijfels. Die zien voor het kernvak Engels hogere cijfers op het CE, maar ervaren geen vaardigheidsstijging bij de leerlingen. Zij menen vooral een versimpeling van de examens na 2012 te zien². Wie heeft er nu gelijk? Zijn de examens met opzet ‘lichter’ gemaakt om de nieuwe uitslagregels een succes te laten worden? Laten we eens wat veelgehoorde aannames, vragen en feiten op een rij zetten.

“Ze zouden eens een docent naar de examens moeten laten kijken!”

Onderwijservaring is belangrijk. Dat beseft zowel het College voor Toetsen en Examens (CvTE), het bestuursorgaan dat in opdracht van het Ministerie van OC en W verantwoordelijk is voor de examens, als het Cito, het kennis- en onderzoeksinstituut dat de examens in opdracht maakt. Aan de basis van elk examen staan daarom docenten die lesgeven in eindexamenklassen van het vak en het schooltype van het centrale examen waar ze bij betrokken zijn. Deze docenten leveren het bronmateriaal aan, maken de opgaven en assisteren en adviseren een aan het Cito verbonden toets-deskundige. Een door het CvTE aangestelde vaststellingscommissie, die ook bestaat uit vakdocenten en een voorzitter heeft die werkzaam is in het meest nabijgelegen hogere onderwijs, bewaakt het niveau en de kwaliteit. Zowel tussentijds als bij de uiteindelijke samenstelling van een compleet examen fungeert deze vaksectie als controleur en inhoudelijk eindverantwoordelijke³. Al het materiaal gaat bovendien langs een onafhankelijke externe screener (een vakdeskundige die noch aan het Cito, noch aan het CvTE verbonden is), langs een neerlandicus, en, bij moderne vreemde talen, ook nog langs een externe ‘native speaker’ van de doeltaal. Een groep ervaren en actieve docenten, die geregeld vernieuwd wordt, afkomstig van verschillende scholen uit verschillende delen van het land, in verschillende rollen en onafhankelijk van elkaar, heeft dus een zeer bepalende invloed op de inhoud en niveau van het examen.

“Ik kan toch zien aan de prestaties van mijn leerlingen dat ze gemakkelijker zijn!”

Het is maar de vraag of individuele docenten wel genoeg kunnen zien om een oordeel te kunnen vellen. De waargenomen prestaties worden immers bepaald door enerzijds de vaardigheid van de groep kandidaten en anderzijds de moeilijkheidsgraad van het afgenomen examen. Om te weten hoe de examens zich verhouden ten opzichte van elkaar, over de jaren heen, moet men wel beseffen dat deze beide variabelen een rol spelen. Omdat het streven van het CvTE is dat het diploma, over de jaren heen eenzelfde waarde vertegenwoordigt, onderzoekt Cito in opdracht van het CvTE elk jaar hoe de populatie zich ontwikkelt en wat de relatieve moeilijkheidsgraad is van de examens. Ieder jaar worden er pre-tests en post-tests afgenomen met daarin een set ankeropgaven in combinatie met delen van het nieuwe examen⁴. De gegevens die dit oplevert stelt het CvTE en Cito in staat de moeilijkheid van ieder examen los te zien van de vaardigheid van de actuele examenpopulatie.

Van het CE HAVO Engels verzamelt Cito sinds 1995 deze aanvullende gegevens: al meer dan 20 jaar is de moeilijkheidsgraad van jaar tot jaar anders (zie afbeelding 1), maar deze onvermijdelijke variatie⁵ is, met behulp van de N-term, te corrigeren.

¹ zie “Eindexamen 2013: meer geslaagden en hogere cijfers” op rijksoverheid.nl

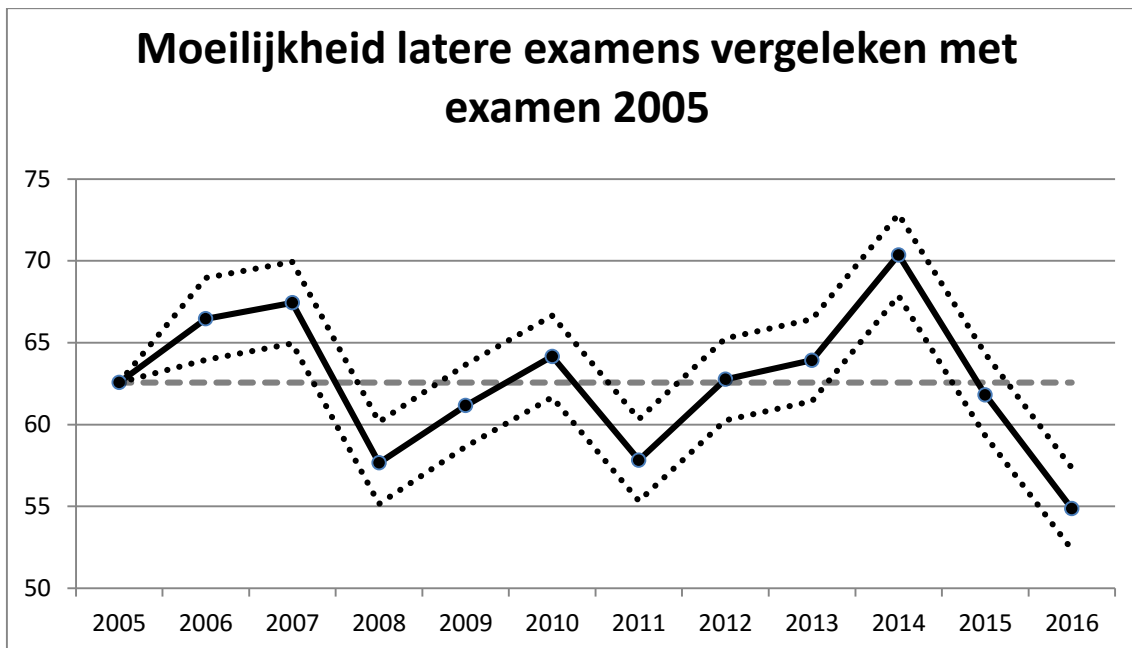
² Na afname van het CE mogen docenten hun inschatting geven over de moeilijkheidsgraad van het examen. Kijk bij ‘examenverslagen-waardering per examen’ op cito.nl voor de resultaten.

Overigens, in 2016 vond men het examen HAVO Engels vooral te moeilijk. In combinatie met de toegekende N-term van 1.1 kwam het gemiddelde cijfer in 2016 uit op 6,8, dicht bij het gemiddelde cijfer gescoord in de jaren er direct voor, de examenjaren waarvan men vond dat de leerlingen een (te) gemakkelijk examen hadden gehad.

³ voor meer informatie over de totstandkoming van de examens, zie: <https://www.examenblad.nl/onderwerp/ik-heb-een-examenklas> en http://www.cito.nl/onderwijs/voortgezet%20onderwijs/centrale_examens/de-juiste-examenvraag

⁴ De psychometrische afdeling van het Cito controleert de moeilijkheidsgraad en de vaardigheidsontwikkelingen van de leerjaren met behulp van onder andere een pre- of post-test. Voor deze pre- of post-tests worden ankeropgaven gebruikt: niet eerder gepubliceerde opgavensets die al meerdere jaren worden voorgelegd aan bepaalde groepen leerlingen. Door deze ankeropgaven te combineren met nieuwe examenopgaven krijg je een nauwkeurig beeld van de nieuwe opgaven in vergelijking met de ankeropgaven.

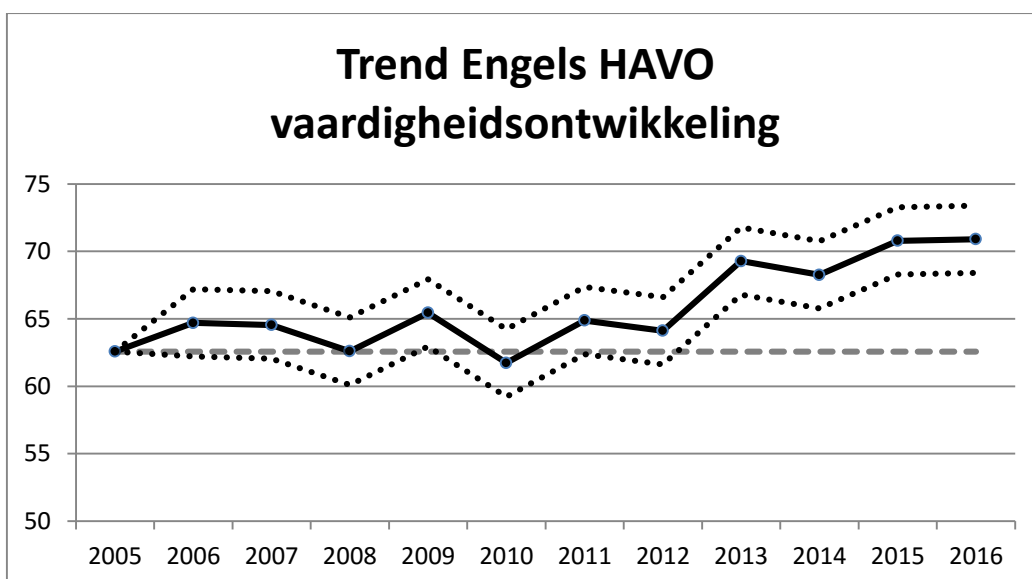
⁵ De examens zijn meteen na afname openbaar en door iedereen in te zien en te beoordelen. Aan deze praktijk van openheid zijn we in Nederland zeer gehecht, maar het heeft wel tot gevolg dat er ieder jaar weer compleet nieuwe examens moeten worden ingezet. Deze examens zullen nooit exact even moeilijk zijn als het examen van het jaar ervoor, maar wel altijd voldoen aan wat in het examenprogramma beschreven is. De syllabus is een operationalisatie van het examenprogramma (zie examenblad.nl).



afbeelding 1:

Visualisering van de fluctuatie in moeilijkheid van de examens HAVO Engels van 2005 tot en met 2016: door weergave van hoe de examenpopulatie van 2005 volgens statistische berekeningen op deze examens zou scoren, vergeleken met hoe ze op hun eigen examen hebben gepresteerd. De rechte grijze stippellijn toont hoe goed delichting van 2005 het examen 2005 heeft gemaakt. De zwarte lijn is de weerspiegeling van de feitelijke moeilijkheid van de verschillende examens. Het toont het percentage goed dat de lichting 2005 zou scoren op dit examen. Hoe lager het percentage, hoe moeilijker het examen. De stippellijnen naast de zwarte doorlopende lijn geven aan in hoeverre er onzekerheid kan bestaan over de uitkomsten.

Zoals al eerder gezegd, niet alleen de moeilijkheid van het examen kan fluctueren, ook de groep leerlingen die examen doet, zal qua vaardigheid niet elk jaar van exact hetzelfde niveau zijn. Afbeelding 2 laat zien hoe de verschillende jaarpopulaties presteren op het examen van 2005. Berekeningen op basis van meerjarige vergelijkingen van post-test resultaten tonen aan dat de leerlingen, los van hoe moeilijk of gemakkelijk een examen in een bepaald jaar is geweest, steeds beter presteren. We zien een duidelijk stijgende lijn, vooral na 2012, vermoedelijk ten minste deels samenhangend met de veranderde uitslagregel. Als we willen dat de cijfers over jaren vergelijkbaar zijn, zullen we deze ontwikkeling van de examenpopulatie moeten terugzien in de cijfers.



afbeelding 2:

De prestaties van de leerlingen van de jaarpopulaties 2005 tot en met 2016 (volgens statistische berekeningen) op het examen 2005. De rechte stippellijn (tussen 60 en 65) laat zien wat de jaar-laag van 2005 op het examen van 2005 scoorde. De populaties uit latere jaren met een punt boven de lijn scoorden beter, onder de lijn minder, waarbij de

verschillen in moeilijkheid tussen de verschillende examens zijn verdisconteerd. De stippellijnen naast de zwarte doorlopende lijn geven aan in hoeverre er onzekerheid kan bestaan over de uitkomsten.

Docenten beschikken niet over grote hoeveelheden aanvullende data, maar zien alleen hun eigen leerlingen. Ze voelen goed aan dat de examens van jaar tot jaar niet helemaal van gelijk niveau zijn. En ze zien dat hun leerlingen hogere cijfers scoren, en concluderen op basis daarvan dat de examens gemakkelijker zijn geworden. Analyses op grond van de afname gegevens en aanvullende data laten, wat betreft de leesvaardigheid Engels, echter een duidelijk ander beeld zien. Wat aan de stijging ten grondslag ligt is weten we niet precies, maar wat we wel weten is dat naar aanleiding van de nieuwe examenregelingen er door de meeste scholen verschillende maatregelen zijn genomen zoals onder andere strengere selectie in voor-eindexamenjaren, minder gemakkelijkere bevordering bij tegenvallende rapportcijfers voor met name de kernvakken, uitbreiding van lesuren voor de kernvakken en examentrainingen⁶. Daarnaast bieden steeds meer scholen (vroeg) TTO-onderwijs aan en is de rol van het Engels, ook buiten het klaslokaal, via internet, games etc. de laatste jaren alleen maar toegenomen. Bedenk echter wel dat het CE alleen gaat over leesvaardigheid en dat het dus niet zo hoeft te zijn dat eenzelfde ontwikkeling is waar te nemen bij de andere, in het schoolexamen getoetste, vaardigheden.

“Via manipulatie met de N-term krijgen ze precies wat ze willen.”

De bij het eindexamen behaalde score-punten worden via een formule (de N-term) omgezet naar een cijfer op een schaal van 1 tot 10⁷. Die N-term kan variëren van 0 tot 2, waarbij een hogere N-term bij eenzelfde aantal gehaalde scorepunten resulteert in een hoger cijfer. Er is door het CvTE gekozen voor een variabele N-term om recht te doen aan de eerder besproken aanwezige moeilijkheidsverschillen bij de examens. De flexibiliteit van de N-term heeft dus inderdaad een duidelijk doel: het opvangen en corrigeren van verschillen tussen de jaren om de waarde van het diploma over meerdere jaren vergelijkbaar te houden. Zolang de eisen met betrekking tot het gewenste vaardigheidsniveau niet zijn veranderd zou het onrechtvaardig zijn voor de nieuwe generatie leerlingen als hun hogere vaardigheidsniveau niet gewaardeerd wordt met een hoger cijfer.

Een van de aannames die nogal eens rondgaat in docentenkamers is dat de methodiek ter bepaling van de N-term vrij simpel is: de verantwoordelijken willen ieder jaar een vergelijkbaar percentage onvoldoendes, dus dat is het gegeven van waaruit de N-term wordt vastgesteld. Een blik op het overzichtsstaatje voor het CSE HAVO Engels (zie tabel hieronder) maakt duidelijk dat deze aanname niet klopt.

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
gem p waarde	.63	.68	.62	.58	.64	.63	.60	.64	.71	.76	.70	.63
N-term	0.5	0	0	1.3	0.4	0.4	0.7	0.3	0.3	0	0.5	1.1
% onvoldoende	29.0	25.0	21.0	24.0	28.0	32.8	30.6	28.4	16.2	12.1	12.9	16.0
gem cijfer	6.1	6.5	6.5	6.2	6.5	6.2	6.1	6.1	6.7	6.9	6.8	6.8

tabel:

Van 2005 tot en met 2012 fluctueert het percentage onvoldoendes van 21% tot ruim 32%. Vanaf 2013 varieert het van 16% tot 12%, een lager percentage dan de jaren voor de aanscherping van de zak/slaagregeling, maar nog steeds is het percentage onvoldoendes niet gelijk van jaar tot jaar.

Wat wel een belangrijke rol speelt bij de normering is de gemiddelde p-waarde (het door de leerlingen behaalde percentage van de maximaal haalbare score). Deze gemiddelde p-waarde is niet extern te manipuleren; de leerlingen scoren per jaar nu eenmaal wat ze scoren. Dat laat onverlet dat de p-waarde zonder extra informatie nog tamelijk nietszeggend is: een examen met een p-waarde van .70 (gemiddeld is dan dus 70% van de te scoren punten gehaald) is niet vanzelfsprekend een gemakkelijker examen dan

⁶ zie Terugblikken op: <https://www.examenblad.nl/examen/engels-havo-2/2013?topparent=v41h1h4i9qd> en <https://www.examenblad.nl/examen/engels-vwo-2/2014?topparent=v41h1h4i9qe>

⁷ Hoe de N-term tot stand komt wordt uitgelegd in de notitie “De Normeringssystematiek van de centrale examens VO” (als pdf te downloaden: www.examenblad.nl/onderwerp/normering-centrale-examens/2017).

een examen uit een eerder jaar dat toen een lagere gemiddelde p-waarde had. Immers, het kan ook zo zijn dat de populatie vaardiger is geworden⁸.

Pas nadat alle data statistisch zijn geanalyseerd, met als doel te bepalen waar het examen zich bevindt in vergelijking met eerdere examens wat betreft moeilijkheidsgraad en hoe de populatie zich heeft ontwikkeld ten opzichte van eerdere jaarlagen, wordt duidelijk welke N-term resulteert in uitkomsten die vergelijkbaar zijn met eerdere jaren. Er wordt ook bekeken of het examen in zijn geheel en de individuele opgaven apart valide en betrouwbaar zijn. De toets- en itemanalyses (TIA) bij examens zijn openbaar en te vinden op de site van het Cito⁹. Als docenten problemen met opgaven melden, dan wordt hier altijd goed naar gekeken door de toetsdeskundige en de vaststellingscommissie van het CvTE. In enkele gevallen kan dit ook leiden tot een aanpassing van de N-term. Uitgangspunt is te allen tijde dat een kandidaat niet door een foute vraagstelling of andere onvolkomenheid gedupeerd mag worden.

“Meten is weten.”

Wat betreft het eindexamen HAVO Engels, gemaakt door docenten en met het doel leesvaardigheid te toetsen, wijzen alle gegevens dezelfde kant op: de vaardigheid van de kandidaten is sinds 2013 significant hoger dan deze was in de jaren ervoor. De hoge cijfers en lage percentages onvoldoendes die we de laatste jaren bij de examens HAVO Engels zien, worden niet veroorzaakt door het inzetten van gemakkelijkere examens of het toekennen van te hoge N-termen. De uitspraak van de staatssecretaris dat er beter gepresteerd wordt op de examens sluit aan bij de uitkomsten van jarenlange, uitgebreide onderzoeken, analyses en berekeningen.¹⁰

Natalie Schols, Cito

met dank aan Maarten de Groot, voor de grafieken en redactionele ondersteuning.

⁸ Dit is waar het artikel van Age Letor op Saragossa.nl (sargasso.nl/n-term-eindexamens-te-soepel/) niet klopt. Zie ook het antwoord van het CvTE op deze blog: www.examenblad.nl/nieuws/20161129/reactie-van-het-cvte-op-de-blog/2017

⁹ Te vinden op cito.nl, bijvoorbeeld bij: voortgezet onderwijs-centrale examens-schriftelijke examens havo en vwo. Bij elk apart examen per jaar staan de examens zelf, de correctievoorschriften, en zodra ze gereed zijn, de TIA's

¹⁰ Voor de examens Engels op VWO niveau zien we vergelijkbare ontwikkelingen.