



Het landelijk
gemiddelde
voor Frans en
Duits is in
2023 een 6,3

ALLES STAAT VOORAF VAST

Dit communiceert de clustermanager talen VWO/HAVO van het CvTE in 2016

Examenblad.nl

Officiële mededeling

Eindexamen 2016: normering en resultaten centrale examens Duits, Engels en Frans havo/vwo

Geachte heer, mevrouw,

Elk jaar ontvangen wij vragen van docenten over het verschil in gemiddeld cijfer voor de centrale examens Duits, Engels en Frans havo en vwo. Al enkele jaren is het gemiddeld cijfer van examenkandidaten Engels havo en vwo hoger dan het gemiddeld cijfer dat kandidaten halen voor de centrale examens Duits en Frans havo en vwo. Anders dan soms wordt gedacht, betekent dit niet dat de centrale examens Engels havo en vwo in de loop van de jaren makkelijker zijn geworden of dat het College voor Toetsen en Examens deze examens soepeler normeert.

Voor de normering van de centrale examens Duits, Frans en Engels wordt de moeilijkheidsgraad van het nieuwe te normeren examen door een pre- of een posttest vergeleken met een referentie-examen. Deze werkwijze maakt het mogelijk om populatie-onafhankelijk te normeren en ook vaardigheidsverschillen tussen de populatie examenkandidaten van het ene jaar te vergelijken met die van een ander jaar. Zo heeft het CvTE een duidelijke stijging van de vaardigheid van kandidaten Engels havo en vwo kunnen vaststellen sinds de invoering van de kernvakkenregeling in 2013. Deze vaardigheidsstijging is de reden van het hogere gemiddeld cijfer voor deze centrale examens in het huidige examenjaar.

Bij de centrale examens Duits en Frans havo en vwo ziet het CvTE een vaardigheidsstijging zoals bij Engels niet. Daarom is het gemiddeld cijfer voor Duits havo en vwo ook dit jaar duidelijk lager dan het gemiddeld cijfer voor Engels havo en vwo. Daarbij speelt dat er bij Duits havo en vwo in het begin van deze eeuw sprake is geweest van een vaardigheidsdaling, terwijl de eisen voor de centrale examens Duits havo en vwo niet zijn bijgesteld; het niveau is gehandhaafd. Inmiddels is al jaren de vaardigheid van examenkandidaten Duits vrij constant, maar de daling in vaardigheid in de eerste acht jaar van deze eeuw is daarmee niet opgeheven.

De cijfers voor de centrale examens Frans havo en vwo laten een ander beeld zien, hoewel ook hier net als bij Duits geen sprake is van een duidelijke vaardigheidsstijging. Met de laagste W-term binnen de afgesproken bandbreedte met 0,0 als ondergrens was het niet mogelijk om met de examens Frans havo en vwo 2016 vergelijkbare eisen aan examenkandidaten te stellen als in voorgaande jaren. Het hoge gemiddelde cijfer voor het centraal examen Frans vwo 2016 betekent dus niet dat er sprake is van een grote vaardigheidsstijging van kandidaten Frans vwo. Als we met dit examen gelijke eisen hadden kunnen stellen, dan hadden kandidaten Frans vwo gemiddeld een cijferpunt lager gescoord.

Met de centrale examens Frans havo en vwo 2016 is geen kandidaat benadeeld. De beschreven situatie bij Frans is echter niet wenselijk. De eerste gesprekken over maatregelen om dit voor de toekomst te voorkomen, zijn in dit kader inmiddels met deskundigen van het CvTE en Cito gevoerd.

Wij hopen u met dit bericht inzicht te hebben gegeven over de wijze waarop de centrale examens Duits, Engels en Frans worden genormeerd.

Met vriendelijke groet,

Freya Martin
clustermanager talen havo/vwo
College voor Toetsen en Examens

Deze mailing is verzonden aan de directeuren, (plaatsvervangend) examensecretarissen en de geregistreerde docenten Duits, Engels en Frans vwo, havo.

Zo kwam ik er achter

“Met de laagste N-term binnen de afgesproken bandbreedte met **0,0 als ondergrens** was het niet mogelijk om met de examens Frans havo en vwo 2016 vergelijkbare eisen aan examenkandidaten te stellen als in voorgaande jaren. Het hoge gemiddelde cijfer voor het centraal examen Frans vwo 2016 betekent dus niet dat er sprake is van een grote vaardigheidsstijging van kandidaten Frans vwo. **Als we met dit examen gelijke eisen hadden kunnen stellen, dan hadden kandidaten Frans vwo gemiddeld een cijferpunt lager gescoord.**”

<https://nieuwsbrief.examenblad.nl/wv/vk57gs4rlgyq/8o3ahkblja4vmg53>

Op 24 januari 2020 zegt het CvTE Frans en Duits HAVO een 6,3 toe



Zie hier:

Een versoepeling van 0,5 cijferpunt voor Duits en Frans havo sluit aan bij de resultaten van ons onderzoek en de prestatie-eis van andere niet-kernvakken in het havo met het gemiddeld cijfer 6,3. Met deze maatregel beogen wij de prestatie-eis weer in overeenstemming te brengen met de vaardigheid van de leerlingen. Tevens hopen wij dat hiermee ruimte ontstaat voor het inhoudelijk debat over de toekomst van deze vakken in het havo.

Hoe pakt die Kernvakregeling uit?

In 2019 was het landelijk gemiddelde Nederlands VWO een 6.2 en dat van Wiskunde B een 7. De normering bij Nederlands was 0.6 en bij Wiskunde B 1.8. Een normaal examen streeft naar een norm van 1. Bij Nederlands zijn er dus 0.4 punten afgetrokken en bij Wiskunde B heeft iedereen gratis 0.8 punten erbij gekregen. Eigenlijk is het dus zo dat:

$6.2 + 0.4 = \mathbf{6.6}$ het ongemanipuleerde cijfer is voor Nederlands

$7.0 - 0.8 = \mathbf{6.2}$ het eigenlijke cijfer voor Wiskunde B is.

Kinderen kregen dus collectief strafpunten voor Nederlands en konden hun nagestreefde cum laude op hun buik schrijven, en dat is wreed want Nederlands was beter gemaakt dan Wiskunde B.

Gratis bier en wiskunde B certificaat

Als je gaat spelen met de cijfers rond de normering en de gemiddelden dan zie je dat 41% van de VWO leerlingen wiskunde B in 2000 zakten. Dat was vóór de invoering van de vermaledijde kernvakregeling. In 2022 zakten op Wiskunde B nog maar 26%. Leuk voor de kandidaten, maar de normering was 1.8. Hoeveel waren er gezakt als ze geen 0.8 gratis punten hadden gekregen? Toen 41% zakte was de normering 1.3. Zij kregen maar 0.3 punt cadeau.

Beetje pervers is het wel

De perverse prikkel van de kernvakregeling is dat scholen meer tijd in die drie vakken steken en dat dus de andere vakken wegzakken. Ik was vorige week bij het *Levende Talen* congres en hoorde van de mevrouw van het *Cito of CvTE**) dat de kandidaten Engels het tegenwoordig zo ontzettend goed deden. Ik moest lachen en zei dat het eindexamen VWO Engels zo makkelijk is dat het in de helft van de tijd gemaakt kan worden dan die ervoor staat en dat het zo eenvoudig is dat je het geblinddoekt en pijltjesgooiend kunt halen met een normering voor watjes. Dat was niet waar, zei ze:

Wanneer denkt u dat de kernvakregeling is ingegaan als u kijkt naar het aantal onvoldoendes en naar de gemiddelde score? 2013?

Afdeling	Examenjaar	Vak	Nterm	Gemiddeld_cijfer	Percentage_onvoldoende
VWO	2011	Engels	0,4	6,1	26
VWO	2007	Engels	0,1	6,2	30
VWO	2008	Engels	0,5	6,2	28
VWO	2012	Engels	0,6	6,3	24
VWO	2009	Engels	1,1	6,3	28
VWO	2010	Engels	1,1	6,3	27
VWO	2000	Engels	0,2	6,4	20
VWO	2003	Engels	0,4	6,4	22
VWO	2002	Engels	0,7	6,4	22
VWO	2004	Engels	1,1	6,4	25
VWO	2005	Engels	0,9	6,5	23
VWO	2006	Engels	0,4	6,6	21
VWO	2001	Engels	0,5	6,6	22
VWO	2014	Engels	0,4	6,8	12
VWO	2021	Engels	0,4	6,9	10
VWO	2019	Engels	0,7	6,9	11
VWO	2013	Engels	1,1	6,9	13
VWO	2018	Engels	1,3	6,9	11
VWO	2022	Engels	0,4	7	12
VWO	2015	Engels	0,5	7	11
VWO	2016	Engels	1,4	7	11
VWO	2017	Engels	1,5	7	13

Minder grappig bij Frans

Afdeling	Examenjaar	Vak	Nterm	Gemiddeld_cijfer	Percentage_onvoldoende
VWO	2016	Frans	0	7,1	7
VWO	2005	Frans 12	0,6	6,5	19
VWO	2012	Frans	1,5	6,5	21
VWO	2001	Frans	0,3	6,4	22
VWO	2006	Frans 12	0	6,5	22
VWO	2010	Frans	0,7	6,3	22
VWO	2013	Frans	0,8	6,3	24
VWO	2008	Frans 12	0,6	6,3	25
VWO	2015	Frans	0,5	6,3	25
VWO	2018	Frans	0,6	6,2	25
VWO	2019	Frans	1,1	6,3	25
VWO	2004	Frans 12	0	6,2	27
VWO	2007	Frans 12	0,6	6,2	27
VWO	2017	Frans	0,5	6,2	27
VWO	2022	Frans	0,6	6,3	27
VWO	2011	Frans	0,5	6,2	28
VWO	2021	Frans	1,2	6,4	28
VWO	2003	Frans	0,3	6,1	29
VWO	2009	Frans 12	0,3	6,2	29
VWO	2000	Frans	0,3	6,2	30
VWO	2014	Frans	0,3	6,1	31
VWO	2002	Frans	0,1	6	36

Wederom gefilterd op het percentage onvoldoendes. Je mag dus zakken op Frans maar niet op Engels. Frans is geen kernvak...

2016 was een anomalie. Er is zelfs een brief over gekomen dat de normering helaas niet -1 kon zijn Die 7 was een fout (zie sheet 1)

Wat bij Engels de norm is, was bij Frans een helaas niet te repareren fout. Waarom? Omdat de cijfers van tevoren zijn vastgesteld door Den Haag. Voor Frans en Duits Havo is het gelukkig voortaan een 6.3, dat was de laatste jaren 5.9. Wel fijn dat het streven nu in ieder geval een voldoende is... Het was jarenlang heel naar om steeds te moeten zien dat een voldoende voor Frans of Duits nauwelijks haalbaar was.

In de Groene Amsterdammer van 19 oktober 2022 lees ik in deel 3 van De leescrisis in het onderwijs*) , een voortreffelijk artikel van Yra van Dijk en Marie-José Klaver:

“Het centraal eindexamen Nederlands is een berucht struikelblok, waar veel zwakke leerlingen op zakken en waardoor sterke leerlingen een cum laude mislopen. Op het VWO is het laagste gemiddelde cijfer van alle schoolvakken dat voor Nederlands: een 6,2. Ter vergelijking: voor wiskunde B is het gemiddelde een 7,0. Dat zegt iets over de motivatie van leerlingen voor het vak Nederlands, maar het zegt vooral iets over het examen zelf.”

Ik pak het citaat uit de Groene er bij en pas het aan:

Het centraal eindexamen Nederlands is een berucht struikelblok, waar veel zwakke leerlingen op zakken en waardoor sterke leerlingen hun cum laude ONTERECHT mislopen. Op het VWO is het laagste gemiddelde cijfer van alle schoolvakken dat voor Nederlands: een 6,2. DAT ZOU ONGEMANIPULEERD EEN 6.6 ZIJN.

($6.2 + 0.4 = 6.6$ het ongemanipuleerde cijfer voor Nederlands) N 0.6

Het ligt niet aan de motivatie van de leerlingen. Het ligt aan het systeem

Ter vergelijking: voor wiskunde B is het gemiddelde een 7,0. ONGEMANIPULEERD ZOU DAT EEN 6.2 ZIJN.

Dat zegt NIETS over de motivatie van leerlingen voor het vak Nederlands, maar het zegt vooral iets over het examen DE NORMERING zelf.

Kijk even naar 2017. Frans Havo kreeg een heel strafpunt want N-term 0 (1 is neutraal) Die 5.9 was ongecorrigeerd een 6.9...

38% kreeg een onvoldoende. Het examen was te goed gemaakt.

Afdeling	Examenjaar	Vak	Nterm	Gemiddeld_cijfer	Percentage_onvoldoende
HAVO	2000	Frans	0,4	6,8	15
HAVO	2001	Frans	0,5	5,9	31
HAVO	2002	Frans	1,2	5,8	37
HAVO	2003	Frans	1,4	6,2	27
HAVO	2004	Frans 12	0,6	6	34
HAVO	2005	Frans 12	1	6,1	28
HAVO	2006	Frans 12	0,4	5,8	37
HAVO	2007	Frans 12	1,1	6,1	29
HAVO	2008	Frans 12	0,1	6	30
HAVO	2009	Frans	0,8	6,1	31
HAVO	2010	Frans	0,4	6,2	25
HAVO	2011	Frans	0,3	5,8	43
HAVO	2012	Frans	1,3	6,1	28
HAVO	2013	Frans	1	6,1	31
HAVO	2014	Frans	0,4	6	31
HAVO	2015	Frans	0	6,3	22
HAVO	2016	Frans	0	6,3	19
HAVO	2017	Frans	0	5,9	38
HAVO	2018	Frans	0,4	5,9	34
HAVO	2019	Frans	1,2	5,8	40
HAVO	2021	Frans	1,4	6,3	24
HAVO	2022	Frans	1,1	6,3	23

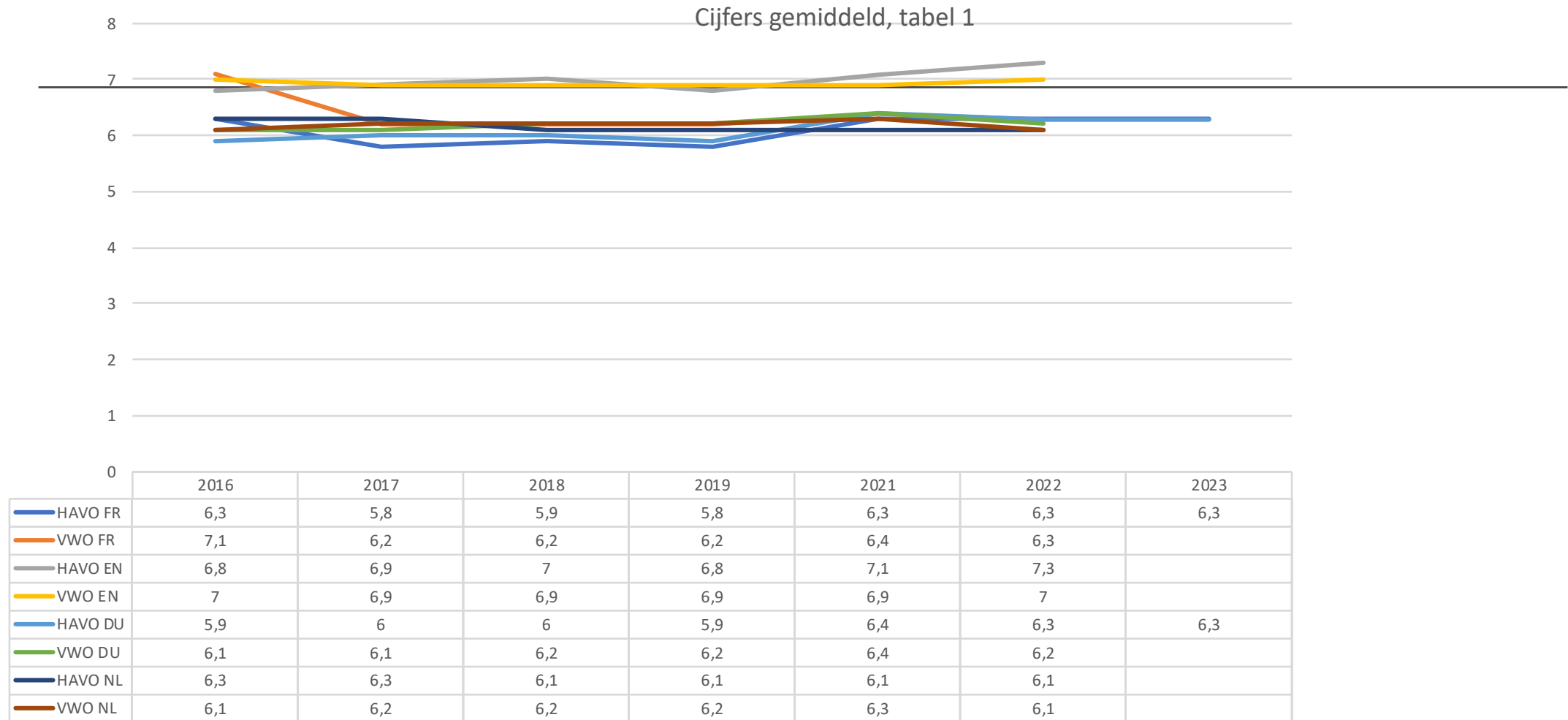
Bij Havo Engels dat jaar was er niks aan de hand.

Iedere leerling kreeg 0.4 bonuspunt zodat het gemiddelde op een 6.9 kwam, slecht 15% was onvoldoende.

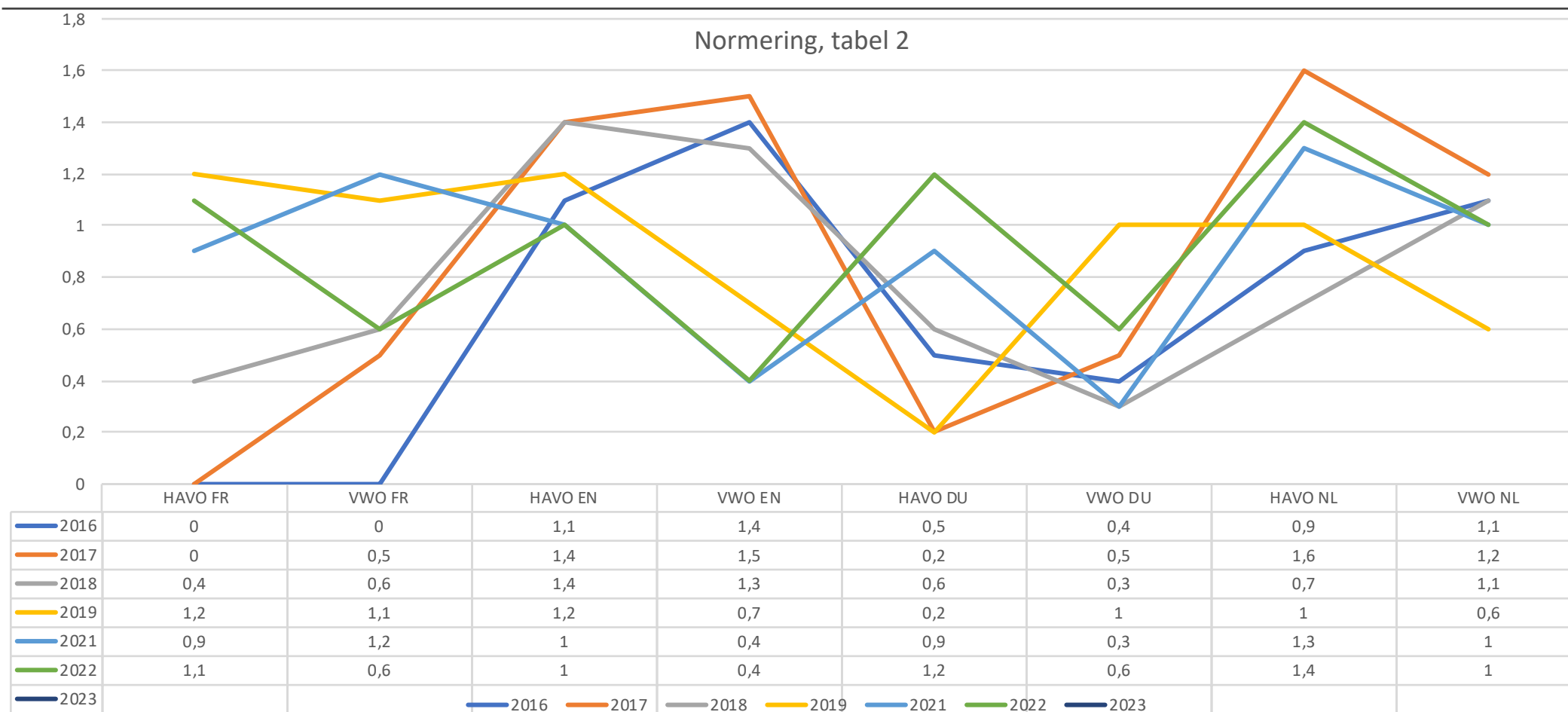
Frans Havo was anders eigenlijk ook een 6.9 geweest.

Afdeling	Examenjaar	Vak	Nterm	Gemiddeld_cijfer	Percentage_onvoldoende
HAVO	2000	Engels	0,7	6,6	18
HAVO	2001	Engels	0,8	6,5	20
HAVO	2002	Engels	0,8	6,3	26
HAVO	2003	Engels	0,6	6,1	27
HAVO	2004	Engels	0,6	6,4	22
HAVO	2005	Engels	0,5	6,1	29
HAVO	2006	Engels	0	6,2	25
HAVO	2007	Engels	0	6,2	21
HAVO	2008	Engels	1,3	6,5	24
HAVO	2009	Engels	0,4	6,2	28
HAVO	2010	Engels	0,4	6,1	33
HAVO	2011	Engels	0,7	6,1	31
HAVO	2012	Engels	0,4	6,2	24
HAVO	2013	Engels	0,3	6,7	16
HAVO	2014	Engels	0	6,9	12
HAVO	2015	Engels	0,5	6,8	13
HAVO	2016	Engels	1,1	6,8	16
HAVO	2017	Engels	1,4	6,9	15
HAVO	2018	Engels	1,4	6,9	15
HAVO	2019	Engels	1,2	6,8	16
HAVO	2021	Engels	1	7,1	11
HAVO	2022	Engels	1	7,3	7

Keurige rechte lijntjes. Engels àltijd een 7, de rest een 6



En zo krijg je dan die nette lijntjes



Frans en Duits trekken ze nu gelijk maar dat zijn ze niet...

(Bron Levende Talen's eigen Frank Claessen)

VMBO GL TL 2019 1 ^e tijdvak	Frans	Duits	Engels
Aantal teksten	16	12	12
met aantal woorden	3998	2909	4039
Aantal vragen	45	39	42
met aantal woorden	2006	1907	2183
Totaal te lezen woorden	6004	4816	6222
Frans heeft 24,66 % meer woorden dan Duits			

VMBO GL TL 2019 2 ^e tijdvak	Frans	Duits	Engels
Aantal teksten	17	12	14
met aantal woorden	4028	2935	4094
Aantal vragen	43	39	42
met aantal woorden	1835	1632	2111
Totaal te lezen woorden	5863	4567	6205
Frans heeft 28,37 % meer woorden dan Duits			

HAVO 2019 1 ^e tijdvak	Frans	Duits	Engels
Aantal teksten	13	12	13
met aantal woorden	4496	4501	4745
Aantal vragen	45	41	40
met aantal woorden	2285	1668	2013
Totaal te lezen woorden	6781	6169	6758
Frans heeft 9,92 % meer woorden dan Duits			

HAVO 2019 2 ^e tijdvak	Frans	Duits	Engels
Aantal teksten	13	11	13
met aantal woorden	4822	4581	4948
Aantal vragen	45	42	41
met aantal woorden	2344	1869	2250
Totaal te lezen woorden	7166	6450	7198
Frans heeft 11,10 % meer woorden dan Duits			

VWO 2019 1 ^e tijdvak	Frans	Duits	Engels
Aantal teksten	12	10	12
met aantal woorden	5075	4306	5213
Aantal vragen	41	42	35
met aantal woorden	2663	2192	2007
Totaal te lezen woorden	7738	6498	7220
Frans heeft 19,08 % meer woorden dan Duits			

VWO 2019 2 ^e tijdvak	Frans	Duits	Engels
Aantal teksten	12	10	11
met aantal woorden	5016	4432	4803
Aantal vragen	42	41	41
met aantal woorden	2419	1812	2355
Totaal te lezen woorden	7435	6244	7158
Frans heeft 19,07 % meer woorden dan Duits			

Er is een breder probleem en dat is de algemene kwaliteit van de eindexamens bij de talen. Al in 2017 schreef Michel Pijpers <http://www.deleraar.nl/blogpagina/het-begon-zo-onschuldig>

“De *Rit* geeft aan hoe goed een vraag het verschil aangeeft tussen ‘goede’ en ‘slechte’ kandidaten. Eenvoudig gezegd gaat het bij de *Rit-waarde* om de vraag of de kandidaten met een hoge score voor de toets als geheel ook de moeilijke vragen beter gemaakt hebben dan de kandidaten met een lage score. Als vooral zwakke leerlingen een ‘moeilijke’ opgave goed doen, mag je twijfelen over de betrouwbaarheid van de vraag (Goosen, 2015). En gelukkig biedt ook hier de site van Cito zelf uitkomst. In Toets- en itemanalyse met TIA (Goldebeld, 1992) noteert P. Goldebeld op pagina 49 het volgende schema:

tabel 3

Rit-waarde	Itembeoordeling
0.40 en hoger	zeer goed
0.30 – 0.39	goed
0.20 – 0.29	twijfelgeval
0.19 en lager	slecht

6.2 Normen voor Rit-waarden

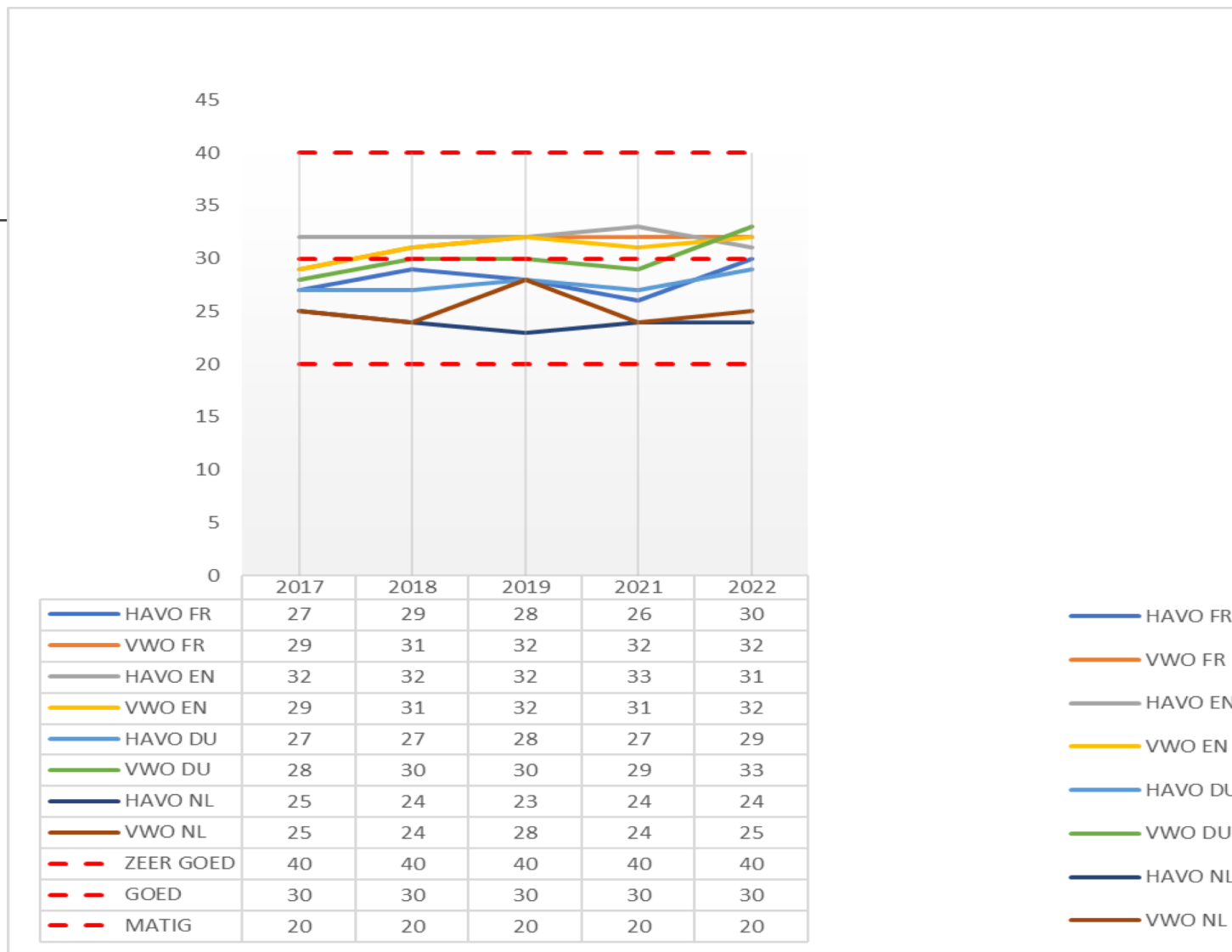
Absolute normen voor Rit-waarden zal men in de literatuur niet gauw aantreffen. Zoals bekend kan een Rit-waarde variëren tussen -1 en +1 waarbij we in de psychometrie streven naar een zo hoog mogelijke Rit, Ebel en Frisbie (1986) hebben wel een beoordeling opgesteld voor de Rit-waarden.

Tabel 2. Normen voor –waarden -waarde itembeoordeling

0.40 en hoger	zeer goed
0.30 - 0.39	goed
0.20 - 0.29	twijfelgeval
0.19 en lager	slecht

Omdat de grootte van de afhankelijk is van het aantal items in een toets moet men strikt genomen bovenstaande normen alleen hanteren bij die gecorrigeerd zijn voor toets lengte.

Gemiddelde RIT-waarden, tabel 4, het gros hangt tussen matig tot goed



Het discriminerend vermogen van een item geeft de mate aan waarin op grond van de scores op dat item hoog- en laagscorende leerlingen (op de hele toets) onderscheiden kunnen worden

De rit-waarde komt uit de statistiek en is vooral bedoeld ter ondersteuning bij het maken van toetsen. Op de website van Cito staat een tabel waarin betekenis wordt gegeven aan rit-waarden. Daar staat dat een rit-waarde van onder de 20 onvoldoende is. (...) In deze situatie lijkt het beter om de vragen met een rit-waarde van onder de 20 niet op te nemen in het examen.

https://www.cito.nl/-/media/files/nieuws/vakinhoudelijke-publicaties/de_waarde_van_de_rit-waarde.pdf?la=nl-NL

Is dit nieuws? Nee. Dit schreef Monique Marreveld op Didactief in 2014 al: **Cijfers op bestelling?**

“Nieuw onderzoek naar de examencijfers op de kernvakken in 2013 en 2014 doet vermoeden dat de normering aangepast is. Is dat om beleidsmakers te plezieren?

RTL Nieuws meldde half december nog – op basis van eigen onderzoek – dat middelbare scholieren in 2012 en 2013 beter scoorden op het centraal examen dan in 2011. Het College voor Toetsen en Examens had half juni de vlag al uitgestoken en gemeld dat vwo-ers voor de kernvakken Nederlands, Engels en wiskunde in 2014 zelfs iets beter scoorden dan de jaren hiervoor.

(...) <https://didactiefonline.nl/artikel/cijfers-op-bestelling>

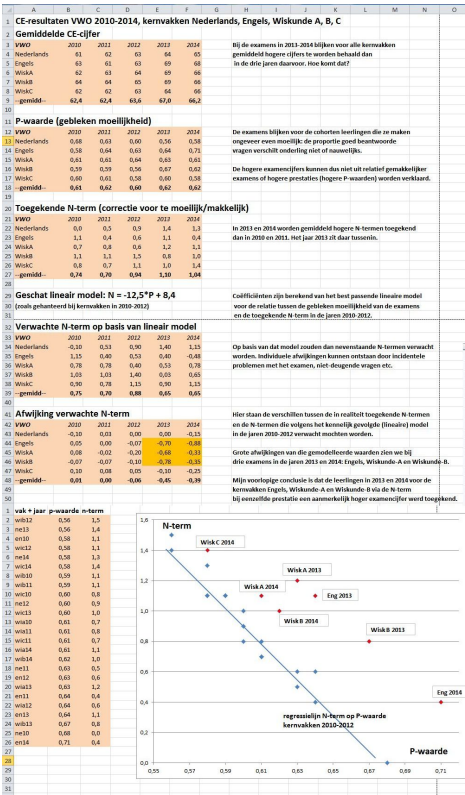
Gratis punten

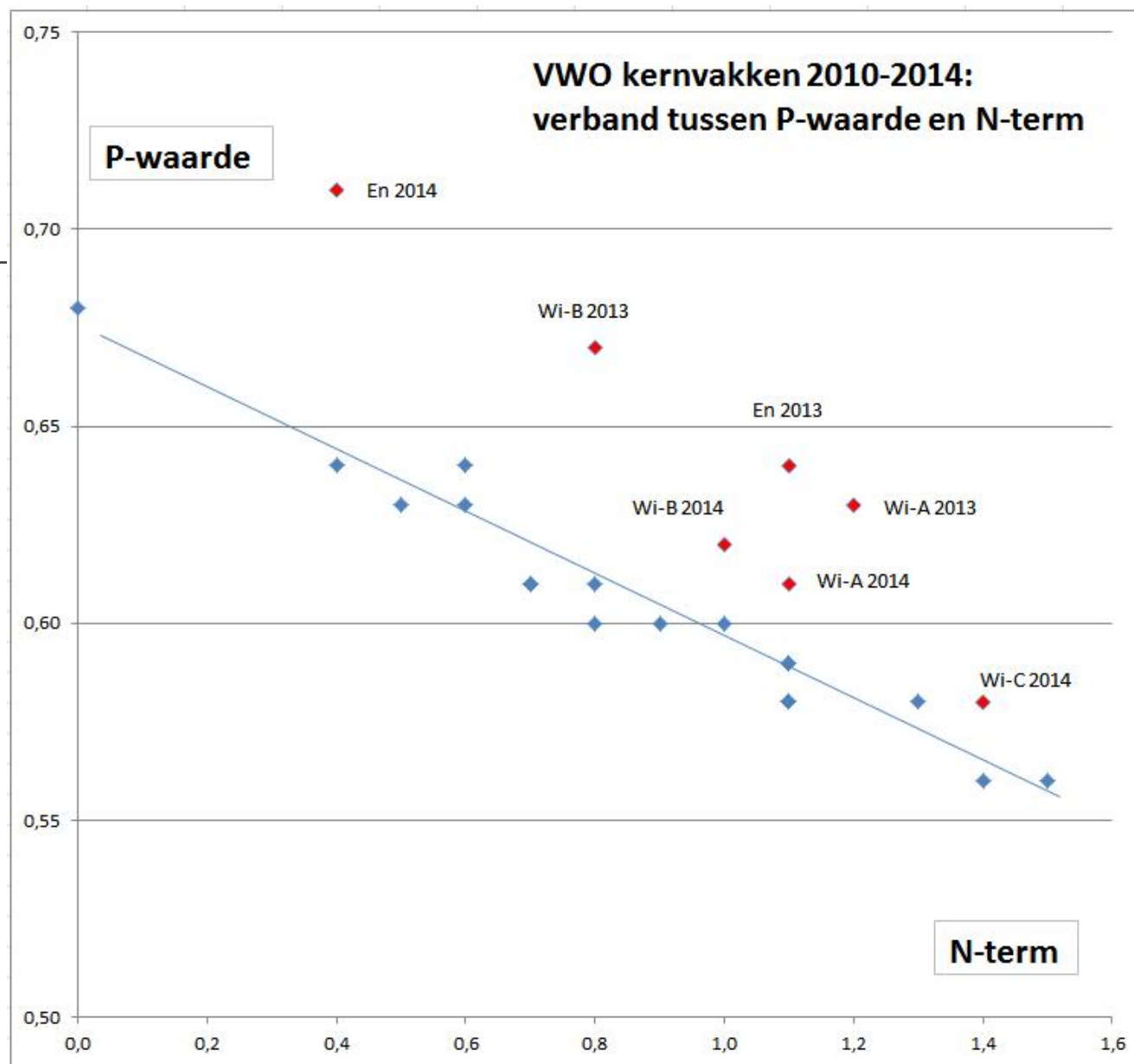
Onderzoek van Michel Couzijn, docent aan de Universiteit van Amsterdam en het Pieter Nieuwland College, doet vermoeden dat er nog een andere verklaring is. Hij zette de N-termen (de correctie die het Cito toepast op een gemaakt examen) eens op rijtje over de periode 2010-2014 voor de kernvakken Nederlands, Engels en wiskunde. **Als een examen te moeilijk of juist te makkelijk blijkt (en daardoor te veel leerlingen dreigen te zakken of ten onrechte slagen) past het Cito de normering aan.** Daardoor kunnen examenkandidaten gemiddeld enkele tienden tot een heel punt winnen of verliezen vergeleken met de standaardnormering. **Het middel is bedoeld om de examens van jaar op jaar ongeveer even moeilijk te maken.** (Idem)

Maar Couzijn kwam, tien jaar geleden al, tot een opmerkelijke ontdekking.

Hij ging voor de kernvakken in de jaren 2010, 2011 en 2012 na bij welke gebleken moeilijkheid er een bepaalde corrigerende N-term was toegekend. In technische termen vond hij een mooie lineaire relatie tussen de p-waarden en de N-termen. Vervolgens bleek dat drie examens van kernvakken zowel in 2013 als in 2014 sterk van die norm afweken. De examens voor Nederlands en wiskunde-C houden zich in 2013 en 2014 redelijk aan de eerder aangehouden norm. Maar in 2013 blijken examenkandidaten er opeens voor wiskunde A en B en voor Engels zo'n zeven-tiende punt bij te hebben gekregen, en in 2014 bij Engels acht-tiende en bij wiskunde A en B drie-tiende. (Idem)

Dit ligt geprint voor jullie





Minder gezakten

Waren die examens dan zoveel moeilijker? Nee, zegt Couzijn, over alle kernvakken gemeten was de moeilijkheidsgraad vergelijkbaar met de voorgaande jaren. Onderling varieerde die wat; zo bleek Nederlands iets moeilijker dan voorheen, maar Engels en wiskunde-B gemakkelijker. Waar het om gaat is dat er voor drie examens uitzonderlijk hoge N-termen zijn toegekend in relatie tot hun gebleken moeilijkheid, en de verklaring daarvoor is onbekend.

Neem als voorbeeld het examen vwo Engels in 2011 en 2013. In beide gevallen maakten kandidaten eenzelfde percentage vragen goed (64%). Maar in 2011 werd daarbij een N-term van 0,4 gehanteerd en in 2013 maar liefst 1,1. Die zeven tiende punt verschil leidde tot heel wat hogere examencijfers.

Ook voor het aantal zakkers en geslaagden lijkt deze aanpassing consequenties te hebben gehad. In 2011 zakte 26% van de kandidaten voor het examen Engels, in 2013 voor het examen met dezelfde p-waarde maar 12%. Dat scheelt zo'n 5500 leerlingen die misschien wel zijn geslaagd in plaats van gezakt voor zo'n cruciaal vak, dankzij de aangepaste N-term. (Idem)

Kernvakregeling; de wassen neus van de eeuw

“Zo heeft het CvTE een duidelijke stijging van de vaardigheid van kandidaten Engels havo en vwo kunnen waarnemen sinds de invoering van de kernvakkenregeling in 2013. Deze vaardigheidsstijging is de reden van het hogere gemiddeld cijfer voor deze centrale examens in het huidige examenjaar.”

<https://nieuwsbrief.examenblad.nl/wv/vk57gs4rlgyq/8o3ahkblja4vmg53>

N-terminen

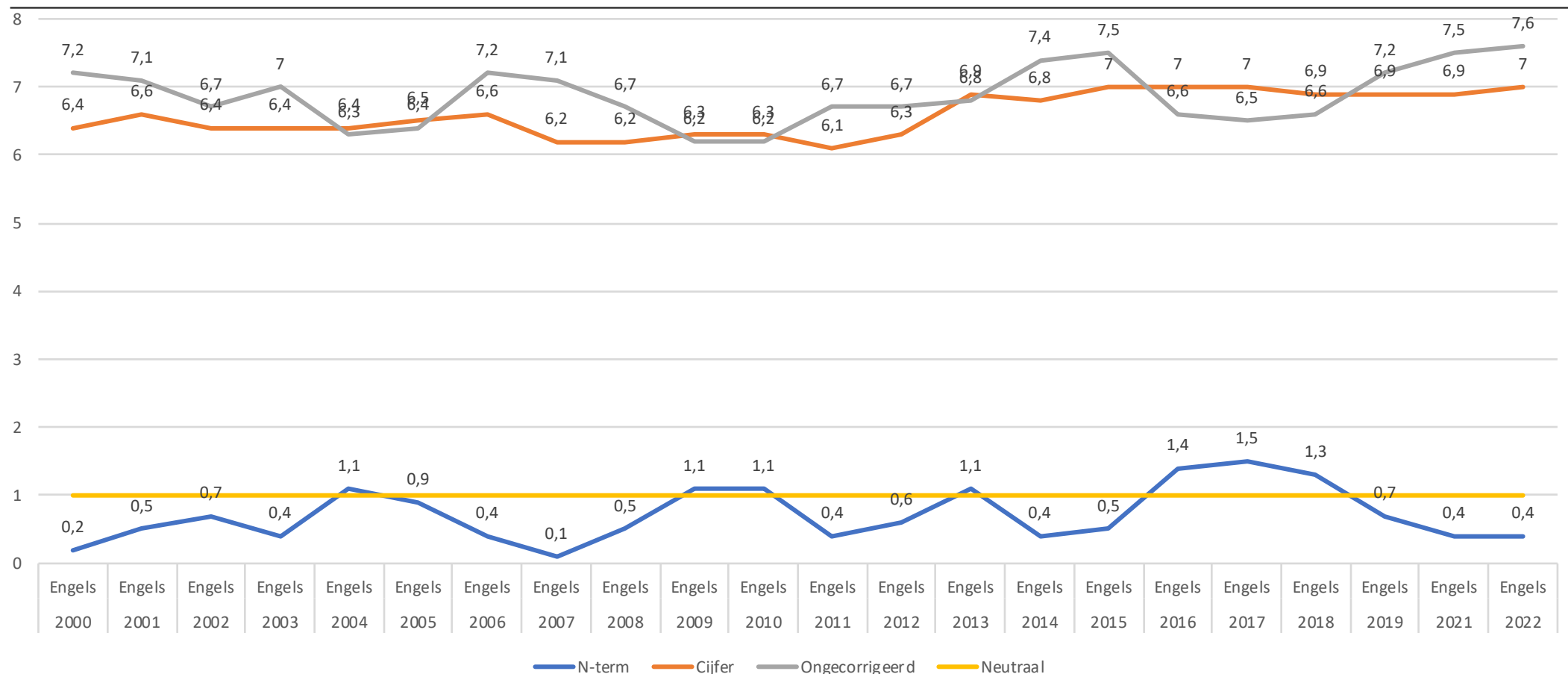
De hierna volgende cijfers komen uit Arjan van der Meij's

<https://ntermen.nl/ntermen.php>

Alle mogelijke fouten zijn van mij!

VWO Engels kernvak

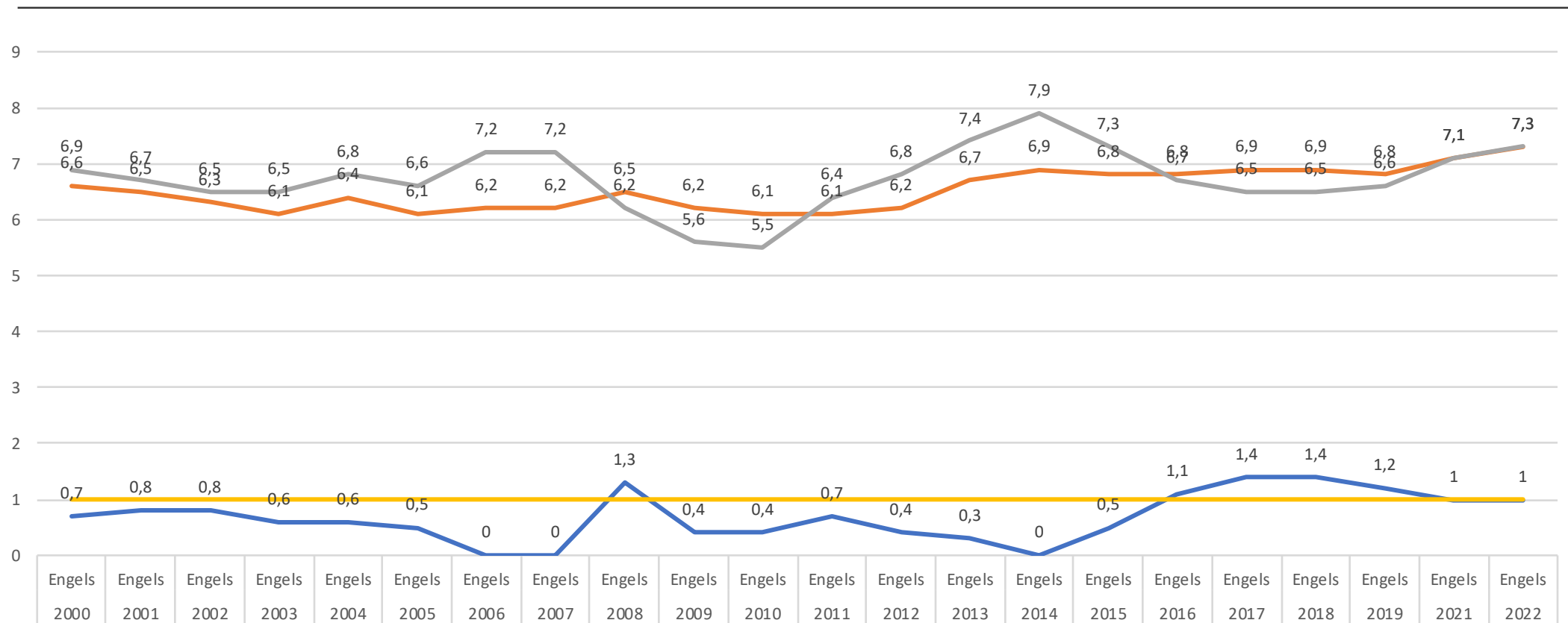
6,595455 6,872727



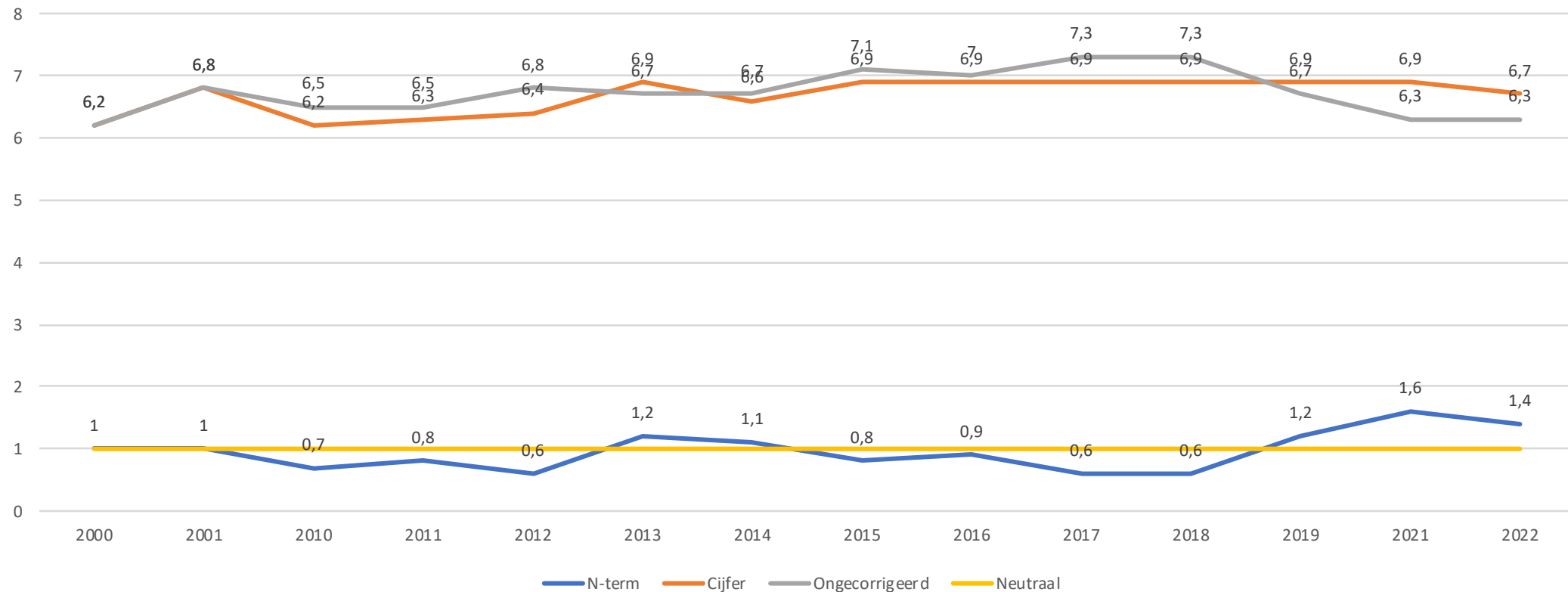
HAVO Engels kernvak

6,531818 6,736364

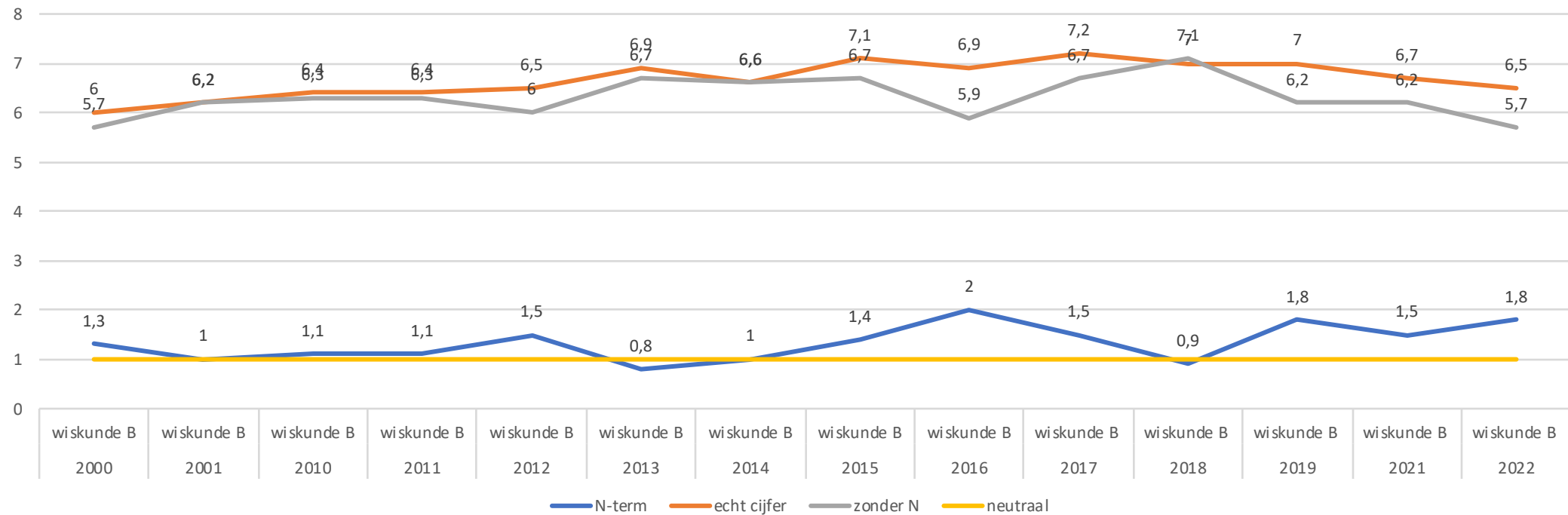
Scoort bijna altijd hoger dan wat ze krijgen, grote niveau verschillen, stijgt sinds 2013 naar 7



VWO Wiskunde A kernvak is vrij stabiel maar sinds 2013 een 7



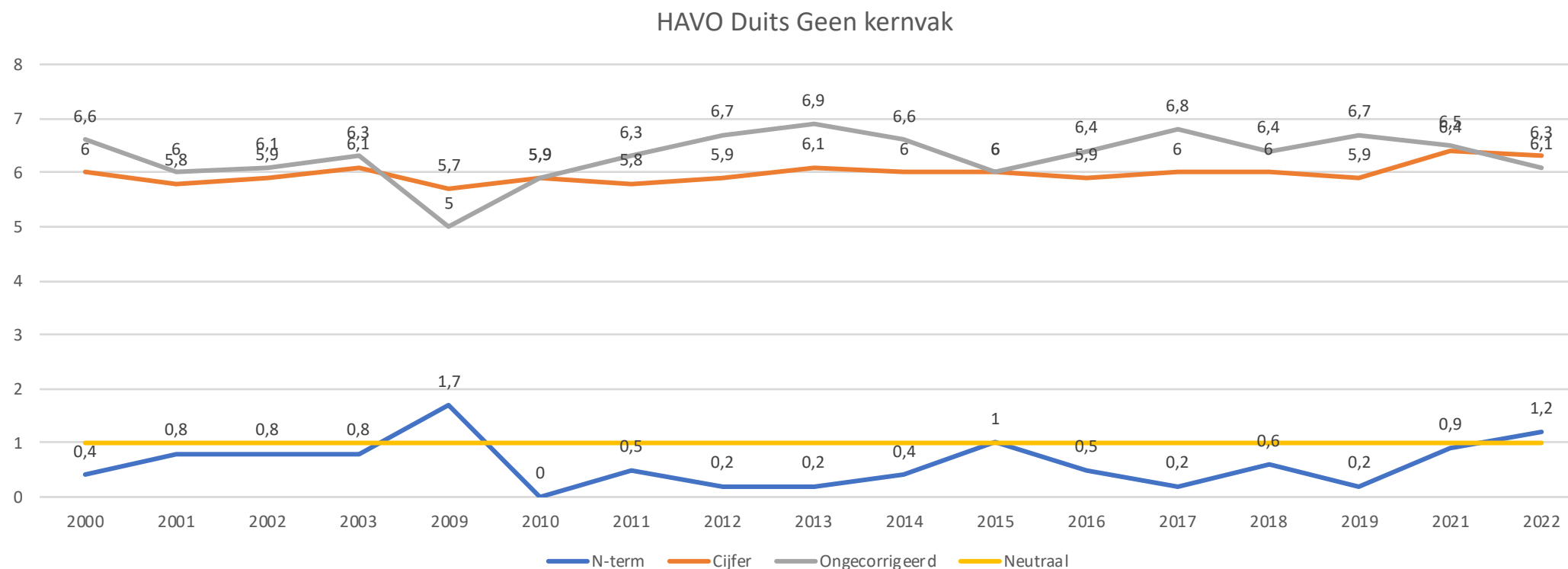
VWO Wiskunde B Kernvak is quasi systematisch opgehoogd naar een 7



Duits HAVO, geen kernvak

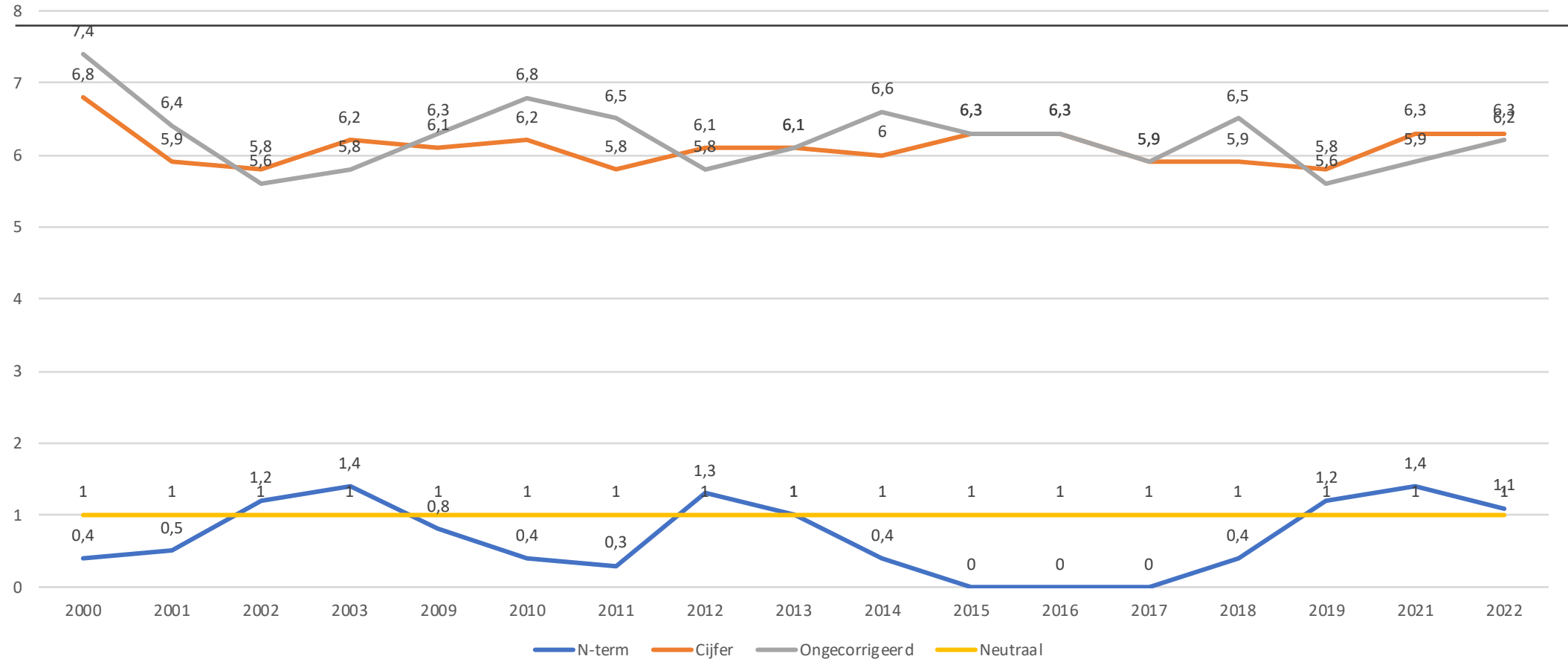
Stabiel op 6 ook al scoren de kandidaten hoger

5,982353 6,311765



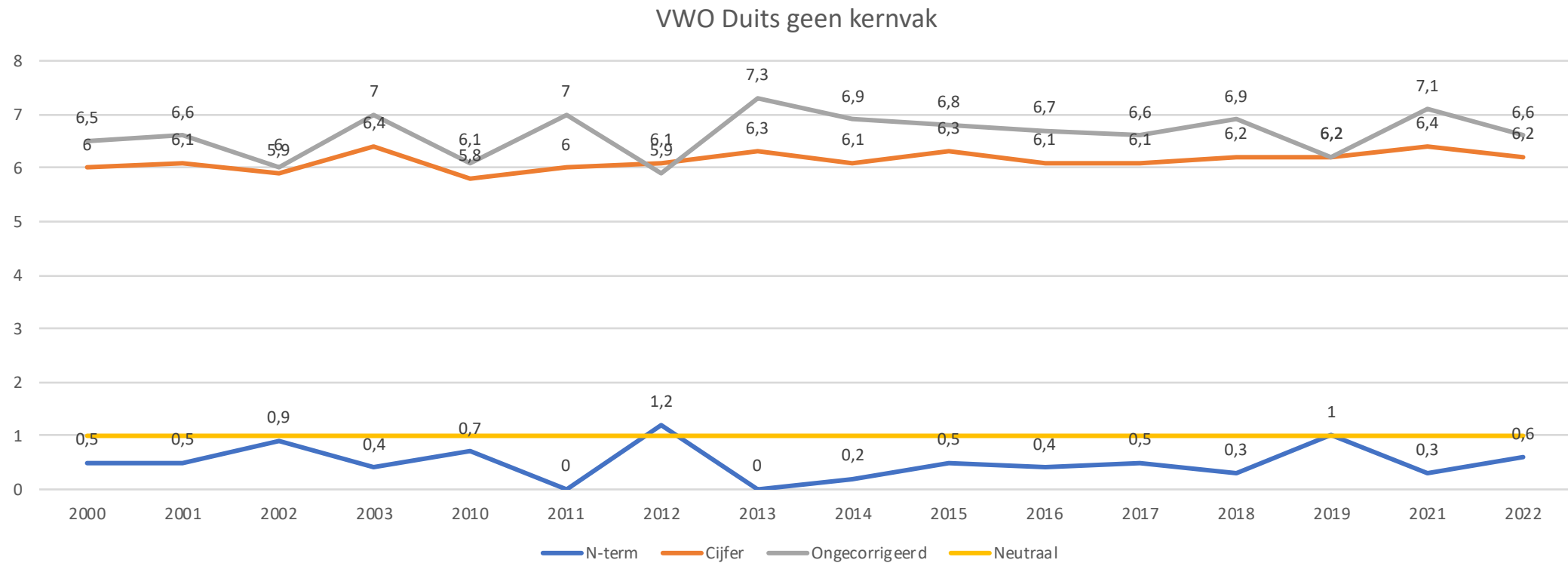
HAVO Frans geen kernvak

6,105882 6,235294



VWO Duits geen kernvak

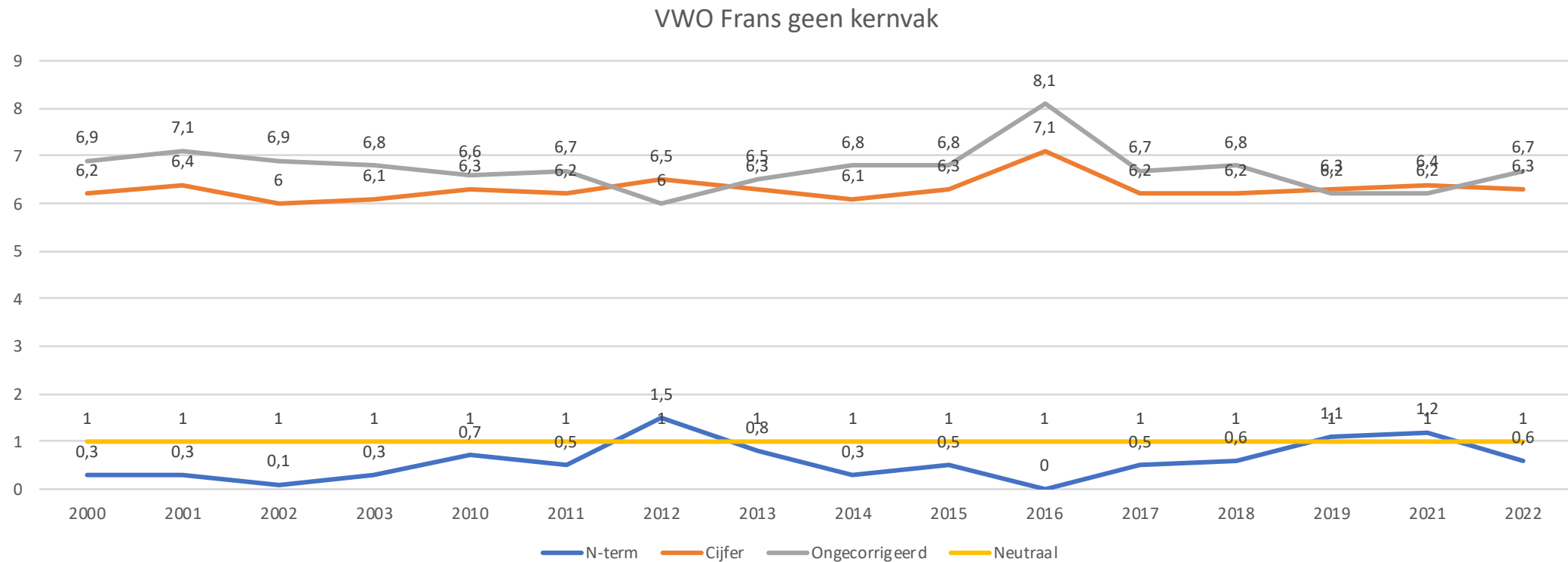
5,982353 6,311765



VWO Frans geen kernvak

scoort bijna altijd een 7 maar krijgt bijna altijd een 6

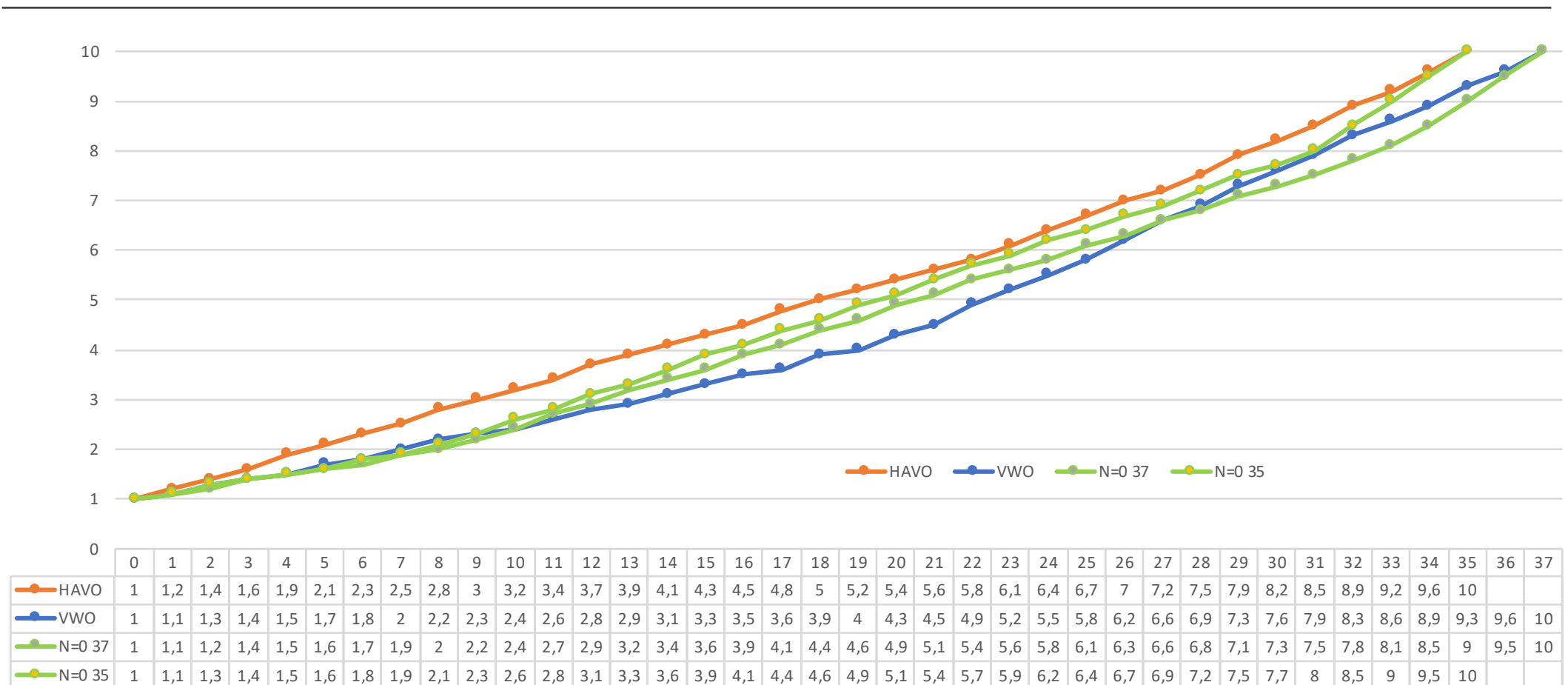
6,30625 6,7375



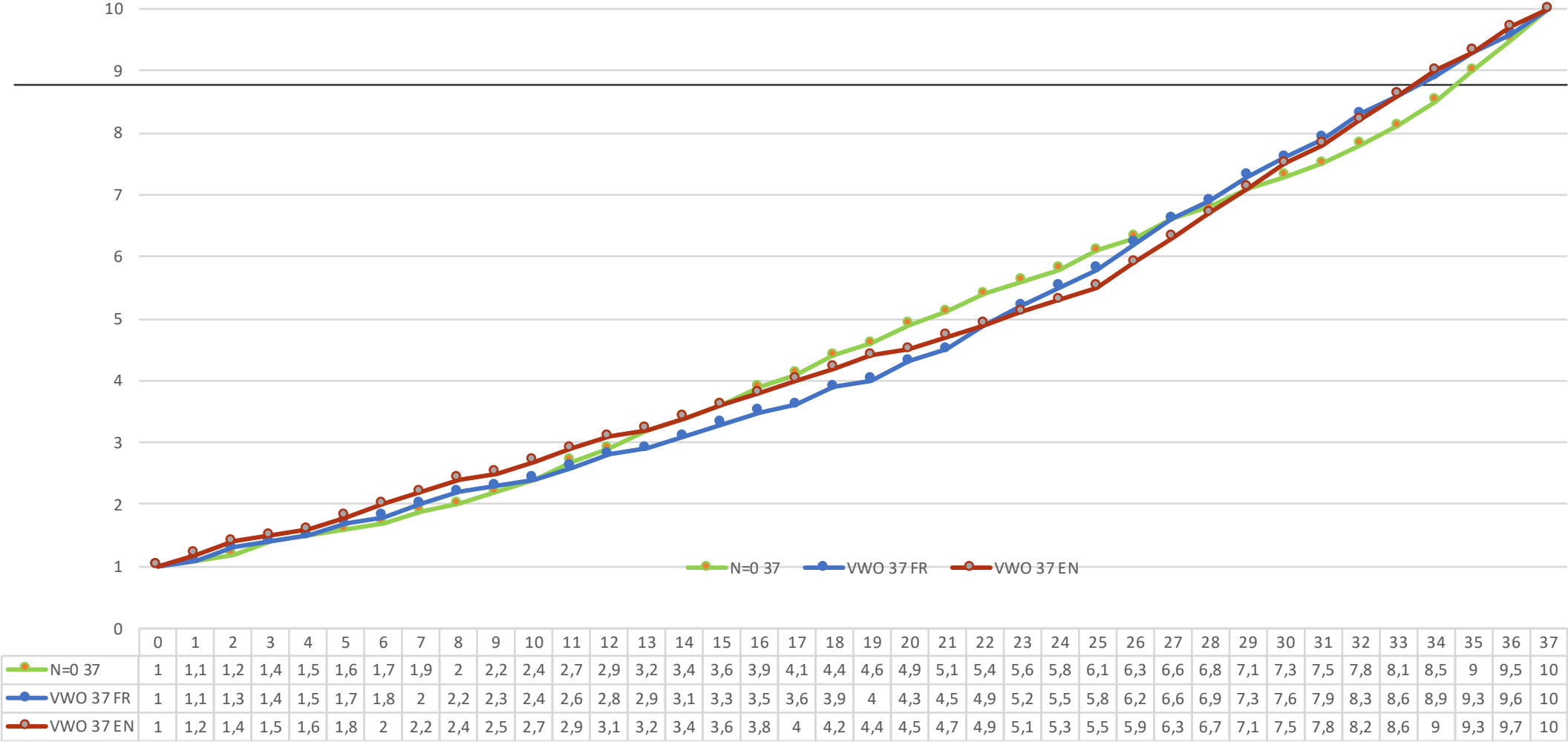
Conclusie

Noch het CvTE, noch Mombarg leveren een verklaring voor de door Couzijn geïdentificeerde afwijkende N-termen voor Engels, wiskunde-A en wiskunde-B. Met enige argwaan kan men daarom vermoeden dat deze N-termen zijn aangepast op basis van beleidswensen. Immers, sinds de aanscherping van de exameneisen volgde OCW het verloop van de examens met argusogen. Veel deskundigen verwachtten vooraf een slachting op havo en vwo, sinds verplicht is geworden dat alle CE-cijfers gemiddeld een 5,5 moeten zijn en kandidaten voor de kernvakken Nederlands, Engels en wiskunde nog maar één onvoldoende (en dan niet lager dan een 5) op de eindlijst mogen hebben. Met de aanpassing van de N-termen kon dat risico in ieder geval behoorlijk verkleind worden. (Idem)

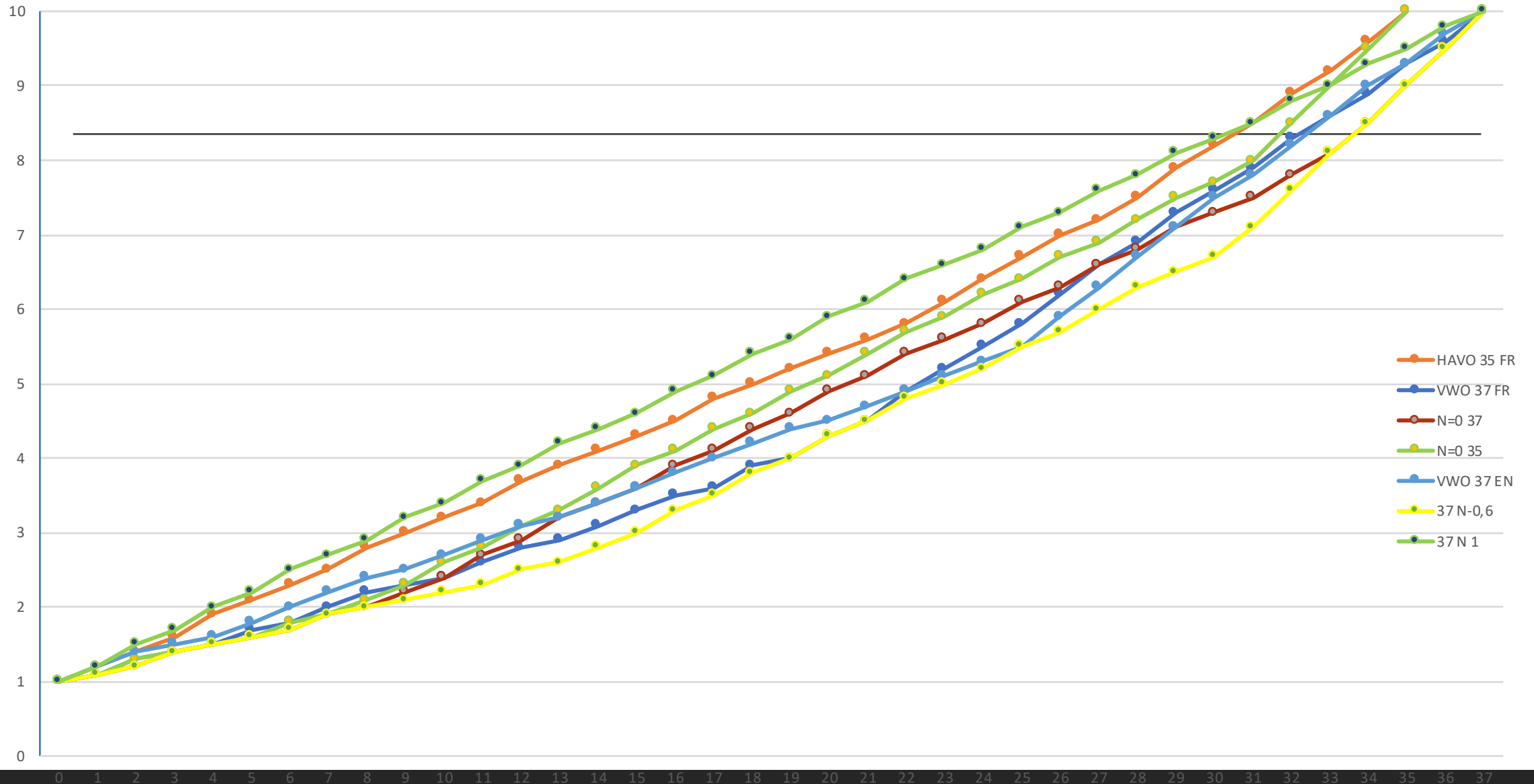
Bij de **luistertoetsen** Frans spelen ook rare dingen, Havo was makkelijk maar krijgt een soepeler N dan Vwo die tot 7 onder de zwaarst mogelijke N zit



VWO Engels is in het onvoldoende deel soepeler genormeerd dan VWO Frans, pas bij 7 krijgt het, net als Frans, hoger dan N0



De normale $N = 1$ is niet gebruikt maar **min nul komma zes** wel....
Dat is 1,6 punt aftrek! Rond het snijvlak voldoende onvoldoende.



Bij Frans krijgen de lage scores minder punten dan bij Engels

Reactie Cito:

“De methode heeft als effect dat leerlingen met lage scores niet zoveel hogere cijfers krijgen per extra scorepunt. **Dat is verdedigbaar omdat er ook met veel gokken altijd wel een paar vragen goed beantwoord kan worden** (maar nog steeds een lage totaalscore). De waardering neemt echter toe bij hogere scores. Daarom zijn de cijfersprongen bij een extra scorepunt hier ook groter.”

Bij Engels gokken ze blijkbaar niet want daar krijgen ze hogere cijfers voor de lage scores....

Dit jaar is er versoepeld bij Frans vwo, om op acceptabele landelijke cijfers uit te komen

We normeren nooit strenger, we versoepelen wanneer nodig

In principe leggen we dus bij elke toets de lat bij dezelfde vaardigheid als vorige jaren. Als we na de afnameweek zien dat dat prima cijfers oplevert, dan is duidelijk dat we die normering één op één over kunnen nemen. Als de toets goed gemaakt is, dan zal het landelijke gemiddelde hoger uitvallen, en/of is het percentage leerlingen dat een onvoldoende haalt lager. Maar: als we zien dat met de normering het landelijke gemiddelde opeens lager dan een 6,0 is, en/of dat meer dan 40% van de leerlingen een onvoldoende gaat scoren, dan grijpen we in. We versoepelen dan de normering: we leggen de lat iets lager, zodat er weer acceptabele, landelijke cijfers verschijnen. Dit zullen we dan wel vermelden bij het publiceren van de normering.

<https://www.cito.nl/-/media/files/voortgezet-onderwijs/kijk-en-luistertoetsen/adviesnormen/cito-toelichting-adviesnormen-klt-2023.pdf?la=nl-nl>

Je past een N-term toe op de normering van de Kijk- en luistertoetsen. Ik herhaal dat een N-term één methode van normeren is, niet dé methode. Bij de Kijk- en luistertoetsen gebruiken we geen N-term, waardoor het niet mogelijk is de formule die daar geldt los te laten op de scoretabel van welke *Kijk- en luistertoets* dan ook. Er zal dan altijd discrepantie tussen het scoreverloop zijn.

Schema 55

C1-3: De N-termen (minus 1) en cesuur(verschuivingen) duiden op het aantal tienden dat het cijfer wordt aangepast in beide perioden

Normhandhaving in 1990-1995 versus 2015-2021

	In 1990-1995...	In 2015-2021...
Norm-handhaving	Middels een cesuur(verschuiving) die minstens 0 en hoogstens 1,0 is	Middels een normeringsterm (N) die minstens 0 en hoogstens 2,0 is
Formule voor cijfer (1 t/m 10)¹	$\text{Cijfer} = 9 \cdot \frac{\text{Behaalde score (S)}}{\text{Lengte Scoreschaal (L)}} + 1 + \text{cesuur}$	$\text{Cijfer} = 9 \cdot \frac{\text{Behaalde score (S)}}{\text{Lengte Scoreschaal (L)}} + N$
Interpretatie cijferformule	De cesuur(verschuiving) bepaalt het aantal tienden dat het cijfer wordt opgehoogd (maximaal 1,0 erbij)	- De N-term is standaard 1 - De N-term kan groter dan 1 zijn (maximaal 2,0, maar dit was bijvoorbeeld bij CE Wiskunde B 2019 2^e tijdvak 2,2) Dit duidt op het aantal tienden dat het cijfer wordt opgehoogd (maximaal 1,0 erbij) - De N-term kan lager dan 1 zijn (minimaal 0,0). Dit duidt op het aantal tienden dat het cijfer wordt verlaagd (maximaal 1,0 eraf)
Gehanteerde N-term in dit onderzoek	- Een berekende N-term op basis van de officieel gepubliceerde cesuur. Hierbij is de berekende N-term gelijk aan 1 + cesuur - Het aantal tienden dat het cijfer wordt aangepast is gelijk aan deze berekende N-term minus 1²	- N-termen zoals officieel gepubliceerd - Het aantal tienden dat het cijfer wordt aangepast is gelijk aan de N-term minus 1²

1. Geldt in het algemeen, maar bij zeer hoge en lage cijfers vindt afvlakking plaats

2. Een N-term van 1,6 betekent dat 0,6 bij het cijfer wordt opgeteld en een N-term van 0,6 betekent dat 0,4 van het cijfer wordt afgetrokken

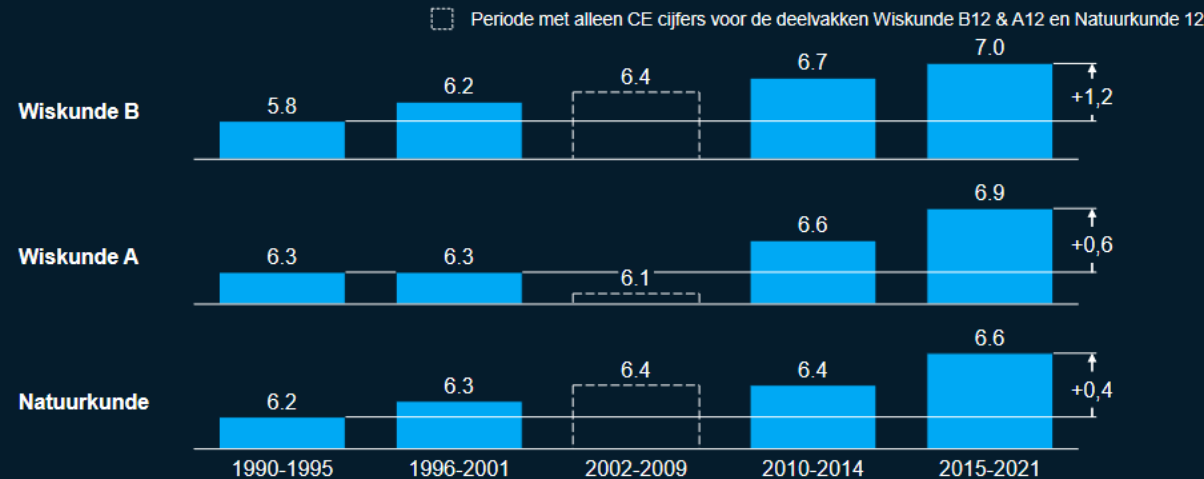
Of je het nu cesuur noemt of N-term, het komt ongeveer op hetzelfde neer schrijft McKinsey

De normhandhaving (nu middels de N-term destijds de cesuur) biedt hiervoor geen verklaring omdat deze nu ongeveer even hoog is als 30 jaar geleden. Dat wil zeggen het aantal tienden dat bij het cijfer wordt opgeteld is vergelijkbaar.

Het Wiskunde McKinseyrapport van Loek Zonnenberg and Paul Rutten p. 41

Schema 1

De afgelopen 30 jaar zijn de vwo Centraal Examen (CE) cijfers voor Wiskunde en Natuurkunde met 0,4-1,2 punt gestegen



Noot 1: Waargenomen stijgingen van 1990-1995 tot en met 2015-2021 zijn statistisch significant. Voor de jaren 2002 t/m 2009 zijn de CE cijfers voor de deelvakken Wiskunde B12 en A12 en Natuurkunde 12 gebruikt. Deze CE cijfers zijn minder representatief, doordat slechts een deel van de populatie die Wiskunde B (B1 of B12), Wiskunde A (A1 en A12) en Natuurkunde (1 en 12) Centraal Examen in dit deelvak deed. Dit waren met name leerlingen van het profiel N&T voor Wiskunde B12 en Natuurkunde 12 en van het profiel E&M voor Wiskunde A12. In 2020 was er geen CE en 2021 was het "Corona" examenjaar

Noot 2: Naast Natuurkunde zijn ook de CE cijfers voor de andere Natuurwetenschappenvakken afgelopen 20-30 jaar gestegen. Op basis van gevonden publiek beschikbare data: Scheikunde van 6,4 in 1990-1994 naar 6,6 in 2015-2021, Biologie van 6,1 in 1990-1994 naar 6,3 in 2015-2021 en Aardrijkskunde van 6,2 in 2000-2005 naar 6,3 in 2015-2021 (voor Aardrijkskunde geen publieke data gevonden voor de periode 1990-1995)

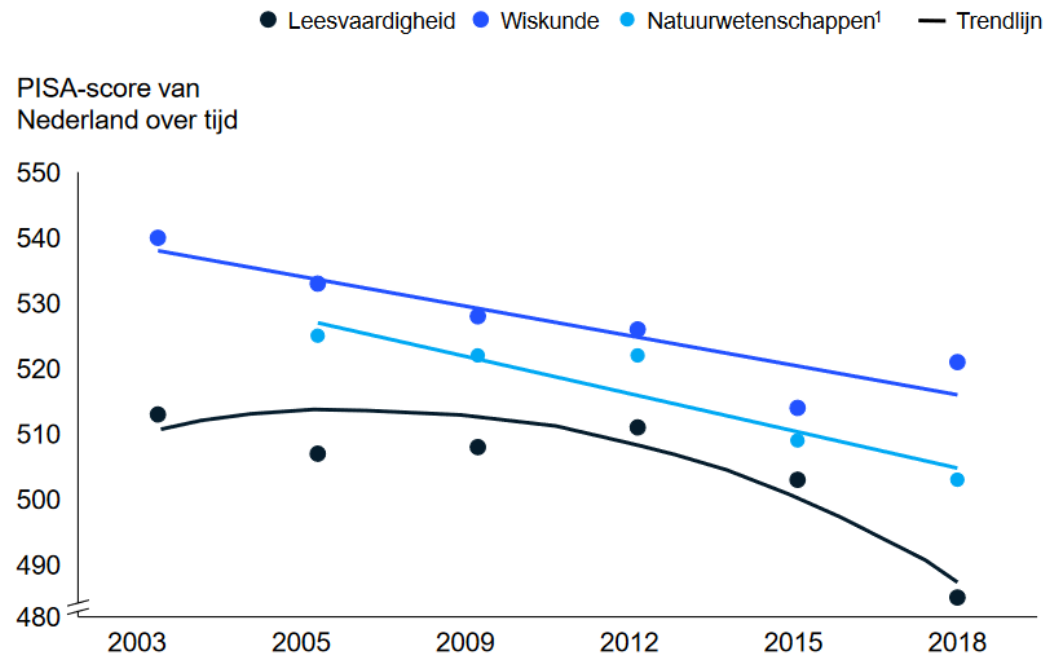
Bron: Examenverslagen voor de periode 2000-2021; Vakbladen Euclides (Wiskunde) en NVOX (Natuurkunde) voor de periode 1990-1999; Analyse door auteurs

Dit tref je aan over leesvaardigheid op p.44

Schema 7

PISA laat over afgelopen 15-20 jaar een significante niveaudaling van 15-jarigen zien over de gehele linie van het funderend onderwijs

Verandering in scores tussen 2003 en 2018, per PISA-onderdeel



1. Combinatie van Natuurkunde, Scheikunde, Biologie en Aardrijkskunde ("Science")

Bron: PISA; McKinsey onderzoek naar de doelmatigheid en toereikendheid van het funderend onderwijs (primair en voortgezet): "Een verstevigd fundament voor iedereen, April 2020" (Paul Rutten et al); Aanvullende analyse door de auteurs

Het rapport Steur zegt (McKinsey p. 50)

- Welke rol speelt de N-term en hoe robuust is deze procedure?

De commissie constateert dat de N-termprocedure een belangrijke rol speelt bij het vaststellen van de examencijfers op het CE over de jaren heen, met beperkte effecten op de hoogte van die cijfers. De commissie is ervan overtuigd dat de N-termprocedure nodig is om op een verantwoorde wijze te examineren en om examencijfers over de jaren heen met elkaar te kunnen vergelijken. Voor een leerling is het van groot belang dat het niet uitmaakt in welk jaar hij of zij het examen maakt. De N-termprocedure is een essentieel correctiemechanisme dat corrigeert voor de moeilijkheid van het examen waardoor de cijfers van jaar tot jaar vergelijkbaar zijn. De daarbij gekozen methodiek (per vak, per niveau) lijkt volgens de commissie binnen de budgettaire ruimte adequaat.⁴

Daarnaast constateert de commissie dat de N-termprocedure over het algemeen robuust is. Dat betekent dat het goed lukt om de examens op een vergelijkbaar moeilijkheidsniveau te houden zodat CE-cijfers over jaren heen goed te vergelijken zijn en dat examenprestaties niet kunstmatig opgehoogd zijn door middel van de N-term. Uit het Cito-onderzoek blijkt dat het effect van de bijstelling van de technische N-term op de uitslag van het examen verwaarloosbaar klein is.

2: De Commissie Steur samengesteld door het NRO concludeerde dat de vergelijking tussen PISA en het CE mank gaat en er geen sprake is van teruglopende onderwijsresultaten (McKinsey p.51)

De commissie is ervan overtuigd dat voor wat betreft de normering het CvTE en het Cito de juiste middelen in handen hebben om de kwaliteit van het CE over de jaren heen te waarborgen en te monitoren, in ieder geval voor de vakken waarbij de meer geavanceerde normhandhavingsmethoden mogelijk zijn. Deze meer geavanceerde methoden worden toegepast bij de meeste examens voor de grootste groepen leerlingen. Indien de wens leeft om de examens van meer vakken op het hoogst mogelijke niveau te normeren met de meer geavanceerde methoden, dan is daarvoor meer budget nodig. Dit vraagt om een politieke beoordeling en afweging.

Bij bovenstaande dient aangetekend te worden dat de complexiteit van de normeringsprocedure maakt dat deze snel vragen kan oproepen. Hoewel dit niet helemaal te vermijden is, is de commissie wel van mening dat een grotere transparantie over de beslissingen in de N-termprocedure (o.a. de vaststelling, de mogelijke bijstellingen en de impact daarvan) bij kan dragen aan het vertrouwen daarin. In het bijgaande Cito-rapport is daar in samenwerking met het CvTE een begin mee gemaakt.

Ik vrees dus dat de commissie Steur en het McKinsey rapport

zich niet hebben verdiept in de Kernvakregeling
die maakt dat die vakken gemiddeld hoger scoren
dankzij

- te makkelijke eindexamens
- gecombineerd met een ultra soepele N-term

Schaf kernvakregeling af

Loek Zonnenberg en Paul Rutten schrijven over de achteruitgang van wis- en natuurkunde (*ND* 1 maart). De centrale examens in die vakken zijn volgens hen aanzienlijk makkelijker geworden. Het valt me op hoe in hun onderzoek de ‘kernvakregeling’ buiten beschouwing is gebleven. Het woord komt welgeteld één keer voor in het rapport, maar kan heel veel verklaren, want leerlingen mogen niet collectief zakken op de kernvakken. Dat zou een schande zijn. Kernvakken zijn Nederlands, wiskunde en Engels. Zakken moet nu dus op de andere vakken. Kijk naar kernvak Engels. Het examen is makkelijker en

korter dan Frans, toch krijgt Frans een zwaardere normering en komt gemiddeld altijd uit rond de zes. Frans leer je woordje voor woordje, Engels hoor je de hele dag gratis en voor niks ... Het examen Engels zou niet beloond moeten worden met een eitje van een examen met een gemiddelde 7 cadeau als normering. Alleen Nederlands schijnt niet versimpeld te zijn, maar daar spelen weer hele andere problemen. Schaf de kernvakregeling af, dan kunnen Wiskunde en Engels weer normaal worden beoordeeld. <