

Normering is geen sinecure

Jacqueline Wooning en Harco Weemink

{intro}

In normale, rustige, stabiele tijden is normering al geen sinecure. Maar komend uit de roerige coronatijd is het dat zeker niet. Dat de normering nu meer vragen oproept, is dan ook niet vreemd. Met dit artikel willen we wat van die vragen wegnemen. Omdat de normeringsvragen die zijn gesteld betrekking hadden op de havo- en vwo-examens, richt dit stuk zich op deze examens.

{/intro}

Recente geschiedenis

Van 2015 tot en met 2018 was er sprake van een stabiele tijd in het voortgezet onderwijs. Er waren over het algemeen geen grote veranderingen en de vaardigheid van leerlingen was redelijk constant over de jaren ^[1]. In 2019 hebben we een nieuwe normeringssystematiek ingevoerd waarin die stabiliteit het uitgangspunt was. In 2020 kregen we te maken met de coronapandemie die veel invloed had op het onderwijs. In 2021 tot en met 2023 hebben we om die reden een aangepaste normering moeten toepassen ^[2]. Deze systematiek had tot gevolg dat we gemiddeld over alle vakken onbedoeld iets soepeler zijn gaan normeren. Dat werd zichtbaar in iets hogere N-termen. Omdat de vaardigheid van leerlingen gemiddeld over alle vakken ook gedaald was ^[3], zag je dit niet terug in de gemiddelde cijfers. Die waren veelal lager dan voor corona. In tabel 1 zie je de getallen voor wiskunde. Je ziet dat de gemiddelde cijfers van 2021 tot en met 2023 behoorlijk fluctueren. Door de aangepaste normering kunnen we niet met zekerheid zeggen of de vaardigheid bij een vak in dezelfde mate fluctueerde.

	wis A vwo		wis B vwo		wis C vwo		wis A havo		wis B havo	
	N-term	gem cijfer	N-term	gem cijfer	N-term	gem cijfer	N-term	gem cijfer	N-term	gem cijfer
2018	0,6	6,9	0,9	7,0	0,6	6,3	1,2	6,4	1,2	6,7
2019	1,2	6,9	1,8	7,0	0,9	6,3	1,0	6,4	1,5	6,7
2020	-	-	-	-	-	-	-	-	-	-
2021 ^[4]	1,6	6,9	1,5	6,7	1,2	6,1	1,2	6,6	1,8	6,2
2022 ^[4]	1,4	6,7	1,8	6,5	0,9	6,3	1,2	6,5	1,7	6,0
2023	1,4	7,0	1,9	6,7	1,2	5,9	1,3	6,5	1,7	6,5

tabel 1 N-termen en gemiddelde cijfers ce's tijdvak 1 – van stabiliteit in 2018 en 2019 naar schommelingen in 2021 tot en met 2023

Met de coronapandemie achter ons, vroeg 2024 om een andere normering. Een normeringssystematiek zoals in 2019 was niet mogelijk omdat er absoluut geen sprake is van stabiliteit zoals je kunt zien in tabel 1. Daarom is een nieuwe systematiek ontwikkeld die we zowel in stabiele als in minder stabiele tijden kunnen toepassen. Kort gezegd komt de nieuwe systematiek erop neer dat we op verschillende manieren (hierna bronnen genoemd) de moeilijkheidsgraad van het examen schatten en de bijbehorende N-term bepalen. Het gewogen gemiddelde van die verschillende N-termen, waarbij de weging van elke N-term afhangt van de standaarddeviatie van die schatting, wordt de *technische N-term* genoemd ^[5].

Het was ook niet mogelijk de lat zonder meer op dezelfde hoogte te leggen als voor corona ^[6]. Daarom hebben we besloten de hoogte van de lat van 2023 als uitgangspunt te nemen voor de normering van 2024. Een peiling onder docenten over die hoogte van de lat gaf ons informatie over de mate waarin die als passend werd ervaren. Deze informatie hebben we meegenomen bij de normering.

In het vervolg van dit artikel lichten we de normering van 2024 toe aan de hand van enkele voorbeelden. De voorbeelden zijn gekozen omdat we over de normering van deze examens vragen hebben gehad via de Examenlijn.

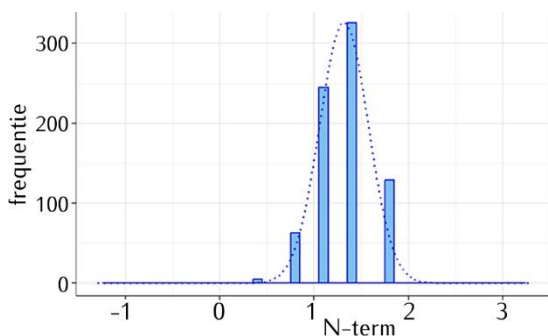
Normering wiskunde A vwo tijdvak 1

afkorting bron	bron van informatie	N-term op basis van bron	standaarddeviatie bron
D	docentenoordelen	1,32	0,26
E	schatting o.b.v. examendata	2,21 (gedempt 2,11)	0,07
H	historische N-termen	1,10	0,74
V	oordelen vaststellingscommissie	1,43	0,60

tabel 2: N-term per bron met bijbehorende weging vwo wiskunde A tijdvak 1

Opvallend is dat bron E een beduidend hogere N-term aangeeft dan de andere bronnen. Voor bron E is gebruikge- maakt van de afnamegegevens en de resultaten van een pretest ^[7]. De pretest wees uit dat het examen van 2024 fors moeilijker is (1,1 cijferpunt) dan het examen van 2023. Daarnaast bleek de vaardigheid van de leerlingen flink hoger dan in 2023 (0,5 cijferpunt). In het verleden werden pretestresultaten gedempt omdat heel grote vaardigheidsverande- ringen van jaar op jaar eigenlijk niet waarschijnlijk zijn. Dit jaar ligt dat mogelijk iets anders omdat we verwachtten dat de vaardigheid toe zou nemen na de coronajaren. Deze vaardigheidstoename viel erg op ten opzichte van alle an- dere vakken en ook leverde de pretest een uitkomst op die sterk afweek van de andere bronnen bij dit examen. Daarom hebben we dit resultaat gedempt gevolgd. Bron E levert hierdoor een schatting van de N-term van 2,1.

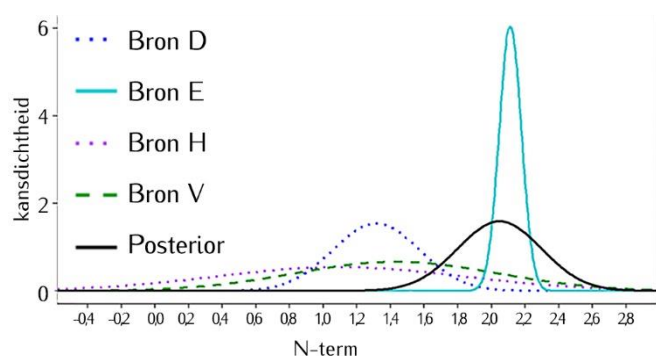
Ook uit het docentenoordeel (768 docenten – zie figuur 1) en het oordeel van de vaststellingscommissie over de moei- lijkeitsgraad komt naar voren dat het examen moeilijker is dan het examen van 2023. Alleen niet zoveel moeilijker zo- als volgt uit bron E. Hierbij moet worden opgemerkt dat docenten niet kunnen aangeven hoeveel moeilijker zij het exa- men vinden. Zij kunnen alleen aangeven of zij het examen ‘iets moeilijker’ of ‘veel moeilijker’ vinden. Dit wordt door ons gekwantificeerd als een 0,7 cijferpunt hogere N-term in geval van ‘veel moeilijker’ en een 0,3 cijferpunt hogere N- term in geval van ‘iets moeilijker’. Wanneer een examen veel moeilijker is, zoals hier het geval lijkt te zijn, zou dit kun- nen leiden tot een onderschatting van de moeilijkheidsgraad van het examen.



figuur 1 docentoordeelen over de moeilijkheid vwo wiskunde A tijdvak 1

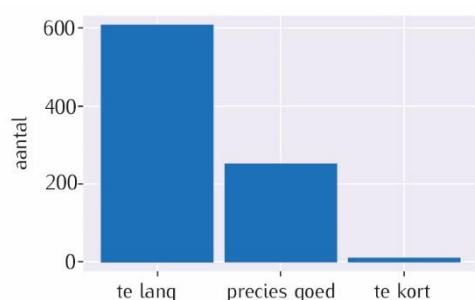
De historische N-term is, als gevolg van de korte historie na de coronapandemie, alleen gebaseerd op het examen in 2023. De standaarddeviatie van deze bron is dan ook het grootst van allemaal aangezien op basis van één N-term uit het verleden moeilijk te voorspellen valt hoe hoog toekomstige N-termen zullen zijn.

Van de verschillende schattingen van N-termen wordt een gewogen gemiddelde bepaald (de posterior) waarbij de we- ging omgekeerd evenredig is met de standaarddeviatie in het kwadraat. Gezien de kleine standaarddeviatie van bron E, gebaseerd op de afnamegegevens van 16.497 leerlingen, in combinatie met de pretest, weegt deze bron het zwaarst. De technische N-term komt hierdoor uit op 2,0.



figuur 2 Weergave N-termen bij gebruikte bronnen vwo wiskunde A tijdvak 1

Bij de inhoudelijke analyse kwam aan het licht dat dit examen mogelijk te lang is geweest. De docenten beoordeelden dit examen in meerderheid ook als 'te lang'.



figuur 3 Docentenoordeel examenlengte vwo wiskunde A tijdvak 1

Als een examen te lang is geweest, leidt dat tot een toename aan overgeslagen vragen (in Wolf gemarkeerd met een N). Dit leidt tot lagere p-waarden per vraag en dus lijkt het examen, op basis van deze getallen, moeilijker. Het effect dat een te lang examen heeft op de prestatie van een leerling wordt dus al deels automatisch gecompenseerd door een hogere technische N-term. De verwachting is echter dat een leerling die net een voldoende kan halen meer last heeft van de tijdnoed en daarvoor niet volledig wordt gecompenseerd via de technische N-term. Daarom voeren we een analyse uit op de overgeslagen vragen. Dit levert een zogenaamde *tijdnoodcorrectiefactor* op. Deze tijdnoodcorrectiefactor geeft aan hoeveel last van tijdnoed een leerling die net een voldoende kan halen bovenop de compensatie via de technische N-term nog gehad zal hebben. De tijdnoodcorrectiefactor voor dit examen was 0,09.

Dit leidt tot een definitieve N-term van $2,0 + 0,09 = 2,1$. De leerlingen in de steekproef halen daarbij een gemiddeld cijfer van 7,1.

Normering andere wiskunde-examens

Voor de andere wiskunde-examens geldt eenzelfde werkwijze. Alleen voor wiskunde C wordt geen pretest uitgevoerd vanwege het kleine aantal leerlingen bij dit examen. Voor vakken die geen pretest of andere vorm van equivalering hebben, wordt bepaald wat de vaardigheidsontwikkeling is op basis van de vaardigheidsontwikkeling bij vakken uit dezelfde schoolsoort die wel een vorm van equivalering hebben. Dit doen we onder de aanname dat als leerlingen bij die vakken een vaardigheidsontwikkeling laten zien, dit bij niet-geëquivalenteerde vakken waarschijnlijk ook het geval zal zijn. Deze gemiddelde vaardigheidsontwikkeling wordt meegenomen in bron E.

Als we tabel 1 nu uitbreiden met de resultaten van 2024 dan ziet dat er als volgt uit:

	wis A vwo		wis B vwo		wis C vwo		wis A havo		wis B havo	
2023	1,4	7,0	1,9	6,7	1,2	5,9	1,3	6,5	1,7	6,5
2024	2,1	7,1	2,1	6,7	0,8	6,0	1,1	6,3	1,4	6,7

tabel 3: N-termen en gemiddelde cijfers ce's tijdvak 1 2023 en 2024 – uitbreiding tabel 1

Normering wiskunde C tijdvak 1

Het examen wiskunde C, zie tabel 4, kende een referentie-N-term van 1,2 en een referentiegemiddeld-cijfer van 5,9. De gemiddelde vaardigheidsontwikkeling voor vakken uit dezelfde schoolsoort was +0,13. In combinatie met de afnamegegevens voor wiskunde C levert dat voor bron E een N-termschatting van 0,8. De docenten beoordelen dit examen als iets moeilijker dan het examen van 2023. Aangezien het aantal docentenoordelen bij een wat kleiner vak als wiskunde C niet zo groot is, is de standaarddeviatie wat groter waardoor dit oordeel minder zwaar meeweegt. Dit resulteert uiteindelijk in een technische N-term van 0,8. Bij de inhoudelijke analyse kwamen verder geen problemen aan het licht die aanleiding gaven de N-term nog op te hogen, waardoor de N-term is vastgesteld op 0,8 wat leidt tot een gemiddeld cijfer van 6,0 op basis van de steekproef.

afkorting bron	bron van informatie	N-term op basis van bron	standaarddeviatie bron
D	docentoorndelen	1,33	0,29
E	schatting o.b.v. examendata	0,83	0,04
H	historische N-termen	1,20	0,74
V	oordelen vaststellingscommissie	0,87	0,60

tabel 4: N-term per bron met bijbehorende weging vwo wiskunde C tijdvak 1

Het examen wiskunde C kent een overlap met het examen wiskunde A. In het verleden werd hiervan gebruikgemaakt bij de normering. Met de invoering van de nieuwe examenprogramma's is hiermee gestopt omdat het de bedoeling was dat wiskunde C veel meer een eigen karakter kreeg. We hebben deze gegevens na de normering wel bekeken. Daarbij zien we dat de vragen in de overlap tussen beide examens door leerlingen met wiskunde C aanzienlijk slechter zijn gemaakt dan door leerlingen met wiskunde A, zie tabel 5. Daarnaast zien we dat leerlingen bij wiskunde C ongeveer gelijk presteren op het examen als geheel als leerlingen bij wiskunde A. Als we dat koppelen aan de lagere vaardigheid op de overlap, dan betekent dat dat het examen wiskunde C fors makkelijker is dan het examen wiskunde A. Deze analyse ondersteunt de uitkomsten van de normeringstechniek voor het examen wiskunde C in vergelijking met wiskunde A.

		wiskunde A		wiskunde C	
opgave	max score	vraag	p-waarde	vraag	p-waarde
Veldleeuweriken	2	13	0,69	4	0,59
	4	14	0,50	5	0,40
	4	15	0,38	6	0,31
Meerlingen	4	9	0,62	21	0,55
	3	10	0,69	22	0,44
hele examen			0,45		0,45

tabel 5: Overlap wiskunde A vwo en wiskunde C vwo tijdvak 1

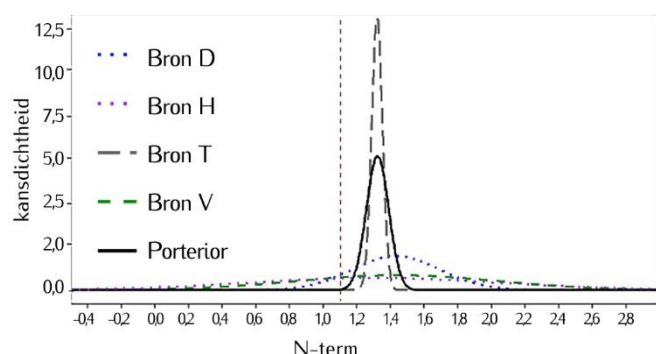
Normering tijdvak 2

De normeringssystematiek in tijdvak 2 is gelijk aan die in tijdvak 1, alleen wordt bron E vervangen door bron T. Bron T levert een schatting van de N-term gebaseerd op de prestatieverbetering van de herkansende leerlingen die in tijdvak 1 een onvoldoende hebben gehaald ^[9]. Een ander verschil met tijdvak 1 is dat voor tijdvak 2 het examen van 2023 niet direct het referentiepunt vormt, maar indirect via het examen van tijdvak 1. We vergelijken daarom ook de moeilijkheidsgraad van het examen in tijdvak 2 met het examen in tijdvak 1.

Wiskunde A havo in tijdvak 2

Voor het examen wiskunde A havo is in tijdvak 1 een N-term van 1,1 vastgesteld. Het examen in tijdvak 2 werd moeilijker gevonden dan het examen in tijdvak 1. Dit bleek uit alle bronnen. Gezien de betrouwbaarheid van bron T en het

feit dat de andere bronnen daar dichtbij lagen, is het gewogen gemiddelde van de bronnen gelijk aan de geschatte N-term op basis van bron T, namelijk 1,3.



figuur 4 Weergave N-termen bij gebruikte bronnen havo wiskunde A tijdvak 2

Kan dat zomaar: N-termen hoger dan 2,0?

Via de Examenlijn en op sociale media vroegen mensen zich af of N-termen boven de 2,0 wel toegestaan zijn. In de *Regeling omzetting scores in cijfers VO* ^[8] staat inderdaad dat de N-term waarden tussen 0,0 en 2,0 kan aannemen. Waar we 0,0 als harde ondergrens hanteren, kan het College voor Toetsen en Examens naar boven afwijken als een N-term van 2,0 of hoger gerechtvaardigd is op grond van de verzamelde informatie over de moeilijkheidsgraad. Zouden we in een dergelijk geval strak vasthouden aan de grenzen in de regeling, dan zouden we daarmee de leerlingen benadelen. N-termen van 2,0 en hoger zijn wel altijd aanleiding om nog eens kritisch te kijken wat er met het examen en/of de normering ervan aan de hand is om dit in de toekomst te voorkomen. In 2024 zijn in tijdvak 1 voor drie examens N-termen hoger dan 2,0 vastgesteld, namelijk voor wiskunde A vwo, wiskunde B vwo en geschiedenis vmbo bb.

Prestatie-eisen

Een reden waarom sommige mensen ontevreden zijn over de hoge N-termen kan zijn dat de hoogte van de N-term niet past bij de ervaren moeilijkheidsgraad van het examen. Men vindt dan eigenlijk de prestatie-eis niet passend. De prestatie-eis is het aantal punten dat een leerling moet halen om een voldoende te halen in relatie tot de moeilijkheidsgraad van het examen en dat wat een leerling heeft moeten leren voor het examen. Aan deze onvrede kunnen verschillende dingen ten grondslag liggen. Al voor corona kregen we signalen dat de lat wel wat hoger mocht bij enkele wiskundevakken. Nu is de lat mogelijk nog wat lager dan voor die tijd. Behalve dat een prestatie-eis haalbaar moet zijn – er moeten immers voldoende leerlingen over de lat kunnen springen – moet deze ook uitdagen om te leren.

Ook in 2024 is docenten de vraag voorgelegd of de N-term voor het examen 2024, en daarmee de prestatie-eis, als passend werd ervaren. De opbrengst van deze vragenlijst wordt, in combinatie met andere verzamelde informatie, meegenomen bij het vaststellen van de normering van 2025.

We zullen deze vraag de komende jaren herhalen om zo toe te werken naar prestatie-eisen die passen bij de examens en de leerlingen van dat moment. Jouw bijdrage daaraan is zeer welkom!

Noten

- [1] Uitzondering waren de exacte vakken waar nieuwe examenprogramma's voor werden ingevoerd in deze tijd.
- [2] Zie ook: Normering 2021: Anders dan anders, Paul van der Molen, *Euclides* 97-2
- [3] Zie *Vaardigheid Examenkandidaten 2023*
<https://www.examenblad.nl/system/files/2023-12/Rapport%20Vaardigheid%20Examenkandidaten%202023.pdf>
- [4] In 2021 en 2022 was er een extra tijdvak op de scholen en bestond de mogelijkheid de examens te spreiden over tijdvak 1 en tijdvak 2. De N-termen en gemiddelde cijfers in de tabel zijn die van tijdvak 1. De gemiddelde cijfers veranderen niet als we de leerlingen toevoegen die in tijdvak 2 het examen voor het eerst maakten. De reden hiervoor is dat deze groep naar verhouding zeer klein is.
- [5] Zie *Toelichting op de normering 2024*
<https://www.examenblad.nl/system/files/2024-03/toelichting-op-de-normering-2024.pdf>
- [6] Zie *Prestatie-eisen normering centrale examens vo*

<https://www.youtube.com/watch?v=eUC9VLZyGVk>

[7] Zie *De N-term maken we samen*

<https://www.youtube.com/watch?v=EZxx6BdzXmw&t=25s>

[8] zie: <https://wetten.overheid.nl/BWBR0037590/2024-03-09>

Over de auteurs

Jacqueline Wooning is programmamanager normering bij het College voor Toetsen en Examens. E-mailadres: info@cvte.nl

Harco Weemink is manager Normering & Digitalisering bij Stichting Cito. E-mailadres: harco.weemink@cito.nl