



## **Regeling van het College voor Toetsen en Examens van 9 februari 2024, nr. CvTE- 40.00438, houdende wijziging van Regeling omzetting scores in cijfers VO in verband met het vervallen van de maatregelen die werden ingesteld om de gevolgen van COVID-19 voor de normering van de centrale examens weg te nemen**

Het College voor toetsen en examens,

Gelet op artikel 2, tweede lid, aanhef en onderdeel e, van de Wet College voor toetsen en examens;

Gezien de goedkeuring van de Minister voor Primair en Voortgezet Onderwijs van 27 februari 2024, nr. 44148146;

Besluit:

### **ARTIKEL I**

Bijlage 1 van de Regeling omzetting scores in cijfers VO wordt vervangen door de bijlage bij deze regeling.

### **ARTIKEL II**

Deze regeling treedt in werking met ingang van de dag na de datum van uitgifte van de Staatscourant waarin zij wordt geplaatst.

Deze regeling zal met de toelichting in de Staatscourant worden geplaatst.

*Het College voor toetsen en examens,  
de voorzitter,  
J.H. van der Vegt*



## BIJLAGE BIJ ARTIKEL I

### Bijlage 1. Wijze van omzetting van scores in cijfers bij centrale examens VO als bedoeld in artikel 1

De instructies die het CvTE toepast om tot de N-termen te komen.

#### 1. Uitgangspunt

De normering van de centrale examens dient er zorg voor te dragen dat in opeenvolgende jaren aan kandidaten bij eenzelfde centraal examen (vak en schooltype/leerweg) dezelfde eisen worden gesteld. Omdat het technisch onmogelijk is onze examens van jaar op jaar precies even moeilijk te maken, zorgt de keuze van de normeringsterm (N-term) daarvoor. Het proces van normering is erop gericht, gegeven dit uitgangspunt, de juiste N-term te bepalen.

In supplement I staan de formules die bij de omzetting van score naar cijfer worden gehanteerd.

Voor ieder centraal examen stelt het College voor toetsen en examens (CvTE), zodra dat mogelijk is, referentiegegevens vast; een referentie-examen met een referentie-N-term en een referentiepopulatie met een referentie-gemiddeld cijfer en referentie-percentages onvoldoende.

- Het referentie-examen met de referentie-N-term vormt, in combinatie met de inhoud van de syllabus, de prestatie-eis op het centraal examen.<sup>1</sup> Bij een nieuw examen zou het, bij gelijke vaardigheid, even moeilijk moeten zijn om hetzelfde cijfer te halen als op het referentie-examen. Het CvTE monitort of handhaven van de prestatie-eis op een centraal examen reëel en mogelijk is en stelt deze zo nodig bij.<sup>2</sup>
- De referentiepopulatie is de groep examenkandidaten uit één of meerdere examenjaren. Het referentiegemiddeld cijfer geeft aan welk gemiddeld cijfer de referentiepopulatie gehaald zou hebben op het referentie-examen bij de referentie-N-term. De prestaties van nieuwe kandidaten kunnen vergeleken worden met die van de referentiepopulatie.

Bij een stabiel vaardigheidsniveau wordt de N-term bij een nieuw examen zo gekozen dat het gemiddeld cijfer zo dicht mogelijk ligt bij het referentie-gemiddeld cijfer. In een periode waarin het vaardigheidsniveau stabiel is, kunnen de referentiegegevens gebaseerd worden op meerdere jaren en in een periode waarin het vaardigheidsniveau niet stabiel is, zullen de referentiegegevens gebaseerd zijn op het meest recente geschikte centraal examen.

#### 2. Bepalen technische N-term

Voor ieder centraal examen worden afnamegegevens verzameld; scores per item per kandidaat. Daarnaast wordt informatie verzameld over de moeilijkheidsgraad van dat examen in vergelijking met het referentie-examen. Daarvoor kan gebruik gemaakt worden van:

- Pretest, posttest of anchor in package<sup>3</sup>
- Vergelijkende standaardbepalingen<sup>4</sup>
  - Uitgevoerd door de vaststellingscommissie van het CvTE
  - uitgevoerd door een groep experts onder leiding van Stichting Cito
  - uitgevoerd door correctoren van de staatsexamens
- Methoden om vaardigheidswijziging in een cluster van vakken te schatten<sup>5</sup>
- Docentenoordeel<sup>6</sup>
- Historische N-termen<sup>7</sup>
- vergelijking tijdvak 2 met tijdvak 1 op basis van de resultaatverbetering van kandidaten die in het eerste tijdvak een onvoldoende hadden

<sup>1</sup> De prestatie-eis is de prestatie die een leerling moet leveren om een voldoende op een centraal examen te halen en daarmee te laten zien dat de leerstof beschreven in de syllabus in voldoende mate wordt beheerst.

<sup>2</sup> Als blijkt dat het handhaven van de prestatie-eis niet meer mogelijk of reëel is, benadert het CvTE de Minister met een analyse waarin wordt onderbouwd dat de eisen niet meer te handhaven zijn en of wellicht maatregelen kunnen worden genomen (bijvoorbeeld in het onderwijs) waardoor de vaardigheid van de kandidaten weer in lijn komt met de eisen.

<sup>3</sup> Bij een pre- of posttest is de vergelijking gebaseerd op resultaten van leerlingen die onderdelen uit een examen en ankeropgaven hebben gemaakt. Bij anchor in package maken opgaven uit een eerder examenjaar deel uit van het huidige examen. De term package komt van de digitale examens waar de verschillende varianten van het examen tezamen package genoemd worden.

<sup>4</sup> Bij een vergelijkende standaardbepaling is het oordeel over de moeilijkheidsgraad gebaseerd op een door een groep experts uitgevoerde vergelijking van de moeilijkheidsgraad van de twee examens.

<sup>5</sup> Op basis van de waargenomen vaardigheidsontwikkeling op examens met een pretest, posttest, anchor in package of standaardbepaling, wordt de vaardigheidsontwikkeling geschat voor vergelijkbare vakken zonder dergelijke equivaleringsmethoden.

<sup>6</sup> Examendocenten worden gevraagd de moeilijkheidsgraad van het door hen nagekeken examen te vergelijken met een examen uit een eerder examenjaar.

<sup>7</sup> De N-termen die de afgelopen jaren zijn vastgesteld op het betreffende centrale examen.



De N-term wordt bepaald aan de hand van een gewogen oordeel over alle beschikbare bronnen voor het betreffende centraal examen. Op deze manier wordt gezorgd dat een verandering in vaardigheidsniveau gesignaleerd wordt en de prestatie-eis gelijk blijft ook als de vaardigheid van de examenpopulatie verschilt van de referentiepopulatie<sup>8</sup>. Het gemiddeld cijfer wordt hoger bij een toegenomen vaardigheidsniveau en lager bij een gedaald vaardigheidsniveau.

### **2.1. Voorlopige N-term in tijdvak 2**

Voor het examen in het tweede tijdvak geldt in beginsel de N-term van het eerste tijdvak. Aan de hand van verzamelde informatie over de moeilijkheidsgraad van het examen in het tweede tijdvak in vergelijking met het examen in het eerste tijdvak, wordt nagegaan of deze N-term zou leiden tot hogere eisen aan de kandidaat dan in het eerste tijdvak. Als dat het geval is, wordt de N-term zo vastgesteld dat de eisen gelijk zijn.

### **2.2. Centraal examen met een nieuw examenprogramma**

Bij de normering van een examen met een nieuw examenprogramma kan vaak geen vergelijking gemaakt worden met een referentie-examen. In dat geval kan gebruik gemaakt worden van niet-vergelijkende standaardbepalingsmethoden.

In geval van een pilot kan een examen op basis van het nieuwe programma (het pilotexamen) gelijktijdig worden afgenomen met een examen op basis van het oude programma (het reguliere examen). Als de oude en nieuwe stof elkaar deels overlappen en de overlap eenzelfde positie in het programma inneemt, wordt op grond van de resultaten van de kandidaten op de vragen over de overlappende stof de vaardigheid van de kandidaten bepaald. Deze informatie kan gebruikt worden om de moeilijkheidsgraad van de vragen over de niet-overlappende stof en vervolgens het gehele examen te bepalen.

## **3. Compensatie voor fouten of onvolkomenheden via de N-term**

Als een examen een onvolkomenheid bevat die kandidaten aanwijsbaar kan benadelen, krijgen correctoren nadere instructies door een aanvulling op het correctievoorschrift. De aanvulling kan inhouden dat alle punten voor (een deel van) de vraag aan alle kandidaten moeten worden toegekend ongeacht of er wel of geen antwoord gegeven is, en ongeacht het gegeven antwoord. In dat geval is sprake van neutralisatie van (een deel van) de vraag. De vraag draagt de facto niet meer bij aan het behaalde resultaat op het centraal examen.

Voor de nadelige gevolgen voor een kandidaat van een door het CvTE vastgestelde fout of onvolkomenheid in een examen of correctievoorschrift, compenseert het CvTE via de N-term indien:

- a. de fout of onvolkomenheid niet eerder is hersteld via een erratum op de opgaven of via een aanvulling op het correctievoorschrift, bedoeld in artikel 10 van de Regeling beoordelingsnormen en bijbehorende scores centraal examen VO;
- b. herstel via een aanvulling op het correctievoorschrift gelet op het tijdstip waarop de fout wordt vastgesteld naar het oordeel van het CvTE leidt tot een te groot risico op onjuistheden bij het vaststellen van scores op examens; en
- c. de N-term nog niet is vastgesteld.

Deze correctieprocedure via de N-term geldt voor alle tijdvakken en vindt in voorkomende gevallen plaats op grond van de in supplement II bij deze bijlage opgenomen formules voor het compenseren voor een fout of onvolkomenheid. Hierdoor wordt voorkomen dat de score van een kandidaat in een te laag cijfer wordt omgezet.

### **Supplement I. De formules voor de omzetting van score naar cijfer**

#### **1. Uitgangspunten**

Het systeem voor de omzetting van score naar cijfer is gebaseerd op de volgende vier uitgangspunten:

1. Elk gescoord punt draagt altijd bij tot een hoger examencijfer (afrondding daargelaten);
2. Een score van 0% correspondeert altijd met examencijfer 1,0;
3. Een score van 100% correspondeert altijd met examencijfer 10,0;
4. Over een zo breed mogelijk centraal interval van de scoreschaal is er (afrondding daargelaten) sprake van een evenredige stijging van score- en cijferpunten die onafhankelijk is van de normering.

<sup>8</sup> Dit is een groep kandidaten uit een of meer eerdere examenjaren, waarvan gegevens over het vaardigheidsniveau bekend zijn.



Hierbij wordt onder de score verstaan: de zuivere score, dus uitsluitend de punten die aan de kandidaat zijn toegekend voor goede antwoordelementen.

## 2. Het normeringsvoorschrift

Het normeringsvoorschrift bestaat uit twee onderdelen:

- de hoofdrelatie: de formule die, voor de overgrote meerderheid van de kandidaten, het berekeningsvoorschrift geeft voor het omzetten van score naar cijfer;
- vier grensrelaties: vier formules die (bij andere N-termen dan 1,0) voorkomen dat kandidaten met zeer lage of zeer hoge scores een cijfer zouden krijgen dat in strijd is met bovengenoemde vier uitgangspunten.

## 3. De hoofdrelatie

De hoofdrelatie geeft aldus het examencijfer als functie van de score:

$$C = 9,0 * (S/L) + N \quad (1)$$

waarin:

C = het cijfer voor het centraal examen.

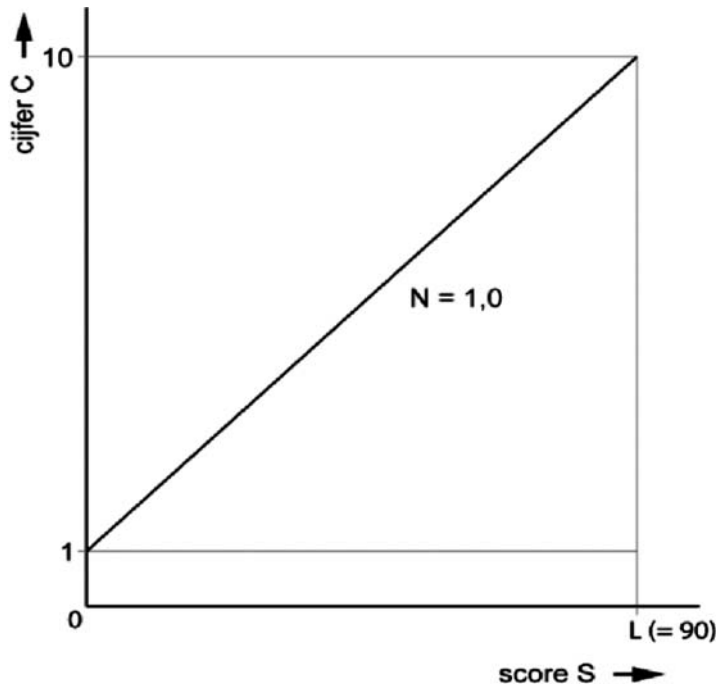
S = de score, dat wil zeggen de zuivere aan de kandidaat toegekende score.

L = de lengte van de scoreschaal, zoals vastgelegd in het correctievoorschrift;

N = de normeringsterm, een getal met één decimaal liggend tussen de waarden:  $N = 0,0$  en  $N = 2,0^9$ , vast te stellen door het CvTE middels een normeringsbeslissing.

Zijn zowel L als N bekend, dan leidt invullen van de score S direct tot het examencijfer C.

De hoofdrelatie is gevisualiseerd in figuur 1:



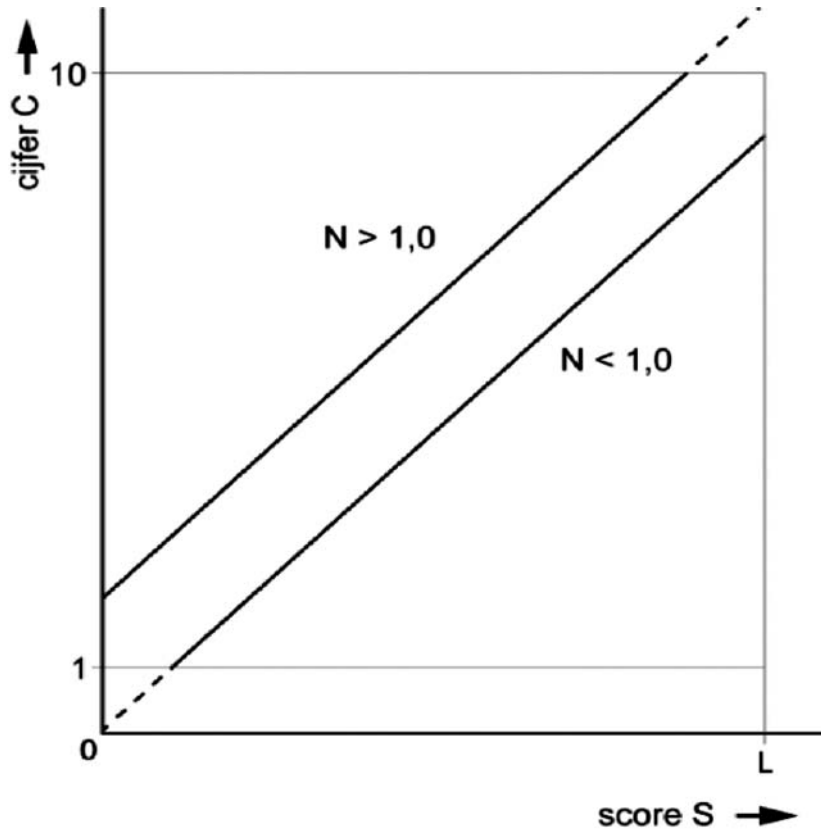
Figuur 1

<sup>9</sup> Incidenteel kan een N-term hoger dan 2,0 worden vastgesteld als dit gegeven de moeilijkheidsgraad van het examen nodig is om recht te doen aan de prestaties van de kandidaten.

#### 4. De grensrelaties

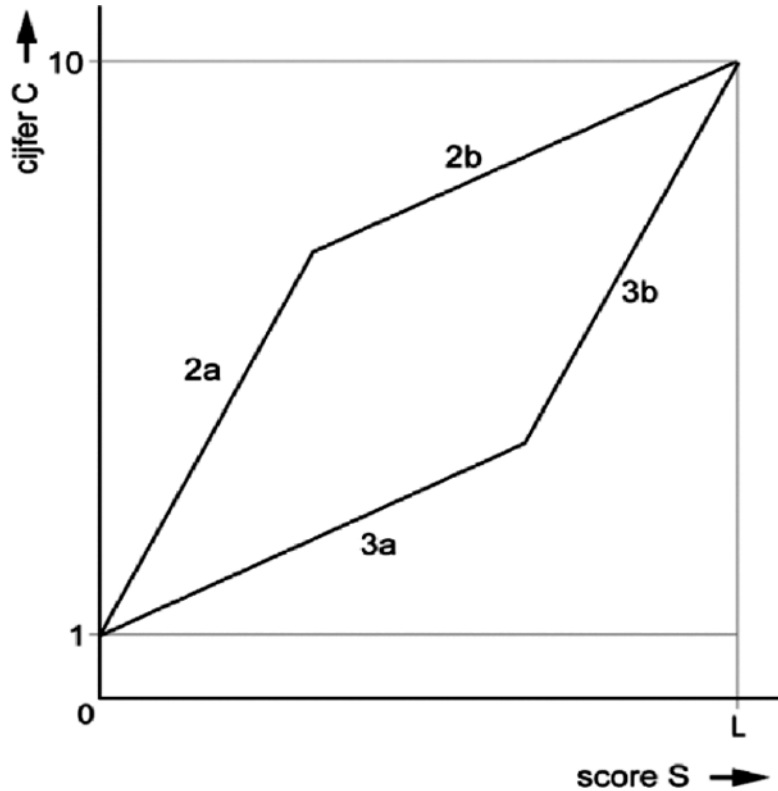
Als de normeringsterm groter of kleiner is dan 1,0 wordt niet automatisch voldaan aan de eerder genoemde vier uitgangspunten. Bij N-termen groter dan 1,0 zou een score van 0 een cijfer hoger dan 1,0 opleveren en de maximale score hoger dan 10,0. Het omgekeerde gebeurt bij N-termen kleiner dan 1,0. Hiermee is niet voldaan aan de uitgangspunten 2 en 3.

Dit is in beeld gebracht in figuur 2:



Figuur 2

Dit wordt opgelost door middel van een systeem van zogeheten grensrelaties. Het principe van grensrelaties is gevisualiseerd in figuur 3.



Figuur 3

Samen vormen de vier lijnstukken 2a, 2b, 3a en 3b een gebied waarbinnen alle toegestane score-cijfercombinaties moeten liggen. Dreigt bij toepassing van de hoofdrelatie – formule (1) – een score-cijfercombinatie buiten deze grenzen te vallen, dan moet voor de desbetreffende score dat cijfer vervangen worden door het cijfer berekend met de corresponderende grensrelatie.

De grensrelaties worden gevormd door de volgende vier formules:

$$C = 1,0 + S * (9/L) * 2 \quad (2a)$$

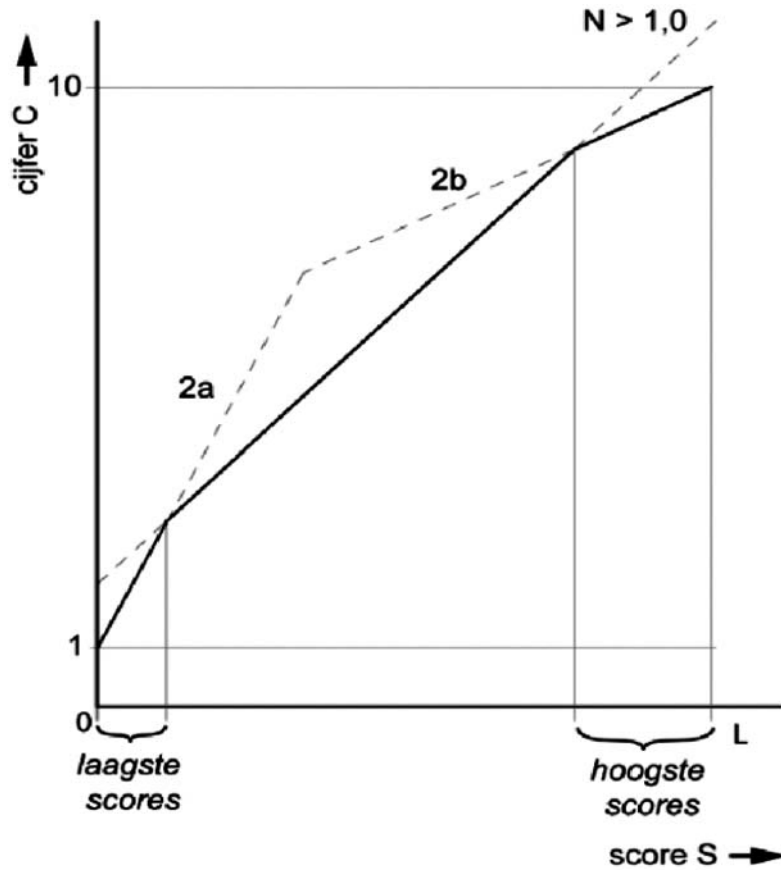
$$C = 10,0 - (L-S) * (9/L) * 0,5 \quad (2b)$$

$$C = 1,0 + S * (9/L) * 0,5 \quad (3a)$$

$$C = 10,0 - (L-S) * (9/L) * 2 \quad (3b)$$

Bij  $N > 1,0$  geldt voor de laagste scores de formule (2a) en voor de hoogste scores de formule (2b).

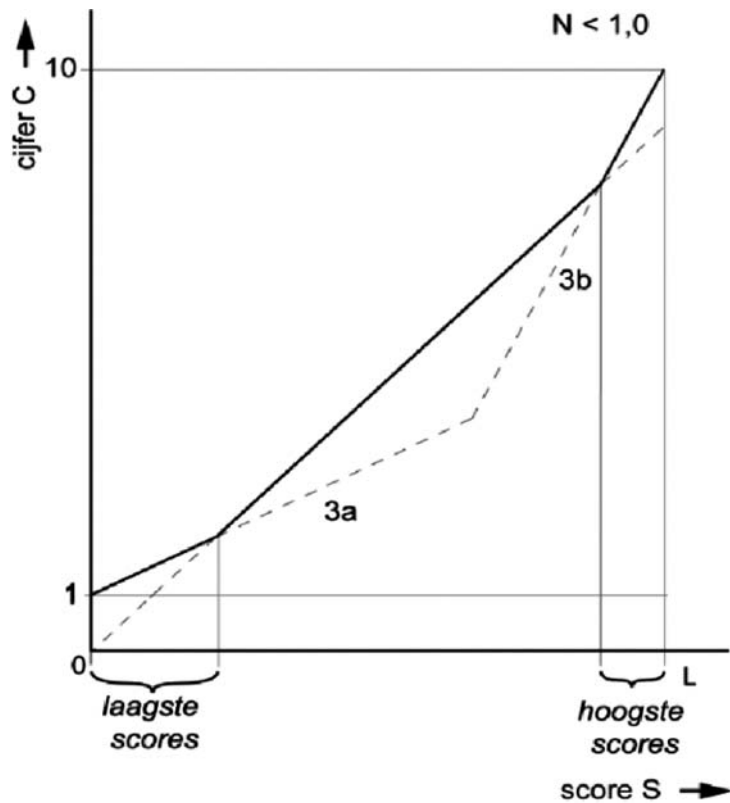
In figuur 4 is dit gevisualiseerd.



Figuur 4

Bij  $N < 1,0$  geldt voor de laagste scores de formule (3a) en voor de hoogste scores de formule (3b).

In figuur 5 is dit gevisualiseerd.



Figuur 5

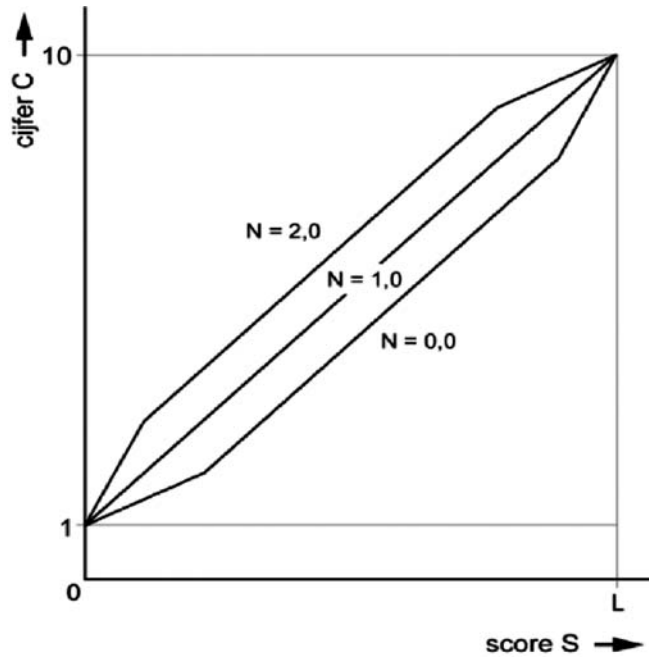
### 5. Samenvattend

Samenvattend ziet het er als volgt uit:

Bij een waarde voor de normeringsterm van  $N = 1,0$  treedt het systeem van grensrelaties niet in werking en resulteert een score-cijfertransformatie die grafisch wordt gerepresenteerd door de rechte lijn van Fig.1, de lijn die in Fig. 6 is gelabeld met: 'N=1,0'.

Bij alle andere waarden van  $N$  zijn de grensrelaties wel van belang. In figuur 6 zijn als voorbeelden de twee gevallen in beeld gebracht die respectievelijk corresponderen met  $N = 2,0$  en  $N = 0,0$ . Deze leveren als score-cijfertransformaties de twee dubbel-geknikte lijnen op (gelabeld met 'N=2,0' en 'N=0,0').





Figuur 6

### Supplement II. De formules voor het compenseren voor een fout of onvolkomenheid

De N-term die zou zijn vastgesteld als de desbetreffende fout of onvolkomenheid in een examen of correctievoorschrift niet door het CvTE was vastgesteld, wordt verhoogd met  $9 * P_{vrg} * M_{vrg} / L$ , waarbij deze uitkomst wordt afgerond op één decimaal.

In deze formule staat  $P_{vrg}$  voor de P-waarde van de onvolkomen vraag en  $M_{vrg}$  voor de maximaal haalbare score op deze vraag. Een P-waarde van 0,63 betekent dat de kandidaten gemiddeld 63% van  $M_{vrg}$  behaald hebben. L staat voor de lengte van de scoreschaal ofwel de maximaal haalbare score op het gehele examen.

Het uitgangspunt bij deze werkwijze is dat de kandidaat die geen punten heeft kunnen scoren op de onvolkomen vraag, voldoende wordt gecompenseerd om geen nadeel te ondervinden van de fout. Aan het derde tijdvak nemen zo weinig kandidaten deel dat de P-waarden geen relevante informatiebron vormen. De N-term die zonder de fout zou zijn vastgesteld wordt verhoogd met  $9 * M_{vrg} / L$ . Deze compensatie is wellicht te hoog, maar zo wordt voorkomen dat kandidaten benadeeld worden door de fout.



## TOELICHTING

### Algemeen

In 2021 en 2022 golden diverse examenmaatregelen om examenkandidaten tegemoet te komen tijdens de coronapandemie. Daarnaast werd de normering in 2021, 2022 en 2023 aangepast om te kunnen omgaan met de gevolgen van de coronapandemie. Met het afschaffen van de examenmaatregelen in 2023 is een fase aangebroken waarin het logisch is om terug te gaan naar een normering waarmee de prestatie-eis weer gehandhaafd wordt op een vergelijkbare wijze als voor corona. Samen met Stichting Cito heeft het College voor toetsen en examens de normeringssystematiek doorontwikkeld om beter te kunnen omgaan met veranderingen. Deze aangepaste systematiek geldt vanaf 2024.

### ***Meer bronnen gebruikt voor technische N-term***

De belangrijkste verandering in de systematiek is dat alle beschikbare bronnen gebruikt worden bij het bepalen van de technische N-term, ongeacht het aantal kandidaten, het soort examen of het tijdvak waarin een examen is afgenomen. Hierbij krijgen bronnen met een grotere betrouwbaarheid een zwaarder gewicht dan bronnen met een kleinere betrouwbaarheid. De beschikbare bronnen zijn gelijk aan de bronnen die voor de coronapandemie werden gebruikt, aangevuld met het docentoordeel en de historische N-term die in de coronajaren 2021 t/m 2023 werden gebruikt. De bronnen zijn dus niet nieuw maar worden in de nieuwe systematiek allemaal gecombineerd om tot een technische N-term te komen.

### ***Alle examens dezelfde normering***

Ongeacht het aantal kandidaten, het type examen of het tijdvak is de wijze waarop de technische N-term bepaald wordt hetzelfde; voor elk examen wordt de technische N-term gebaseerd op een gewogen oordeel van alle beschikbare bronnen van informatie. Er zijn wel normeringsbronnen die niet voor ieder examen beschikbaar zijn. Zo kan het voorkomen dat door wijzigingen in een examen (programma) geen historische N-term beschikbaar is, is een vergelijking tussen tijdvak 2 en tijdvak 1 vanzelfsprekend alleen beschikbaar voor examens in tijdvak 2 en is een oordeel over de moeilijkheidsgraad door correctoren van de staatsexamens alleen beschikbaar voor examens in tijdvak 3 en de aangewezen vakken in tijdvak 2. Maar in de basis is de normeringssystematiek voor alle examens gelijk.

### ***Van referentie-examen naar referentiegegevens***

In de nieuwe normeringssystematiek wordt expliciet gebruik gemaakt van referentiegegevens in plaats van alleen een referentie-examen. Allereerst sluit dit beter aan bij de variatie aan 'centrale examens die er is. Voor de profielvakken is er ieder jaar een blauwe en een rode versie die beide door een deel van de examenpopulatie gemaakt worden en waarvan er één wordt aangewezen als referentie-examen. Om iets te kunnen zeggen over de kenmerken van de examenpopulatie van een bepaald jaar is het van belang om te kijken naar de gehele populatie en niet naar de deelpopulatie die een bepaalde versie heeft gemaakt. Er kan dan niet gesproken worden van dé examenpopulatie die hét referentie-examen heeft gemaakt. Iets vergelijkbaars geldt voor de digitale examens voor vmbo bb en kb waar per jaar negen (of vier) varianten van het examen worden afgenomen. Een splitsing tussen een referentie-examen met referentie-N-term enerzijds en een examenpopulatie met referentiegemiddeld cijfer en referentiepercentage onvoldoende anderzijds maakt ook zichtbaar dat er twee aspecten zijn om naar te kijken:

- Het overzetten van de prestatie-eis van een referentie-examen op een nieuw examen. Dit bepaalt de N-term van het nieuwe examen.
- De vaardigheid van de nieuwe populatie in vergelijking met het referentie-examen. Dit bepaalt het gemiddeld cijfer dat de nieuwe populatie haalt.

De referentiegegevens kunnen, net als in 2019, betrekking hebben op meerdere jaren (en dus ook meerdere examens) als er sprake is van een stabiele situatie. Dit is het geval als er geen wijzigingen zijn in onderwijs, examens en de examenpopulatie. Er is dan sprake van continuïteit voor wat betreft de samenstelling van de examens én de vaardigheid van de examenpopulatie. Als dat niet het geval is, kan alleen naar het meest recente jaar worden gekeken. Als er bijvoorbeeld sprake is van een nieuw examenprogramma kan soms helemaal niet worden vergeleken met een recent jaar en moet een nieuwe standaard gezet worden, ofwel een nieuwe prestatie-eis vastgesteld worden. Door op deze manier met referentiegegevens om te gaan, kan beter worden ingespeeld op veranderingen.



---

### **Supplement I: De formules voor de omzetting van score naar cijfer**

De inhoud van supplement I is niet veranderd, maar de tekst is hier en daar geactualiseerd en ingekort.

### **Supplement II: Formules voor compensatie via de N-term in geval van fouten in een examen**

De formule voor tijdvak 2 is gelijkgetrokken met de formule voor tijdvak 1. Reden is dat de formule voor tijdvak 1 beter aansluit bij de wijze waarop de N-term wordt vastgesteld, zeker nu de normerings-systematiek in de basis gelijk is voor alle tijdvakken. Alleen voor tijdvak 3 is de formule nog afwijkend. Dat heeft als oorzaak dat voor deze examens, afgenomen door de Staatsexamens vo onvoldoende afnamegegevens beschikbaar zijn om p-waarden te kunnen bepalen met voldoende informatiewaarde.

*Het College voor toetsen en examens,  
de voorzitter,  
J.H. van der Vegt*