

Ministerie van Onderwijs, Cultuur en Wetenschap
Minister voor Basis- en Voortgezet Onderwijs en Media
Drs. A. Slob
Postbus 16375
2500 BJ 'S-GRAVENHAGE



NATIONAAL REGIEORGAAN
ONDERWIJSONDERZOEK (NRO)
POSTADRES:
POSTBUS 93461 2509 AL DEN HAAG
BEZOEKADRES:
LAAN VAN NIEUW OOST INDIË 300
2593 CE DEN HAAG
T 070 344 09 11 W WWW.NRO.NL

Datum: 28 maart 2019

Betreft: Aanbieding bevindingen ontwikkeling examenresultaten VO n.a.v. motie Van Meenen c.s.

Geachte heer Slob,

Met genoegen bied ik u hierbij de bevindingen aan van onze commissie naar aanleiding van uw verzoek om een antwoord te formuleren op de vragen zoals gesteld in de motie Van Meenen c.s. over de ontwikkeling van de PISA- en examenresultaten in het voortgezet onderwijs in de afgelopen tien jaar.

Naar aanleiding van de door de Tweede Kamer aangenomen motie heeft u het Cito verzocht om nader onderzoek te doen in nauwe samenwerking met het College voor Toetsen en Examens (CvTE). Om de onafhankelijkheid van het onderzoek te waarborgen en een antwoord op de Kamervragen te formuleren, heeft het Nationaal Regieorgaan Onderwijsonderzoek (NRO) op uw verzoek een commissie ingesteld. Onze commissie heeft vanuit die rol regie gevoerd over het onderzoek. Het NRO voerde het secretariaat voor de commissie. Wij hebben als commissie richting gegeven aan het onderzoek en ons een oordeel gevormd over het resultaat; daarvan doen wij in deze brief verslag. Bij de wegging hebben naast het Cito-onderzoek gesprekken een rol gespeeld die wij gevoerd hebben met het CvTE, de VO-raad, de Inspectie van het Onderwijs en een aantal panels van bestuurders en schoolleiders. Het geheel van onderzoek en gevoerde gesprekken vormt een goede basis voor deze brief.

De commissie bestond uit dr. Iris Egberink (RUG/COTAN), prof. dr. Peter van der Heijden (UU) en dr. Tim Mainhard (UU) als inhoudsdeskundigen en ondergetekende als voorzitter. Mr. Niels Rijke (NRO) ondersteunde de commissie als secretaris. Van de zijde van het Cito was ir. Paul van der Molen projectleider van het onderzoek, waarvan wij u de rapportage hierbij integraal als bijlage aanbieden.¹

De motie Van Meenen c.s.² stelt – door ons ontleed – de volgende vragen:

Over de uitgangssituatie:

1. Zijn de prestaties van Nederlandse leerlingen voor PISA voortdurend dalend op alle niveaus en alle domeinen?
2. Is het gemiddelde cijfer op het centraal examen (CE) gelijkblijvend tot stijgend (in samenhang met de gewijzigde kernvakkenregeling)?
3. Blijven de cijfers op het schoolexamen (SE) gelijk?

¹ P. van der Molen, S. Schouwstra, R. Feskens & M. van Onna, *Vaardigheidsontwikkelingen volgens PISA en examens*, Cito 2019.

² Motie Van Meenen c.s., 13 februari 2018, TK 2017-2018, 31289, nr. 359.

Nader te onderzoeken:

4. Welke rol speelt de N-term en hoe robuust is deze procedure?
5. Hoe verhouden de metingen van PISA, CE en SE zich tot elkaar en wat betekent dit voor het antwoord op vraag 6?
6. Wat is de reële ontwikkeling van de vaardigheden van scholieren in de afgelopen 10 jaar (minimaal voor Engels en wiskunde)?
7. In hoeverre bevorderen/belemmeren de N-term en andere factoren het zicht op het beeld als bij 6 bedoeld?

Te beantwoorden hoofdvragen:

- I Kunnen de kwaliteit van de examens en de ontwikkeling van de vaardigheden van scholieren optimaal worden vastgesteld en gevolgd?
- II Wat valt er – met de nuance vanuit het antwoord op hoofdvraag I – te zeggen over de ontwikkeling van die vaardigheden in de afgelopen 10 jaar?

In deze brief gaan wij achtereenvolgens in op de voorgaande vragen en verwijzen wij voor een uitgebreide onderbouwing naar het Cito-rapport. Wij besluiten deze brief met een conclusie, waarin wij de hoofdvragen zullen beantwoorden, en we maken een aantal slotopmerkingen over het belang van de centrale examens.

Beantwoording vragen

1. Zijn de prestaties van PISA voortdurend dalend op alle niveaus en alle domeinen?

De commissie constateert dat de prestaties van Nederlandse leerlingen voor PISA niet voortdurend dalen op alle niveaus en alle domeinen, maar wel voor bepaalde niveaus op een aantal domeinen, ook vergeleken met andere OESO-landen. Het algemene beeld van de ontwikkeling van prestaties van Nederlandse leerlingen voor PISA is deels stabiel en deels dalend.

Toelichting

PISA ('Programme for International Student Assessment') is een internationaal vergelijkend onderzoek dat wordt uitgevoerd in opdracht van de Organisatie voor Economische Samenwerking en Ontwikkeling (OESO). PISA evalueert iedere drie jaar de kennis en vaardigheden van 15-jarigen voor de domeinen 'leesvaardigheid', 'wiskundige geletterdheid' en 'natuurwetenschappelijke geletterdheid', waarbij de nadruk ligt op het toepassen ervan in alledaagse situaties. In Nederland is de deelname voor scholen en leerlingen vrijwillig en zijn er geen consequenties verbonden aan de prestaties van de leerlingen. Daarmee is PISA een zogeheten 'low stakes' toets, er staat voor leerlingen weinig op het spel.

Voor het onderzoek van het Cito zijn de gegevens van de periode 2003 tot en met 2015 gebruikt, dat wil zeggen gegevens van vijf achtereenvolgende PISA afnames.³ De gegevens van de Nederlandse scholieren zijn zowel vergeleken met het gemiddelde van de in het betreffende jaar deelnemende OESO-landen als bekeken per schooltype en leerweg (vmbo-bb, vmbo-kb, vmbo-gt, havo en vwo). Daarbij zijn tevens de vernieuwingen en wijzigingen die in de loop der jaren in PISA plaatsvonden in ogenschouw genomen.

De gegevens voor het domein 'Leesvaardigheid' laten zien dat voor de havo en het vwo de scores van de Nederlandse scholieren over het geheel genomen nauwelijks veranderd zijn in de periode 2003-2015. In het vmbo (vmbo-bb, vmbo-kb en vmbo-gt) is een neerwaartse trend te zien. Het betreft een daling van ongeveer 40 punten voor zowel vmbo-bb als kb. Deze daling is aanzienlijk als in ogenschouw wordt genomen dat het verschil tussen de opeenvolgende schooltypen en leerwegen (vmbo-bb/vmbo-kb/vmbo-gt/havo/vwo) ongeveer 50 punten is. Voor vmbo-gt is de daling ongeveer 30 punten.

³ De gegevens van de PISA afname van 2018 zullen in december 2019 bekend worden gemaakt en zijn daarom niet meegenomen in dit onderzoek.

Voor het domein 'Wiskundige geletterdheid' laten de resultaten van Nederlandse scholieren voor alle schooltypen en leerwegen bijna bij iedere volgende PISA-meting een lichte daling zien. Over een periode van 12 jaar leidt dit tot een aanzienlijke daling, die per schooltype en leerweg ligt tussen 30 en 50 punten. De PISA scores van de OESO-landen als geheel dalen ook, maar Nederland daalt relatief harder. Wanneer gekeken wordt naar de laatste twee PISA-metingen, dat wil zeggen in 2012 en 2015, dan zien we een zich voortzettende daling bij vmbo-bb, havo en vwo (bij vmbo-kb en vmbo-gt blijven de scores dan gelijk). In 2015 werd een digitale afname van PISA standaard (daarvoor waren afnames op papier). Naar het oordeel van de commissie kan het voor wiskundige domeinen een verschil maken of opgaven op papier of digitaal worden afgenomen; wellicht dat dat hier een rol heeft gespeeld.

De gegevens voor het domein 'Natuurwetenschappelijke geletterdheid' laten zien dat voor alle schooltypen en leerwegen de scores van de Nederlandse scholieren in de periode 2003-2012 zijn gedaald, al is de daling geringer dan bij 'Wiskundige geletterdheid'. Deze neerwaartse trend zet zich in 2015 verder door voor het vmbo (vmbo-bb, vmbo-kb en vmbo-gt), terwijl de gemiddelde scores voor havo en vwo gelijk blijven. Bij deze daling voor 'Natuurwetenschappelijke geletterdheid' merken we op dat de ontwikkeling in Nederland vergelijkbaar is met die van de OESO als geheel en dat in 2015 ook voor dit domein een grondige inhoudelijke herziening heeft plaatsgevonden die een mogelijke verklaring voor de daling zou kunnen zijn.

Voor uitgebreide informatie: paragraaf 2.1 van het Cito-rapport.

2. Is het gemiddelde cijfer op het centraal examen (CE) gelijkblijvend tot stijgend (in samenhang met de gewijzigde kernvakkenregeling)?

Wij hebben als commissie geconstateerd dat in de periode van 2008 tot 2017 het over alle vakken gemiddelde cijfer op het CE voor het vmbo-bb en vwo aanzienlijk is gestegen (resp. 0,38 en 0,22 punt), voor de overige schooltypen en leerwegen is er een lichtere stijging.

De resultaten voor Engels zijn voor alle schooltypen en leerwegen overwegend stijgend, voor Nederlands en wiskunde verschilt dit per schooltype/leerweg en is de ontwikkeling grilliger. Kijkend naar de verschillen in gemiddeld cijfer tussen 2008 en 2017, valt op dat de resultaten voor Nederlands en wiskunde voor vmbo-bb in 2017 hoger zijn dan in 2008. Voor wiskunde zijn de resultaten in 2017 op het vwo aanzienlijk hoger dan in 2008 terwijl de resultaten voor vmbo-kb, vmbo-gt en havo lager zijn.

In ons antwoord op vraag 5 gaat de commissie in op de vele wijzigingen die in de afgelopen tien jaar het beeld van het CE en de behaalde resultaten hebben beïnvloed.

Voor uitgebreide informatie: paragraaf 2.2 van het Cito-rapport.

3. Blijven de cijfers op het schoolexamen (SE) gelijk?

De commissie constateert dat in de periode 2008-2017 het gemiddelde cijfer voor alle vakken van het schoolexamen tezamen stabiel is. Tussen 2008 en 2012 daalde het gemiddelde cijfer licht, en daarna trad een lichte stijging op. Hierbij merken we wel op dat de schoolexamens per school anders ingericht kunnen zijn, vastgelegd in het programma van toetsing en afsluiting (PTA). Zowel de ordening in het PTA als de validiteit van de toetsen en evaluaties loopt uiteen.

De drie kernvakken afzonderlijk wijken over de periode 2010-2017 iets af van dit beeld. Voor het schoolexamen Engels zien we voor het vmbo een daling tot 2012 voor alle schooltypen en leerwegen. Daarna zien we een stijging, waarbij de resultaten in 2017 iets hoger liggen dan in 2010. Bij het schoolexamen Nederlands is voor alle schooltypen en leerwegen een lichte daling zichtbaar in de periode 2010-2013, waarna het gemiddeld cijfer in de periode tot 2017 stabiliseert. Voor wiskunde stijgen in de periode 2010-2017 de gemiddelde schoolexamencijfers voor havo en vwo (met uitzondering van een lichte daling in 2017 voor havo op het schoolexamen wiskunde A), voor het vmbo (alle drie de leerwegen) zijn de veranderingen in het wiskundecijfer geringer.

Voor uitgebreide informatie: paragraaf 2.2.3 en bijlage 1 van het Cito-rapport.

4. 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.

Toelichting

Om de vergelijkbaarheid van CE-cijfers over jaren heen te bewerkstelligen wordt de N-termprocedure gebruikt.⁵ Examenversies zijn namelijk nooit precies even moeilijk en populaties van leerlingen hebben nooit precies dezelfde vaardigheid. De N-term drukt het verschil in moeilijkheid van het examen uit in vergelijking met eerdere jaren, waarbij tevens rekening gehouden kan worden met een eventueel verschil in vaardigheid tussen populaties van leerlingen in opeenvolgende jaren. De N-term wordt gebruikt om een eventueel verschil zodanig te verdisconteren, dat de cijfers over de jaren heen vergelijkbaar zijn en de leerlingen een zo eerlijk mogelijk cijfer krijgen dat hun vaardigheidsniveau weerspiegelt. Hoewel dit principe van vergelijkbaarheid eenvoudig is, is het bepalen van het verschil in moeilijkheid van examenversies en/of het verschil in vaardigheid van leerlingenpopulaties dat niet. Dit heeft te maken met de complexiteit van het meten van vaardigheden. De zogeheten normhandhavingsonderzoeken leveren informatie over de vaardigheden van leerlingen.

Op dit moment wordt gebruik gemaakt van drie typen normhandhavingsonderzoek, namelijk 'Anchor in Package' (bij de digitale examens vmbo-bb en vmbo-kb), 'pretest en posttest' (hoofdzakelijk bij vakken met gelijkblijvende onderwerpen in het examenprogramma en met veel examenleerlingen) en 'standaardbepaling'. Grofweg gezegd, is 'Anchor in Package' drie keer zo nauwkeurig als 'pretest en posttest' en is 'pretest en posttest' weer drie keer zo nauwkeurig als 'standaardbepaling'. Echter, 'Anchor in Package' en 'pretest en posttest' zijn ook de meest bewerkelijke procedures en vereisen gegevens van veel examenleerlingen. Bovendien is bij 'Anchor in Package' geheimhouding van delen van het examen vereist, dat als bezwaarlijk wordt gezien. De methoden 'pretest en posttest' kosten veel tijd en geld om te realiseren.

In het algemeen geldt dat hoe meer leerlingen deelnemen aan een examen voor een bepaald vak, des te robuuster de procedure is die gevolgd kan worden. Daarnaast is het type vak en de mate waarin de inhoud van examens wijzigt bepalend voor de mogelijkheden qua te volgen procedure. Zo geldt dat voor een vak als geschiedenis, waarvan de inhoud regelmatig sterk wijzigt, een minder robuuste procedure gevolgd moet worden met een grotere onzekerheidsmarge. Tevens bepaalt de budgettaire ruimte de mogelijkheden voor de te volgen procedure.

Op basis van de informatie uit de normhandhavingsonderzoeken, de afnamegegevens van de examenleerlingen en de 'Regeling omzetten scores in cijfers' wordt de zogeheten technische N-term bepaald. De commissie constateert dat er over de tijd geen grote verschillen zijn tussen de technische N-termen voor de examennormering. Dat betekent dat het goed lukt om de examens op een vergelijkbaar moeilijkheidsniveau te houden en dat examenprestaties niet kunstmatig opgehoogd zijn door middel van de technische N-term.

⁴ De normeringsmethodiek bij het CE is eerder in meer detail onderzocht door H.J.A. Hoijtink en K. Sijtsma (*Metten onder druk. Advies aan de CEVO inzake de normering van eindexamens voortgezet onderwijs*, 2009) en C.A.W. Glas en R.R. Meijer (*Commentaar op review normering 2009-2014: rapportage van kwaliteitsonderzoek naar normering*, CvTE 2014).

⁵ Het uitgangspunt van vergelijkbare examencijfers over de jaren heen vloeit voort uit de 'Regeling omzetten scores in cijfers centrale examens en rekentoets VO 2016' (laatst gewijzigd per februari 2019).

Daarnaast constateert de commissie dat in 84% van de 2044 onderzochte centrale examens zonder negatieve technische N-term in de periode 2008 tot en met 2018, de technische N-term ook de definitieve N-term werd. Bij het vaststellen van de definitieve N-term kunnen de reacties van leerlingen, docenten en vakverenigingen op de examens (o.a. kritiek op een vraag) ook meegenomen worden en leiden tot het bijstellen van de N-term. De gemiddelde bijstelling van technische naar definitieve N-term bij de overige 16% van de onderzochte examens is +0,09. Dit betekent dat de gemiddelde verwachting is dat een leerling bij één van alle afgelegde examenvakken door toedoen van de bijstelling van de technische N-term 0,1 cijferpunt hoger krijgt. Hoewel bijstelling van de technische N-term relatief vaak voorkomt op havo en vwo en bij de exacte vakken, is de gemiddelde bijstelling voor havo, vwo en exacte vakken respectievelijk +0,07, +0,12 en +0,10. Dit is in de orde van grootte van de gemiddelde bijstelling voor alle schooltypen, leerwegen en vakken samen. Het effect van de bijstelling op de uitslag van het examen is daarmee verwaarloosbaar klein te noemen en geeft aan dat de N-term procedure als geheel (dat wil zeggen de technische N-term inclusief een eventuele bijstelling door het CvTE tot een definitieve N-term) robuust is.

Voor uitgebreide informatie: paragraaf 3.1 tot en met 3.4 van het Cito-rapport.

5. Hoe verhouden de metingen van schoolexamen, centraal schriftelijk en PISA zich tot elkaar en wat betekent dit voor het antwoord op vraag 6?

Wij constateren als commissie dat de examens (zowel SE als CE) en PISA niet goed te vergelijken zijn vanwege verschillen tussen gemeten vaardigheden, de leerlingenpopulatie die (deels) afwijkt, de belangen die op het spel staan voor leerlingen en hun daarmee samenhangende motivatie, en daarbij aansluitend de toegenomen nadruk op het CE. Vanwege die toenemende nadruk op het CE is het SE meer op het CE gaan lijken, zoals de commissie Ten Dam heeft geconcludeerd, waardoor de focus binnen het curriculum meer is komen te liggen op het ontwikkelen en toetsen van vakkennis en vakspecifieke vaardigheden binnen de examendomeinen.⁶ Eerder merkten we al op dat de resultaten op het CE een vergelijking over de jaren goed mogelijk maken; door de verschillende invulling die scholen aan de schoolexamens kunnen geven is een vergelijking over de jaren heen voor het SE minder betrouwbaar.

Toelichting

Waar PISA en het CE grofweg dezelfde thema's beslaan (zoals wiskunde of natuurkunde) hebben zij een andere focus, dat wil zeggen zij toetsen andere type vaardigheden. PISA richt zich met name op de vraag in hoeverre leerlingen in staat zijn om 'alledaagse' problemen adequaat op te lossen, terwijl de Nederlandse examens zich meer op kennis en technische of vakspecifieke vaardigheden van een vak richten. Bijvoorbeeld bij leesvaardigheid richt PISA zich op reflectie en functioneel gebruik van verschillende soorten teksten ('lezen om te leren'), terwijl het CE zich nadrukkelijk richt op een abstracte tekstanalyse en de beheersing van een tekstanalytisch begrippenapparaat.

Daarnaast is er (deels) een leeftijdsverschil tussen de leerlingenpopulatie die PISA maakt en die eindexamen doet. De leeftijd van de leerlingenpopulatie die PISA maakt is 15 jaar. Er is deels overlap van deze populatie met 15-jarige leerlingen die vmbo-examen doen, maar andere leerlingen zijn 16, 17 of 18 jaar als ze examens doen en hebben dan meestal een onderwijsaanbod in de bovenbouw gevolgd dat (sterk) afwijkt van het aanbod in de onderbouw.

Verder is er bij de examens (zowel SE als CE) sprake van een zogenaamde 'high stakes' toets, terwijl bij PISA sprake is van een 'low stakes' toets. PISA-toetsen hebben voor leerlingen geen consequenties, terwijl de examens (SE en CE) de verdere mogelijkheden voor doorstroom naar het vervolgonderwijs bepalen. Tevens lijken Nederlandse leerlingen voor PISA minder gemotiveerd dan leerlingen in andere OESO-landen. Er zijn echter geen meerjarencijfers over de motivatieontwikkeling van Nederlandse leerlingen voor PISA beschikbaar om dit te onderbouwen.

⁶ G.T.M. ten Dam, S. Waslander, A.A. Béguin & M.J.M. van Leeuwen, *Een volwaardig schoolexamen*, Commissie Kwaliteit Schoolexaminering 2018.

Wat betekent dit voor het volgen van de ontwikkeling van de vaardigheden van leerlingen? PISA en de eindexamens meten andere vaardigheden binnen vergelijkbare domeinen in verschillende fases van de loopbaan van de leerlingen. Bij PISA rijst de vraag in hoeverre de mogelijk lage motivatie van de leerlingen een betrouwbaar beeld van hun vaardigheden in de weg staat. Voor het SE geldt dat de validiteit en inrichting van het schoolexamen tussen scholen kan verschillen waardoor uitspraken op nationaal niveau over de vaardigheidsontwikkeling niet goed mogelijk zijn ('optelprobleem').

Voor uitgebreide informatie: hoofdstuk 4 en 6 van het Cito-rapport.

6. Wat is de reële ontwikkeling van de vaardigheden van scholieren in de afgelopen 10 jaar (minimaal voor Engels en wiskunde)?

De commissie baseert zich voor de beantwoording van deze vraag primair op de resultaten van het CE en PISA. Wij constateren dat de vaardigheden van Nederlandse leerlingen zoals getoetst in het CE de afgelopen jaren over het geheel stabiel zijn of stijgen. Specifiek voor Engels betreft dit overwegend een stijging voor alle leerlingen terwijl de vaardigheidsontwikkeling voor wiskunde een divers beeld laat zien voor de verschillende schooltypen en leerwegen. Tegelijkertijd blijven de PISA-resultaten deels stabiel en ten dele dalen ze. Dit laatste wijst mogelijk op een verminderde vaardigheid om algemene kennis op dagelijkse problemen toe te passen.

Toelichting

Bij vraag 1 constateren wij als commissie dat er relatief gezien sprake is van een sterkere daling van de Nederlandse prestaties voor PISA dan de daling van het OESO gemiddelde bij wiskundige geletterdheid en in 2015 bij natuurwetenschappelijke geletterdheid. Hiervoor zijn twee mogelijke redenen te geven. Ten eerste lijkt het de commissie aannemelijk dat een versterkte focus op de voorbereiding op de Nederlandse examens hier een rol speelt, met name de nadruk van het onderwijs op vakkennis en vakspecifieke vaardigheden, terwijl PISA algemene 'geletterdheid' toetst. Dit zou betekenen dat er in het Nederlandse onderwijs minder nadruk op het oplossen van alledaagse, meer praktische problemen is komen te liggen, zeker in de bovenbouw van het VO. Hoewel dit voor vmbo-leerlingen plausibel lijkt, is het de vraag in hoeverre dit ook voor havo en vwo leerlingen speelt die ten tijde van het PISA-onderzoek nog aan de bovenbouw moeten beginnen. Ten tweede lijkt het de commissie aannemelijk dat de PISA-uitkomsten geen goed beeld geven van de vaardigheden van de leerlingen, omdat PISA een 'low stakes' toets is. Daarbij leeft binnen de commissie de vraag in hoeverre leerlingen door de jaren heen 'low stakes' toetsen anders zijn gaan benaderen, dat wil zeggen minder hun best zijn gaan doen op dergelijke toetsen waar voor leerlingen weinig op het spel staat.

Bij vraag 4 constateert de commissie dat voor het CE voor veel vakken, waar veel of alle leerlingen aan deelnemen, gebruik kan worden gemaakt van de meest geavanceerde normeringsmethoden. Dit betekent dat in die situaties met meer zekerheid een conclusie getrokken kan worden over de vaardigheidsontwikkeling van leerlingen dan voor examens waarbij dit niet het geval is (bijvoorbeeld bij de beroepsgerichte vakken in het vmbo). De commissie is van mening dat de robuuste normering van examens, zoals die plaatsvindt met de N-term, essentieel is om de daadwerkelijke trend van de vaardigheden van Nederlandse leerlingen te kunnen volgen. In het verlengde van ons antwoord op vraag 2 constateren wij dan ook dat de vakspecifieke kennis en vaardigheden van de leerlingen, afgaande op de CE-resultaten, gelijk is gebleven of gestegen.

Echter, bij de vaardigheidsontwikkeling op het CE dient opgemerkt te worden, dat het de commissie aannemelijk lijkt, in het verlengde van de conclusies van de commissie Ten Dam, dat met het toenemend belang van het eindexamen en specifiek het CE, het SE meer en meer is gaan dienen als een voorbereiding op het CE met een overlap van de vaardigheden die daarmee gemeten worden. Dit betekent dat de leerlingen vaardiger geworden zijn op een smaller gebied, zoals gedefinieerd door de domeinen van het CE.

Voor uitgebreide informatie: hoofdstuk 2 en paragraaf 3.5 van het Cito-rapport.

7. In hoeverre bevorderen/belemmeren de N-term en andere factoren het zicht op het beeld als bij 6 bedoeld?

De commissie concludeert dat de N-term het zicht op de reële ontwikkeling van vaardigheden niet belemmert maar juist bevordert. De N-term zorgt ervoor dat cijfers over de jaren heen een goede weerspiegeling zijn van de vaardigheidsontwikkeling van de leerlingen. Hierbij dient opgemerkt te worden dat de nauwkeurigheid van de normering afhangt van de gebruikte normhandhavingmethode. Grofweg gezegd, is 'Anchor in Package' drie keer zo nauwkeurig als 'pretest en posttest' en is 'pretest en posttest' weer drie keer zo nauwkeurig als 'standaardbepaling'.

De in de afgelopen 10 jaar genomen beleidsmaatregelen (o.a. verscherping van exameneisen) belemmeren enigszins het zicht op de reële ontwikkeling van vaardigheden, omdat niet duidelijk is hoe de vaardigheden zich zouden hebben ontwikkeld zonder deze interventies. In 2012 werd de zak-/slaagregeling aangepast en moesten leerlingen gemiddeld minimaal een 5,5 voor het CE halen. Deze eis werd in 2013 verder aangescherpt (kernvakkenregeling); er mocht voor havo/vwo voor de vakken Nederlands, Engels en wiskunde nog maximaal één 5 worden gehaald (voor SE en CE gezamenlijk). Vanaf 2014 moest in het vmbo minimaal een 5 voor Nederlands behaald worden om te kunnen slagen (SE plus CE). Deze wijzigingen hebben tot meer aandacht van scholen geleid voor het onderwijzen van de specifieke vaardigheden die nodig zijn voor een leerling om aan de exameneis te kunnen voldoen (waaronder 'teaching to the test'). Daarbij spelen examentrainingen die binnen en buiten de scholen steeds vaker worden aangeboden eventueel ook een rol. Scholen hebben zich meer op het CE gericht, waarbij het SE meer op het CE is gaan lijken.⁷ Het belang voor scholen en leerlingen van goede CE-uitslagen is duidelijk toegenomen, terwijl het belang voor en de motivatie van leerlingen voor de PISA-toets laag wordt ingeschat.

De vraag is of het beeld ook beïnvloed wordt door wijzigingen in de diverse leerlingenpopulaties die aan het examen hebben deelgenomen. Deze vraag kan niet eenduidig worden beantwoord vanuit de cijfers van het Cito-rapport omdat hier diverse tegengestelde effecten spelen. Wel valt het op dat in de jaren 2010-2012 scholen lijken voor te sorteren op de verscherpte exameneisen (met meer zittenblijvers en afstromers); daarna vallen deze effecten weg, mogelijk doordat men beter weet waar men aan toe is in de aangepaste situatie. Ook valt het op dat sinds 2015 het aantal leerlingen dat niet is aangemeld voor het centraal examen licht stijgt (met uitzondering van vmbo-bb waar het tot 2016 duidelijk daalt en daarna in 2017 weer toeneemt).

Voor uitgebreide informatie: paragraaf 2.2.1, hoofdstuk 3 tot paragraaf 3.5, hoofdstuk 5 en paragraaf 6.2 van het Cito-rapport.

Conclusie

De commissie komt in deze conclusie tot beantwoording van de hoofdvragen.

- I. Kunnen de kwaliteit van de examens en de ontwikkeling van de vaardigheden van scholieren optimaal worden vastgesteld en gevolgd?

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 normhandhavingmethoden 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.

⁷ Zie ook G.T.M. ten Dam, S. Waslander, A.A. Béguin & M.J.M. van Leeuwen, *Een volwaardig schoolexamen*, Commissie Kwaliteit Schoolexaminering 2018.

Wat betreft het systematisch monitoren van de ontwikkeling van de vaardigheden van VO-leerlingen is het Nederlandse CE een belangrijk instrument. Het is een toets van goede kwaliteit. Hoewel PISA ook een toets van goede kwaliteit is, heeft de commissie twijfel over de validiteit van PISA met betrekking tot het meten van vaardigheden. Het is een 'low stakes' toets en niet uitgesloten moet worden dat PISA hierdoor geen goed beeld geeft van vaardigheden van leerlingen. Ook moet worden benadrukt dat het CE en PISA verschillende vaardigheden van leerlingen in kaart brengen. Het feit dat de ontwikkeling van de resultaten van deze twee soorten toetsen van elkaar kan afwijken, zegt daarom niets over de vaardigheidsontwikkeling van de leerlingen in het algemeen. Uiteraard kan een daling op een van beide toetsen op zich gezien reden zijn om nader onderzoek te doen naar de kwaliteit van het aangeboden onderwijs dat de betreffende vaardigheden moet aanleren. Daarbij is het wel van belang rekening te houden met de motivatie van de leerlingen voor de toets en de leeftijd waarop de toets wordt afgenomen.

II. Wat valt er – met de nuance vanuit het antwoord op hoofdvraag I – te zeggen over de ontwikkeling van die vaardigheden in de afgelopen 10 jaar?

De ontwikkelingen rondom de examens in het voortgezet onderwijs en de afnames van PISA in de afgelopen tien jaar laten zien dat de resultaten op het CE stabiel zijn gebleven en op een aantal onderdelen zijn verbeterd. Het SE laat een stabiel beeld zien met een paar afwijkingen, meestal naar boven. Tegelijkertijd zijn de PISA-scores gelijk of (duidelijk) afgenomen, ook in vergelijking met de overige OESO-landen.

Onze conclusie is dan ook dat de vaardigheden van leerlingen qua niveau, zoals getoetst door het CE, in de afgelopen tien jaar gelijk gebleven zijn of zijn toegenomen. Dit wijst overigens niet op een toename van de vaardigheden over de volle breedte, want dan zouden ook de resultaten op het SE zijn toegenomen (alhoewel de validiteit van het SE niet van eenzelfde niveau is als het CE). Een vergelijking met de (deels gelijkblijvende en deels dalende) PISA-scores is niet mogelijk omdat het hier om een 'low stakes' toets gaat, deels afgenomen bij een andere leeftijdsgroep en in een andere fase van hun opleiding, en die een andere vraagstelling als basis heeft.

Wij als commissie komen tot de volgende verklaring voor de betere leerlingenresultaten op het CE. Wij baseren ons daarbij op het Cito-onderzoek en de gevoerde gesprekken en panels. Nader onderzoek kan dit beeld aanvullen of nuanceren. Overkoepelende verklaring is het toegenomen belang van het CE.

- De lat voor het slagen is door het beleid van de overheid (aanpassing zak-/slaagregeling) hoger komen te liggen (met sterk effect op de resultaten van het CE).
- De experts die wij spraken, gaven aan dat binnen de vakverenigingen van leraren de neiging bestaat om het belang van een vak af te meten aan de status als examenvak met een daarbij passende examensyllabus die – bij herziening - vaker zwaarder dan lichter wordt gemaakt.
- Het SE wordt te weinig ingezet om de eigenheid van het onderwijs van de school te honoreren maar functioneert steeds meer als een voorbereiding op het CE.⁸ Ook het bewaken van een maximale marge tussen het gemiddelde SE-cijfer en CE-cijfer per opleiding heeft hier vermoedelijk aan bijgedragen.⁹
- De focus op het examen en de examencijfers in het dagelijkse onderwijs in de bovenbouw van het VO is toegenomen, onder andere blijkend uit de toename van examentrainingen in schoolverband of op particulier initiatief ('teaching to the test').
- De examenresultaten zijn in de afgelopen jaren door de openbaarheid van schoolprestaties én het toezicht van de Inspectie van het Onderwijs (risicogericht vanaf 2007 met nadruk op de leeropbrengsten) op schoolniveau zwaar gaan wegen.
- Ouders en leerlingen hechten onverminderd belang aan het behalen van een diploma ten behoeve van de vervolgopleidingen van de eindexamenkandidaten. Dit wordt versterkt door het belang van goede cijfers bij loting en intake voor een deel van de opleidingen in het hoger onderwijs.

⁸ Zie de conclusies van de commissie Ten Dam in opdracht van de VO-raad van december 2018: G.T.M. ten Dam, S. Waslander, A.A. Béguin & M.J.M. van Leeuwen, *Een volwaardig schoolexamen*, Commissie Kwaliteit Schoolexaminering 2018.

⁹ Zie ook de analyse van de Onderwijsraad (*Toets wijzer. Naar een eigen(tijdse) wijze van toetsen en examineren*, 2018).

Tenslotte

Al met al zijn de resultaten behaald op het CE in de afgelopen tien jaar stabiel gebleven of verbeterd. De verbetering wijst op een toegenomen vaardigheidsniveau van de leerlingen; de leerlingen bezitten op een aantal gebieden meer kennis en vaardigheden binnen de gedefinieerde examendomeinen.

Tijdens dit onderzoek en de gesprekken met betrokken partijen heeft de commissie de stellige indruk gekregen dat de sterke focus op het examen, en de start van het SE in de voorexamenklassen, naast betere resultaten tegelijk een versmalling van het onderwijs heeft opgeleverd met een te grote nadruk op het behalen van goede toetsresultaten ('teaching to the test'). Ook de Onderwijsraad heeft recent op dit risico gewezen in haar advies "Toets wijzer" waarin ze spreekt van een grote nadruk op beslissende, centraal ingerichte en vooral kwantitatieve toetsing.

Het beeld waarbij het CE het onderwijs in de bovenbouw stevig in zijn greep heeft met het risico van verschraling werd door veel gesprekspartners bevestigd. Het belang van het CE als objectief ijkpunt voor het behaalde opleidingsniveau door leerlingen is niettemin evident. Onze vraag is echter of het hier een 'natuurlijke' afsluiting van een opleiding betreft of dat het examen tegenwoordig zijn schaduw zozeer vooruit werpt, dat het examen het bovenbouwonderwijs in het VO in hoge mate is gaan bepalen. Leraren, vakgroepen en schoolleiders lijken weinig ruimte te zien om het onderwijs in te richten naar hun eigen visie om zo bijvoorbeeld meer aandacht te besteden aan bredere vorming of vakoverstijgend werken. Hiervoor zou een andere invulling van het SE goede mogelijkheden bieden in lijn met hoe het SE bedoeld is. Ook zou er voor de scholen ruimte moeten zijn om meer aandacht te besteden aan de vaardigheidsontwikkeling van leerlingen op het terrein van meer alledaagse vraagstukken, zoals door PISA getoetst.

Wij willen u van harte aanbevelen om door middel van nader onderzoek in het VO na te gaan hoe groot het gewicht en de uitstraling is van het CE voor het dagelijkse onderwijs (inclusief de invulling van het SE) en - als dit gewicht erg groot is, zoals wij veronderstellen - te overwegen of dit consequenties zou moeten hebben voor de aard, de invulling en de omvang van het SE en het CE in het voortgezet onderwijs.

Namens de commissie,

Hoogachtend,



Drs. H.G.J. (Rick) Steur
Voorzitter

BIJLAGE: P. van der Molen, S. Schouwstra, R. Feskens & M. van Onna, *Vaardigheidsontwikkelingen volgens PISA en examens*, Cito 2019.