Algemene Bestuursdienst
Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

PO-Eindtoets 2018-2019

Review incident



Colofon
ABDTOPConsult

Muzenstraat 97
2511 WB DEN HAAG
www.abdtopconsult.nl

Dick Heerschop

In samenwerking met:
Judith van Dongen

ABDTOPConsult
Dichtbij en onafhankelijk

De consultants van ABDTOPConsult zijn lid van de topmanagementgroep (TMG) van de Algemene Bestuursdienst en worden benoemd door de Ministerraad. Ze zijn rijksbreed en interbestuurlijk inzetbaar voor interimopdrachten, projecten en onafhankelijke advisering bij complexe en (politiek) gevoelige zaken.

PO-Eindtoets 2018-2019

Review incident

25 juli 2019

Inhoud

Samenvatting		5
1	Inleiding en verantwoording	9
2	De setting: een duaal stelsel	11
2.1	Actoren	11
2.2	Sturing en toezicht	12
3	Implementatie: eindtoetsen in de praktijk	13
3.1	Proces: toetsafname in de praktijk	13
3.2	Kansengelijkheid en vergelijkbaarheid	14
3.2.1	Ankeropgaven	15
3.2.2	Brede toetsadviescategorieën	15
3.3	Proces ‘komen tot vergelijkbaarheid’	16
4	Het incident: Eindtoets PO 2018-2019	18
4.1	Waar ging het mis?	18
4.2	Damagecontrol	19
4.2.1	Onafhankelijke toetsing	19
4.2.2	Crisisaanpak	20
4.3	Impact	20
5	Tot slot, advies voor de toekomst	22
Bijlagen		23
Bijlage 1: Overzicht gehouden interviews		23

Samenvatting

Met het inwerkingtreden van de Wijzigingswet op het primair onderwijs zijn met ingang van het schooljaar 2014-2015 drie belangrijke zaken veranderd. De eindtoets is verplicht geworden, het schooladvies is leidend geworden bij de overgang naar het voortgezet onderwijs en de markt voor eindtoetsen is gestimuleerd. Naast een door de overheid aangeboden toets kunnen scholen ook een op de markt toegelaten toets van een andere partij kiezen (duale systeem).

Zowel OCW als het toenmalige College voor Examens (nu College voor Toetsen en Examens) waren in de aanloop naar het schooljaar 2014-2015 ingesteld op het uitvoeren van één centrale eindtoets. De politieke werkelijkheid maakte echter dat onvoorbereid voorzien moest worden in een markt van meerdere toetsaanbieders. Met veel inzet en moeite van alle betrokkenen is een proces ingericht om tot een markt van kwalitatief goede en vergelijkbare eindtoetsen te komen. Gesignaleerde knelpunten zijn gaandeweg pragmatisch opgelost. Het heeft echter bij aanvang ontbroken aan een analyse voor deze specifieke situatie en het opstellen van een implementatieplan.

Wel is bij aanvang de Expertgroep geformeerd en is er een inhoudelijk werkoverleg tot stand gekomen met alle betrokken partijen waar stap voor stap en jaar na jaar met veel energie en betrokkenheid is en wordt gewerkt aan noodzakelijke verbeteringen. Dit inhoudelijk gedreven overleg heeft echter onvoldoende oog voor planning en voortgangsbewaking en kent onvoldoende verbinding met de managementagenda die bij zo'n grote wijziging als paraplu voor samenhang kan zorgen.

Het bovenstaande schetst een beeld van een kwetsbaar proces met kans op fouten en onvoorziene gebeurtenissen. Het incident met de eindtoets 2019 toont ook aan dat de huidige governance kwetsbaar is. OCW heeft vastgesteld dat de oorzaak van het incident ligt bij een interpretatiefout van de Expertgroep. De Expertgroep heeft het "brede adviescategorie-scenario" anders geïnterpreteerd voor de private toetsaanbieders dan dat CvTE heeft gedaan voor de CET en is besproken in het afstemmingsoverleg. Hierdoor is in de vertaling van de cesuren van de CET naar de andere eindtoetsen een verschil ontstaan bij het vaststellen wanneer een leerling van de een naar de andere toetscategorie over gaat. Dit heeft als gevolg gehad dat bij de private eindtoetsen leerlingen in de meeste gevallen een hoger toetsadvies hebben ontvangen dan de leerlingen die de CET hebben gemaakt.

In de door ons uitgevoerde review hebben wij geen andere verklaringen gevonden voor de oorzaken van het incident, dan de door OCW aangedragen en gecommuniceerde verklaringen.

Wel is ons opgevallen dat betrokkenen niet eenduidig de exacte fout (interpretatiefout, afleesfout, verkeerd gebruik van een tabel) kunnen duiden. Het ontbreken van transparantie in zowel het gevolgde proces als de onderliggende berekeningen van de Expertgroep en Cito is hier mede oorzaak van. De onderliggende berekeningen worden onder meer vanwege de concurrentiegevoeligheid niet uitgewisseld tussen CvTE en Cito enerzijds en de Expertgroep en de private toetsaanbieders anderzijds.
De communicatie vanuit OCW ten tijde van het incident, de wijze waarop OCW regie heeft gevoerd en het vervolgproces heeft ingericht zijn toereikend geweest en is ook door alle betrokkenen gewaardeerd.

Ten behoeve van de komende jaren (2019-2021) doen wij de volgende aanbevelingen ter verbetering van de huidige governance en het aanbrengen van een aantal proceswaarborgen om de kans op herhaling van soortgelijke incidenten te verkleinen.

Versterk de governance

Het ontbreken van integrale regie over de partijen heen en onvoldoende waarborgen voor kwaliteitsbewaking zijn debet aan de kwetsbaarheid van de huidige governance.

Aanbeveling 1: Duale opzet vraagt om integrale regie
Zorg bij het bestaande afstemmingsoverleg voor een geformaliseerde procesgang met bijbehorende planning en heldere besluitvormingsmomenten met adequate vastlegging.
Leg het mandaat voor de essentiële besluitvorming bij het bovenliggende managementniveau, dus bij OCW bij de directeur PO. Richt daartoe een regietafel in met als deelnemers het verantwoordelijk management van de betrokken partijen en de voorzitter van de Expertgroep. Beschouw het afstemmingsoverleg als het voorbereidende gremium voor deze regietafel.

Extra kwaliteitsborging

Aanbeveling 2: Kwaliteitsbewaking door scheiding van werkzaamheden
Laat, als onderdeel van voornoemde besluitvorming, de resultaten van zowel de Expertgroep als die van het CvTE controleren door een onafhankelijke derde partij, bijvoorbeeld RCEC. Voer deze controle uit voordat de resultaten met scholen en leerlingen worden gecommuniceerd.

Wij adviseren om de Expertgroep niet alleen de resultaten van de vier externe aanbieders te laten toetsen maar ook die van de CET. Omgekeerd kan CvTE naast de CET-resultaten ook de resultaten van de vier externe aanbieders controleren.
Zowel CvTE als de Expertgroep beschikken dan over de volledige dataset wat de betrouwbaarheid van de berekeningen ten goede komt.

Om de benodigde tijd voor het uitvoeren van de genoemde controles te creëren, dient er te worden gekozen voor één centraal moment waarop alle toetsaanbieders de uitslagen naar de scholen communiceren.

Aanbeveling 3: organiseer interne kwaliteitscontrole binnen de Expertgroep

De werkwijze van de Expertgroep Toetsen PO is kwetsbaar. Deze onafhankelijke adviescommissie is in het leven geroepen om jaarlijks de kwaliteit van de eindtoetsen te beoordelen en de minister te adviseren over toelating en/of intrekking van private eindtoetsen. De Expertgroep bestaat uit een voorzitter en twee experts op het gebied van psychometrie en onderwijskunde. In de afgelopen vier jaar heeft de Expertgroep er taken bij gekregen die niet in het instellingsbesluit zijn opgenomen.

Laat de voorzitter van de Expertgroep zorg dragen voor een interne kwaliteitscontrole binnen de Expertgroep, voordat de resultaten naar andere partijen gaan. Met de, naar wij begrepen, inmiddels goedgekeurde personele uitbreiding kan dat eenvoudig met een ‘vier ogen principe’ worden vormgegeven.

Rond het proces t.a.v. de Eindtoets 2019 goed af

Uit de interviews en de second en third opinion blijkt dat er nog een aantal losse eindjes zijn die nadere uitwerking behoeven.

Aanbeveling 4: Doe alsnog een finale check op juistheid van resultaten eindtoets 2019

Laat RCEC onderzoek uitvoeren naar de berekening van de resultaten over 2019 nu de leerlingpopulatie compleet is. In de uitgevoerde second opinion was de dataset incompleet, omdat een aantal private toetsaanbieders nog data van de inhaalperiode zouden verkrijgen. Op deze wijze kan zekerheid worden verkregen dat onderliggende berekening van zowel Cito als de Expertgroep correct is. Dit als finale check op de juistheid van de berekening ten behoeve van de eindtoets 2019.

Aanbeveling 5: laat een objectieve analyse van gebruikte programmatuur uitvoeren

Zowel CvTE (via CITO) als de Expertgroep maken gebruik van softwaretoepassingen die in eigen beheer zijn ontwikkeld en waarin de rekenregels als algoritmen zijn opgenomen. Wij hebben niet onderzocht of deze softwareproducten voldoen aan gangbare criteria met betrekking tot testbaarheid, documentatie, beveiliging, AVG etc. Aangezien op de uitkomsten moet kunnen worden vertrouwd bevelen wij een onafhankelijke analyse van de software aan.

Aanbeveling 6: Communicatie OCW continueren

Er is veel waardering voor de rol van OCW in de afhandeling van dit incident.

Wij constateren dat er behoefte is, bij de private toetsaanbieders en scholen, aan meer duidelijkheid over de stand van zaken naar aanleiding van het incident en wat er nu gaat gebeuren om herhaling te voorkomen.

Stuur ter ondervanging van mogelijke imagoschade voor de private toetsaanbieders (waarom waren de uitslagen van de CET goed en van de private toetsen fout) vanuit OCW bij de start van het schooljaar naar de scholen en andere belanghebbenden een informerende brief over de stand van zaken n.a.v. incident en welke maatregelen daaruit voortvloeien om herhaling te voorkomen. In de herfst maken schoolbesturen de jaarlijkse keuze voor een eindtoets bij inschrijving in december-januari.

1 Inleiding en verantwoording

Plusminus 11% van de leerlingen uit groep 8 heeft een onjuist toetsadvies op de eindtoets 2019 gekregen. Er is een fout ontstaan bij het vertalen van de behaalde score van een leerling op de eindtoets naar een toetsadvies voor het best passende brugklastype. Dit onjuiste toetsadvies is in de meeste gevallen te hoog uitgevallen. Het betreft de leerlingen die de eindtoets afkomstig van de vier private toetsaanbieders hebben gemaakt, niet de leerlingen die de Centrale Eindtoets (CET – CvTE/Cito) hebben gemaakt.

OCW heeft vastgesteld dat de toetsaanbieders niets te verwijten valt, dat de (menselijke) fout ligt bij de onafhankelijke commissie Expertgroep Toetsen PO (**hierna:** Expertgroep). Deze is belast met de beoordeling van de kwaliteit van de commerciële eindtoetsen en daaropvolgend advisering aan de minister over de toelating van eindtoetsen, evenals het geven van een eigenstandig kwaliteitsoordeel over de inhoudelijke validiteit, betrouwbaarheid en normering van de toetsen in leerlingvolgsystemen. In het proces om tot de vergelijkbaarheid van de eindtoetsen te kunnen komen, is een fout gemaakt. Na een herberekening die door twee externe partijen is gecontroleerd, dit om zowel zekerheid over de cijfers te verkrijgen als een inschatting te kunnen maken van de omvang van het probleem, hebben de toetsaanbieders alsnog de correcte toetsadviezen aan de scholen gecommuniceerd. Van de 19.854 vastgestelde incorrecte adviezen is het overgrote deel (18.288) te hoog uitgevallen. 1566 leerlingen hebben een te laag toetsadvies gekregen[1].

Opdracht

Op verzoek van de directeur-generaal Primair en Voortgezet Onderwijs hebben wij een review uitgevoerd op de bovengenoemde bevindingen van OCW ten aanzien van de oorzaak van het incident. Daarnaast is ons verzocht om te komen tot aanbevelingen om kwetsbaarheden in de huidige (duale) inrichting te kunnen voorkomen in de vervolgjaren.

Aanpak

Wij hebben kennis genomen van de volgende documentatie; recente kamerbrieven, verslagen van het afstemmingsoverleg, rapportages en adviezen m.b.t. de vergelijkbaarheid van toetsen, het komen tot ankeropgaven en gemeenschappelijke cesuren, evenals de second en third opinion op de herberekening. Tevens hebben wij de rapportages van de Expertgroep ontvangen en signaleringsberichten van betrokkenen aan OCW. Daarnaast hebben wij interviews gehouden met de betrokken actoren, zie Bijlage 1 voor een overzicht van de geïnterviewden.

———————

[1] Kamerbrief, “Onjuiste toetsadviezen deel leerlingen bij afname eindtoets 2019”, 21 mei 2019, pag. 3

Op basis hiervan hebben wij een beschrijving van de huidige inrichting en werkwijze opgenomen in H2 en H3. Onze bevindingen en aanbevelingen zijn verwoord in H4 en H5.

Afbakening

Dit onderzoek is gericht op de praktische werking van de huidige duale inrichting. De nieuwe governance die (op zijn vroegst) wordt voorzien vanaf het schooljaar 2021-2022 valt buiten scope. Gedurende de looptijd van dit onderzoek heeft de Minister van OCW zijn visie voor de toekomst van de doorstroomtoets in het primair onderwijs gepresenteerd en tevens een aantal waarborgen in de aanloop daartoe kenbaar gemaakt om verdere incidenten te voorkomen. Dit is ook veelvuldig in interviews ter sprake gekomen. Wij hebben in de slotparagraaf suggesties opgenomen die verder reiken dan de huidige duale inrichting.

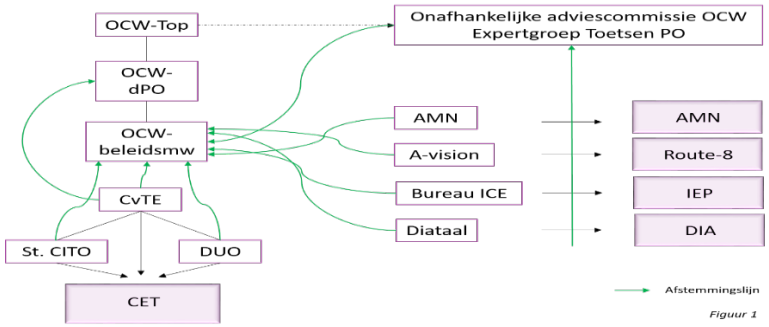
2 De setting: een duaal stelsel

Met ingang van het schooljaar 2014-2015 is de wijzigingswet Eindtoetsing PO van kracht. Deze wet regelt onder meer dat alle leerlingen in groep 8 verplicht een eindtoets moeten maken. De eindtoets geeft een objectief beeld van het kennisniveau van de leerling in relatie tot de referentieniveaus Nederlandse taal en rekenen. Het toetsadvies van de eindtoets fungeert als objectief gegeven naast het advies van school over het best passende instroomniveau op het voortgezet onderwijs.
Hierbij is het schooladvies leidend, dit wordt voorafgaand aan de afname van de toets reeds bekend gemaakt. Leerlingen worden op basis van het schooladvies toegelaten op het voortgezet onderwijs. Indien het toetsadvies hoger is dan het schooladvies is de school verplicht om het eigen advies te heroverwegen, andersom is niet toegestaan.

Tijdens de behandeling van deze wet in de Tweede Kamer is bij amendement nadrukkelijk gekozen voor keuzevrijheid voor scholen om voor een toets te kunnen kiezen die bij hen past. Dit ter voorkoming van een verplichte afname van de Centrale Eindtoets (**hierna:** CET), die namens de overheid onder verantwoordelijkheid van het College voor Toetsen en Examens (**hierna:** CvTE) en in samenwerking met Stichting Cito en Dienst Uitvoering Onderwijs (**hierna:** DUO) is ontwikkeld en wordt uitgevoerd. Daarbij heeft de zorg van de Tweede Kamer ten aanzien van het verder versterken van de monopolypositie van Cito ook een rol gespeeld[2]. Zodoende is het duale systeem ontstaan; een hybride stelsel waarbij wordt gewerkt met zowel een publieke overheidstoets (de CET) en eindtoetsen van private partijen. Inmiddels kiest iets meer dan de helft van de scholen voor een private eindtoets in plaats van de CET. De IEP-eindtoets en Route-8 hebben daarbij het grootste afnamebereik.

2.1 Actoren

Zowel OCW als CvTE (het toenmalige College voor Examens) hebben aangegeven destijds niet voorbereid te zijn geweest op een markt voor eindtoetsen. Ten behoeve van het realiseren van het duale systeem is toentertijd in rap tempo een governance ingericht die de kwaliteitsborging moest zeker stellen. Deze is als volgt vormgegeven, zie figuur 1.



Figuur 1

[2] Kamerstuk 33 157, nr. 38, "Gewijzigd amendement van de leden Schouten en Rog".

In dit duale systeem is CvTE verantwoordelijk voor de kwaliteitsborging van de publieke eindtoets CET. De toelating en kwaliteitsborging van eindtoetsen van private partijen is belegd bij een onafhankelijke adviescommissie, de Expertgroep Toetsen PO. De Expertgroep bestaat uit een voorzitter en twee experts op het gebied van psychometrie en onderwijskunde. Deze beoordeelt jaarlijks de kwaliteit van de eindtoetsen en adviseert de minister over toelating en/of intrekking van private eindtoetsen.

Stichting Cito verdient in dit kader een nadere toelichting. Voor 1999 was het als centraal toetsinstituut een publiekrechtelijke instelling. In 1999 volgt op verzoek van de overheid een privatisering, waarna in 2004 wordt gekozen voor een hybride organisatiestructuur. Stichting Cito vervult de wettelijke en publiekrechtelijke taken. Cito B.V. ontwikkelt commerciële producten voor het onderwijs en het bedrijfsleven die in de markt worden gezet (bv. leerlingvolgsystemen etc.). Deze verdeling zorgt onder meer voor het scheiden van subsidiestromen aan de stichting en marktinkomsten van de B.V. In dit rapport hebben wij het uitsluitend over Stichting Cito (**hierna:** Cito).

2.2 Sturing en toezicht

Het departement heeft de kwaliteitsbewaking en regie op onderwijsinhoudelijke, organisatorische en psychometrische kwaliteit in het duale stelsel op afstand geplaatst en bij twee partijen belegd. Enerzijds bij de Expertgroep voor de toetsen van de private aanbieders en anderzijds bij het CvTE voor de publieke CET.

Er is geen formele afstemming georganiseerd tussen CvTE, Expertgroep en OCW. Wel is onder voorzitterschap van OCW een zes wekelijks breed (vooral inhoudelijk) afstemmingsoverleg ingericht. Naast de bovenstaande actoren hebben tevens de Inspectie van het Onderwijs (**hierna:** IvhO), de PO-en VO-Raad (sectororganisaties voor primair en voortgezet onderwijs) en maatschappelijke partners daarin zitting.

Op het niveau van beleidsmedewerkers wordt door OCW één maal per jaar een gesprek met de voorzitter van de Expertgroep gevoerd op basis van het jaarverslag. Er wordt niet inhoudelijk getoetst op de kwaliteit van het geleverde werk. Binnen de Expertgroep is geen proces ingericht om de kwaliteit van het geleverde werk nog apart te controleren. Ook wordt de IvhO niet ingeschakeld om hier uitspraken over te doen. De inspectie geeft aan dit ook niet als onderdeel van haar taak te zien.
Wel heeft de inspectie een rol bij het toezicht op de CvTE. Als toezichthouder is het haar taak om het functioneren van het CvTE jaarlijks te beoordelen.

3 Implementatie: eindtoetsen in de praktijk

Zowel OCW als het toenmalige College voor Examens (nu College voor Toetsen en Examens) waren in de aanloop naar het schooljaar 2014-2015 ingesteld op het uitvoeren van één centrale eindtoets. De politieke werkelijkheid maakte echter dat onvoorbereid voorzien moest worden in een markt van meerdere toetsaanbieders.

De door OCW ingestelde Expertgroep beoordeelt bij aanmelding op de markt de eindtoets van de marktaanbieder, op basis van een beoordelingskader dat mede is gebaseerd op de door CvTE vastgestelde Algemene Toetswijzer. Deze beoordeling omvat de onderwijskundige en organisatorische kwaliteit van de eindtoets. Tevens wordt door COTAN (Commissie Testaangelegenheden Nederland) een psychometrische beoordeling uitgevoerd. Op basis van deze beoordelingen adviseert de Expertgroep de minister over toelating van de betreffende eindtoets op de markt. Toetsen worden voor vier jaar toegelaten op de markt, mits zij een jaarlijks positieve beoordeling toegekend krijgen van de Expertgroep. In het schooljaar 2014-2015 worden naast de CET (beoordeeld door CvTE) twee private toetsaanbieders toegelaten; Route 8 en de IEP-eindtoets. In het daaropvolgende jaar geldt dat ook voor Diataal en AMN.[3]

In diverse interviews is opgemerkt dat er met name bij aanvang kwalitatief grote verschillen tussen de eindtoetsen van de private toetsaanbieders zitten. Om aan de politieke wens tot het komen van een markt te kunnen voldoen, is in meerdere interviews aangegeven dat de Expertgroep potentiële toetsaanbieders aan de voorkant van het toelatingsproces is gaan adviseren. Dit om te kunnen komen tot een positief advies voor toelating bij beoordeling op de criteria.

3.1 Proces: toetsafname in de praktijk

Uit onder meer de verslagen van het afstemmingsoverleg is te herleiden hoe de betrokkenen zijn gekomen tot een uniform afnameproces van de verschillende eindtoetsen. Dat ziet er als volgt uit:

- Scholen zijn wettelijk verplicht voor 1 maart alle leerlingen te voorzien van een schooladvies. Op basis daarvan kunnen leerlingen zich ook inschrijven voor het voortgezet onderwijs.
- De afname van de eindtoets vindt plaats in de periode 15 april tot 15 mei. Sommige aanbieders hanteren nog een inhaalperiode nadien.
- Alle eindtoetsresultaten worden direct door de private toetsaanbieders na afname naar de Expertgroep gestuurd. Vanaf de eindtoets 2019 gebeurt dat via een door de Expertgroep ontwikkelde software applicatie genaamd Lexter.

———

[3] De eindtoets CESAN wordt toegelaten voor 2015-2016 en 2016-2017. Wegens het niet behalen van de minimale afnamepopulatie van 1000 leerlingen om in aanmerking te komen voor subsidie trekt CESAN zich in het daaropvolgende jaar terug. Derhalve hebben wij CESAN niet meegenomen in dit onderzoek.

- De Expertgroep checkt per toetsaanbieder met behulp van ankeropgaven of de scores op de referentieniveaus consistent zijn c.q. of de normering klopt.
- De Expertgroep adviseert de toetsaanbieders over de normering van de referentieniveaus, de weging van de diverse onderdelen en de cesuren van de toetsadviescategorieën.
- De toetsaanbieders rapporteren voor eind mei aan de scholen over de resultaten (standaardscores, behaalde referentieniveaus en het toetsadvies). Hier worden door de toetsaanbieders verschillende termijnen aangehouden. Een van de toetsaanbieders is bijvoorbeeld in staat om binnen 5 werkdagen na afname de uitslagen terug te rapporteren.
- Scholen bekijken het toetsadvies in relatie tot het eerder gegeven schooladvies. Heroverweging is verplicht indien het toetsadvies hoger is dan het schooladvies.
- Beide adviezen worden door de scholen geregistreerd in BRON (van DUO) ten behoeve van stelselinformatie voor de IvhO en de sturingsinformatie ten behoeve van de scholen zelf. Kanttekening hierbij is dat BRON vervuiling kent omdat een heroverwogen schooladvies niet altijd door scholen wordt aangepast in BRON of op een verkeerde plek wordt geregistreerd in BRON.
- Het voortgezet onderwijs baseert zich voor toelating van een leerling op het schooladvies, dat is immers leidend. Het voortgezet onderwijs kan het toetsadvies gebruiken voor het bepalen in welke brugklas een leerling wordt geplaatst.

3.2 Kansengelijkheid en vergelijkbaarheid

Kansengelijkheid, het beginsel dat iedereen met dezelfde talenten dezelfde kans zou moeten kunnen maken om deze verder te ontwikkelen onafhankelijk van sociaal-economische status, sekse etc., is één van de ambities van onder meer dit kabinet. Voor kinderen is het van belang dat het toetsadvies onafhankelijk is van het type eindtoets dat wordt afgenomen. Meerdere typen toetsen betekent echter dat de onderlinge vergelijkbaarheid niet automatisch is gegarandeerd, zelfs niet bij een uniform afnameproces.

De Expertgroep constateert in de eerste afnamejaren dat, ondanks de verplichting dat alle toetsen taal-en rekenonderdelen bevatten die gerelateerd zijn aan de verplichte referentieniveaus, de resultaten van de toetsen verschillen met betrekking tot de behaalde referentieniveaus[4]. Dit als gevolg van een verschillende ijking aan de referentieniveaus. Een zorgelijke ontwikkeling, gezien de impact op de legitimiteit van de toetsen en de kansengelijkheid. Deze zorg heeft ook de IvhO aan OCW kenbaar gemaakt[5]. De IvhO benut de resultaten van de eindtoets voor het waarborgen van de kwaliteit van het onderwijs op het niveau van scholen en het stelsel. De voornaamste zorg van de Inspectie is dat zonder de onderlinge vergelijkbaarheid van de toetsen, de resultaten onvoldoende betrouwbaar zijn.

———————

[4] Expertgroep, rapportage vergelijking eindtoetsen, december 2016.
[5] Nota, zicht op onderwijsresultaten primair onderwijs, IvhO, 19-10-2018

De Expertgroep adviseert in december 2016 om de psychometrische verantwoording van de eindtoetsen te uniformeren ten bate van de rapportages, de gebruikte statistische methoden en de daarbij behorende software (de zogeheten blue print voor toetsverantwoording). De Expertgroep adviseert om een anker[6] in te voeren die zorgt voor een voldoende vergelijkbare normering op de referentieniveaus[7]. Evenals het aanhouden van gelijke adviescategorieën in de eindtoetsen en het daarvoor maken van afspraken over de onderliggende cesuren (grenswaarden) van de onderscheiden onderwijsniveaus.[8] Uniforme normering vergroot de validiteit van de eindtoetsen, waarop de Expertgroep alle eindtoetsen mede beoordeelt, evenals de onderlinge vergelijkbaarheid.

3.2.1 Ankeropgaven

In samenwerking met OCW, CvTE, Cito en de private toetsaanbieders worden door de Expertgroep ankeropgaven ontwikkeld om de ijking aan de referentieniveaus te uniformeren. Als gevolg van dit gezamenlijke anker kunnen alle aanbieders qua normering op één schaal worden geplaatst[9], waardoor de resultaten onderling kunnen worden vergeleken. De Expertgroep heeft vervolgens op basis van het gemeenschappelijk anker en data afkomstig uit eerdere toetsen van de private toetsaanbieders de nieuwe cesuren voor referentieniveaus berekend. Voor de toets van 2018 zijn de private toetsaanbieders vrij geweest om het advies over de te hanteren cesuren voor de referentieniveaus van de Expertgroep al dan niet over te nemen. Dit resulteerde echter daardoor in verschillen ten aanzien van de referentieniveaus, wat de vergelijkbaarheid van toetsen beïnvloedde. Voor de toets van 2019 geldt dan ook een verplichting tot overnemen van het advies.

3.2.2 Brede toetsadviescategorieën

Met dezelfde doelstelling, het vergroten van de onderlinge vergelijkbaarheid van de afnameresultaten, wordt op verzoek van OCW onderzoek gedaan naar een herindeling van de toetsadviescategorieën[10], die moet leiden tot de meest optimale schooladviezen die de meeste kansen geven voor een leerling. Er wordt geconstateerd dat Toetsaanbieders verschillende toetsadviescategorieën hanteren[11]. Daarbij worden naast de verplichte ankeropgaven ook andere kennis en vaardigheden getest die door de toetsaanbieders worden meegewogen. De weging van de diverse onderdelen kan daarbij verschillen per toetsaanbieder, wat eveneens van invloed is op de resultaten van de verschillende toetsen en de vergelijkbaarheid.

———————

[6] Ankeropgaven: een gezamenlijke set van opgaven voor lezen, rekenen en taalverzorging die alle toetsaanbieders in de eigen eindtoets moeten opnemen.
[7] Referentieniveaus: omschrijven wat leerlingen moeten kennen en kunnen voor rekenen en taal. Er wordt onderscheid gemaakt in fundamentele niveaus en streefniveaus.
[8] Expertgroep, rapportage vergelijking eindtoetsen, december 2016, pag. 3
[9] Een procedure om scores van verschillende toetsen op dezelfde schaal te brengen.
[10] Toetsadviescategorie: het resultaat van de eindtoets dat een specifiek instroomniveau voor het voortgezet onderwijs beschrijft, bijvoorbeeld havo/vwo of vwo.
[11] Expertgroep, rapportage vergelijkbaarheid eindtoetsen en invoering van gezamenlijk anker, december 2018, pag. 9-14

Cito heeft, in opdracht van OCW, vijf scenario's opgesteld op basis van haar eigen doorstroomonderzoek[12] om te komen tot een uniform model voor de eindtoets. In de gezamenlijkheid van het afstemmingsoverleg is besloten tot het terugbrengen van tien toetsadviescategorieën naar zes; het "brede categorieën-scenario". Dat wil zeggen vijf meervoudige adviescategorieën, te weten pro/bb, bb/kb, kb/gt, gt/havo, havo/vwo en de enkelvoudige toetsadviescategorie vwo. Het grootste discussiepunt bij dit besluit betrof het verdwijnen van de enkelvoudige praktijkonderwijs- en de havo categorie. Doorslaggevend bij de keuze voor dit scenario is de positieve bijdrage aan kansengelijkheid en de stimulans voor scholen om ook in brede categorieën het school- en eindadvies te geven.

Het gekozen scenario c.q. model is toegepast in de eindtoets 2019 voor de normering van de toetsadviezen. Dit model, gebaseerd op de normering van Cito voor de CET, heeft de Expertgroep vervolgens doorgerekend om de effecten op de toetsadviezen van de private toetsaanbieders inzichtelijk te maken. Daarnaast heeft de Expertgroep zich gebogen over de vraag waar de cesuren van de toetsadviescategorieën van de private toetsen komen te liggen bij een verschuiving van tien naar zes brede toetsadviescategorieën. Oftewel wanneer gaat een leerling "over" van bijvoorbeeld gl/havo naar havo/vwo. Die achterliggende berekening gaat uit van een vertaling van de cesuren van de zes toetsadviescategorieën van de CET naar de cesuren voor de private toetsaanbieders. Hierbij is rekening gehouden met de weging van de verschillende toetsonderdelen ten einde deze voor alle toetsen te equivaleren. Dit is vastgelegd in de "Rapportage vergelijkbaarheid eindtoets en invoering van het gezamenlijk anker", van de hand van de Expertgroep en gedateerd op december 2018.

3.3 Proces 'komen tot vergelijkbaarheid'

Terugkijkend op het proces om te komen tot vergelijkbaarheid van de diverse eindtoetsen, wordt unaniem door ondervraagden aangegeven dat dit op constructieve wijze en in goede samenwerking is gebeurd. Wel wordt opgemerkt door de private toetsaanbieders dat er voor CvTE en Cito andere procedures en afspraken gelden, waarop zij nauwelijks zicht hebben. Ditzelfde geldt ook voor inzicht in de onderliggende berekeningen van de Expertgroep en CvTE met betrekking tot de benutte modellen om te komen tot vergelijkbaarheid.

Om tot de vergelijkbaarheid te komen heeft OCW diverse onderzoeken laten uitvoeren door zowel de Expertgroep als Cito. Die onderzoeken liggen aan de basis van essentiële besluitvorming die heeft plaatsgevonden gedurende dit proces.

———————

[12] Doorstroomonderzoek: brengt in kaart op welk niveau leerlingen drie jaar na afname van de CET functioneren en de correlatie met het gegeven toetsadvies.

Het brede afstemmingsoverleg wordt door de meesten als inhoudelijk nuttig ervaren, zij het zeer informeel en weinig zakelijk. Daarnaast heeft de Expertgroep inhoudelijke werkgroepen opgezet om in gezamenlijkheid met experts uit het veld te kunnen komen tot bijvoorbeeld de ankeropgaven.

Voor alle betrokken toetsaanbieders geldt dat iedereen afhankelijk is van de goede wil en het commitment van collega-toetsaanbieders die zich in een concurrerende markt bevinden.

4 Het incident: Eindtoets PO 2018-2019

4.1 Waar ging het mis?

Om na de exercitie van tien-naar-zes toetsadviescategorieën te komen tot het vaststellen van de cesuren van de zogenoemde zes brede toetsadviescategorieën maakt de Expertgroep een vertaling van de cesuren zoals Cito deze gebruikt voor de zes toetsadviescategorieën van de CET. Er vindt onderling mailverkeer plaats. Op 14 januari 2019 wordt in Arnhem op hoofdlijnen het advies besproken, zoals geformuleerd in de eerder genoemde "Rapportage vergelijkbaarheid eindtoetsen en invoering van het gezamenlijk anker". De rapportage laat echter maar beperkt de onderliggende berekeningen zien achter het gekozen model noch inzicht in de wijze waarop de cesuren van CET tot stand zijn gekomen die aan de basis van het model liggen. Cito verschaft daartoe in maart een notitie op hoofdlijnen. De onderliggende berekeningen blijven in deze notitie buiten beschouwing.

In het rapport is door de Expertgroep een doorrekening uitgevoerd op de gegevens van de Eindtoets 2018. De private toetsaanbieders signaleren dat er verschuivingen optreden bij het nieuw gekozen model ten aanzien van de percentages van de verschillende toetsadviescategorieën bij nagenoeg alle toetsen[13]. Dit laatste blijkt niet uit de officiële verslaglegging van het afstemmingsoverleg maar wordt wel in diverse interviews benadrukt. De verklaring daarvan wordt gezocht in het gebruik van de brede toetsadviescategorieën waardoor er verschuivingen op zijn getreden.

Direct na de afname van de toets wordt er door A-vision (voert als eerste partij medio april een toetsafname uit) aan de bel getrokken. Scholen hebben zich gemeld bij A-vision omdat de toetsadviezen over het algemeen beduidend hoger liggen dan het eerder gegeven schooladvies. Navraag bij de Expertgroep leidt tot bevestiging dat de juiste gegevens zijn gebruikt. De verschillen worden wederom verklaard door de verschuiving van tien-naar-zes toetsadviescategorieën ("brede categorieën-scenario"). Als vervolgens verderop in april en begin mei met uitzondering van Cito alle martkpartijen, gealarmeerd door scholen, zich melden bij de Expertgroep en OCW wordt een en ander uitgezocht en geconstateerd dat er een fout is gemaakt door de Expertgroep bij de berekening van de cesuren voor de toetsadviescategorieën van de private toetsen die alle vier de private toetsaanbieders raakt.

Uit de verslaglegging blijkt dat de doorrekening van het model en de voorgestelde cesuren in het eerste kwartaal van 2019 verschillende malen worden besproken, maar dat leidt niet tot aanpassing van de cesuren.

———————

[13] Zoals geconstateerd kan worden bij vergelijking van tabellen 7 en 8 in de "Rapportage vergelijkbaarheid eindtoetsen en invoering van gezamenlijk anker", Expertgroep, december 2018.

OCW heeft na afloop van het incident vastgesteld dat de Expertgroep het “brede categorieën-scenario” anders heeft geïnterpreteerd dan CvTE voor de CET heeft gedaan, waardoor de tabel waarin de vertaling van de cesuren van de CET naar de private eindtoetsen verkeerd door de Expertgroep is toegepast. Uit de interviews blijkt dat CvTE noch Cito op voorhand hebben gekeken naar de doorrekening en de voorgestelde nieuwe cesuren, simpelweg vanwege het feit dat dit niet hun toets betrof. Uit de interviews wordt duidelijk dat alle betrokkenen zich ervan bewust zijn geworden dat er geen controle heeft plaatsgevonden op de door de Expertgroep uitgevoerde doorvertaling naar de cesuren van de marktaanbieders.

Ons is opgevallen dat de betrokkenen verschillende opvattingen hebben bij de precieze toedracht van de fout. Betrokkenen kunnen niet eenduidig de exacte fout (interpretatiefout, afleesfout, verkeerd gebruik van een tabel) benoemen. Ditzelfde geldt ook voor het antwoord op de vraag of er al dan niet een controle door CvTE of Cito is uitgevoerd op de berekeningen van de Expertgroep. Het ontbreken van transparantie in zowel het gevolgde proces als de onderliggende berekeningen van de Expertgroep en Cito is hier mede oorzaak van.

4.2 Damagecontrol

Op het zogeheten “crisisoverleg” op 14 mei wordt gezamenlijk geconstateerd dat de Expertgroep een andere interpretatie heeft gehanteerd van het brede advies scenario dan Cito. Ook wordt geconstateerd bij toelichting op de werkwijze van de CET, dat voor de berekening van de toetsadviezen van de CET de cesuren zijn gehanteerd zoals op basis van het doorstroomonderzoek in hetzelfde gremium is besproken. Met als conclusie dat de cesuren voor de private toetsaanbieders afwijken en daartoe hogere toetsadviezen tot gevolg hebben dan de toetsadviezen van de CET.

Er wordt in hetzelfde overleg besloten tot een herberekening van de normering van de vier private eindtoetsen. Nu echter aan de hand van de verdeling van toetsadviezen van de CET als referentie voor de normering.[14]

4.2.1 Onafhankelijke toetsing

Om inzicht te krijgen in de ordegrote van het probleem (hoeveel leerlingen per toets zijn getroffen door de fout) en absolute zekerheid te krijgen over de herberekening, verzoekt OCW zowel Cito als RCEC (Research Centre voor Examinering en Certificering) om de herberekening te controleren. Cito vraagt daartoe alle data op bij de andere in de markt concurrerende partijen, wat ongemakkelijk voelt voor de private partijen maar door allen als noodzakelijk wordt beschouwd in deze situatie. Hierdoor kan Cito een volledige herberekening uitvoeren en constateren dat deze correct is uitgevoerd. RCEC heeft met name de procesgang bekeken; wat ging mis, hoe is het opgelost en hoe wordt de oplossing gekwalificeerd.

———————

[14] Verslag breed afstemmingsoverleg, d.d. 14 mei 2019

De gebruikte programmatuur is daarbij niet beoordeeld. RCEC deelt de conclusie dat het probleem mede door de onduidelijke procesgang is opgetreden.
Er heeft geen controle plaatsgevonden op de onderliggende berekeningen van de vertaling van de cesuren van de CET naar die van de private eindtoetsen.

Wij concluderen dat het “brede categorieën-scenario” door de Expertgroep bij haar berekeningen anders is geïnterpreteerd dan CvTE/Cito voor de CET heeft gedaan. Hierdoor is de tabel, die het model van de brede categorieën en de bijbehorende cesuren weergeeft voor de private eindtoetsen, door de Expertgroep anders toegepast dan dat CvTE/Cito voor de CET heeft gedaan. Door het ontbreken aan checks and balances in het proces is dit niet tijdig gesignaleerd, ondanks dat de resultaten van de berekening zoals opgenomen in de betreffende tabel wel met alle betrokkenen zijn gedeeld.

4.2.2 Crisisaanpak

Alle ondervraagden delen de mening dat het proces na constatering van de fout, het inrichten van het vervolgproces en de eenduidige communicatie naar scholen, goed is verlopen. De toetsaanbieders zijn met een eenduidige boodschap naar buiten getreden, waarbij OCW nadrukkelijk de toetsaanbieders heeft gevrijwaard. De regie die OCW in die fase heeft genomen, is door alle direct betrokken partijen als sterk ervaren. Het heeft tevens het gemis aan eenduidige regie in het overall proces benadrukt, waar normaliter de Expertgroep en CvTE elk op een deel van het proces regie voeren.

4.3 Impact

Het is evident dat dit incident tot veel onrust bij toetsaanbieders, scholen, leerlingen en ouders heeft geleid en bij de eerste twee genoemde tot veel extra werk. Desalniettemin zijn scholen te spreken over de communicatie van de toetsaanbieders en hun ondersteuning bij het informeren van de ouders en leerlingen. Daarnaast heeft OCW een digitaal informatiepunt ingericht op rijksoverheid.nl en heeft zij gezamenlijk met toetsaanbieders op online fora vragen afgedaan. Dat is mede bepalend geweest voor de geconstateerde “rustige landing in het veld”, zoals in menig interview naar voren is gebracht.

Wat daarbij van doorslaggevende betekenis voor scholen is geweest, is dat zij met de herberekening van de toetsadviezen zich gesterkt hebben gevoeld in hun eigen professionele opinie. Het schooladvies en het toetsadvies moeten vanzelfsprekend in hoge mate correleren. De bijstelling van het aantal toetsadviezen, wat voor sommige toetsaanbieders 40-60% van hun totale toetsadviezen betrof, zorgde voor een bevestiging van het professionele oordeel van de docent. Het aantal heroverwegingen van het eigen schooladvies door scholen lijkt op basis van een eerste indicatie mee te vallen.

Private toetsaanbieders vrezen echter wel imago-schade. Vooral de vraag “waarom de CET niet?” en de signalen die uitdrukking geven aan het al dan niet hebben van vertrouwen in de (commerciële) toets(en), leiden tot zorg. Ondanks de vrijwaring van de toetsaanbieders door OCW, zal uit de inschrijving voor 2020 pas blijken hoe groot de impact is voor de private toetsaanbieders.

5 Tot slot, advies voor de toekomst

De minister heeft op 28 juni 2019 de nieuwe opzet c.q. inrichting aan de Tweede Kamer gezonden. Deze nieuwe opzet lijkt volgens de geïnterviewden aan een belangrijk deel van de geconstateerde knelpunten tegemoet te komen. De verleiding is dan ook groot om hier de vlucht naar voren te kiezen en in het huidige proces al werkzaamheden op een ander wijze te beleggen. Toch leert de afgelopen vier jaar dat dit niet verstandig is. Het voorstel voor de nieuwe inrichting betreft een wijziging die een zorgvuldige uitwerking van de implementatie vraagt om te voorkomen dat je in de jaren daarna wederom voor verrassingen komt te staan.

Ondanks dat deze toekomstige opzet buiten de scope van onze opdracht valt, willen wij een aantal handvatten meegeven om de voorziene forse wijziging in de werking van het stelsel vorm te geven:

- kies een projectmatige aanpak; maak eerst een plan en ontwerp de projectbesturing;
- zorg *vooraf* voor een doordacht en breed gedragen implementatiescenario met daarin een gedeelde opvatting over het ontwerp van de nieuwe werkwijze met duidelijkheid over ieders rol, taken en bevoegdheden in de toekomst;
- zorg daarbij voor stevige departementale bestuurlijke verankering van de implementatiestrategie en monitor de voortgang. Overweeg om de nu te vormen regietafel onder leiding van directeur PO tevens in te zetten als stuurgroep voor deze verandering.

Houdt bij de inrichting van het nieuwe voorstel voor ogen dat de huidige combinatie van werkzaamheden binnen de Expertgroep (beoordeling van de kwaliteit van de eindtoetsen, advisering, equivalering) en bij de CvTE (toetsaanbieder en kwaliteitsbewaking) in de toekomst moet worden voorkomen. Overweeg om de accreditatie voor toelating op de markt bij een onafhankelijke partij te beleggen.

Tot slot, voor de inspectie zijn de toetsresultaten in BRON van groot belang voor haar toezicht op de onderwijskwaliteit op school - en stelselniveau. De huidige werkwijze van aanlevering per school leidt tot onvoldoende betrouwbare gegevens mede door de foutgevoeligheid en de onnauwkeurigheid in aanpassing van heroverwogen schooladviezen. Met zoveel toeleveranciers ontstaan ook meer risico's in de uitwisseling. Aangezien de data nu centraal beschikbaar komt, is het aan te bevelen om te onderzoeken of er eenvoudigere (geautomatiseerde) methoden zijn om de gegevens in BRON te zetten.

Bijlagen

Bijlage 1: Overzicht gehouden interviews

Naam	Organisatie	Functie
Angelique Egberink	A-Vision	Manager Route 8
Anja Sinke	Basisschool De Omnibus	Intern begeleider
Anton Beguin	Stichting Cito	Directeur
Arnold Brouwer	RCEC	Directeur
Astrid Lubbers	A-Vision	Directeur
Bert Bulder	Inspectie van het Onderwijs	Directeur Kennis
Cees Glas	Expertgroep	Lid, psychometricus
Cees van der Vleuten	Expertgroep	Voorzitter
Cosmina Crisan	AMN	Directeur
Hilde Hacquebord	Diataal	Directeur
Janneke Helsloot	Bureau ICE	Manager IEP-eindtoets
Kees Theunis	Inspectie van het Onderwijs	Afdelingshoofd PO
Maarten Engelsman	OCW	Hoofd Kwaliteit en Leraren
Marc Spierings	CvTE	Directeur, secretaris college
Marjolein van Oijen	Basisschool De Stelberg	Intern begeleider
Michiel Colenbrander	Bureau ICE	Directeur
Ria Westerdorp	OCW	Directeur Primair Onderwijs
Rosalie Kropman	AMN	Manager AMN

Dit is een uitgave van:

ABDTOPConsult
Postbus 20011
2500 EA Den Haag

abdtc@minbzk.nl

www.abdtopconsult.nl