



RAPPORTAGE VERGELIJKBAARHEID EINDTOETSEN

December 2016

W.H.M. Emons, C.A.W.Glas, P.K. Berding-Oldersma

Inhoudsopgave

Samenvatting	3
1. Inleiding	4
1.1 Aanleiding	4
1.2 Gebruikte Data	5
1.3 Leeswijzer	5
1.4 Begrippenlijst	5
2. Achtergrondkenmerken van Leerlingen en Scholen per Toets	6
2.1 Achtergrond en Werkwijze.....	6
2.2 Resultaten.....	6
3. Vergelijking Toetsadviezen	8
3.1 Vergelijking op basis van de Toetsadviezen	8
3.1.1 Achtergrond en Werkwijze	8
3.1.2 Resultaten	9
3.2 Overeenstemming tussen Schooladvies, Toetsadvies en Herzieningen	11
3.2.1 Achtergrond en Werkwijze	11
3.2.2. Resultaten	12
4. Vergelijking op Basis van Referentieniveaus	14
4.1 Achtergrond en Werkwijze.....	14
4.2 Resultaten.....	15
5. Vergelijking Toetsadviezen en Beheersing Referentieniveaus	16
5.1 Achtergrond en Werkwijze.....	16
5.2 Resultaten.....	17
6. Vergelijking van Overstappende Scholen	18
6.1 Achtergrond en Werkwijze.....	18
6.2 Resultaten.....	18
7. Conclusies en aanbevelingen	19
7.1 Conclusies.....	19
7.2 Aanbevelingen	20
8. Literatuurlijst.....	21
Bijlage 1. De Weging naar Leerlinggewichten en Regio's	22
Bijlage 2. Het samenvoegen van Schooladviezen	27

Samenvatting

De wetwijziging 'Eindtoetsing PO' (Stb. 2014, 13) verplicht basisscholen om vanaf het schooljaar 2014/2015 bij alle leerlingen in groep 8 een eindtoets af te nemen met als minimale eis dat de toets taal- en rekenonderdelen bevat die gerelateerd zijn aan de verplichte referentieniveaus. Zowel bij de afname in 2015 als in 2016 zijn, bij een globale bestudering van de afnamedata, aanwijzingen gevonden, die wijzen op onderlinge verschillen in de resultaten met betrekking tot de behaalde referentieniveaus bij de drie tot dan toe toegelaten eindtoetsen (de Centrale Eindtoets, de Route 8 Eindtoets en IEP Eindtoets). Het ministerie OCW heeft de Expertgroep verzocht om de beschikbare data van de afname van 2016 nader te onderzoeken om vast te stellen in welke mate de eindtoetsen van elkaar verschillen.

Bij dit onderzoek wordt gebruik gemaakt van de afnamedata van 2016 welke door DUO zijn vrijgegeven. Daarnaast zijn de wetenschappelijke verantwoordingen van de Centrale Eindtoets, Route 8 en IEP (CvTE, 2015; Langeveld et al., 2016; Egberink et al., 2016) bestudeerd om deze data te duiden. De aanbieders van deze eindtoetsen zijn tevens gevraagd de afnamedata van 2016 te duiden.

De eerste vraag die in dit rapport aan de orde komt is of de verschillende eindtoetsen verschillende populaties bedienen. Dit is onderzocht voor de regionale spreiding van de toetsen en de gewichtenleerlingen in de verschillende toetspopulaties. Met betrekking tot gewichtenleerlingen zijn geen belangrijke verschillen tussen de toetsen gevonden. Bij de regionale spreiding valt op dat Route 8 relatief vaker afgenomen wordt in regio's Noord en Zuid en minder vaak in regio's West en Oost, terwijl de IEP relatief vaak in Regio West wordt afgenomen. De Centrale Eindtoets is in alle regio's evenredig gespreid.

Analyse van de toetsadviezen laat zien dat Route 8 relatief lager adviseert dan de andere toetsen (meer vmbo, minder vwo). Bij de IEP worden iets meer vwo adviezen gegeven, maar dat effect was veel kleiner dan het tegenovergestelde effect zoals gevonden bij Route 8. Conditioneren op regio en leerlinggewichten veranderden de gevonden effecten niet. Met andere woorden, regio en leerlinggewicht konden de gevonden verschillen niet verklaren. De analyse van de samenhang tussen toets- en schooladvies versterken het beeld van de niveauverschillen tussen de toetsen..

Met betrekking tot de referentieniveaus wordt geconstateerd dat de percentages leerlingen die bepaalde niveaus haalden bij de Centrale Eindtoets en de IEP vergelijkbaar zijn. Bij Route 8 is het percentage leerlingen dat niveau 1F niet behaald steeds beduidend hoger. Conditioneren op regio of leerlinggewicht verandert hier niets aan, dus ook hier konden regio en leerlinggewicht de gevonden verschillen niet verklaren. De uitzonderingspositie van Route 8 bij de referentieniveaus is duidelijker dan bij de toetsadviezen. Vergelijking van de referentieniveaus met de schooladviezen bevestigt dit verder.

In combinatie met de uitkomsten van het rapport **Scenario's voor ijking van de eindtoetsen op de referentieniveaus** wordt aanbevolen op korte termijn de psychometrische verantwoordingen van de eindtoetsen te uniformeren, met betrekking tot rapportages, gebruikte statistische methoden en de daarbij behorende software. Daarnaast dient er nader onderzoek gedaan te worden naar de verschillen tussen de verschillende cesuren voor de schooladviezen en referentieniveaus door alle toetsen (dus ook de AMN, CESAN en Dia eindtoets) op een gezamenlijke schaal te kalibreren, en door onderzoek te doen naar de betrouwbaarheid van die ankering. Op de langere termijn moet er gewerkt worden aan een gezamenlijk anker die in alle eindtoetsen op wordt genomen als deel van de eindtoets. Deze aanbevelingen zullen de vergelijkbaarheid en de ijking van de eindtoetsen verbeteren.

1. Inleiding

1.1 Aanleiding

De wetwijziging 'Eindtoetsing PO' (Stb. 2014, 13) verplicht basisscholen om vanaf het schooljaar 2014/2015 bij alle leerlingen in groep 8 een eindtoets voor taal en rekenen af te nemen. De eindtoets die een basisschool gaat gebruiken, moet van goede kwaliteit zijn. Daarom is door de minister een onafhankelijke commissie (de Expertgroep Toetsen PO) aangewezen die de minister adviseert over het toelaten van 'andere' toetsen als eindtoets voor het PO. De term 'andere' heeft betrekking op het feit dat een onderdeel van de genoemde wetwijziging is dat de overheid de publieke taak heeft om een centrale eindtoets beschikbaar te stellen. Het College voor Toets en Examens (CvTE) is belast met deze taak. In opdracht van het CvTE wordt door stichting Cito de Centrale Eindtoets ontwikkeld.

In 2014 zijn er, op basis van een positief advies van de Expertgroep Toetsen PO (hierna Expertgroep) twee andere eindtoetsen in het PO toegelaten, dit betreft Route 8, ontwikkeld en uitgegeven door A-VISION B.V., en IEP Eindtoets, ontwikkeld en uitgegeven door Bureau ICE. Beide toetsen zijn in 2015 opnieuw beoordeeld, waaruit eveneens een positief advies volgde, waarna de minister beide toetsen ook in 2015/2016 heeft toegelaten. Route 8 is een adaptieve toets (elke leerling maakt verschillende opgaven die zijn afgestemd op zijn/haar niveau) die afgenomen wordt via een computer (digitale afname), de Centrale Eindtoets en IEP zijn lineaire toetsen (elke leerling maakt dezelfde opgaven in dezelfde volgorde). De IEP wordt op papier afgenomen (schriftelijke afname). Voor de Centrale Eindtoets kunnen de scholen kiezen voor een schriftelijke of digitale afname.

Voor schooljaar 2016/2017 hebben de scholen naast de drie eerder genoemde eindtoetsen ook de keuze uit de Dia-Eindtoets, ontwikkeld en uitgegeven door Diataal B.V., de CESAN Eindtoets, ontwikkeld en uitgegeven door SM&C Internet services b.v., en de AMN Eindtoets, ontwikkeld en uitgegeven door AMN. Het voorliggende rapport heeft enkel betrekking op de IEP, de Centrale Eindtoets en Route 8.

Eindtoetsen in het basisonderwijs kennen twee primaire doelen. Ten eerste, het geven van advies welke schooltype voortgezet onderwijs het beste bij de leerling past (toetsadvies). Dit advies geldt als tweede objectieve gegeven naast het advies van de school. Indien het toetsadvies aangeeft dat de leerling een hoger niveau schoolniveau aan kan dan dat het schooladvies aangeeft, dan dient de school het oorspronkelijke schooladvies te *heroverwegen*. Het schooladvies kan, maar hoeft niet, op basis van het toetsadvies naar boven worden bijgesteld. Het schooladvies is uiteindelijk doorslaggevend voor de toelating tot het voortgezet onderwijs.

Het tweede primaire doel is om te bepalen in hoeverre leerlingen de referentieniveaus beheersen. Referentieniveaus geven voor Nederlandse taal en rekenen het gewenste niveau aan op verschillende momenten in het onderwijs (Stb. 2010, 265). De referentieniveaus worden uitgesplitst in de fundamentele niveaus (1F, 2F, 3F en 4F) en de streefniveaus (1S, 2S en 3S). Voor het onderdeel Nederlandse Taal worden vier domeinen onderscheiden: Mondelinge Taalvaardigheid, Lezen, Schrijven en Begrippenlijst en Taalverzorging. Van deze domeinen is toetsing met een eindtoets verplicht. Op beide domeinen wordt afzonderlijk een uitspraak gedaan per referentieniveau. Bij de referentieniveaus Rekenen worden de vier domeinen Getallen, Meten en Meetkunde, Verbanden en Verhoudingen onderscheiden en samengevoegd tot één uitspraak per referentieniveau. Alle eindtoetsen rapporteren op de referentieniveaus Lezen 1F en 2F, Taalverzorging 1F en 2F en Rekenen 1F en 1S. De IEP rapporteert daarnaast ook 2F rekenen.

Beide primaire uitkomsten worden door de onderwijsinspectie meegewogen bij de evaluatie van individuele scholen en zijn ook van belang voor stelsevaluaties van het primair onderwijs als geheel.

De keuzevrijheid voor eindtoetsen heeft er voor gezorgd dat er binnen het basisonderwijs verschillende eindtoetsen worden gebruikt. Hoewel de drie eindtoetsen zijn geconstrueerd op basis van verschillende uitgangspunten, met inachtneming van de vastgestelde kaders (Algemeen deel Toetswijzer (CvTE, 2014), Beoordelingskader Expertgroep PO (Folmer et al., 2014) en Beoordelingsstelsel van de COTAN

(Evers et al. 2010), is hun gemeenschappelijk doelstelling het doen van uitspraken met betrekking tot het best passende brugklatype in het voortgezet onderwijs en de beheersing van de referentieniveaus. Een belangrijke vraag daarbij is de mate waarin verschillende eindtoetsen vergelijkbare schooladviezen geven en vergelijkbare rapportages van de referentieniveaus hanteren. Daartoe heeft de staatssecretaris van OCW de Expertgroep PO gevraagd om de resultaten van de verschillende eindtoetsen uit 2015-2016 naast elkaar te zetten en de vergelijkbaarheid met betrekking tot toetsadviezen en bepaling beheersing referentieniveaus nader te onderzoeken. Dit rapport is het resultaat van dit onderzoek.

1.2 Gebruikte Data

De tabellen zijn gebaseerd op toets-gegevens van de toetsafname 2016 zoals deze in BRON bij DUO geregistreerd staan. De niet-deelnemers en deelnemers met ontheffingen zijn niet meegenomen in de resultaten.

1.3 Leeswijzer

Dit rapport is als volgt opgebouwd. Eerst worden de achtergrondkenmerken van de leerlingen en scholen per toets gerapporteerd. Daarna volgen de resultaten per toets-doel, dat wil zeggen, de resultaten van de vergelijking van schooladviezen, gevolgd door de resultaten van de vergelijking van de bepaling van de referentieniveaus. Tot slot zullen de resultaten voor de verschillende toets-doelen tegen elkaar afgezet worden en worden enkele algemene conclusies getrokken.

1.4 Begrippenlijst

Toetsadvies: doorstroomadvies met betrekking tot het best passende brugklatype in het voortgezet onderwijs dat door de eindtoets wordt gegeven.

Schooladvies: doorstroomadvies met betrekking tot het best passende brugklatype in het voortgezet onderwijs dat door de basisschool van de betreffende leerling wordt opgesteld.

Eindtoets PO: de eindtoets die binnen het PO afgenomen dient te worden.

2. Achtergrondkenmerken van Leerlingen en Scholen per Toets

2.1 Achtergrond en Werkwijze

In dit hoofdstuk bespreken we voor elke toets de achtergrondkenmerken van de leerlingen en scholen die aan de betreffende eindtoets hebben deelgenomen (toetspopulaties). Doordat scholen zelf mogen kiezen welke eindtoets zij afnemen, kunnen de toetspopulaties op belangrijke achtergrondkenmerken verschillen. Deze verschillen in achtergrondkenmerken kunnen mogelijk deels verschillen tussen toetsadviezen en de behaalde referentieniveaus verklaren. Bijvoorbeeld, als een bepaalde toets relatief vaker bij leerlingen met laag opgeleide ouders wordt afgenomen dan andere toetsen, dan zou dat een verklaring kunnen zijn voor lagere percentages leerlingen met vwo-advies. Dit zijn verschillen die niet zijn toe te schrijven aan de specifieke toets en dus daadwerkelijke verschillen tussen de toetsen vertekenen. Door middel van weging kunnen verschillen tussen eindtoetsen die te verklaren zijn door verschillen in achtergrondkenmerken er uit worden gefilterd. De gewogen resultaten geven dan een beter beeld van de verschillen tussen eindtoetsen dan de ongewogen scores.

Een belangrijke achtergrondvariabele is leerling-gewicht. Voor elke leerling wordt bij de toelating tot een school een gewicht vastgesteld. Er kunnen drie verschillende gewichten toegekend worden: 0,0, 0,3, en 1,2. Het gewicht wordt bepaald op basis van het opleidingsniveau van de ouders. Hoe hoger het gewicht, hoe lager het opleidingsniveau van de ouders. Gewichtenleerlingen zijn leerlingen met gewicht 0,3 of 1,2. Om de resultaten te wegen naar leerlinggewicht kan naar de verdeling van gewichten over scholen gekeken worden, en/of er kan bij elke leerling nagegaan worden hoeveel gewichtenleerlingen er gemiddeld op zijn/haar school zitten. Er wordt een onderscheid gemaakt in drie categorieën van scholen: scholen met minder dan 1% gewichtenleerlingen, tussen 1% en 10% gewichtenleerlingen en meer dan 10% gewichtenleerlingen. In dit onderzoek worden beide operationalisaties van leerlinggewicht meegenomen.

2.2 Resultaten

Tabellen 1 en 2 geven de achtergrondgegevens weer. De eerste tabel beschrijft de achtergrondkenmerken op het niveau van de leerlingen en de tweede tabel beschrijft de kenmerken op schoolniveau.

De opvallendste verschillen tussen eindtoetsen zijn te vinden in de regionale spreiding van de deelnemende scholen. In vergelijking met de landelijke spreiding wordt Route 8 relatief vaker afgenomen in regio's Noord en Zuid en minder vaak in regio's West en Oost. De afnames van de IEP concentreren zich juist in Regio West, terwijl de andere regio's evenredig zijn vertegenwoordigd. De afnames van de Centrale Eindtoets volgen de landelijke spreiding. Dit ligt in de lijn van de verwachtingen aangezien de Centrale Eindtoets een marktaandeel van 77% heeft.

Naast de regionale spreiding zien we ook verschillen in een aantal andere achtergrondkenmerken zoals schoolgrootte en urbanisatiegraad. Deze verschillen hangen deels samen met regionale verschillen. Het gevolg is ook dat de prestaties van leerlingen kunnen verschillen over de regio's. Om te voorkomen dat daadwerkelijke verschillen tussen de eindtoetsen worden vertekend door regionale verschillen, worden de resultaten ook gewogen naar rato van de regionale verdeling. De methode van weging wordt toegelicht bij de bespreking van de resultaten (met gedetailleerde informatie in Bijlage 1). Door te wegen naar regio worden mogelijke vertekeningen door verschillen in de overige kenmerken die met regionale verschillen samenhangen deels ondervangen.

Met betrekking tot de verdeling van gewichtenleerlingen laten Tabellen 1 en 2 zien dat de toetspopulaties goed vergelijkbaar zijn. De IEP wordt procentueel wat vaker bij leerlingen afgenomen van een school met weinig gewichtenleerlingen (8,4%), de Centrale Eindtoets verhoudingsgewijs wat minder (5,3%). Echter de percentages leerlingen van scholen met meer dan 10% gewichtenleerlingen is vergelijkbaar. Als we kijken naar de verdeling van de gewichtenleerlingen zelf (Tabel 1; onderste paneel), dan zien we nauwelijks verschillen tussen de toetspopulaties.

Tabel 1: Achtergrondkenmerken Deelnemende Leerlingen per Toets

	Centrale Eindtoets	Route 8	IEP	Totaal
Aantal Leerlingen				
<i>N</i>	139,287	11,814	29,759	180,860
<i>%</i>	77,0%	6,5%	16,5%	100%
Regio^a				
Noord	8,6%	20,8%	13,6%	10,2%
Oost	24,7%	16,6%	15,2%	22,6%
Zuid	21,0%	27,2%	13,4%	20,1%
West	45,8%	35,4%	57,8%	47,1%
Gewichtenleerlingen (binnen scholen)				
< 1%	5,3%	7,5%	8,4%	6,0%
1-10%	66,8%	62,4%	62,1%	65,7%
>10%	27,9%	30,2%	29,5%	28,3%
Gewichtenleerlingen (op leerling-niveau)				
0.00	90,3%	90,7%	90,7%	90,4%
0.30	5,3%	5,4%	5,4%	5,3%
1.20	4,4%	3,9%	3,9%	4,3%

a: Noord: Groningen, Friesland, Drenthe; Oost: Flevoland, Gelderland, Overijssel; Zuid: Brabant, Limburg; West: Noord- en Zuid-Holland, Zeeland en Utrecht.

Tabel 2: Achtergrondkenmerken Deelnemende Scholen per Toets

	Centrale Eindtoets	Route 8	IEP	Totaal
Aantal Scholen				
<i>N</i>	4,956	466	1,119	6,543
<i>%</i>	75,7%	7,1%	17,1%	100,0%
Regio				
Noord	12,7%	28,3%	19,5%	15,0%
Oost	26,9%	17,6%	17,4%	24,6%
Zuid	19,8%	21,9%	12,1%	18,6%
West	40,6%	32,2%	51,0%	41,8%
Denominatie				
Openbaar	35,8%	41,4%	41,2%	37,2%
Bijzonder	64,2%	58,6%	58,8%	62,8%
Schoolgrootte				
<100	17,8%	23,4%	21,2%	18,7%
100-200	32,6%	33,0%	32,1%	32,5%
200-300	26,6%	25,8%	24,9%	26,2%
>300	23,1%	17,8%	21,8%	22,5%
Gewichtenleerlingen				
< 1%	6,6%	7,9%	9,9%	7,3%
1-10%	60,7%	55,6%	56,1%	59,5%
>10%	32,7%	36,5%	34,0%	33,2%
Urbanisatiegraad				
G4	10,6%	3,6%	5,6%	9,3%
G5-G37	20,2%	21,5%	21,3%	20,5%
>G37	69,2%	74,9%	73,1%	70,3%
Impulsgebied				
Ja	75,7%	73,0%	73,2%	75,1%
Nee	24,3%	27,0%	26,8%	24,9%

G4: vier grootste gemeentes; G5-G37: 5 tot 37 grootste steden.

De percentages op schoolniveau (Tabel 2) laten dezelfde trends zien. Merk op dat de percentages met betrekking tot gewichtenleerlingen op leerlingenniveau (Tabel 1; middenpaneel) niet precies overeenkomen met de percentages gewichtleerlingen in Tabel 2. Dit komt omdat schoolgrootte varieert; dus niet elke school met een bepaald percentage gewichtenleerlingen levert evenveel leerlingen.

In het vervolg van deze rapportage zijn de resultaten gewogen naar de verdeling van het percentage gewichtenleerlingen per school (Tabel 1, middenpaneel), voor zover de cijfers hiervoor voor handen waren. Hierbij is gewogen naar de verdeling van de percentages leerlingen binnen een school met 0%, 1-10%, of >10% gewichtenleerlingen (zie Tabel 1, middenpaneel). Echter, omdat de verdeling van de gewichtenleerlingen vergelijkbaar is, is de verwachting dat deze weging nauwelijks een effect heeft op de uitkomsten.

3. Vergelijking Toetsadviezen

Een van de primaire einddoelen van de Eindtoets PO is het geven van een advies over welke type voortgezet onderwijs het beste bij de leerling past. Het advies op basis van de eindtoets, vanaf nu aangeduid met toetsadvies, dient daarbij als tweede onafhankelijke gegeven naast het advies van de school, hierna aangeduid met schooladvies. Indien het toetsadvies hoger is dan het schooladvies, dan moet het oorspronkelijke schooladvies *heroverwogen* worden. Dit kan mogelijk leiden tot herziening maar de school is daartoe niet verplicht. Het mogelijk herziene schooladvies is uiteindelijk bepalend. In dit hoofdstuk wordt de verdeling van de toetsadviezen bij de verschillende eindtoetsen met elkaar vergeleken. Daarnaast wordt er gekeken in hoeverre de toetsadviezen overeenstemmen met de schooladviezen en naar de percentages herzieningen. Alvorens in te gaan op de resultaten zal de achtergrond worden geschetst waartegen de resultaten bekeken dienen te worden.

3.1 Vergelijking op basis van de Toetsadviezen

3.1.1 Achtergrond en Werkwijze

Om vast te stellen welke toetsadviezen horen bij welke toetsprestaties dient normeringsonderzoek plaats te vinden. De normeringen zijn bij IEP en Route 8 zodanig bepaald dat de toetsadviezen zoveel mogelijk consistent zijn met de schooladviezen. Voor de normering van de Centrale Eindtoets ten aanzien van schooladviezen wordt gebruik gemaakt van gegevens uit doorstroomonderzoek en door middel van ankering. De schooladviezen worden gebruikt om de normering te valideren. Aan normeringsonderzoek worden doorgaans hoge eisen gesteld, bijvoorbeeld door de COTAN (Evers, A. et al., 2010). Er dienen toetsgegevens te worden verzameld door middel van (proef)afname van de te normeren eindtoets in een steekproef die groot genoeg is en representatief voor de populatie. Op basis van de beschikbare gegevens worden vervolgens de scoringsregels vastgesteld om te komen tot een toetsadvies. Hiervoor kunnen schooladviezen van de leerlingen die ook de toets hebben gemaakt worden gebruikt, of men kan afgaan op de verdeling van schooladviezen op nationaal niveau. Indien de normering is gebaseerd op nationale gegevens, is de representativiteit van de gebruikte normeringssteekproef ten opzichte van de landelijke verdeling essentieel. Indien de normering van de toetsadviezen uitsluitend is gebaseerd op de schooladviezen van de leerlingen in de eigen steekproef, is de eis van landelijke representativiteit wat minder stringent zolang de steekproef maar groot genoeg is, de steekproef leerlingen over het hele vaardigheidsspectrum omvat en de schooladviezen volledig onafhankelijk van het toetsadvies tot stand zijn gekomen. Hierbij wordt wel de aanname gemaakt dat de schooladviezen als zodanig landelijk vergelijkbaar zijn. Indien men niet beschikt over schooladviezen maar (ook) gebruik maakt van de nationale gegevens dient de representativiteit van de steekproef strikt te worden aangetoond. Eventueel kan door middel van weging de representativiteit worden verbeterd.

Hoewel er strenge kwaliteitseisen zijn aan het normeringsonderzoek, zijn er verder geen voorgeschreven regels hoe elke eindtoets het advies voor vervolgonderwijs moet vaststellen. De verschillende

eindtoetsen bevatten verschillende onderdelen, die bovendien verschillend worden meegewogen, waardoor toetsadviezen van de eindtoetsen niet op dezelfde wijze tot stand komen. Deze keuzes kunnen mogelijk leiden tot verschillen in de verdeling van de toetsadviezen tussen de toetsen. Daarnaast worden de toetsen genormeerd op een verschillende criteriummaten. Voor de IEP en Route 8 is dat het schooladvies, voor de Centrale Eindtoets zijn dat primair de doorstroomgegevens en ankering, maar secundair worden schooladviezen meegenomen. Op basis hiervan is het niet aannemelijk dat de verschillen heel groot zijn, tenzij de wijze waarop het schooladvies tot stand komt samenhangt met de gebruikte toets. Deze samenhang zou bijvoorbeeld kunnen ontstaan als het schooladvies wordt bijgestuurd op basis van eerdere resultaten van de eindtoets waardoor toets en schooladvies naar elkaar toegroeien.

De toetsaanbieders hebben de vrijheid om voor het toetsadvies zelf een indeling van de schoolcategorieën te kiezen. Dit heeft als gevolg dat voor de afnames 2015 en 2016 elke eindtoets verschillende categorieën hanteerde voor het toetsadvies. Daarnaast gaven de Centrale Eindtoets en IEP gecombineerde adviezen (bijv. havo/vwo) en Route 8 deed dat in 2016 nog niet. Het ontbreken van een standaardisering van de gerapporteerde schoolcategorieën bemoeilijkt de vergelijkbaarheid van de eindtoetsen en het onderzoek naar deze vergelijkbaarheid. Om op basis van de huidige gegevens tot een vergelijkbare indeling van de toetsadviezen te komen is voor de volgende aanpak gekozen.

Allereerst is gekozen om aan te sluiten bij de gangbare indeling in het voortgezet onderwijs in drie algemene niveaus: vmbo, havo, en vwo. Normaliter wordt het vmbo verder onderverdeeld in beroepsgerichte opleidingen en theoretische opleidingen (leerwegen), te weten: vmbo-t/vmbo-tl (theoretische leerweg), vmbo-g/vmbo-gl (gemengde leerweg), vmbo-k/vmbo-kbl (kaderberoepsgerichte leerweg), en vmbo-b/vmbo-bbl (basisberoepsgerichte leerweg). In dit onderzoek is er voor gekozen om binnen de categorie vmbo alleen nog een onderscheid te maken tussen beroepsgerichte opleidingen en theoretische leerweg. Vervolgens zijn de oorspronkelijke categorieën per toets samengevoegd tot de bovengenoemde categorieën (vmbo pr, vmbo tl, havo, en vwo). De procedure en de gebruikte data zijn beschreven in Bijlage 2.

3.1.2 Resultaten

Tabel 3 laat per toets de percentages van de verschillende toetsadviezen zien. In de tabel staan de percentages zowel ongewogen als gewogen gewichtenleerlingen en regio. De verschillen tussen de ongewogen en gewogen percentages zijn zeer gering; door weging verandert het percentage maximaal met 0.9%. Dit betekent dat regionale verschillen in de toetspopulaties verschillen tussen de eindtoetsen niet verklaren. We zullen ons daarom beperken tot het bespreken van de ongewogen percentages.

Kijken we naar de percentages per eindtoets, dan zien we dat de percentages adviezen voor vmbo, havo en vwo voor Centrale Eindtoets en IEP vergelijkbaar zijn. Het grootste verschil heeft betrekking op het vwo-advies; dat ligt 1.3% hoger voor de IEP dan voor de Centrale Eindtoets. Wel zien we verschillen in de percentages met betrekking tot de adviezen voor vmbo-bg en vmbo-tl. In vergelijking met de IEP geeft de Centrale Eindtoets 8% vaker het advies vmbo-bg en 7.4% minder vaak het advies vmbo-tl.

De percentages voor Route 8 wijken echter substantieel af van de andere twee toetsen. Het percentage leerlingen met vwo advies ligt beduidend lager (12.6 % voor Route 8 en ongeveer 26% voor de andere twee toetsen) en dat voor het vmbo advies een stuk hoger (56.6% voor Route 8 en ongeveer 44% voor de Centrale Eindtoets en IEP). Het percentage havo-adviezen is voor Route 8 vergelijkbaar met de andere twee eindtoetsen. Dit betekent overigens niet dat daarmee het havo-advies van Route 8 vergelijkbaar is met de andere twee eindtoetsen, maar kan ook komen omdat Route 8 over de hele linie "strenger" is in het geven toetsadviezen.

Tabel 3: Vergelijking Toetsadviezen: Ongewogen en Gewogen naar Gewichten en Regio

Advies	Centrale Eindtoets	ROUTE8	IEP
Ongewogen Percentages			
Vmbo	44,5%	56,6%	43,9%
<i>BG</i>	25,9%	27,6%	17,9%
<i>TL</i>	18,6%	29,1%	26,0%
Havo	30,3%	30,8%	29,4%
Vwo	25,3%	12,6%	26,6%
Gewogen naar Gewichtenleerlingen			
Vmbo	44,5%	56,5%	43,9%
<i>BG</i>	25,9%	27,4%	17,9%
<i>TL</i>	18,5%	29,1%	26,0%
Havo	30,3%	30,9%	29,4%
Vwo	25,3%	12,6%	26,6%
Gewogen naar Regio			
Vmbo	44,5%	55,7%	44,0%
<i>BG</i>	25,9%	27,1%	17,9%
<i>TL</i>	18,5%	28,7%	26,2%
Havo	30,2%	31,5%	29,4%
Vwo	25,3%	12,8%	26,6%

Tabel 4 laat de resultaten per regio zien. Uit deze resultaten blijkt dat er bij de Centrale Eindtoets en de IEP geen samenhang is tussen regio en toetsadvies. Het effect dat Route 8 meer vmbo adviezen geeft dan de andere twee toetsen en minder vwo adviezen is ook hier zichtbaar, maar het effect is het sterkst in de regio Noord.

Tabel 4: Vergelijking Toetsadviezen: Uitgesplitst naar Regio

Advies	Noord	Oost	Zuid	West	Totaal
Centrale Eindtoets					
Vmbo	44,8%	45,8%	42,2%	44,7%	44,5%
<i>BG</i>	25,7%	26,6%	23,7%	26,6%	25,9%
<i>TL</i>	19,1%	19,2%	18,5%	18,1%	18,6%
Havo	30,7%	30,6%	31,1%	29,6%	30,3%
Vwo	24,5%	23,6%	26,6%	25,7%	25,3%
Route 8					
Vmbo	60,9%	53,4%	57,3%	55,1%	56,6%
<i>BG</i>	30,5%	24,0%	27,2%	27,8%	27,6%
<i>TL</i>	30,5%	29,5%	30,1%	27,3%	29,1%
Havo	27,8%	34,3%	30,0%	31,5%	30,8%
Vwo	11,3%	12,3%	12,8%	13,4%	12,6%
IEP					
Vmbo	42,7%	43,5%	45,5%	43,9%	43,9%
<i>BG</i>	16,8%	16,9%	18,7%	18,2%	17,9%
<i>TL</i>	26,0%	26,6%	26,8%	25,7%	26,0%
Havo	29,8%	29,5%	28,6%	29,5%	29,4%
Vwo	27,4%	27,0%	25,9%	26,6%	26,6%

Om na te gaan of de toetsadviezen van de verschillende eindtoetsen samenhangen met leerlinggewicht is gekeken naar de verdeling van de toetsadviezen uitgesplitst naar gewicht (**Tabel 5**). Hoewel de

algemene trends hetzelfde zijn als in Tabel 4 geven de resultaten een iets gedifferentieerder beeld. We zien, bijvoorbeeld, dat bij de IEP van de leerlingen met gewicht 1.2 er 3.0% leerlingen meer een vwo-advies krijgen dan bij de Centrale Eindtoets. Om mogelijke interactie tussen toetsadvies en leerlinggewicht met de eindtoets nog beter in kaart te brengen is er ook gekeken naar de verschillen in de sterkte van de samenhang tussen toetsadvies en gewicht. De samenhang (uitgedrukt als Cramer's V) was gelijk aan 0.125 voor de Centrale Eindtoets, 0.110 voor Route 8, en 0.105 voor de IEP. We zien (volgens verwachting) dat er een zwakke samenhang is tussen leerling-gewicht en toetsadvies en de sterkte van de samenhang is vergelijkbaar over de eindtoetsen. Met andere woorden, de mate waarin het toetsadvies samenhangt met leerlinggewicht is voor de ene toets niet veel anders dan voor de andere toets.

Tabel 5. Toetsadvies uitgesplitst naar Percentage Gewichtenleerlingen Binnen School en Gewichten op Niveau van Leerlingen.

Advies	Gewicht binnen School			Per Leerling		
	0	1-10%	>10%	0	0.3	1.2
Centrale Eindtoets						
(n =)	7,393	92,980	38,914	125,637	7,440	6,128
Vmbo	32,5%	41,6%	53,6%	41,6%	71,3%	69,8%
BG	16,4%	23,1%	34,5%	23,3%	51,0%	49,9%
TL	16,1%	18,5%	19,2%	18,4%	20,3%	19,9%
Havo	32,9%	31,4%	27,0%	31,3%	20,4%	21,2%
Vwo	34,6%	27,0%	19,3%	27,1%	8,3%	9,1%
Route 8						
(n =)	885	7,367	3,562	10,707	637	460
Vmbo	48,7%	53,3%	65,3%	54,1%	81,2%	80,7%
BG	22,8%	24,4%	35,2%	25,1%	53,1%	49,8%
TL	25,9%	28,9%	30,2%	29,0%	28,1%	30,9%
Havo	34,9%	33,2%	24,8%	32,5%	14,8%	14,1%
Vwo	16,4%	13,5%	9,9%	13,4%	4,1%	5,2%
IEP						
(n =)	2,487	18,491	8,781	26,697	1,556	1,216
Vmbo	35,6%	41,6%	51,2%	41,6%	66,9%	64,6%
BG	11,2%	16,3%	23,0%	16,1%	35,8%	33,6%
TL	24,4%	25,2%	28,2%	25,5%	31,1%	31,1%
Havo	32,2%	30,1%	27,3%	30,2%	21,8%	23,2%
Vwo	32,2%	28,4%	21,5%	28,2%	11,3%	12,1%

3.2 Overeenstemming tussen Schooladvies, Toetsadvies en Herzieningen

3.2.1 Achtergrond en Werkwijze

In deze paragraaf worden de toetsadviezen afgezet tegen het oorspronkelijke schooladvies. Iedere leerling in het primair onderwijs heeft een oorspronkelijk schooladvies gekregen voorafgaand aan de afname van de eindtoets. We gaan er van uit dat het oorspronkelijke schooladvies gemiddeld een goede indicatie is van het best passende brugklatype. Onder deze aanname kunnen we evalueren in hoeverre de toetsadviezen van de drie eindtoetsen structureel afwijken ten opzichte van de oorspronkelijke schooladviezen en, belangrijker nog, of de mate waarin toetsadviezen afwijken van schooladviezen verschillend zijn voor de verschillende eindtoetsen.

Een tweetal belangrijke kanttekeningen zijn daarbij op zijn plaats. Ten eerste kunnen scholen andere aspecten (bijv. functioneren in de klas, interesses) mee wegen in het schooladvies. Het schooladvies en het toetsadvies zijn geen volledig uitwisselbare uitkomsten; het toetsadvies is juist bedoeld als *tweede* onafhankelijk gegeven. Dus we verwachten geen perfecte samenhang, maar als de discrepanties tussen toetsadviezen en schooladviezen erg uiteenlopen, zowel per toets als tussen de toetsen, dan is dat een

belangrijk punt van aandacht. Ten tweede gaan we er vanuit dat de scholen die voor de ene eindtoets kiezen bij gelijke leerlingen niet structureel hoger of lager adviseren dan de scholen die voor een andere eindtoets kiezen. Met andere woorden, de wijze van totstandkoming van het schooladvies mag niet samenhangen met de keuze voor een eindtoets. Deze aanname is belangrijk omdat dergelijke structurele verschillen in advisering tussen groepen scholen mogelijk deels worden gemaskeerd door de wijze waarop de toetsadviezen zijn genormeerd. De grenswaarden voor een eindtoets zijn bepaald aan de hand van de schooladviezen van leerlingen uit een proefafname¹ of operationele afname van de toets. De grenswaarden zijn zodanig gekozen dat de overeenstemming tussen toets en schooladvies of doorstroomgegevens maximaal is. De proefafnames vinden plaats op scholen uit de eigen toetspopulatie met als gevolg dat de grenswaarden afgestemd zijn op de specifieke kenmerken van de eigen toetspopulatie in een low-stakes afnameconditie. Dit betekent dat een vergelijkbare overeenstemming tussen toets- en schooladvies op zich geen voldoende voorwaarde is om structurele verschillen in toetsadviezen tussen de eindtoetsen uit te sluiten.

Bij het vergelijken van toetsadviezen en schooladviezen doet zich ook het probleem voor dat de gebruikte categorieën voor de schooladviezen in de beschikbare data niet consistent zijn met de toetsadviezen, die op hun beurt ook weer variëren tussen de eindtoetsen. Daarnaast geeft Route 8 in 2016 geen meervoudig advies. Dit bemoeilijkt de vergelijkbaarheid van de overeenstemming tussen het school- en toetsadvies.

Om de overeenstemming tussen schooladvies en toetsadvies inzichtelijk te maken is de volgende werkwijze gehanteerd. Allereerst zijn de toetsadviezen samengevoegd tot de genoemde basiscategorieën (vbmo-bg, vmbo-tl, havo, & vwo) met de methode die hierboven en in Bijlage 2 besproken is. Ten tweede, om ervoor te zorgen dat er voldoende observaties van een combinatie van toetsadvies en categorie schooladvies beschikbaar zijn om verschillen tussen eindtoetsresultaten niet op basis van toevalligheden te laten ontstaan, zijn de categorieën voor de schooladviezen ook samengenomen. Daartoe is met volgende clusters van oorspronkelijke schooladviezen gewerkt: [vmbo-bb; vmbo-bb/kb; vmbo-kb; vmbo-kb/gt], [vmbo-gt; vmbo-gt/havo; havo]; [havo/vwo; vwo]. Vervolgens is gekeken in hoeverre het toetsadvies *aantoonbaar* afwijkt van het schooladvies. Daartoe is de volgende overeenstemmingstabel gehanteerd:

Tabel 6: Overeenstemmingstabel Toetsadvies en Schooladvies

Toetsadvies	Oorspronkelijk Schooladvies		
	vmbo-bb t/m vmbo-kb/gt	vmbo-gt t/m havo	havo/vwo en vwo
Vmbo	Consistent	Toetsadvies Lager	Toetsadvies Lager
<i>BG</i>	<i>Consistent</i>	<i>Toetsadvies Lager</i>	<i>Toetsadvies Lager</i>
<i>TL</i>	<i>Consistent</i>	<i>Consistent</i>	<i>Toetsadvies Lager</i>
Havo	Toetsadvies Hoger	Consistent	Consistent
Vwo	Toetsadvies Hoger	Toetsadvies Hoger	Consistent

Dit betekent, bijvoorbeeld, dat het meervoudig schooladvies vmbo-gt t/m havo en toetsadvies vmbo-tl en havo als consistent worden gezien. Deze aanpak maakt het mogelijk om overeenstemming tussen toets- en schooladvies voor de verschillende eindtoetsen met elkaar te vergelijken, maar heeft als nadeel dat verfijndere verschillen tussen schooladviezen en toetsadviezen niet meer zichtbaar zijn. Bij de resultaten zullen zowel de gecombineerde als de niet gecombineerde adviezen worden besproken.

3.2.2. Resultaten

Tabel 7 laat de overeenstemming tussen school- en toetsadviezen zien. We zien dat (volgens deze indeling) Route 8 procentueel de hoogste overeenstemming heeft (namelijk 86,4%), gevolgd door IEP (85,4%) en Centrale Eindtoets (83,5%). Opvallend is dat bij de Route 8 en IEP de percentages over- of

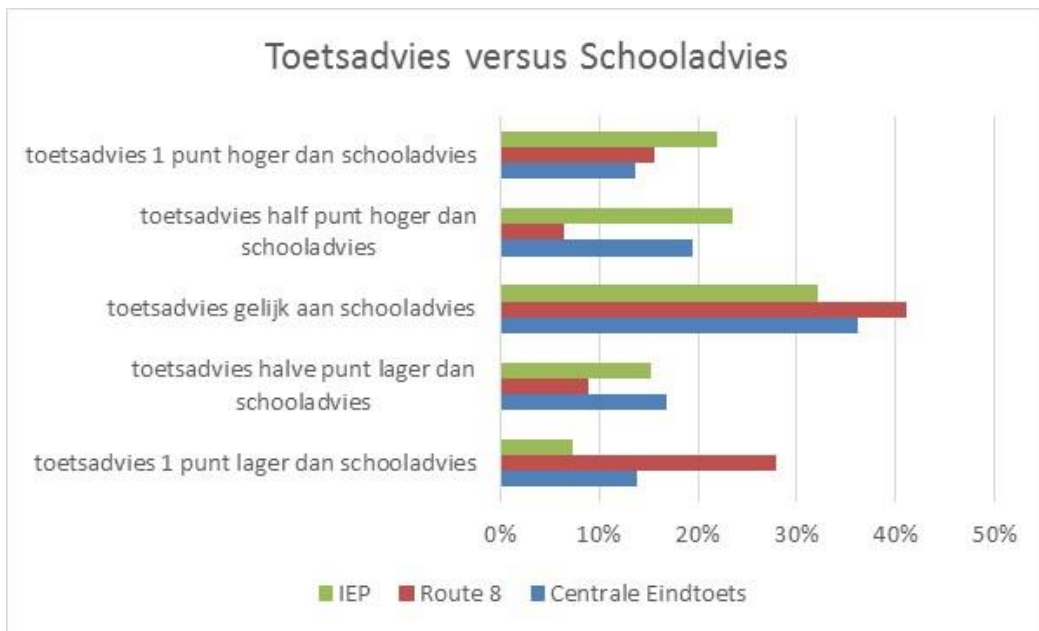
¹ Vanaf 2017 zal de IEP ook de gegevens ten aanzien van het schooladvies verzamelen onder de leerlingen van de operationele afname.

onder-advisering niet gelijkmatig verdeeld is. Onder leerlingen bij wie het toetsadvies afwijkt van het schooladvies is er bij Route 8 relatief vaker sprake van een lager toetsadvies en bij de IEP is er relatief vaker sprake van een hoger toetsadvies. Bij de Centrale Eindtoets zien we een gelijkmatige verdeling van hogere en lagere adviezen.

Tabel 7: Vergelijking Schooladvies en Toetsadviezen

Eindtoets	Toetsadvies in Vergelijking tot Schooladvies		
	Toetsadvies lager	Vergelijkbaar	Toetsadvies Hoger
Centrale Eindtoets	8,5%	83,5%	8,0%
Route 8	10,5%	86,4%	3,1%
IEP	4,6%	85,4%	10,0%

Een nuancering bij Tabel 7 is op zijn plaats. Omdat de oorspronkelijke categorieën zijn samengevoegd is een aantal gedetailleerde verschillen uitgemiddeld. Eén daarvan is dat bij Route 8 zo'n 7% van de leerlingen het advies voor een beroepsgerichte vmbo opleiding krijgt terwijl de leerkracht een theoretische leerweg adviseert. Een gedetailleerde overzicht van herzieningen is te vinden in onderstaand figuur die ontleend is aan de brief over de eindtoetsresultaten en de overgang van leerlingen naar het voortgezet onderwijs in schooljaar 2015-2016 van de staatssecretaris van OCW aan de Tweede Kamer. Figuur 1 geeft een iets genuanceerder beeld omdat de schooladviezen en eindtoetsadviezen niet gecombineerd zijn, zodat, bijvoorbeeld, het verschil tussen een meervoudig advies en een enkelvoudig advies kan resulteren in een verschil van 0,5.



Figuur 1: Verschillen tussen adviezen

In **Figuur 1** zien we dat Route 8 de hoogste overeenstemming laat zien volgens de meest strikte definitie (d.w.z., geen enkel verschil tussen toets- en schooladvies); hanteren we een iets minder strikte definitie en spreken we van overeenstemming als het schooladvies en toetsadvies niet meer dan een 0.5 punt afwijken dan is de Centrale Eindtoets het meest in overeenstemming het schooladvies en Route 8 het

minst. Kortom, resultaten zijn afhankelijk voor wat als 'in overeenstemming met elkaar' wordt gezien. Wel lijkt de algemene trend dat het toetsadvies van Centrale Eindtoets ongeveer even vaak hoger als lager zit dan het schooladvies, Route 8 vaker lager (maar omdat Route 8 een grovere indeling gebruikt kom je sneller in de extremen terecht) en IEP vaker hoger, alleen de percentages die in dit onderzoek zijn gepresenteerd zijn anders omdat de er een andere indeling is gebruikt.

Nadere bestudering van de gegevens laat een ander opmerkelijke resultaat zien die niet direct uit Tabel 7 of Figuur 1 af te lezen. Namelijk dat schooladviezen en toetsadviezen soms behoorlijk uiteen kunnen lopen. Er zijn leerlingen die schooladvies vmbo-bg vmbo-tl krijgen, maar vervolgens op de toets uitkomen op havo of zelfs vwo advies. Het is gissen naar de specifieke oorzaken voor dergelijke discrepanties, en het betreft geen grote aantallen leerlingen (dit zijn conditionele percentages, dus percentage leerlingen met vmbo schaaladvies die hoog toetsadvies krijgt), maar het feit dat er zulke grote verschillen kunnen bestaan onderschrijft de noodzaak van een betrouwbaar objectief tweede gegeven naast het schooladvies waarbij mogelijk ook achtergrondkenmerken meegenomen zijn.

Tabel 8 laat de percentages herzieningen zien. De IEP leidt relatief gezien het vaakst tot herziening van het oorspronkelijk schooladvies. Dit past in het beeld dat de IEP relatief vaker een hoger brugklatype adviseert dan de school. Bij de IEP zal het schooladvies dus vaker heroverwogen *moeten* worden en naar verwachting zal het vaker leiden tot herziening. Overigens dient opgemerkt te worden dat voor Route 8 de percentages herzieningen in Tabel 8 hoger liggen dan de genoemde percentages in Tabel 7 waarbij de toets een hoger advies geeft dan de leerkracht (3,1% tegenover 5,6 herzieningen). Dit komt omdat de percentages in Tabel 7 zijn gebaseerd op grovere definitie van afwijkingen tussen school- en toetsadvies dan op basis waarvan de herzieningen in Tabel 8 zijn gebaseerd. Als we echter kijken in Figuur 1 waar de percentages afwijkingen op fijnmazige indeling is gebaseerd, dan zien we dat de percentages afwijkingen veel hoger zijn, waarbij Route 8 vaker wat lager zit, en de IEP hoger.

Tabel 8: Percentages Herzieningen Schooladvies op Basis van Toetsadvies

	Nee	Ja
Centrale Eindtoets	93,2%	6,8%
Route 8	94,4%	5,6%
IEP	91,2%	8,8%

4. Vergelijking op Basis van Referentieniveaus

4.1 Achtergrond en Werkwijze

In dit hoofdstuk vergelijken we de behaalde referentieniveaus per toets. In 2010 is de Wet Referentieniveaus voor Nederlandse Taal en Rekenen in werking getreden (Stb. 2010, 265). Referentieniveaus geven aan wat een leerlingen moet kunnen op bepaalde momenten in zijn of haar schoolloopbaan. Voor uitstromende leerlingen van PO worden de niveaus 1F (fundamenteel niveau) en 2F/1S voor taal en 1F, 1S (streefniveau), en 2F voor rekenen onderscheiden. De niveaus 1S en 2F voor lezen en taalverzorging vallen samen. Voor rekenen is dat niet het geval. Referentieniveau 1 wordt gezien als het fundamentele niveau aan het einde van groep 8. Niveau 1F en 1S zijn van toepassing op het einde van het basisonderwijs. Het is de bedoeling dat zo veel mogelijk leerlingen aan het einde van het basisonderwijs een zo hoog mogelijk niveau bereiken, maar ten minste niveau 1F.

Voor deze rapportage is van belang dat bij de afnames in 2015-2016 niet elke eindtoets exact dezelfde meetpretenties had ten aanzien van bepaling van de referentieniveaus 2F en 1S voor het domein Rekenen. Route 8 en CET rapporteerden geen niveau 2F voor rekenen maar wel 1S. IEP rapporteerde voor rekenen naast het streefniveau 1S ook 2F. In tegenstelling tot het domein Taal zijn de twee niveaus

niet a priori gelijkwaardig. Analyses van de IEP (Langeveld et al., 2016) laten zien dat beide niveaus een gelijkwaardig beroep doen op iemands rekenvaardigheden (zie bijvoorbeeld de grensscores op de IEP).. Aangezien het hier gaat om streefniveaus, en omdat 2F/1S ook voor rekenen zeer waarschijnlijk dicht bij elkaar liggen, zullen we bij de duiding van de behaalde referentieniveaus 2F en 1S als één gezamenlijk (streef)categorie beschouwen zodat de resultaten tussen de toetsen te vergelijken zijn.

4.2 Resultaten

In de **Tabel 9** staan de percentages leerlingen die een bepaald referentieniveau hebben behaald², zowel ongewogen als gewogen volgens de methode van bijlage 1. Omdat de verschillen niet veranderen na controle voor regio en gewichtenleerlingen, beperken we ons weer tot de bespreking van de ongewogen resultaten. Het eerste paneel in de tabel geeft de behaalde referentieniveaus voor Lezen weer. Bij de Centrale Eindtoets en de IEP behaalt een hoog percentage van de leerlingen het hoogste niveau (respectievelijk 76,2% en 75,8% voor Lezen). Voor Route-8 behaalt een laag percentage van de leerling het hoogste niveau (voor Lezen is dit slechts 39,6%). Dit patroon zet zich voort in de twee overige domeinen.

Tabel 9: Vergelijking op Behaalde Referentieniveaus per Toets

	Centrale Eindtoets			Route 8			IEP ³		
	(n = 139.287)			(n = 11.814)			(n = 29.759)		
Niveau =	<1F	1F	2F/1S	<1F	1F	2F/1S	<1F	1F	2F/1S
Lezen									
Totaal	1,2%	22,6%	76,2%	18,6%	41,7%	39,6%	0,9%	23,3%	75,8%
Gewogen:									
Gewicht	1,1%	22,6%	76,2%	18,5%	41,7%	39,8%	0,6%	23,1%	76,3%
Regio	1,2%	22,6%	76,2%	18,0%	41,8%	40,2%	0,9%	23,4%	75,8%
Taalverzorging									
Totaal	4,2%	39,2%	56,6%	11,7%	50,1%	38,2%	3,8%	30,9%	65,4%
Gewogen:									
Gewicht	4,2%	39,2%	56,6%	11,7%	50,1%	38,3%	3,8%	30,9%	65,3%
Regio	4,2%	39,2%	56,6%	11,7%	49,3%	39,2%	3,7%	30,9%	65,4%
Rekenen									
Totaal	12,8%	43,4%	43,8%	32,1%	54,0%	13,9%	8,9%	46,7%	44,4%
Gewogen:									
Gewicht	12,8%	43,4%	43,8%	32,0%	54,0%	14,0%	8,9%	47,7%	44,4%
Regio	12,8%	43,4%	43,8%	31,9%	54,1%	13,9%	8,9%	46,7%	44,4%

<1F is onder fundamenteel, 1F is fundamenteel niveau 1, 1S is streefniveau 1, 2F is fundamenteel niveau 2

Het meest opvallend zijn de resultaten voor Route 8. De percentages leerlingen die onder 1F scoren voor elke vaardigheid (lezen, taalverzorging, rekenen) zijn erg hoog. Volgens deze gegevens zou 18,6% van uitstromende leerlingen die Route 8 hebben gemaakt niet het gewenste niveau voor lezen hebben, 11,7% niet het gewenste fundamentele niveau voor taalverzorging hebben, en 32,1% niet voor rekenen. Deze percentages liggen voor de Route 8 beduidend hoger dan de percentages voor de Centrale

² In de brief van de Staatsecretaris OCW aan de kamer (Vergaderjaar 2016-2017; 31289-31293-nr340 Figuur 2, blz. 3) worden de cumulatieve percentages genoemd terwijl in Tabel 9 de percentages zijn uitgesplitst. Hierdoor lijken de percentages anders, maar zijn dan niet. Bijvoorbeeld, in de kamerbrief wordt een percentage van 87% genoemd voor 1F leerlingen genoemd. Dit percentage bevat echter ook de 44% 1S leerlingen. Dit percentage is dan ook gelijk aan de afgeronde percentage van 43.4% + 43.8% = 87.2% in Tabel 9.

³ Percentages ten aanzien van referentieniveaus voor Taalverzorging bij de IEP zijn gebaseerd op een interne anker.

Eindtoets en IEP. Dit hoge percentage leerlingen dat 1F niet behaald, is opvallend omdat 1F het minimaal gewenste niveau is om naar vervolgonderwijs te gaan. De percentages voor Centrale Eindtoets en IEP zijn redelijk vergelijkbaar. De percentages verschillen het meeste voor het domein rekenen. De Centrale Eindtoets lijkt iets strenger te zijn dan IEP; voor de IEP halen relatief meer leerlingen 1F of hoger dan de Centrale Eindtoets.

Tabel 10 geeft de percentages met betrekking tot behaalde referentieniveaus uitgesplitst naar regio. De percentages variëren over de regio's, maar er is geen systematische trend waar te nemen. Er is dus nauwelijks samenhang tussen regio en prestaties ten aanzien van de referentieniveaus. Dit gebrek aan samenhang verklaart het minimale effect van wegen naar regio.

Tabel 10: Behaalde Referentieniveaus Per Toets Uitgesplitst naar Regio

	Centrale Eindtoets			Route 8			IEP		
	<1F	1F	1S/2F	<1F	1F	1S/2F	<1F	1F	1S/2F
Taal: Lezen									
Totaal	1,2%	22,6%	76,2%	18,6%	41,7%	39,6%	0,9%	23,3%	75,8%
Noord	0,9%	22,7%	76,3%	22,0%	42,4%	35,6%	0,8%	22,5%	76,7%
Oost	1,1%	23,3%	75,6%	15,5%	42,5%	41,9%	0,6%	23,1%	76,3%
Zuid	0,9%	20,7%	78,4%	18,7%	40,9%	40,4%	1,0%	24,3%	74,8%
West	1,3%	23,2%	75,5%	18,1%	41,6%	40,3%	0,9%	23,3%	75,8%
Taalverzorging									
Totaal	4,2%	39,2%	56,6%	11,7%	50,1%	38,2%	3,8%	30,9%	65,4%
Noord	4,7%	41,3%	54,0%	13,5%	52,8%	33,8%	3,4%	28,0%	68,6%
Oost	4,2%	40,3%	55,4%	9,9%	47,7%	42,3%	3,2%	29,7%	67,1%
Zuid	3,8%	37,8%	58,4%	11,2%	52,3%	36,5%	3,7%	32,2%	64,1%
West	4,3%	38,8%	56,9%	12,0%	47,9%	40,0%	4,0%	31,6%	64,4%
Rekenen									
Totaal	12,8%	43,4%	43,8%	32,1%	54,0%	13,9%	8,9%	46,7%	44,4%
Noord	12,3%	43,1%	44,7%	34,5%	52,7%	12,8%	7,8%	47,6%	44,5%
Oost	13,0%	44,7%	42,3%	29,7%	55,7%	14,6%	8,2%	46,7%	45,2%
Zuid	11,0%	43,3%	45,6%	30,7%	54,5%	14,8%	9,7%	47,0%	43,2%
West	13,6%	42,8%	43,5%	33,0%	53,4%	13,6%	9,2%	46,5%	44,4%

< 1F is onder functioneel, 1F is functioneel, 1S is Streefniveau, 2S is Functieniveau 2

5. Vergelijking Toetsadviezen en Behersing Referentieniveaus

5.1 Achtergrond en Werkwijze

Hoewel in de toetswijzer (CvTE, 2014a) geen bepalingen staan over het beheersen van bepaalde referentieniveaus als voorwaarde voor bepaalde schooladviezen, zijn de toets-onderdelen waarop referentieniveaus worden gerapporteerd direct of indirect mede van invloed op de totstandkoming van het toetsadvies. Dit betekent per definitie dat leerlingen met een hoger toetsadvies ook de referentieniveaus beter beheersen. De referentieniveaus voor de eerste toets-versies (afname april 2015) zijn geoperationaliseerd aan de hand van de toets-onafhankelijke openbare referentie- en ankersets in een headstart onderzoek. Deze referentiesets zijn in opdracht en onder auspiciën van het CvTE ontwikkeld door Stichting Cito. Voor het bepalen van de referentieniveaus voor de afname in 2016 zijn items uit de eigen toets uit 2015 (of itembank), of een eigen ontwikkelde anker gebruikt. De gebruikte werkwijze is te vinden in de wetenschappelijke verantwoordingen van elke toets (CvTE, 2015; Langeveld et al., 2016; Egberink et al., 2016).

Omdat de referentieniveaus in beginsel zijn geijkt op basis van toets-onafhankelijke referentie-items bieden deze referentieniveaus een geschikt uitgangspunt om ook de vergelijkbaarheid van de toetsadviezen te evalueren. Aangenomen wordt dat de vastgestelde cesuren op een adequate wijze zijn overgebracht op de drie eindtoetsen. Indien de ijking inderdaad adequaat heeft plaatsgevonden dan

zouden gegeven de toetsadviezen de percentages leerlingen die de verschillende referentieniveaus beheersen vergelijkbaar moeten zijn over de verschillende toetsen, waarbij wordt aangenomen dat de toetspopulaties vergelijkbaar zijn. De eindtoetsen maken echter verschillende keuzes in de operationalisering van de referentieniveaus, de keuze van items uit de referentiesets en de wijze waarop wordt geankerd aan de openbare referentie- en ankersets in termen van steekproef, statistisch model en pretestdesign. Deze keuzes kunnen ertoe leiden dat er verschillen ontstaan in beheersing van de referentieniveaus bij eenzelfde toetsadvies. Het doel van dit hoofdstuk is om daar wat meer zicht op te krijgen.

5.2 Resultaten

Tabel 11 laat per niveau van toetsadvies de percentages leerlingen die een bepaald referentieniveau behalen zien. Resultaten bevestigen vergelijkbaarheid van de Centrale Eindtoets en IEP, en de afwijkende resultaten voor Route 8. Voor de Centrale Eindtoets en IEP zien we ook redelijke overeenstemming tussen het advies en gewenste referentieniveaus van het brugklatype dat de leerling wordt geadviseerd; leerlingen die naar het vwo gaan beheersen bijvoorbeeld bijna allemaal streefniveau 1S. De resultaten laten zien dat Route 8 in dit opzicht streng is: de schooladviezen waren al relatief laag, en gegeven deze adviezen zijn de ingeschatte referentieniveaus nog lager. Opvallend is, dat bij Route 8 weinig leerlingen met een havo-advies het streefniveau 1S/2F halen. Bijvoorbeeld bij rekenen haalt slechts 12,8% van deze leerlingen het streefniveau 1S/2F, terwijl dat bij de Centrale Eindtoets en de IEP 56,8% is. Opvallend is ook dat er bij de IEP en Centrale Eindtoets enkele leerlingen zijn met een vwo-advies die op een vaardigheid *onder* 1F scoren. En bij elke eindtoets zien we ook dat er vwo-adviezen zijn voor leerlingen die enkel 1F behalen, (voor rekenen bijvoorbeeld IEP 5,8%, Centrale Eindtoets < 4,1% en Route 8 21,4%). Dit kunnen ook leerlingen zijn die 1F net niet gehaald hebben, maar op basis van prestaties op andere vlakken een vwo-advies hebben gekregen.

Tabel 11: Vergelijking Referentieniveaus en Toets-adviezen (2F/S samengevoegd)

	Centrale Eindtoets			Route 8			IEP		
	<1F	1F	1S/2F	<1F	1F	1S/2F	<1F	1F	1S/2F
Taal – Lezen									
Vmbo	2,6%	46,2%	51,2%	30,3%	46,3%	20,5%	2,0%	46,9%	51,2%
BG	4,3%	63,0%	32,6%	45,4%	44,6%	10,0%	4,7%	67,1%	28,2%
TL	0,1%	22,7%	77,2%	15,9%	53,7%	30,4%	0,1%	33,0%	66,9%
Havo	0,0%	6,2%	93,8%	4,3%	37,9%	57,8%	0,0%	8,3%	91,7%
Vwo	0,0%	0,8%	99,2%	1,3%	17,3%	81,5%	0,0%	0,9%	99,1%
Taal – Taalverzorging									
Vmbo	9,4%	70,1%	20,5%	19,0%	59,3%	21,7%	8,4%	56,8%	34,7%
BG	15,8%	76,7%	7,5%	27,7%	58,5%	13,7%	17,8%	66,1%	16,2%
TL	0,5%	60,8%	38,7%	10,7%	60,1%	29,2%	2,0%	50,5%	47,5%
Havo	0,1%	24,3%	75,7%	3,3%	45,1%	51,6%	0,2%	18,1%	81,7%
Vwo	0,0%	2,8%	97,2%	0,3%	20,3%	79,3%	0,0%	2,2%	97,7%
Rekenen									
Vmbo	28,5%	66,2%	5,4%	56,2%	43,7%	0,1%	20,2%	74,0%	5,8%
BG	45,2%	54,0%	0,7%	89,0%	10,9%	0,1%	46,1%	53,5%	0,4%
TL	5,0%	83,2%	11,8%	25,1%	74,8%	0,1%	2,4%	88,0%	9,6%
Havo	0,0%	42,7%	56,8%	1,1%	86,2%	12,7%	0,0%	43,1%	56,8%
Vwo	0,0%	4,1%	95,8%	0,0%	21,4%	78,5%	0,0%	5,8%	94,2%

Bovenstaande gegevens zijn gebaseerd op gewogen uitsplitsing van de gecombineerde categorieën. De gegevens in bovenstaande tabel wijken daardoor af van de percentages in de Rapportage Referentieniveaus 2015-2016 (Hoofdstuk 3) van het CvTE (CvTE, 2016).

6. Vergelijking van Overstappende Scholen

6.1 Achtergrond en Werkwijze

In dit hoofdstuk vergelijken we de toetsadviezen van 2015 en 2016 voor de scholen die zijn overgestapt naar een andere eindtoets. Als we er van uit gaan dat de gemiddelde prestatieniveaus van op de overstappende scholen niet substantieel is veranderd, dan kunnen we de resultaten van de betreffende eindtoetsen met elkaar vergelijken. Anders dan de andere onderzoeksmethoden vergelijken we hier de ene eindtoets van het vorige jaar met de andere eindtoets van dit jaar.

In totaal zijn 709 scholen overgestapt naar een andere eindtoets. Verreweg de meeste (693 scholen; 97.7%) betreft scholen die zijn overgestapt van de Centrale Eindtoets naar of Route 8 (213 scholen) of IEP (480 scholen). Slechts een gering aantal scholen is overgestapt van de Route 8 of IEP naar een andere eindtoets. Deze aantallen zijn te klein om betrouwbare conclusies te kunnen trekken. Daarom worden alleen de cijfers voor de scholen die van de Centrale Eindtoets naar Route 8 of IEP overstappen worden met elkaar vergeleken. Helaas is het daarmee niet mogelijk om de Route 8 en IEP direct met elkaar te vergelijken. Vergelijking van deze twee toetsen is wel mogelijk via de Centrale Eindtoets.

Bij het vergelijken van de resultaten dient weer rekening gehouden te worden met het feit dat de verschillende eindtoetsen verschillende categorieën voor de adviezen hanteren. Gekozen is om de basale indeling vmbo/havo/vwo te hanteren.

6.2 Resultaten

Tabel 12 laat de percentages adviezen zien per toets, per jaar. Twee zaken vallen op. Ten eerste zijn voor scholen die zijn overgestapt van Centrale Eindtoets in 2015 naar Route 8 in 2016 de resultaten vergelijkbaar. Dit is opvallend omdat de andere resultaten suggereren dat Route 8 juist strenger is of relatief vaker wordt afgenomen in een minder vaardige populatie. Dus je zou verwachten dat met het overstappen van CET naar Route 8 het percentage vwo adviezen daalt en vmbo adviezen stijgt. Het eerste is deels het geval. Voor vmbo zien we juist een lichte daling. Het tweede opvallende resultaat zijn de percentages voor de IEP. Die laten zien dat de IEP vaker tot een vwo-advies heeft geleid dan de Centrale Eindtoets die er voor werd gebruikt. Dit resultaat is consistent met eerdere resultaten waarbij we zagen dat de IEP bij rekenen en lezen relatief hogere percentages 1F en 2F niveaus liet zien en relatief hogere adviezen, die, zoals uit deze vergelijking blijkt, vaak hoger uitkomen dan het advies van de leerkrachten die tot voor kort voornamelijk met de Eindtoets Basisonderwijs (citotoets; voorloper van de Centrale Eindtoets) werkten. Verschillen zijn mogelijk toe te schrijven aan toevallige variatie in de vaardigheid van cohorten leerlingen over de jaren heen en er is sprake van zelfselectie (scholen kiezen er zelf voor om over te stappen van toets). Niettemin is er sprake van een forse verschuiving waarbij een effect van de eindtoets niet kan worden uitgesloten. Omdat verder cijfers over de variatie van toetsadviezen binnen de scholen over jaren heen niet beschikbaar waren kunnen we daar verder geen uitspraken over doen.

Tabel 12: Vergelijking Toets-Adviezen Overstappende Scholen

	Centrale Route 8	Eindtoets -> Route8- 2016	Centrale Eindtoets -> IEP	Centrale Eindtoets -> IEP
	Centrale Eindtoets 2015	- 2016	Centrale Eindtoets -2015	IEP-2016
# leerlingen	5664	5589	11724	11849
Vmbo	55,1%	54,6%	55,5%	40,8%
Havo	28,9%	31,6%	28,9%	30,2%
Vwo	16,0%	13,7%	15,5%	29,1%

Voetnoot: #scholen: 213 & 480.

7. Conclusies en aanbevelingen

7.1 Conclusies

In het schooljaar 2014-2015 zijn er, naast de Centrale Eindtoets, twee andere eindtoetsen in het PO toegelaten (Route 8 en de IEP Eindtoets). Zowel bij de afname in 2015 als in 2016 is gebleken dat de resultaten op drie eindtoetsen onderling verschillen. Er zijn verschillen in de niveaus van de adviezen voor vervolgonderwijs, in de mate waarin het toetsadvies overeenkomt met het advies van de basisschool en in de resultaten van de gemeten referentieniveaus. Omdat niet duidelijk was of deze verschillen zijn toe te schrijven aan de niveauverschillen van de populaties van de verschillende eindtoetsen, of aan de verschillen tussen de eindtoetsen zelf, heeft het ministerie van OCW aan de Expertgroep Toetsing PO twee opdrachten te geven. Ten eerste om de gevonden verschillen tussen de eindtoetsen nader te duiden. En ten tweede om het ministerie OCW te adviseren hoe de vergelijkbaarheid van de eindtoetsen en de ankerings aan de referentieniveaus kan worden verbeterd. Het voorliggende rapport behelst de eerste opdracht, de tweede opdracht heeft geresulteerd in een tweede rapport (Glas et al., 2016). Beide rapporten hangen samen en kunnen als een samenhangend geheel worden gelezen.

De eerste vraag die in dit rapport aan de orde komt is of de verschillende eindtoetsen ook verschillende populaties bedienen. Dit is onderzocht voor de regionale spreiding van de toetsen en de gewichtenleerlingen in de verschillende toetspopulaties. Met betrekking tot gewichtenleerlingen zijn geen belangrijke verschillen tussen de toetsen gevonden. Bij de regionale spreiding valt op dat Route 8 relatief vaker afgenomen wordt in regio's Noord en Zuid en minder vaak in regio's West en Oost, terwijl de IEP relatief vaak in Regio West wordt afgenomen. De Centrale Eindtoets is in alle regio's evenredig gespreid. Verschillen op andere achtergrondvariabelen, zoals de denominatie van de school, schoolgrootte en urbanisatiegraad zijn goed verklaarbaar uit de regionale spreiding van de toetsen en leveren geen aanvullende opvallende verschillen op. Bij het onderzoek naar de vraag of de achtergrondvariabelen de verschillende uitkomsten van de toetsen kunnen verklaren, is steeds gewogen naar regio en leerlinggewicht. Echter, omdat de toetsen alleen op de variabele regio verschillen, werd alleen daar een mogelijke verklaring verwacht.

Analyse van de toetsadviezen liet het eerder geschetste beeld zien. Route 8 adviseerde relatief lager dan de andere toetsen (meer vmbo, minder vwo). Bij de IEP werden iets meer vwo adviezen gegeven, maar dat effect was veel kleiner dan het tegenovergestelde effect zoals gevonden bij Route 8. Conditioneren op regio en leerlinggewichten veranderden de gevonden effecten niet. Met andere woorden, regio en leerlinggewicht konden de gevonden verschillen niet verklaren.

De analyse van de samenhang tussen toets- en schooladvies versterken het beeld van de niveauverschillen tussen de toetsen. Route 8 gaf vaker een lager advies dan de school, en de IEP gaf vaker een hoger advies. De Centrale Eindtoets adviseert relatief even vaak hoger dan de school, als dat het lager adviseert dan de school.

Dit beeld wordt verder bevestigd door een analyse van veranderingen in adviezen bij een overstap van de Centrale Eindtoets naar Route 8 of naar de IEP. Het aantal vwo adviezen bij een overstap naar Route 8 daalt van 16,0% naar 13,7%, en het aantal vwo adviezen stijgt van 15,5% naar 29,1% bij een overstap naar de IEP. Het gaat hierbij echter respectievelijk om 213 en 480 scholen, die bovendien zelf de keuze hebben gemaakt om over te stappen, dus deze resultaten zijn minder betrouwbaar dan de resultaten die uitgerekend zijn op alle relevante eindtoetsleerlingen en dienen daardoor met de nodige terughoudendheid worden bekeken.

Met betrekking tot de referentieniveaus moet geconstateerd worden dat de percentages leerlingen die bepaalde niveaus haalden bij de Centrale Eindtoets en de IEP vergelijkbaar waren. Bij Route 8 was het

percentage leerlingen dat niveau 1F niet behaalde steeds beduidend hoger. Conditioneren op regio of leerlinggewicht veranderde hier niets aan, dus ook hier konden regio en leerlinggewicht de gevonden verschillen niet verklaren. Opgemerkt dient te worden dat de uitzonderingspositie van Route 8 bij de referentieniveaus duidelijker was dan bij de toetsadviezen. Vergelijking van de referentieniveaus met de schooladviezen bevestigde dit verder. De resultaten laten bijvoorbeeld zien dat bij Route 8 weinig leerlingen met een havo-advies het streefniveau 1S/2F halen. Bij rekenen haalt slechts 12,8% van deze leerlingen het streefniveau 1S/2F, terwijl dat bij de CET en de IEP respectievelijk 56,8% en 56,8% is.

De algemene conclusie van het onderzoek is dat

1. Er verschillen tussen verschillende eindtoetsen zijn in de adviezen voor vervolgonderwijs die niet verklaard kunnen worden door verschillen tussen de toetspopulaties, althans niet uit de voor dit onderzoek beschikbare achtergrondvariabelen;
2. Dit ook geldt voor de referentieniveaus, met de opmerking dat de verschillen hier veel groter zijn.

Bij deze conclusies moet worden aangetekend dat in 2016 voor de Centrale Eindtoets de ankering van de referentieniveaus is gedaan op data uit de operationele (high-stakes) afname, en dat zowel Route 8 als de IEP maatregelen hebben getroffen om hun procedures voor het vaststellen van cesuren voor adviezen voor vervolgonderwijs en referentieniveaus te verbeteren. Beide aanbieders identificeren het feit dat hun proeftoetsing plaatsvindt in een "low-stakes" situatie die geen goede afspiegeling is van de "high-stakes" echte toetssituatie tot onnauwkeurigheid blijkt te leiden. Beide hebben in hun toetsverantwoordingen voor 2017 dan ook procedures voorgesteld waarbij nieuwe opgaven tegelijk met een operationele eindtoets worden afgenomen. Hierdoor worden nieuwe opgaven dus gepretest in een "high-stakes" situatie. De Expertgroep staat positief tegenover dit initiatief.

7.2 Aanbevelingen

Naar aanleiding van de hierboven geformuleerde conclusies en de in het additionele rapport **Scenario's voor ijking van de eindtoetsen op de referentieniveaus** (Glas et al., 2016) onderzochte mogelijkheden worden de volgende aanbevelingen gedaan ter vergroting van de vergelijkbaarheid van de eindtoetsen.

1. Het uniformeren van de psychometrische verantwoording van de eindtoetsen, met betrekking tot rapportages, gebruikte statistische methoden en de daarbij behorende software.
2. Een onderzoek naar de verschillen tussen de verschillende cesuren voor de schooladviezen en referentieniveaus door alle toetsen (dus ook de AMN, CESAN en Dia eindtoets) op een gezamenlijke schaal te kalibreren, en door onderzoek te doen naar de betrouwbaarheid van die ankering.
3. In alle eindtoetsen een gezamenlijk anker op te nemen als deel van iedere eindtoets.

Voor een gedetailleerde beschrijving van deze voorstellen wordt verwezen naar het rapport **Scenario's voor ijking van de eindtoetsen op de referentieniveaus** van de Expertgroep (Glas et al., 2016). Tenslotte leidt het hier gerapporteerde onderzoek tot nog een tweetal additionele aanbevelingen ter vergroting van de vergelijkbaarheid van de eindtoetsen.

1. Het standaardiseren van de door de eindtoetsen gebruikte schoolcategorieën. Hoewel hier al stappen in worden ondernomen is het van belang hier op korte termijn met alle toetsaanbieders overeenstemming over te krijgen.
2. Nader overleg om te komen tot een onderzoeksvoorstel over de plaatsing van de cesuur voor de overeengekomen schoolcategorieën, conform de voorstellen in het rapport **Scenario's**

voor ijking van de eindtoetsen op de referentieniveaus van de Expertgroep (Glas et al., 2016).

8. Literatuurlijst

- CvTE (2014a). *Algemeen deel Toetswijzer voor eindtoets PO. Inhoudelijke kwaliteitseisen aan eindtoetsen PO*. Utrecht: CvTE.
- CvTE (2015). *Wetenschappelijke verantwoording Centrale Eindtoets 2014-2015*. Utrecht: CvTE.
- CvTE (2016). *Rapportage referentieniveaus 2015-2016*. Utrecht: CvTE.
- Egberink, A., Verweij, J. & Termaat, B.R. (2016). *Verantwoording van ROUTE 8*. Apeldoorn: A-VISION Holding B.V.
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN beoordelingsstelsel voor de kwaliteit van tests*. Utrecht: Nederlands Instituut voor Psychologen.
- Folmer, E., Boswinkel, N., Prenger, J. & Vorle, R. van de (2014). *Beoordelingskader onderwijskundige en organisatorische aspecten andere eindtoetsen*. Utrecht: Expertgroep toetsen PO
- Glas, C.A.W., Emons, W.H.M. Emons & Berding-Oldersma, P.K. (2016). *Scenario's voor ijking van de eindtoetsen op de referentieniveaus*. Utrecht: Expertgroep Toetsen PO
- Langeveld, E.A., Bezdan, E., Binsbergen, M., Haitjema, T., Laarhuis, R., Ebert-Flikweert, L., Van den Ouden, L., Helsloot, J. & Bonouvrie, N. (2016). *IEP Eindtoets verantwoording 2016-2017*. Culemborg: Bureau ICE.
- Ministerie van Onderwijs, Cultuur en Wetenschappen (2010). *Besluit referentieniveaus Nederlandse taal en rekenen*. Den Haag: Ministerie van OCW.
- Ministerie van Onderwijs, Cultuur en Wetenschappen (2014). *Toetsbesluit PO*. Den Haag: Ministerie van OCW.
- Wools, S. & Béguin, A. (2013). *Handleiding referentieset*. Arnhem: Cito.

Bijlage 1. De Weging naar Leerlinggewichten en Regio's

Tabel 1, 2 en 3 worden in dit onderzoek gecorrigeerd door te wegen ten opzichte van de verdeling van gewichtenleerlingen op een school in de hele populatie. Aan de hand van een klein, fictief voorbeeld wordt dit nader toegelicht. Eerst splitsen we de resultaten per eindtoets op naar percentage gewichtenleerlingen. Tabel 1.1 toont een dergelijke opsplitsing voor een eindtoets.

Tabel 1.1. Fictief voorbeeld van opsplitsen van vmbo-bb-toetsadvies naar percentage gewichtenleerlingen voor één eindtoets

Brugklastype	Percentage gewichtenleerlingen		
	0%	10%	20%
VMBO	10%	11%	12%

In Tabel 1.1 zien we het percentage leerlingen op een eindtoets met het advies vmbo-bb opgesplitst in drie groepen: Scholen met 0%, 10% of 20% gewichtenleerlingen. Stel dat de hele populatie bestaat uit de helft scholen zonder gewichtenleerlingen, een kwart scholen met 10% gewichtenleerlingen en een kwart scholen met 20% gewichtenleerlingen, dan krijgt deze eindtoets het gewogen percentage leerlingen met een vmbo-bb-toetsadvies gelijk aan $0,5 * 10\% + 0,25 * 11\% + 0,25 * 12\% = 10,75\%$. Op deze wijze kunnen aanpassingen van Tabel 1, 2 en 3 worden geconstrueerd die inzicht geven hoe de eindtoetsresultaten zich tot elkaar verhouden gecontroleerd voor verschillen in percentage gewichtenleerlingen. Hieronder zijn de tabellen weergegeven die ten grondslag liggen aan Tabel 4 en Tabel 10 hierboven.

Tabel 1.2: Verband Adviezen en Percentages Gewicht-leerlingen

	0%		1%-10%		>10%		Totaal		
	Aantal	%	Aantal	%	Aantal	%	Aantal	Ongewogen Gewogen	
Centrale Eindtoets									
Vmbo	2,403	32.5%	38,669	41.6%	20,874	53.6%	61,945	44.47%	44.46%
BG	1,209	16.4%	21,492	23.1%	13,406	34.5%	36,107	25.92%	25.92%
TL	1,194	16.1%	17,177	18.5%	7,468	19.2%	25,838	18.55%	18.54%
Havo	2,436	32.9%	29,202	31.4%	10,514	27.0%	42,151	30.26%	30.25%
Vwo	2,555	34.6%	25,110	27.0%	7,526	19.3%	35,191	25.26%	25.28%
Totaal	7,393	100.0%	92,980	100.0%	38,914	100.0%	139,287	100.0%	100.0%
Route 8									
Vmbo	431	48.7%	3,930	53.3%	2,327	65.3%	6,688	56.61%	56.47%
BG	202	22.8%	1,800	24.4%	1,253	35.2%	3,255	27.55%	27.38%
TL	229	25.9%	2,130	28.9%	1,074	30.2%	3,433	29.06%	29.08%
Havo	309	34.9%	2,445	33.2%	883	24.8%	3,637	30.79%	30.91%
Vwo	145	16.4%	992	13.5%	352	9.9%	1,489	12.60%	12.62%
Totaal	885	100.0%	7,367	100.0%	3,562	100.0%	11,814	100.0%	100.0%
IEP									
Vmbo	885	35.6%	7,688	41.6%	4,498	51.2%	13,071	43.92%	43.96%
BG	279	11.2%	3,020	16.3%	2,023	23.0%	5,323	17.89%	17.93%
TL	606	24.4%	4,668	25.2%	2,475	28.2%	7,749	26.04%	26.03%
Havo	802	32.2%	5,559	30.1%	2,396	27.3%	8,757	29.43%	29.41%
Vwo	800	32.2%	5,243	28.4%	1,887	21.5%	7,931	26.65%	26.64%
Totaal	2,487	100.0%	18,491	100.0%	8,781	100.0%	29,759	100.0%	100.0%

Tabel 1.3: Verband Adviezen en Percentages Leerling-gewicht

	0.0		0.3		1.2		Totaal		
	Aantal	%	Aantal	%	Aantal	%	Aantal	Ongewogen	Gewogen
Centrale Eindtoets									
Vmbo	52,313	41.60%	5,307	71.30%	4,275	69.80%	61,895	44.50%	44.40%
<i>BG</i>	29,222	23.30%	3,794	51.00%	3,058	49.90%	36,074	25.90%	25.92%
<i>TL</i>	23,091	18.40%	1,513	20.30%	1,217	19.90%	25,821	18.50%	18.57%
Havo	39,312	31.30%	1,518	20.40%	1,297	21.20%	42,128	30.30%	30.28%
Vwo	34,011	27.10%	615	8.30%	556	9.10%	35,182	25.30%	25.32%
Totaal	125,637	100.00%	7,440	100.00%	6,128	100.00%	139,205	100.00%	100.0%
Route 8									
Vmbo	5,792	54.10%	517	81.20%	371	80.70%	6,680	56.60%	56.69%
<i>BG</i>	2,683	25.10%	338	53.10%	229	49.80%	3,250	27.50%	27.66%
<i>TL</i>	3,109	29.00%	179	28.10%	142	30.90%	3,430	29.10%	29.03%
Havo	3,476	32.50%	94	14.80%	65	14.10%	3,635	30.80%	30.76%
Vwo	1,439	13.40%	26	4.10%	24	5.20%	1,489	12.60%	12.55%
Totaal	10,707	100.00%	637	100.00%	460	100.00%	11,804	100.00%	100.0%
IEP									
Vmbo	11,229	41.60%	1,041	66.90%	786	64.60%	13,057	43.90%	43.94%
<i>BG</i>	4,345	16.10%	557	35.80%	408	33.60%	5,311	17.90%	17.91%
<i>TL</i>	6,884	25.50%	484	31.10%	378	31.10%	7,746	26.00%	26.04%
Havo	8,132	30.20%	339	21.80%	282	23.20%	8,754	29.40%	29.45%
Vwo	7,606	28.20%	176	11.30%	148	12.10%	7,929	26.70%	26.60%
Totaal	26,967	100.00%	1,556	100.00%	1,216	100.00%	29,739	100.00%	100.0%

Tabel 1.4: Verband Adviezen en Regio's

	Noord		Oost		Zuid		West		Totaal		
	Aantal	%	Aantal	%	Aantal	%	Aantal	%	Ongewogen	Gewogen	
Centrale Eindtoets											
Vmbo	5,343	44,8%	15,782	45,8%	12,330	42,24%	28,491	44,71%	61,946	44,5%	44,5%
<i>BG</i>	3,062	25,7%	9,171	26,6%	6,917	23,69%	16,957	26,61%	36,107	25,9%	25,9%
<i>TL</i>	2,281	19,1%	6,611	19,2%	5,413	18,54%	11,534	18,10%	25,839	18,6%	18,5%
Havo	3,657	30,7%	10,525	30,6%	9,087	31,13%	18,881	29,63%	42,150	30,3%	30,2%
Vwo	2,921	24,5%	8,136	23,6%	7,776	26,64%	16,358	25,67%	35,191	25,3%	25,3%
Totaal	11,921	100,0%	34,443	100,0%	29,193	100,00%	63,730	100,00%	139,287	100,0%	100,0%
Route 8											
Vmbo	1,494	60,9%	1,050	53,4%	1,838	57,26%	2,306	55,09%	6,688	56,6%	55,7%
<i>BG</i>	747	30,5%	471	24,0%	872	27,17%	1,165	27,83%	3,255	27,6%	27,1%
<i>TL</i>	747	30,5%	579	29,5%	966	30,09%	1,141	27,26%	3,433	29,1%	28,7%
Havo	681	27,8%	674	34,3%	962	29,97%	1,320	31,53%	3,637	30,8%	31,5%
Vwo	278	11,3%	241	12,3%	410	12,77%	560	13,38%	1,489	12,6%	12,8%
Totaal	2,453	100,0%	1,965	100,0%	3,210	100,00%	4,186	100,00%	11,814	100,0%	100,0%
IEP											
Vmbo	1,733	42,7%	1,965	43,5%	1,816	45,51%	7,558	43,94%	13,072	43,9%	44,0%
<i>BG</i>	679	16,7%	762	16,9%	746	18,70%	3,135	18,22%	5,322	17,9%	17,9%
<i>TL</i>	1,053	26,0%	1,203	26,7%	1,071	26,84%	4,422	25,71%	7,749	26,0%	26,2%
Havo	1,209	29,8%	1,330	29,5%	1,142	28,62%	5,076	29,51%	8,757	29,4%	29,4%
Vwo	1,112	27,4%	1,219	27,0%	1,032	25,86%	4,568	26,56%	7,931	26,6%	26,6%
Totaal	4,054	100,0%	4,514	100,0%	3,990	100,00%	17,202	100,00%	29,760	100,0%	100,0%

Tabel 1.5: Referentieniveaus en Percentages Gewicht-leerlingen

	0%		1%-10%		>10%		Ongewogen	Gewogen
	Aantal	%	Aantal	%	Aantal	%		
Lezen: Centrale eindtoets								
<1F	28	0.4%	729	0.8%	851	2.2%	1.15%	1.16%
1F	1,093	14.8%	18,341	19.7%	12,087	31.1%	22.63%	22.64%
2F	6,272	84.8%	73,910	79.5%	25,976	66.8%	76.22%	76.20%
Totaal	7,393	100.0%	92,980	100.0%	38,914	100.0%	100.0%	100.0%
Lezen: Route 8								
<1F	129	14.6%	1,208	16.4%	863	24.2%	18.62%	18.51%
1F	380	42.9%	2,997	40.7%	1,554	43.6%	41.74%	41.65%
2F	376	42.5%	3,162	42.9%	1,145	32.1%	39.64%	39.84%
Totaal	885	100.0%	7,367	100.0%	3,562	100.0%	100.0%	100.0%
Lezen: IEP								
<1F	10	0.4%	134	0.7%	115	1.3%	0.87%	0.87%
1F	363	14.6%	3,901	21.1%	2,666	30.4%	23.29%	23.34%
2F	2,114	85.0%	14,456	78.2%	6,000	68.3%	75.84%	75.79%
Totaal	2,487	100.0%	18,491	100.0%	8,781	100.0%	100.0%	100.0%
Taalverzorging: Centrale Eindtoets								
<1F	191	2.6%	3,349	3.6%	2,306	5.9%	4.20%	4.20%
1F	2,393	32.4%	35,148	37.8%	17,069	43.9%	39.21%	39.20%
2F	4,809	65.0%	54,483	58.6%	19,539	50.2%	56.60%	56.60%
Totaal	7,393	100.0%	92,980	100.0%	38,914	100.0%	100.0%	100.0%
Taalverzorging: Route 8								
<1F	97	11.0%	770	10.5%	520	14.6%	11.74%	11.66%
1F	436	49.3%	3,649	49.5%	1,832	51.4%	50.08%	50.05%
2F	352	39.8%	2,948	40.0%	1,210	34.0%	38.18%	38.29%
Totaal	885	100.0%	7,367	100.0%	3,562	100.0%	100.0%	100.0%
Taalverzorging: IEP								
<1F	56	2.3%	646	3.5%	420	4.8%	3.77%	3.79%
1F	672	27.0%	5,462	29.5%	3,055	34.8%	30.88%	30.88%
2F	1,759	70.7%	12,383	67.0%	5,306	60.4%	65.35%	65.34%
Totaal	2,487	100.0%	18,491	100.0%	8,781	100.0%	100.0%	100.0%
Rekenen: Centrale Eindtoets								
<1F	551	7.5%	10,431	11.2%	6,863	17.6%	12.81%	12.81%
1F	2,755	37.3%	39,982	43.0%	17,734	45.6%	43.41%	43.39%
1S	4,077	55.1%	42,229	45.4%	14,169	36.4%	43.42%	43.44%
2F	10	0.1%	338	0.4%	148	0.4%	0.36%	0.35%
Totaal	7,393	100.0%	92,980	100.0%	38,914	100.0%	100.0%	100.0%
Rekenen: Route 8								
<1F	235	26.6%	2,165	29.4%	1,396	39.2%	32.13%	32.00%
1F	513	58.0%	4,109	55.8%	1,753	49.2%	53.96%	54.05%
1S	137	15.5%	1,073	14.6%	413	11.6%	13.74%	13.78%
Totaal	885	100.0%	7,367	100.0%	3,562	100.0%	100.0%	100.0%
Rekenen: IEP								
<1F	116	4.7%	1,488	8.0%	1,043	11.9%	8.89%	8.93%
1F	1,112	44.7%	8,419	45.5%	4,375	49.8%	46.73%	46.70%
1S	120	4.8%	980	5.3%	415	4.7%	5.09%	5.11%
2F	1,139	45.8%	7,604	41.1%	2,948	33.6%	39.29%	39.26%
Totaal	2,487	100.0%	18,491	100.0%	8,781	100.0%	100.0%	100.0%

Tabel 1.6: Referentieniveaus en Regio's

	Noord		Oost		Zuid		West		Ongewogen	Gewogen
	Aantal	%	Aantal	%	Aantal	%	Aantal	%		
Lezen: Centrale Eindtoets										
<1F	112	.9%	370	1.1%	270	.9%	856	1.3%	1.2%	1.2%
1F	2,710	22.7%	8,020	23.3%	6,035	20.7%	14,756	23.2%	22.6%	22.6%
2F	9,099	76.3%	26,053	75.6%	22,888	78.4%	48,118	75.5%	76.2%	76.2%
Totaal	11,921	100.0%	34,443	100.0%	29,193	100.0%	63,730	100.0%	100.0%	100.0%
Lezen: Route 8										
<1F	539	22.0%	305	15.5%	599	18.7%	757	18.1%	18.6%	18.0%
1F	1,040	42.4%	836	42.5%	1,314	40.9%	1,741	41.6%	41.7%	41.8%
2F	874	35.6%	824	41.9%	1,297	40.4%	1,688	40.3%	39.6%	40.2%
Totaal	2,453	100.0%	1,965	100.0%	3,210	100.0%	4,186	100.0%	100.0%	100.0%
Lezen: IEP										
<1F	33	.8%	27	.6%	39	1.0%	160	.9%	.9%	.9%
1F	913	22.5%	1,041	23.1%	968	24.3%	4,008	23.3%	23.3%	23.4%
2F	3,108	76.7%	3,446	76.3%	2,983	74.8%	13,033	75.8%	75.8%	75.8%
Totaal	4,054	100.0%	4,514	100.0%	3,990	100.0%	17,201	100.0%	100.0%	100.0%
Taalverzorging: Centrale Eindtoets										
<1F	557	4.7%	1,451	4.2%	1,105	3.8%	2,733	4.3%	4.2%	4.2%
1F	4,923	41.3%	13,895	40.3%	11,039	37.8%	24,753	38.8%	39.2%	39.2%
2F	6,441	54.0%	19,097	55.4%	17,049	58.4%	36,244	56.9%	56.6%	56.6%
Totaal	11,921	100.0%	34,443	100.0%	29,193	100.0%	63,730	100.0%	100.0%	100.0%
Taalverzorging: Route 8										
<1F	330	13.5%	195	9.9%	358	11.2%	504	12.0%	11.7%	11.7%
1F	1,294	52.8%	938	47.7%	1,679	52.3%	2,006	47.9%	50.1%	49.3%
2F	829	33.8%	832	42.3%	1,173	36.5%	1,676	40.0%	38.2%	39.2%
Totaal	2,453	100.0%	1,965	100.0%	3,210	100.0%	4,186	100.0%	100.0%	100.0%
Taalverzorging: IEP										
<1F	139	3.4%	144	3.2%	147	3.7%	692	4.0%	3.8%	3.7%
1F	1,135	28.0%	1,341	29.7%	1,284	32.2%	5,429	31.6%	30.9%	30.9%
2F	2,780	68.6%	3,029	67.1%	2,559	64.1%	11,080	64.4%	65.4%	65.4%
Totaal	4,054	100.0%	4,514	100.0%	3,990	100.0%	17,201	100.0%	100.0%	100.0%
Rekenen: Centrale Eindtoets										
<1F	1,465	12.3%	4,464	13.0%	3,248	11.1%	8,668	13.6%	12.8%	12.8%
1F	5,134	43.1%	15,401	44.7%	12,630	43.3%	27,306	42.8%	43.4%	43.4%
1S	5,265	44.2%	14,425	41.9%	13,287	45.5%	27,498	43.1%	43.4%	43.4%
2F	57	.5%	153	.4%	28	.1%	258	.4%	.4%	.4%
Totaal	11,921	100.0%	34,443	100.0%	29,193	100.0%	63,730	100.0%	100.0%	100.0%
Rekenen: Route 8										
<1F	846	34.5%	584	29.7%	985	30.7%	1,381	33.0%	32.1%	31.9%
1F	1,293	52.7%	1,094	55.7%	1,751	54.5%	2,237	53.4%	54.0%	54.1%
1S	304	12.4%	285	14.5%	469	14.6%	565	13.5%	13.7%	13.9%
Totaal	2,453	100.0%	1,965	100.0%	3,210	100.0%	4,186	100.0%	100.0%	100.0%
Rekenen: IEP										
<1F	318	7.8%	369	8.2%	386	9.7%	1,574	9.2%	8.9%	8.9%
1F	1,931	47.6%	2,106	46.7%	1,877	47.0%	7,992	46.5%	46.7%	46.7%
1S	200	4.9%	306	6.8%	169	4.2%	840	4.9%	5.1%	5.1%
2F	1,605	39.6%	1,733	38.4%	1,558	39.0%	6,795	39.5%	39.3%	39.3%
Totaal	4,054	100.0%	4,514	100.0%	3,990	100.0%	17,201	100.0%	100.0%	100.0%

Bijlage 2. Het samenvoegen van Schooladviezen

Het samenvoegen van schooladviezen is gedaan op basis van de verdeling van de toets-scores (i.e., percentielscores weergegeven in **Tabel 2.1**) per toetsadvies en voor de samengestelde adviezen is gebruik gemaakt van de aanname van normaal verdeelde scores binnen de niveaus van Toetsadviezen en lineaire interpolatie.

Tabel 2.1: Percentielscores horend bij Toetsadviezen

Toets Advies	Gemiddelde	Minimum	Maximum	SD
Centrale Eindtoets				
Bbl	3.42	0.07	6.36	1.97
Bbl en kbl	12.40	7.37	16.98	3.18
Kbl	21.95	19.07	24.24	2.15
Gl/tl	31.77	27.30	36.05	3.27
Gl/tl en havo	44.85	39.22	50.40	4.04
Havo	57.41	53.85	60.94	2.89
Havo/vwo	71.86	64.51	79.38	5.16
Vwo	90.50	82.56	97.48	5.39
Route 8				
PrO	0.35	0.01	0.67	0.20
BBL	4.66	0.75	8.36	2.29
KBL	18.09	8.87	26.97	5.46
TL	42.09	28.15	55.81	8.38
HAVO	72.01	57.38	86.96	8.88
VWO	93.70	87.78	100.00	3.64
IEP				
Vmbo-b en vmbo k/b	6.26	1.11	12.14	3.59
Vmbo-k en vmbo-k/t	17.64	12.91	22.34	2.93
Vmbo-t	29.37	23.28	35.10	3.75
Vmbo-t/havo	41.51	36.65	46.43	3.10
Havo	57.23	48.37	66.11	5.74
Havo-vwo	72.96	68.14	78.06	3.13
Vwo	89.24	79.48	96.95	6.11

Ter illustratie van de lineaire interpolatie geven we een voorbeeld hoe we de gecombineerde havo-vwo advies van de Centrale Eindtoets hebben opgesplitst. De lineaire interpolatie is gebaseerd op de verdeling binnen advies havo-vwo en aangrenzende categorieën. Voor de Centrale Eindtoets zagen de verdelingen er als volgt uit:

Advies	Gemiddelde	SD
Havo	57.41	2.89
Havo-vwo	71.86	5.16
Vwo	90.50	5.39

We zien dat de gemiddelde score voor 'havo-vwo' dichter bij havo ligt dan bij vwo. We zullen dan ook proportioneel meer leerlingen toewijzen aan havo dan vwo. De exacte proportie is berekend onder de aanname van een normale verdeling: $P(havo) = P\left(Z < \left(\frac{57.41+90.50}{2}\right) \middle| \mu = 71.86, \sigma = 5.16\right) = 0.6576$. Deze

proportie kan als volgt worden geïnterpreteerd: naar verwachting zit 65.76% van de leerlingen met gecombineerd havo-vwo advies met hun score onder het midden tussen have en vwo adviezen in. Deze leerlingen neigen meer naar havo dan naar vwo en worden daarom bij de havo-adviezen gerekend. De overige gewichten zijn te vinden in onderstaande figuur. Hierbij moet worden opgemerkt dat de gewichten op vier decimalen nauwkeurig zijn bepaald om de berekeningen zou nauwkeurig mogelijk uit te kunnen voeren, maar we pretenderen daarmee niet dat we de werkelijkheid in die mate van nauwkeurigheid benaderen.

Wegingsfactoren voor Samenvoegen Categorieën:

Voor de Centrale Eindtoets:

VMBO	= bbl + (bbl & kbl) + kbl + (gl/tl) + 0.4740 * (gl/tl en havo)
VBMO-BG	= bbl + (bbl & kbl) + kbl
VMBO-TL	= gl/tl + 0.4740 * (gl/tl en havo)
HAVO	= 0.5260 * (gl/tl en havo) + havo + 0.6576 * (havo/vwo)
VWO	= 0.3424 * (havo/vwo) + vwo

Voor ROUTE 8:

VMBO	= PrO + BBL + KBL + TL
VBMO-BG	= PrO + BBL + KBL
VMBO-TL	= TL
HAVO	= HAVO
VWO	= VWO

Voor IEP:

VMBO	= (vmbo-b en vmbo k/b) + (vmbo-k en vmbo-k/t) + vmbo-t + 0.7187 (vmbo-t/havo)
VBMO-BG	= (vmbo-b en vmbo k/b) + 0.5240 * (vmbo-k en vmbo-k/t)
VMBO-TL	= 0.4760 * (vmbo-k en vmbo-k/t) + Vmbo-t + 0.7187 * (vmbo-t/havo)
HAVO	= 0.2813 * (vmbo-t/havo) + havo + 0.5354 * (havo-vwo)
VWO	= 0.4646 * (havo-vwo) + vwo