



College voor Toetsen en Examens

VERANTWOORDING

CENTRALE EINDTOETS PO

Versie : 1.1
Datum: 12 mei 2015

Voorwoord

Met dit verantwoordingsdocument legt het College voor Toetsen en Examens (CvTE) publiekelijk verantwoording af over de kwaliteit van de Centrale Eindtoets. Het gaat zowel om de inhoudelijke als procedurele kwaliteit.

Inhoudelijk: Aan welke criteria moet de Centrale Eindtoets voldoen? En voldoet de Centrale Eindtoets daaraan?

Procedureel: Hoe komt de Centrale Eindtoets tot stand en welke checks en balances worden daarbij ingezet?

Aangezien 2015 het eerste jaar is dat de Centrale Eindtoets onder regie van het CvTE wordt uitgevoerd en er nog geen afnamegegevens van de Centrale Eindtoets in deze nieuwe vorm zijn, bouwt deze verantwoording voort op de wetenschappelijke verantwoording van de Eindtoets Basisonderwijs van Cito. Dit document heeft het karakter van een groeidocument. Jaarlijks wordt dit aangevuld met afnamegegevens van de Centrale Eindtoets en eventuele andere bronnen.

Inhoud

	Voorwoord	3
1	Samenvatting	7
2	Inleiding	10
2.1	Historische achtergrond	11
2.2	Begripsmatige achtergrond	11
3	Uitgangspunten	13
3.1	Doelgroep	13
3.2	Gebruiksdoel en functie	14
3.3	Theoretisch kader	14
3.3.1	Inhoudelijk	14
3.3.2	Psychometrisch	15
3.4	Constructie	19
3.5	Kwaliteitssystemen	21
3.5.1	A3 aanpak CvTE	21
3.5.2	Toetswijzer Eindtoets PO algemeen deel	22
3.5.3	Toetswijzer Eindtoets PO specifiek deel	22
3.5.4	ISO 9001	22
4	Opzet van de Centrale Eindtoets	24
4.1	Opbouw en Structuur	24
4.2	Inhoudelijke verantwoording	25
4.2.1	Taal	25
4.2.2	Rekenen	29
4.3	Kenmerken	35
4.4	Procedures rond afname en rapportage	36
5	Normering	39
5.1	Kenmerken	39
5.2	Kalibratie	41
5.2.1	De stappen in de kalibratie	41
5.2.2	Evaluatie van de kalibratieprocedure	42
5.2.3	Adviescategorieën	42
5.3	Beoordelingsnormen en de bijbehorende scores Centrale Eindtoets PO	44
5.4	Rapportage op referentieniveaus	45
6	Betrouwbaarheid en meetnauwkeurigheid	47
6.1	Betrouwbaarheid	47
6.2	Nauwkeurigheid	47
7	Validiteit	49
7.1	Inhoudsvaliditeit	49
7.2	Begripsvaliditeit	50
7.2.1	Passing van het meetmodel; psychometrische kwaliteit van de opgaven	50
7.2.2	Structuur	50

7.2.3	Equivalentie met eerdere toetsversies	51
7.2.4	Soortgenootvaliditeit; convergente en discriminante validiteit	51
7.2.5	Verschillen tussen relevante subgroepen; samenhang met achtergrondkenmerken	51
7.2.6	DIF-analyses en invariantie van structuur	52
7.3	Criteriumvaliditeit	52
7.3.1	Relatie met het doorstroomadvies van de leerkracht	53
7.3.2	De score op de Centrale Eindtoets en de feitelijke plaatsing en doorstroom	53
8	Toezicht	55
8.1	Beoordelingskader andere eindtoetsen	55
8.2	Extern toezicht door COTAN	55
8.3	Extern toezicht door de onderwijsinspectie	55
8.4	Extern toezicht door audits	56
9	Referenties	58
10	Bijlage Inhoudelijke eisen en indicatoren Centrale Eindtoets	59
10.1	Inleiding	59
10.1.1	Kwaliteitseis 1, Samenhang Taal en Rekenen	59
10.1.2	Kwaliteitseis 2, Onderdeel Taal	59
10.1.3	Kwaliteitseis 3, Onderdeel Rekenen	63
10.1.4	Kwaliteitseis 4, Niveaubepaling ten aanzien van de referentieniveaus	65
10.1.5	Kwaliteitseis 5, Handleiding	65
11	Bijlage Procedurele eisen en indicatoren Centrale Eindtoets	67
11.1	Inleiding	67
11.1.1	Rapportage	67
11.1.2	Leerling	67
11.1.3	Afname	67
11.1.4	Constructie en verantwoording	67

1 Samenvatting

CvTE, Centrale Eindtoets en kwaliteitsborging

Vanaf het schooljaar 2014-2015 is het voor leerlingen van groep 8 in het reguliere basisonderwijs verplicht om een eindtoets te maken. De overheid stelt hiervoor via het College voor Toetsen en Examens (CvTE) de Centrale Eindtoets ter beschikking. Deze wordt in samenwerking tussen het CvTE en stichting Cito ontwikkeld en bouwt voort op de eindtoets basisonderwijs zoals deze voorheen door Cito werd verzorgd.

Het CvTE verantwoordt zich voor de kwaliteit van de Centrale Eindtoets door inzicht te geven in de procedures die hierbij worden gehanteerd. Deze zijn in overeenstemming met de beoordelingssystemen van testen, toetsen en examens zoals deze zijn beschreven door de COTAN en het RCEC. Deze beschrijving wordt jaarlijks aangevuld met een rapportage van de meest recente afnameresultaten. Op deze manier streeft het CvTE een zo groot mogelijke transparantie na.

Het CvTE hanteert als intern kwaliteitssysteem de A3 aanpak volgens het INK-model. Hiervoor is een convenant met de ketenpartners opgesteld en zijn er binnen de keten procesafspraken gemaakt. Daarnaast zijn er externe audits van de processen rondom de constructie van de toets en zowel de papieren als de digitale afname ervan. Ook is er toezicht op de uitvoering van de Centrale Eindtoets door de onderwijsinspectie.

De kwaliteitsbeoordeling van de Centrale Eindtoets loopt parallel met het *Beoordelingskader andere eindtoetsen*, zoals dat door de Expertgroep toetsen PO wordt gehanteerd en dat gebaseerd is op het Toetsbesluit PO en het algemeen deel van de Toetswijzer Eindtoets PO van 2014.

Inhoudelijke kwaliteitsborging

De kwaliteit van de Centrale Eindtoets berust in hoge mate op de procedures die zijn ingericht om de inhoud van de toets te kunnen garanderen. Daarmee is de inhoudsvaliditeit een van de belangrijkste pijlers onder de validiteit van de toets. Inhoudelijk sluit de toets nauwkeurig aan bij de (communale) doelstellingen op het gebied van taal en rekenen die door het gehele basisonderwijs worden nagestreefd. De afname van het onderdeel wereldoriëntatie is facultatief.

De Centrale Eindtoets is een leervorderingentoets, die speciaal ontwikkeld is om te beschrijven wat elk kind op het eind van het basisonderwijs heeft geleerd. Daarmee verschaft de toets onafhankelijke en objectieve informatie voor de keuze van een passend brugklatype. Want leervorderingen in het basisonderwijs zeggen iets over de vraag hoe goed een leerling de verschillende typen van voortgezet onderwijs aankan. In aanvulling daarop kan de schoolrapportage aangeven wat de plaats van de school is in de landelijke verdeling van schoolgemiddelden. Daarnaast kan de analyse van de gemiddelde scores op de onderdelen in het schoolrapport worden gebruikt bij de evaluatie van het onderwijs.

De Centrale Eindtoets bestaat uit twaalf taken die per vakgebied een aantal domeinen bestrijken. Voor taal zijn dat lezen, woordenschat, schrijven en taalverzorging & grammatica, voor rekenen getallen, verhoudingen, meten & meetkunde en verbanden, en voor wereldoriëntatie aardrijkskunde, geschiedenis en natuur & techniek.

De verdeling van de opgaven over de verschillende domeinen sluit aan bij de verdeling in de Toetswijzer en vormt een afspiegeling van de doorsnee onderwijsbesteding voor taal en rekenen. Per domein worden de gewenste typen opgaven gedetailleerd beschreven. Deze beschrijvingen vormen het uitgangspunt voor het werk van constructieteams die bestaan uit leden die ervaring hebben in het basisonderwijs, onder leiding van een toetsdeskundige van Cito. Het CvTE geeft hiertoe jaarlijks een constructieopdracht waarin nauwkeurig staat omschreven hoeveel opgaven, van welke

onderdelen, met welke inhoud en van welk type, geconstrueerd dienen te worden. Alle opgaven worden geproeftoetst bij steekproeven van 300 à 400 leerlingen per opgave. De inhoud van de Centrale Eindtoets wordt uiteindelijk bepaald door drie onafhankelijke vaststellingscommissies voor taal, rekenen en wereldoriëntatie, bestaande uit vakdeskundigen en leerkrachten uit groep 7 en 8. Aan elke vaststellingscommissie is een lid van de toetswijzercommissie PO toegevoegd. De vaststellingscommissies vormen een belangrijke schakel in de borging van de inhoudsvaliditeit. Zij beoordelen elke opgave in de Centrale Eindtoets op vakinhoudelijke gronden (overeenstemming met de toetswijzer en referentieniveaus voor taal en rekenen), het leerplan (methodeonafhankelijkheid), de belevingswereld van het kind en psychometrische overwegingen. Het scheiden van het constructie- en het vaststellingsproces vormt een extra check waarin de inhoudelijke kwaliteit (inhoudsvaliditeit) van de Centrale Eindtoets wordt gewaarborgd.

De Centrale Eindtoets wordt jaarlijks inhoudelijk volledig vernieuwd. De processen van opgavenconstructie, proeftoetsing en samenstelling van de toets omvatten een driejaarlijkse cyclus.

Psychometrische kwaliteitsborging

Naast de processen die gericht zijn op inhoudelijke kwaliteitsborging is het op item response theorie (IRT) gebaseerde meetmodel essentieel. Dit meetmodel zorgt ervoor dat elke deelttoets is samengesteld uit opgaven die de operationalisatie vormen van eenzelfde unidimensionale, onderliggende deelvaardigheid (bijvoorbeeld spellingvaardigheid of begrijpend lezen). Hiertoe wordt per opgave een kalibratieproces doorlopen waarbij verschillende statistische toetsen en grafische weergaveprocedures (van de itemkarakteristieke curves) worden benut om de opgavenbanken samen te stellen. Opgaven die niet voldoen aan de betreffende eisen worden uit de opgavenbank verwijderd. Uiteindelijk wordt ook de toets als geheel op zijn passing in het meetmodel onderzocht. Met behulp van hetzelfde meetmodel wordt er tevens voor gezorgd dat elke nieuwe Centrale Eindtoets niet alleen qua inhoud, maar ook in psychometrische zin volledig equivalent is met eerdere en latere versies. Dit is mede van belang in relatie tot de handhaving van de moeilijkheidsgraad van de toets over tijd.

De inhoudelijke en psychometrische procedures dienen te resulteren in *betrouwbare* toetsen en deelttoetsen. Voor de totaalscore op de Centrale Eindtoets geldt voor de interne consistentie een waarde van .95 als norm, voor de deelscores op rekenen en taal een waarde van .90. Door gebruik te maken van de eigenschappen van het meetmodel kan ook aandacht worden besteed aan de lokale meetfout. De nauwkeurigheid van de toets is het grootst bij lage en gemiddelde vaardigheidsniveaus. De toets wordt jaarlijks *genormeerd* op de afnamedata (in 2015 verzameld bij circa 165000 leerlingen). Daarbij wordt nagegaan of de data landelijk representatief zijn ten aanzien van de kenmerken regio, mate van verstedelijking, schoolgrootte en percentage leerlingen met een leerlinggewicht (op schoolniveau) en sekse en leeftijd (op leerlingniveau). Op basis van deze data wordt voor elke leerling een standardscore berekend en worden percentielscores bepaald die de relatieve positie van de leerling weergeven. De standardscore maakt het mogelijk om de leerling te positioneren in relatie tot een passend brugklatype in het voortgezet onderwijs. In de 'poppetjesgrafiek' op het leerlingrapport wordt per brugklatype aangegeven hoeveel procent van de leerlingen met eenzelfde of lagere score in het betreffende brugklatype is geplaatst. Verder worden aan de leerkracht op grafische wijze ook aanvullende gegevens verstrekt over het percentage kinderen dat met deze score in het betreffende brugklatype niet werd bevorderd naar het tweede leerjaar of werd overgeplaatst naar een lager brugklatype. Ten slotte is er sprake van een meer absolute normering door

de prestatie van de leerling te relateren aan referentieniveaus (1F, 1S of 2F) voor taalverzorging, lezen en rekenen.

In de onderbouwing van de *begrips- en criteriumvaliditeit* wordt aandacht besteed aan modelpassing en kwaliteit van de opgaven, itembias, structuur (intercorrelaties), equivalentie, soortgenoot-, convergente en divergente validiteit, relatie met doorstroomadvies, verschillen tussen relevante subgroepen en samenhang met achtergrondkenmerken. Daarnaast wordt in samenwerking met het CBS toelatings- en doorstroomonderzoek uitgevoerd en gerapporteerd ter onderbouwing van de criteriumvaliditeit (de voorspellende waarde) van de Centrale Eindtoets.

2 Inleiding

Vanaf het schooljaar 2014-2015 is het voor leerlingen* van groep 8 in het reguliere basisonderwijs verplicht om een eindtoets PO te maken. De overheid stelt de Centrale Eindtoets PO (hierna; Centrale Eindtoets) beschikbaar aan scholen. Scholen kunnen ook kiezen voor een andere eindtoets die door de minister is toegelaten. Het College voor Toetsen en Examens (CvTE) is verantwoordelijk voor de uitvoering van de Centrale Eindtoets. Deze wordt in samenwerking met Stichting Cito (hierna; Cito) gemaakt en bouwt voort op de Cito Eindtoets Basisonderwijs.

Alle toegelaten eindtoetsen moeten aan een aantal kwaliteitseisen voldoen. Met dit document legt het CvTE publiekelijk verantwoording af over de kwaliteit van de Centrale Eindtoets. Het gaat zowel om de inhoudelijke als de procedurele kwaliteit. Inhoudelijk: aan welke criteria moet de Centrale Eindtoets voldoen? En voldoet de Centrale Eindtoets daaraan?

Procedureel: Hoe komt de Centrale Eindtoets tot stand en welke *checks and balances* worden daarbij ingezet.

Voor het toetsen van de kwaliteit van de "andere" eindtoetsen is het Beoordelingskader onderwijskundige en organisatorische aspecten andere eindtoetsen, SLO dd. 5 september 2014, opgesteld (hierna: Beoordelingskader andere eindtoetsen). Alhoewel de Centrale Eindtoets niet getoetst wordt aan het Beoordelingskader andere eindtoetsen, volgt deze wel de inhoudelijke aspecten van deze richtlijn.

In dit document worden de activiteiten en procedures beschreven die nodig zijn voor de beoordeling van zowel de inhoudelijke als procedurele kwaliteit van de Centrale Eindtoets PO. Er wordt per hoofdstuk aangegeven aan welke beoordelingscriteria de Centrale Eindtoets moet voldoen. Hierbij wordt in eerste instantie het COTAN Beoordelingssysteem voor de kwaliteit van tests (Evers, Lucassen, Meijer & Sijtsma, 2010) als uitgangspunt genomen. Ook al wordt dit Beoordelingssysteem regelmatig toegepast om didactische toetsen op hun kwaliteit te beoordelen, op sommige punten is het hiervoor wat minder geschikt. Daarom is ten aanzien van enkele onderdelen gekozen voor het RCEC Beoordelingssysteem voor de kwaliteit van toetsen en examens (RCEC, 2011). Dit beoordelingssysteem neemt grotendeels de COTAN-systematiek over, maar wijkt op – voor toetsen en examens cruciale – aspecten daarvan af. Daarmee is het op deze specifieke punten beter toegesneden op de kwaliteitsbeoordeling van leervorderingentoetsen. Het betreft hier met name de onderdelen 'representativiteit' en 'afname van de toets', die verwerkt zijn in het hoofdstuk over de opzet van de Centrale Eindtoets. Representativiteit in termen van inhoudvaliditeit is een van de belangrijkste pijlers waarop de validiteit van de toets rust.

Dit document is gebaseerd op en verwijst naar de beoordelingscriteria zoals eerder beschreven in de laatst uitgebrachte wetenschappelijke verantwoording van de Cito Eindtoets Basisonderwijs (van Boxtel, Engelen & de Wijs, 2011)

- Uitgangspunten van de toetsconstructie
- Beschrijving van de toets
- Betrouwbaarheid en meetnauwkeurigheid
- Begripsvaliditeit

* Sommige leerlingen hoeven de verplichte eindtoets PO niet te maken. Tot deze groep behoren kinderen die zeer moeilijk lerend zijn, meervoudig gehandicapt zijn of korter dan vier jaar in Nederland wonen en de Nederlandse taal niet voldoende beheersen.

- Criteriumvaliditeit

Als (wetenschappelijke) verantwoording volstaat dit document niet. Het geeft weliswaar aan hoe en op welke aspecten kwaliteit wordt nagestreefd, maar in aanvulling daarop is het nodig om met behulp van empirisch onderzoek na te gaan of de nagestreefde kwaliteit ook daadwerkelijk is gerealiseerd. In het najaar volgend op de afname van de Centrale Eindtoets zal hierover op basis van de afnamegegevens aanvullend worden gerapporteerd. Ook in dit opzicht streeft het CvTE een zo groot mogelijke transparantie na.

Ten slotte, moet dit document opgevat worden als een groeidocument. Het geeft de huidige stand van zaken weer in het denken van het CvTE over de kwaliteit van de Centrale Eindtoets. Het CvTE stelt het document bij wanneer er sprake is van nieuwe ontwikkelingen ten aanzien van de Centrale Eindtoets en van nieuwe inzichten in de verantwoording van de kwaliteit ervan. Optimale kwaliteitszorg is immers steeds in beweging: procedures worden aangescherpt, vernieuwd, aangevuld of juist geschrapt. Daarin schuilt ook de kracht van de procedures die in dit document worden beschreven. Ze zijn de resultante van jarenlang nadenken, in de praktijk brengen, evalueren en zo nodig verbeteren.

2.1 Historische achtergrond

De Centrale Eindtoets PO komt voort uit de Eindtoets Basisonderwijs van Cito, die op zijn beurt zijn oorsprong vindt in de Amsterdamse Schooltoets.

Een belangrijk uitgangspunt van de Amsterdamse Schooltoets in 1966 was dat de informatie die de afname van de toets opleverde ook een analyse van de stand van het onderwijs mogelijk moest maken. Dat kon alleen als de toets bij alle leerlingen in het laatste jaar van de basisschool – toen nog Lagere School – zou worden afgenomen. En niet zoals destijds gebruikelijk was alleen bij leerlingen die mogelijk in aanmerking kwamen voor het voorbereidend hoger en middelbaar onderwijs. De inhoud en de moeilijkheid van de toets moesten dus geschikt zijn voor een brede afname.

En bovendien: belangrijke informatie voor de analyse van de stand of de kwaliteit van het onderwijs op een school is de vergelijking van de gemiddelde score met die van andere scholen. Daartoe moest de inhoud van de toets aansluiten bij onderdelen van het curriculum waaraan scholen vrijwel zonder uitzondering ongeveer dezelfde aandacht besteden.

Cito heeft de ontwikkeling van de Amsterdamse Schooltoets in 1970 overgenomen en daarbij vastgehouden aan deze twee uitgangspunten. Overeenkomstig de uitgangspunten is het onderdeel wereldoriëntatie facultatief: de inhoud van dit onderdeel verschilt per school (en denominatie) relatief sterk.

Tussen 1970 en 2014 is de opzet van de Eindtoets Basisonderwijs (zoals de Amsterdamse Schooltoets ging heten) een aantal keren veranderd. In 1973 bijvoorbeeld werd een indirecte meting van de schrijfvaardigheid geïntroduceerd. In 1977 werden standaardscores ingevoerd en werd spelling een verplicht onderdeel van de toets. In 2003 werd begrijpend lezen ondergebracht bij Taal in plaats van bij Informatieverwerking. Informatieverwerking heette vanaf dat moment Studievaardigheden. Er kwamen twee nieuwe onderdelen bij: hanteren van studieteksten en woordenschat.

Met ingang van het schooljaar 2014/2015 hield de Eindtoets Basisonderwijs op te bestaan.

De Centrale Eindtoets kwam ervoor in de plaats. In het hoofdstuk over de opzet van de Centrale Eindtoets vindt u een gedetailleerde beschrijving van de inhoud van deze toets.

2.2 Begripsmatige achtergrond

In het hoofdstuk over validiteit wordt beargumenteerd dat de Centrale Eindtoets geschikt is voor zijn primaire doel, het geven van onafhankelijke informatie voor de

keuze van het best passende vervolgonderwijs. Is daar ook een inhoudelijke verklaring voor? Is het logisch dat een schoolvorderingentoets iets zegt over latere leerresultaten? Het lijkt gerechtvaardigd om te veronderstellen dat leerprestaties uit het verleden iets zeggen over leerprestaties in de toekomst. Verschillen in leerprestaties op de onderdelen van de Centrale Eindtoets kunnen vermoedelijk voor een belangrijk deel worden verklaard door verschillen in eigenschappen als 'intelligentie', 'concentratie', 'motivatie' en 'doorzettingsvermogen'. Via de Eindtoets worden deze eigenschappen indirect en afgewogen gemeten; elke leerling heeft een 'persoonlijke mix'.

Uit diverse onderzoeken komen aanwijzingen voor de houdbaarheid van de veronderstelling hierboven, althans voor zover het om 'intelligentie' gaat. Er zijn hoge tot zeer hoge correlaties gevonden, tussen .72 en .82, tussen intelligentiemetingen als de NIO of de ISI-Intelligentietest en de Eindtoets Basisonderwijs, zowel met de onderdelen van de toets als met de totaalscore (en het daarop gebaseerde advies voor het voortgezet onderwijs). Naar schatting hebben de Eindtoets en intelligentiemetingen tussen de 50% en 60% van hun variantie gemeenschappelijk. Dit is een comfortabel resultaat in twee opzichten: de correlatie is enerzijds substantieel, en toont aan dat schoolse prestaties inderdaad samenhangen met intelligentie (waar een nul-correlatie toch wel bevreesdheid zou wekken), maar anderzijds is de correlatie laag genoeg om een zelfstandig gebruik van de Eindtoets te rechtvaardigen. In het hoofdstuk over validiteit gaan we hier nader op in, maar vooruitlopend kan worden gesteld dat de voorspellende waarde van leervorderingentoetsen wat hoger ligt dan die van intelligentietests. Dat geldt zowel voor correlaties met de adviezen van de leerkracht als voor de feitelijke doorstroming.

Dat leervorderingentoetsen, zoals de Centrale Eindtoets, in dit opzicht beter voorspellen dan intelligentiemetingen is vermoedelijk te danken aan het feit dat leervorderingentoetsen niet alleen iets zeggen over de intelligentie van het kind (zie de genoemde correlaties), maar indirect ook over persoonlijke eigenschappen die het leereffect mede bepalen. Het is waarschijnlijk dat deze eigenschappen in scores op intelligentietests maar zeer ten dele tot uitdrukking komen. Deze persoonlijke eigenschappen bepalen niet alleen mee hoeveel een kind, gegeven zijn intelligentie, in het basisonderwijs opsteekt, maar zij zijn ook van invloed op wat het kind in het voortgezet onderwijs aankan.

3 Uitgangspunten

De Centrale Eindtoets PO is een zogenoemde school- of leervorderingstoets: de toets meet wat een kind in vergelijking met andere kinderen in acht jaar basisonderwijs geleerd heeft. De inhoud van de Centrale Eindtoets sluit nauw aan bij het onderwijsprogramma voor de basisvaardigheden taal en rekenen. De opgaven zijn een operationalisering van doelstellingen die door het gehele basisonderwijs worden nagestreefd, zogenoemde communale doelstellingen. De toets is ontwikkeld om een bijdrage te leveren aan een adequate plaatsing van leerlingen in het vervolgonderwijs bij afsluiting van het basisonderwijs.

De Centrale Eindtoets is een objectieve momentopname bedoeld als ondersteuning van het eerder gegeven advies van de school. Als een leerling de Centrale Eindtoets beter maakt dan de basisschool gezien het schooladvies verwachtte, dan moet de basisschool het schooladvies heroverwegen. De basisschool is verantwoordelijk voor deze heroverweging, in overleg met de ouders/verzorgers.

De inhoud van de Centrale Eindtoets wordt met ingang van 2015 stapsgewijs afgestemd op de referentieniveaus taal en rekenen. In 2015 meet de toets op de onderdelen lezen, taalverzorging en rekenen welk niveau de leerling heeft behaald ten opzichte van de referentieniveaus 1F en 1S (voor rekenen) en 1F en 2F (voor lezen en taalverzorging). Het is niet ondenkbaar dat in volgende jaren ook van andere onderdelen een referentieniveau kan worden bepaald. Deze ontwikkeling zal geleidelijk verlopen, vandaar het woord 'stapsgewijs'.

Desgewenst kan aanvullend ook het onderdeel wereldoriëntatie worden afgenomen. Afnames van dit onderdeel is facultatief; het wordt niet betrokken bij het advies voor de keuze van een passend brugklatype. Het opgavenboekje Wereldoriëntatie is drie jaar op rij te gebruiken.

3.1 Doelgroep

Vanaf het schooljaar 2014 - 2015 is het voor alle leerlingen van groep 8 in het reguliere basisonderwijs verplicht om een eindtoets te maken. In principe doen dus alle leerlingen in groep 8 mee aan de afname van een eindtoets.

Voor onderstaande groepen leerlingen is het maken van een eindtoets wettelijk niet verplicht. Deze leerlingen *mogen* wel deelnemen aan de Centrale Eindtoets:

- leerlingen die zeer moeilijk lerend zijn
- leerlingen die meervoudig gehandicapt zijn
- leerlingen die korter dan vier jaar in Nederland wonen en de Nederlandse taal nog niet voldoende beheersen.

De Centrale Eindtoets wordt aangeboden op twee niveaus: B (Basis) en N (Niveau).

- De Eindtoets B is bestemd voor leerlingen voor wie het verwachte vervolgadvis brugklatype gemengde/theoretische leerweg (GT) of hoger zal zijn.
- De Eindtoets N is bestemd voor leerlingen van wie verwacht wordt dat ze het beste passen in een brugklatype basisberoepsgerichte leerweg (BB) of kaderberoepsgerichte leerweg (KB).

De leerkracht maakt voor iedere leerling een keuze. Daarbij is het nadrukkelijk niet zo dat deze keuze een vorm van voorselectie impliceert, de bedoeling van de keuze is om alle leerlingen zo nauwkeurig mogelijk te toetsen. Omdat B en N elkaar deels overlappen (25% van de opgaven komt in beide toetsen voor) scoort iedere leerling op dezelfde schaal en kan met beide niveaus over dezelfde typen vervolgonderwijs worden geadviseerd.

Het facultatieve onderdeel wereldoriëntatie wordt op één niveau afgenomen en is in principe geschikt voor alle leerlingen. De score op dit onderdeel heeft geen invloed op het advies voor een type vervolgonderwijs.

Voor de meeste (groepen) leerlingen met een speciale ondersteuningsbehoefte voor wie de afname van de reguliere Centrale Eindtoets problemen oplevert zijn er speciale versies van de toets uitgebracht. Er is een gesproken versie van de Eindtoets op cd en daisy-cd, de laatste specifiek geschikt voor de daisyspeler. Voor leerlingen die gewend zijn te werken met dyslexiesoftware is er een cd met een pdf-naar-spraaksynthese (te gebruiken voor diverse soorten dyslexiesoftware). Ook kunnen scholen een vergrote versie (A3 formaat; voor slechtziende leerlingen) van de opgavenboekjes bestellen. Voor blinde leerlingen is op aanvraag een brailleversie beschikbaar. Alle speciale versies zijn zowel voor de Centrale Eindtoets B als voor de Centrale Eindtoets N verkrijgbaar.

3.2 **Gebruiksdoel en functie**

De specifieke functie van de Centrale Eindtoets PO is het geven van onafhankelijke informatie voor de keuze van een passend brugklatype. De toets meet wat een kind in vergelijking met andere kinderen in acht jaar basisonderwijs geleerd heeft (leervorderingen). Leervorderingen zeggen daarnaast iets over de vraag hoe goed een leerling de verschillende typen van het voortgezet onderwijs aankan. Omdat de Centrale Eindtoets in principe bij alle leerlingen in groep 8 wordt afgenomen, kan een discrepantieanalyse van de gemiddelde scores op de verschillende onderdelen in het schoolrapport worden gebruikt bij de evaluatie van het onderwijs. In aanvulling daarop kan de schoolrapportage aangeven wat de plaats van de school is in de landelijke verdeling van schoolgemiddelden en in de verdeling van de schoolgroep waartoe de school behoort.

Met het oog op de optimalisering van onafhankelijkheid van de informatie wordt van de Centrale Eindtoets jaarlijks een nieuwe versie geconstrueerd. De resultaten op de verschillende jaargangen van de toets zijn onderling vergelijkbaar.

3.3 **Theoretisch kader**

3.3.1 *Inhoudelijk*

Binnen de Centrale Eindtoets kunnen op grond van de inhoud twee subtoetsen onderscheiden worden: Taal en Rekenen. Daarnaast is er het (facultatieve) onderdeel Wereldoriëntatie. In het hoofdstuk over de beschrijving van de toets zullen we deze onderdelen in meer detail beschrijven en verantwoorden. Voor dit moment past de vraag hoe met de opzet van de toets zo goed mogelijk tegemoet gekomen wordt aan de uitgangspunten en functies ervan. Het gaat dan vooral om maatregelen die de aansluiting van de toets op de communale inhouden van het onderwijs borgen. Een belangrijke voorwaarde voor een optimaal voorspellend vermogen is immers dat de toetsinhoud een adequate afspiegeling vormt van het gegeven onderwijs.

Aanpak:

- analyseren van kerndoelen en lesmethoden om te controleren of de Centrale Eindtoets goed aansluit bij de inhoud van het onderwijs;
- de opgaven voor de toets worden geconstrueerd aan de hand van de Toetswijzer (2014);
- alle opgaven worden geproeftoetst voor ze in de Centrale Eindtoets worden opgenomen;
- een onderdeel van deze jaarlijkse proeftoets is het verzoek om commentaar van leerkrachten;
- de constructie van opgaven vindt plaats in commissies waarvan de leden ervaring hebben in het basisonderwijs;
- de afzonderlijke opgaven én de toets als geheel worden vastgesteld door onafhankelijke commissies van vakdeskundigen (de vaststellingscommissies). Meer

over de werkwijze van de vaststellingscommissies als schakel in de borging van de inhoudvaliditeit is te vinden in de paragraaf over de constructie van de Centrale Eindtoets.

Verder is het belangrijk om vast te stellen dat de Centrale Eindtoets en zijn voorganger, de Eindtoets Basisonderwijs, door zeer veel scholen afgenomen wordt. Scholen bespreken de gemiddelde scores met het bevoegd gezag. We zijn er dan ook van overtuigd dat scholen de keuze van niet passende inhoud zeker zouden melden. Afgezien van een enkele opmerking over een afzonderlijke opgave komen dergelijke meldingen in de praktijk echter niet voor.

3.3.2

Psychometrisch

Constructie- en equivaleringsprocedures

De Centrale Eindtoets PO bestaat ieder jaar volledig uit nieuw itemmateriaal. De primaire functie van de toets (het geven van een onafhankelijk advies) vereist maatregelen die garanderen dat het construct dat wordt gemeten door elk van de toetsonderdelen constant blijft. Deze constantie heeft niet zozeer betrekking op de moeilijkheid van de opgaven als wel op hun homogeniteit met betrekking tot de onderliggende vaardigheid die ze pretenderen te meten. Hieronder staat beschreven hoe en met welke procedures er jaarlijks voor wordt gezorgd dat er steeds een nieuwe, volledig equivalente versie van de Centrale Eindtoets wordt opgeleverd.

Toetswijzer en constructieopdracht

In de Toetswijzer staat beschreven welk soort opgaven (inhoud en type) per onderdeel een eindtoets moet bevatten, indien het genoemde onderdeel deel uitmaakt van de toets. De Centrale Eindtoets bevat alle verplichte onderdelen uit de toetswijzer. Niet alle in de Toetswijzer beschreven onderdelen zijn verplicht; zo ontbreekt bijvoorbeeld het onderdeel luistervaardigheid in de Centrale Eindtoets. Voor de onderdelen die wél in de Centrale Eindtoets zitten, vormen de beschrijvingen in de Toetswijzer het kader voor de constructieopdracht.

Het College voor Toetsen en Examens (CvTE) geeft elk jaar een constructieopdracht aan Cito. In deze opdracht staat hoeveel opgaven, van welke onderdelen, met welke inhoud en van welk type, geconstrueerd dienen te worden voor de Centrale Eindtoets PO. Let wel: het gaat dan niet om de eindtoets van het jaar erop, maar om de eindtoets van drie jaar later. Het volledige proces van construeren, proeftoetsen en goedkeuren (vaststellen) van de toetsopgaven heeft een doorlooptijd van drie schooljaren.

Opgavenconstructie

Op basis van de constructieopdracht van het CvTE gaan verschillende constructiegroepen (per vakonderdeel) nieuwe opgaven ontwikkelen. Een constructiegroep (cg) bestaat uit drie à vier mensen 'uit het veld': leerkrachten van groep 7 of 8, PABO-docenten en/of IB'ers, onder leiding van een toetsdeskundige van Cito. De afzonderlijke cg-leden leveren conceptopgaven en uitgangsmateriaal (teksten, afbeeldingen) aan, die door de andere leden besproken en becommentarieerd worden in speciale bijeenkomsten. Soms gaat aan de bespreking eerst een schriftelijke commentaarronde vooraf.

De toetsdeskundige van Cito levert zelf geen opgaven aan. Alle opgaven komen van mensen uit de praktijk van het basisonderwijs. De toetsdeskundige levert vanuit zijn/haar expertise op vakinhoudelijk en toetstechnisch gebied wel commentaar op de aangeleverde opgaven. Ook kan hij of zij de opgaven aanpassen als de

commentaarronde of de bespreking daartoe aanleiding geeft. De opgaven die (eventueel in aangepaste vorm) zijn goedgekeurd, worden opgenomen in een opgavenbank.

Opgavenbank

Een opgavenbank is nadrukkelijk niet eenvoudigweg een verzameling opgaven waaruit een toetsconstructeur min of meer naar willekeur een aantal items selecteert om een nieuwe toets samen te stellen. Hieronder wordt beschreven wat de vereisten zijn om van een deugdelijke en psychometrisch goed gefundeerde opgavenbank te kunnen spreken.

- *Unidimensionaal continuüm en latente vaardigheid*
Het algemene uitgangspunt is dat de vaardigheden die in de onderdelen van de Centrale Eindtoets worden gemeten kunnen worden opgevat als een unidimensionaal continuüm (de reële lijn), en dat elke leerling voorgesteld kan worden als een punt op die lijn, met andere woorden: als een getal. Het getal drukt de mate van vaardigheid uit, waarbij een groter getal wijst op een grotere vaardigheid. Het doel van de meetprocedure – het afnemen van de toets – is de plaats van de leerling op dit continuüm zo nauwkeurig mogelijk te bepalen. De uitkomst van de meetprocedure bestaat strikt genomen uit twee grootheden: de eerste is de schatting van de plaats van de leerling op het vaardigheidscontinuüm. De tweede grootheid geeft aan hoe nauwkeurig die schatting is, en heeft dus de status van een standaardfout, te vergelijken met de standaardmeetfout uit de klassieke testtheorie. De antwoorden van een leerling op de items worden beschouwd als indicatoren van de vaardigheid, hetgeen ruwweg betekent dat men verwacht dat alle items in de bank deze zelfde vaardigheid meten. De vaardigheid zelf wordt als niet-observeerbaar beschouwd, en daarom gewoonlijk omschreven als een latente vaardigheid.
- *'Moeilijkheid' in de Item Respons Theorie*
Hoewel items dezelfde vaardigheid meten, kunnen ze toch systematisch van elkaar verschillen. Het belangrijkste verschil tussen de items is hun moeilijkheidsgraad. In de Item Respons Theorie (IRT) die voor het construeren van de opgavenbanken wordt gebruikt, is de moeilijkheid van een item gedefinieerd in termen van de onderliggende vaardigheid, zonder enige verwijzing naar een bepaalde populatie van leerlingen. Zo kan men ook de uitspraak begrijpen dat in de IRT vaardigheid en moeilijkheid op eenzelfde schaal liggen (voor een nadere toelichting op het meetmodel, zie de paragraaf over het gehanteerde meetmodel).
- *Kansmodel en moeilijkheidsgraad*
De ruwe omschrijving van de moeilijkheidsgraad die in de vorige alinea werd gehanteerd (de mate van vaardigheid die nodig is om het item goed te kunnen beantwoorden) behoeft verdere uitwerking. Men zou deze omschrijving kunnen opvatten als een drempel: heeft een leerling die mate van vaardigheid niet, dan is hij niet in staat het item juist te beantwoorden; heeft hij die drempel wel gehaald, dan geeft hij (gegarandeerd) het juiste antwoord. Deze interpretatie houdt in de praktijk geen stand, omdat eruit volgt dat een leerling die een moeilijk item correct beantwoordt geen fout kan maken op een gemakkelijk item. Daarom wordt in de IRT een kansmodel gebruikt: hoe groter de vaardigheid, des te groter de kans dat een item juist wordt beantwoord. De moeilijkheidsgraad van een item wordt dan gedefinieerd als de mate van vaardigheid die nodig is om met een kans van precies een half, een juist antwoord te kunnen produceren.

Proeftoetsonderzoek (kalibratie)

Vóór de opgaven in een toets gebruikt kunnen worden, moet geprobeerd worden de waarden van de moeilijkheidsgraden te achterhalen. Dit gebeurt met een statistische schattingsmethode die wordt toegepast op de antwoorden die in een

proeftoetsonderzoek, bij een steekproef van leerlingen zijn verzameld. Het hele proces van moeilijkheidsgraden schatten en verifiëren of de modelveronderstellingen houdbaar zijn, wordt kalibratie of ijking genoemd. De steekproef van leerlingen die hiervoor wordt gebruikt heet kalibratiesteekproef.

- *Afnamedesigns*

Meestal bevat een opgavenbank (veel) meer items dan een doorsnee toets, zodat het praktisch niet haalbaar is om alle items aan alle leerlingen voor te leggen. Elke leerling in de kalibratiesteekproef maakt dan ook slechts een gedeelte van de items uit de opgavenbank. Om een idee te krijgen:

- elke leerling krijgt ongeveer het aantal opgaven (180 tot 220) dat ook in de uiteindelijke toets zit;
 - elke opgave wordt aan ongeveer 300 tot 400 leerlingen voorgelegd.
- Dit gedeeltelijk voorleggen moet met de nodige omzichtigheid gebeuren. Er is dan sprake van een zogenoemd onvolledig design. Verderop wordt ingegaan op het afnamedesign dat voor de kalibratie is gebruikt, de geïnteresseerde lezer wordt verwezen naar Eggen (1993).

In een proeftoetsonderzoek worden niet alleen nieuwe opgaven aan de leerlingen voorgelegd.

Het proeftoetsonderzoek van 2014, bijvoorbeeld, bevatte itemmateriaal dat:

- deel uitmaakt van de Centrale Eindtoets 2015;
- deel zal uitmaken van de Centrale Eindtoets 2016;
- afkomstig is uit de referentiesets Lezen en Rekenen, om in 2015 voor deze onderdelen te kunnen rapporteren op referentieniveau;
- diende om nieuwe itemformaten uit te testen, die eventueel in latere eindtoetsen kunnen worden gebruikt.

- *Implicaties van gekalibreerde opgavenverzameling*

Als de kalibratie met succes uitgevoerd is, is het resultaat een zogenoemde gekalibreerde itembank. In het kalibratieproces worden de items die niet passen bij de verzameling uit de collectie verwijderd. De opgavenbank bevat voor elk item niet alleen zijn feitelijke inhoud, maar ook zijn psychometrische eigenschappen, en de statistische zekerheid dat alle items dezelfde vaardigheid aanspreken.

Equivaleringsprocedures

Niet alleen dient er zorg voor te worden gedragen dat opeenvolgende versies van de Centrale Eindtoets inhoudelijk vergelijkbaar zijn, ook de behaalde scores moeten over de jaren heen op vergelijkbare wijze kunnen worden geïnterpreteerd: een standaardscore van 538 in 2015 zal dezelfde betekenis moeten hebben als een standaardscore van 538 in (bijvoorbeeld) 2017. Kortom, opeenvolgende versies van de toets moeten equivalent zijn. Er wordt op verschillende manieren geëquivaaleerd:

- In de proeftoetsingen worden subsets van opgaven die in twee opeenvolgende jaren worden gebruikt in combinatie met elkaar afgenomen bij eenzelfde steekproef van leerlingen. Daardoor kunnen de volledige toetsen voor alle toetsonderdelen steeds op één schaal worden gekalibreerd. Op grond hiervan kunnen beide toetsen lineair worden geëquivaaleerd. In feite wordt deze procedure als het ware verdubbeld doordat alle opgaven twee maal in een proeftoetsing worden opgenomen.
- De scores van een groep zogeheten constante scholen worden vanaf 2015, jaar op jaar met elkaar vergeleken. 'Constante scholen' nemen elk jaar deel (1), hebben elk jaar ongeveer evenveel deelnemende leerlingen (2) die bovendien gemiddeld een constant prestatieniveau laten zien (3). Er worden verschillen steekproeven van constante scholen gevormd waarvan de prestatieniveaus met elkaar worden vergeleken.
- Een extra check wordt gevormd door ankeropgaven op te nemen in de werkelijke eindtoetsafname bij een representatieve subset van leerlingen (die zich hier overigens niet van bewust zijn).

Door deze vergelijking wordt extra zekerheid ingebouwd.

Als kwaliteitsmaat geldt dat de op verschillende manieren geëquivaaleerde toetsen onderling qua gemiddelde en standaarddeviatie slechts zeer geringe afwijkingen mogen laten zien (minder dan één scorepunt bij een standaardafwijking van bijna 10). Zie verder van Boxtel et al, 2011, p. 16-19 en p. 66-67.

Ten slotte moet ook de structuur van de toets in termen van de intercorrelaties tussen de onderdelen van jaar tot jaar gelijk blijven. Op grond van eerder verzamelde empirische gegevens mag – als vuistregel - de grootste afwijking tussen twee overeenkomende correlaties omstreeks .05 bedragen en de gemiddelde afwijking tussen overeenkomende correlaties omstreeks .025 (zie van Boxtel et al, 2011, p. 13-14 en 59-60).

Het gehanteerde meetmodel

Bij de constructie van deelttoetsen voor de Centrale Eindtoets wordt gebruik gemaakt van een meetmodel uit de itemresponse theorie (IRT), namelijk het One Parameter Logistic Model (OPLM). In de IRT staat het te meten begrip of de te meten eigenschap centraal. De IRT beschouwt het antwoord op een item als een indicator voor de mate waarin die eigenschap aanwezig is. Het verband tussen eigenschap en itemantwoord is van probabilistische aard en wordt weergegeven in de zogenaamde itemresponsfunctie. Die geeft aan hoe groot de kans is op een correct antwoord als functie van de onderliggende eigenschap of vaardigheid. Belangrijkste (item-)parameters zijn hierbij de moeilijkheidsgraad en het discriminerend vermogen. Een meer technische beschrijving van het meetmodel is te vinden in van Boxtel et al. (2011, p. 17-20). Bij constructie van een toets in overeenstemming met OPLM als meetmodel wordt elke opgave van een toets of deelttoets (in de regel wordt per deelttoets gekalibreerd) opgevat als een operationalisatie van dezelfde vaardigheid, bijvoorbeeld rekenvaardigheid of spellingvaardigheid. Deze onderliggende vaardigheid wordt ook wel 'latente vaardigheid' genoemd. Hiermee wordt aan de inhoudsvaliditeit in termen

van een voor een leerstofdomein en leerjaar (in dit geval eind groep 8) representatieve verzameling opgaven een belangrijke element toegevoegd. Unidimensionaliteit (alle opgaven zijn – empirisch bewijsbaar – op te vatten als een indicator van dezelfde vaardigheid) vormt de grondslag voor de begripsvaliditeit van de toets.

De schattingen worden berekend met het computerprogramma OPLM (Verhelst, Glas en Verstralen, 1995). Dit programma voert ook statistische toetsen uit op grond waarvan kan worden bepaald of het model de gegevens adequaat beschrijft. Omdat een aantal van deze toetsen bijzonder gevoelig is voor een verkeerde specificatie van de discriminatie-indices, zijn de uitkomsten van deze toetsen bruikbaar als modificatie-indices: ze geven een aanwijzing in welke richting deze discriminatie-indices moeten worden aangepast om een betere overeenkomst tussen model en gegevens te verkrijgen. Kalibratie van items volgens het OPLM is dan ook een iteratief proces waarin alternerend (1) de modelfit van items wordt onderzocht door middel van statistische toetsing en (2) de waarden van de discriminatie-indices worden aangepast op grond van de resultaten van deze toetsen. Deze aanpassingen geschieden in de praktijk op basis van een en hetzelfde gegevensbestand. Er kan dus kanskapitalisatie optreden. Indien een steekproef een voldoende grootte heeft, is het effect van deze kanskapitalisatie echter gering (Verhelst, Verstralen en Eggen, 1991). Bij de omvang van de steekproeven waarop bij de constructie van de Centrale Eindtoets wordt gekalibreerd (N=5000 of meer) zijn de genoemde effecten verwaarloosbaar. Meer informatie over de kwaliteitseisen waaraan de passing van het gehanteerde meetmodel moet voldoen is te vinden in het hoofdstuk over de normering.

3.4 Constructie

Voor de Centrale Eindtoets vindt de beoordeling van de eindtoets plaats door de vaststellingscommissies. De vaststellingscommissies verrichten hun werkzaamheden in relatief grote zelfstandigheid, onder de verantwoordelijkheid van de directeur van het CvTE. Vaststellingscommissies zijn verantwoordelijk voor het ontwikkelen van de opgaven (constructieopdracht), evaluaties van de proeftoetsen en het vaststellen van de definitieve Centrale Eindtoets. Leden van een vaststellingscommissie moeten aan een aantal eisen voldoen die het CvTE aan haar leden stelt. Leden van een vaststellingscommissie kunnen hun functie niet combineren met:

- het schrijven voor of redactie voeren van een methode/schoolboek (voor een commerciële uitgeverij);
- het zitting hebben in een constructiegroep van het Cito.

Door hun onafhankelijkheid en expertise vormen de vaststellingscommissies een belangrijke schakel in de kwaliteitszorg, met name waar het de borging van de inhoudsvaliditeit betreft.

De vaststellingscommissies bestaan uit leerkrachten groep 7 en 8 en vakexperts uit de verschillende domeinen die worden getoetst met de Centrale Eindtoets. Om de samenhang met de toetswijzercommissie van het domein te borgen is het wenselijk om waar mogelijk een lid van de toetswijzercommissie PO deel te laten nemen in de vaststellingscommissie. Ieder lid verricht de werkzaamheden voor een schooljaar met het streven om de werkzaamheden te verlengen tot maximaal acht schooljaren. Met ieder lid vindt jaarlijks een functioneringsgesprek plaats waarna besloten wordt of het lid al dan niet de werkzaamheden in een vaststellingscommissie zal voortzetten.

Naast de leden van een vaststellingscommissie is er ook een toetsdeskundige van Cito als adviseur betrokken bij een vaststellingscommissie. Deze toetsdeskundige heeft geen stemrecht in de totstandkoming van de Centrale Eindtoets.

Er zijn drie vaststellingscommissies:

- 1 Commissie taal en taalonderdelen studievaardigheden.
Deze commissie is verantwoordelijk voor de constructie van de taalopgaven in de centrale eindtoets. De commissie stelt in samenhang met de commissie rekenen de centrale eindtoets B en N vast.
- 2 Commissie rekenen en rekenonderdelen studievaardigheden.
Deze commissie is verantwoordelijk voor de constructie van de rekenopgaven in de centrale eindtoets. De commissie stelt in samenhang met de commissie taal de centrale eindtoets B en N vast.
- 3 Commissie wereldoriëntatie.
Deze commissie is verantwoordelijk voor de itemconstructie van de facultatieve eindtoets wereldoriëntatie. Deze commissie stelt de toets vast (op dit moment op één niveau: eindtoets B).

Constructieopdracht

De vaststellingscommissies stellen een constructieopdracht op voor een volgende Centrale Eindtoets. De constructieopdracht is gebaseerd op de Toetswijzer Eindtoets PO, algemeen deel en de Referentieniveaus. De constructieopdracht wordt aangeboden aan de directie van het CvTE. De directie maakt er een definitieve opdracht van welke wordt aangeboden aan de directie van Cito. Na aanneming van de opdracht door de directie van Cito worden de werkzaamheden van de toetsdeskundigen gestart.

In de constructieopdracht zijn de volgende aspecten opgenomen:

- Toetsmatrijs:
De toetsmatrijs geeft de inhoud van de Centrale Eindtoets in een tabel weer. De inhoud, de getoetste vaardigheden en het uitgangsmateriaal (teksten, afbeeldingen) moeten evenwichtig en gevarieerd zijn. Voor elk van deze aspecten apart kan een overzicht gemaakt worden. Het schema dat ontstaat door deze overzichten te combineren, is de toetsmatrijs.
- Toetsrationale:
De toetsrationale is de uitwerking van de toetsmatrijsen. Aan de hand van een toetsrationale op hoofdlijnen, die door de toetsdeskundige van Cito wordt geleverd, beoordeelt de vaststellingscommissie of de uitwerking van de constructieopdracht overeenkomt met de toetsmatrijs. Als bijlage van de toetsrationale is een overzicht op itemniveau opgenomen waaruit het oordeel van de vaststellingscommissie blijkt gedurende het proces van totstandkoming van de definitieve toets. De vaststellingscommissie heeft bij ieder item de keuze uit drie oordelen:
 1. Goedgekeurd
 2. Voorbehoud (afgewezen met opdracht tot aanpassing)
 3. Afgewezen

Bij oordeel 2 en 3 wordt het argument vastgelegd in een bijlage van de toetsrationale. Een item met oordeel 2 wordt een volgende keer wederom behandeld door de vaststellingscommissie.

- Preambule
- Criteria voor uitgangsmaterialen

Werkwijze vaststellingscommissies

Het CvTE en Cito hebben een werkwijze afgesproken voor de constructie en vaststelling van de Centrale Eindtoets. De werkwijze is stapsgewijs beschreven in het Vademecum, het handboek voor leden van de vaststellingscommissies. De werkzaamheden van de

toetsdeskundigen van Cito zijn deels veranderd door de invoering van vaststellingscommissies. De toetsdeskundige blijft de aangewezen persoon om de analyses van de proeftoetsingen te bestuderen. Op basis van deze analyses en inhoudelijke overwegingen selecteert de toetsdeskundige opgaven die voorgelegd worden aan de vaststellingscommissies.

Na de tweede proeftoets zal de toetsdeskundige een voorstel doen met verschillende mogelijke samenstellingen van de definitieve toets. De uiteindelijke keuze om een individuele opgave in de Centrale Eindtoets op te nemen en de samenstelling van de Centrale Eindtoets ligt bij de vaststellingscommissie van het CvTE. Cito stelt daarvoor verschillende mogelijkheden, mét toelichting, aan de commissie voor. De vaststellingscommissie kan aan de hand daarvan nog bijsturen. De vaststellingscommissie beoordeelt iedere opgave aan de hand van vier pijlers:

1. **(vak)Inhoudelijk:** in overeenstemming met referentieniveaus en toetswijzer
2. **Leerplan:** methodeonafhankelijk (de inhoud van een opgave sluit aan bij de kerndoelen, gebruikers van een specifieke methode worden door de opgave bevoordeeld noch benadeeld)
3. **Belevingswereld:** kindgericht en niet gekunsteld
4. **Psychometrie:** P-waarde en discriminatie-index voldoen aan de gestelde eisen (zie paragraaf 3.3).

Het scheiden van het constructie- en het vaststellingsproces is een extra check op het werk. Feitelijk is dit het proces voor het waarborgen van de inhoudelijke kwaliteit (inhoudsvaliditeit) van de Centrale Eindtoets.

3.5 Kwaliteitssystemen

Naast de theoretische inkadering is voor de constructie van de Centrale Eindtoets PO een aantal kwaliteitssystemen relevant. Deze kwaliteitssystemen zeggen iets over de processen die bij de totstandkoming van de toets een rol spelen en hoe deze zijn verankerd in de CvTE organisatie.

Om te borgen dat de Centrale Eindtoets aan alle gestelde eisen voldoet, worden *geen* externe kwaliteitssystemen gehanteerd. Accreditatie van de ontwikkelings- en uitvoeringsprocessen voor de Centrale Eindtoets in termen van de normen die in externe kwaliteitssystemen zijn ondergebracht is dan ook niet aan de orde. Het CvTE heeft een intern kwaliteitssysteem ingericht met eigen eisen en normen, waarbij de strekking van externe kwaliteitssystemen richtinggevend was. In dit kwaliteitssysteem van het CvTE zijn op eigen initiatief wel audits en externe toetsing ondergebracht.

3.5.1 A3 aanpak CvTE

Het CvTE hanteert als intern kwaliteitssysteem de A3 aanpak volgens het INK model. Het INK is een managementmodel dat resultaatgericht denken stimuleert en bedoeld is voor organisaties om een zelfevaluatie uit te voeren. Het model helpt organisaties te focussen op die gebieden waar verbeteringen mogelijk zijn.

Voor de Centrale Eindtoets wordt een vertaling gemaakt van de strategie en beleidsdoelen van het CvTE naar het management van processen, medewerkers en middelen. Hierbij zijn de doelen specifiek gericht op de rol van het CvTE als regisseur van de processen binnen de ketenpartners. Hiervoor is onder andere een convenant met de ketenpartners opgesteld en zijn er ketenprocesafspraken gemaakt.

Convenant ketenpartners landelijke toetsen en examens

Het huidige convenant is op 30 oktober 2012 vastgesteld tussen de ketenpartners en heeft betrekking op de uitvoering van de wettelijke taken van het College voor Toetsen en Examens (toen nog het College voor Examens). Het huidige convenant heeft als doel te kunnen verwijzen naar een gemeenschappelijk begrippenkader en gedachtegoed teneinde de ketensamenwerking te borgen. In dit convenant is een RASCI matrix opgenomen met hierin de uitvoering, verantwoordelijkheden, tekenbevoegdheid en informatieverstrekking van de ketenpartners. Het huidige convenant wordt momenteel uitgebreid met de afspraken ten aanzien van de Centrale Eindtoets.

Ketenprocesafspraken

Op 22 oktober 2014 is in een memo de intentie geuit om te komen tot ketenprocesafspraken die volgens het door het CvTE gehanteerde INK-kwaliteitszorgsysteem, helder, eenduidig en stuurbaar zijn. Hiermee genereert het een transparant processchema met relevante prestatie-indicatoren met behulp waarvan het CvTE als regisseur de sturing binnen de keten kan uitvoeren. Hiermee kan het CvTE zich expliciet proces- en resultaatgericht opstellen.

3.5.2 Toetswijzer Eindtoets PO algemeen deel

Het *Toetsbesluit PO* schrijft voor aan welke voorwaarden alle eindtoetsen moeten voldoen. Dit zijn onder andere voorwaarden ten aanzien van inhoud en kwaliteit. In het *Toetsbesluit PO* is een van de kenmerken van een eindtoets dat de inhoud van de eindtoets gebaseerd is op de toetswijzer van het CvTE. De *Toetswijzer Eindtoets PO* geeft weer aan welke eisen elke eindtoets moet voldoen qua inhoud en vorm. Deze toetswijzer bestaat uit twee delen: een algemeen deel en een specifiek deel. De *Toetswijzer Eindtoets PO algemeen deel* expliciteert de inhoudelijke kwaliteitseisen voor de wettelijk verplichte domeinen uit het Referentiekader taal en rekenen die voor alle eindtoetsen gelden. Daarnaast worden in het algemene deel de kwaliteitseisen beschreven voor de domeinen die optioneel in eindtoetsen kunnen worden opgenomen. Het algemene deel van de toetswijzer biedt daarmee het kader op basis waarvan ontwikkelaars van eindtoetsen hun toets dienen te ontwikkelen.

3.5.3 Toetswijzer Eindtoets PO specifiek deel

De *Toetswijzer Eindtoets PO specifiek deel* is specifiek van toepassing op de Centrale Eindtoets zoals deze onder auspiciën van het CvTE wordt ontwikkeld en aangeboden. Het specifieke deel van de toetswijzer is gebaseerd op de *Toetswijzer Eindtoets PO algemeen deel* en bevat een beschrijving van de inhoud en toetsingsvorm van de Centrale Eindtoets en het niveau waarop getoetst wordt. Het specifieke deel van de toetswijzer wordt gebruikt voor de ontwikkeling van de Centrale Eindtoets.

3.5.4 ISO 9001

De ISO 9001 is de internationale norm voor kwaliteitsmanagementsystemen. De Internationale Organisatie voor Standaardisatie (ISO) staat in voor het opstellen en beheren van duizenden verschillende normen. De meeste hiervan zijn productgericht, maar enkele normen slaan op generieke managementsystemen. Verreweg de bekendste is de ISO 9001. Deze slaat op het managementsysteem voor kwaliteit.

Volgens deze standaard moet het kwaliteitsbeleid op papier staan en moet dit beleid bekend zijn bij alle medewerkers. De organisatie moet zorgen voor het verhogen van de klantentevredenheid, door te voldoen aan de eisen en wensen van de klanten en aan de wettelijke eisen die van toepassing zijn op het product of de dienst van de

organisatie. Daarnaast moet de organisatie de processen beheersen en dit tevens kunnen aantonen.

Nadat een externe audit heeft plaatsgevonden, kan de organisatie een certificaat ontvangen waaruit blijkt dat zij voldoet aan de eisen van de standaard. Dit is een teken voor de organisatie en haar klanten dat de organisatie op een vooraf gestelde en gestructureerde manier aan kwaliteitsmanagement (zorg, beheersing en borging) doet. Dit is echter nog geen garantie voor een kwalitatief goed product. Hiervoor zijn aanvullende inhoudelijke eisen en normen noodzakelijk.

ISO 9001 wordt ook wel versimpeld weergegeven als: "zeg wat je doet, doe wat je zegt en wees in staat dit aan te tonen". Voor de Centrale Eindtoets PO betekent dit dat er een gedetailleerde procesbeschrijving is van de aanpak voor planning, uitvoering, controle en bijsturing. Er is voor gekozen om geen externe toetsing aan de ISO-certificering uit te laten voeren. Hiermee loopt het kwaliteitssysteem van de Centrale Eindtoets parallel met dat van de Centrale Examens.

4 Opzet van de Centrale Eindtoets

4.1 Opbouw en Structuur

De Centrale Eindtoets PO bestaat uit twaalf taken. In de taken staan opgaven op het gebied van taal, rekenen en wereldoriëntatie. Taal en rekenen zijn verplichte onderdelen. Wereldoriëntatie is facultatief en speelt voor de primaire functie van de toets geen rol. Hieronder treft u een overzicht aan van de domeinen per vakgebied (tussen haakjes staat het aantal opgaven).

1. **Taal** **(135)**
 - lezen (45)
 - woordenschat (20)
 - schrijven (20)
 - taalverzorging en grammatica (50)

2. **Rekenen** **(85)**
 - getallen (30)
 - verhoudingen (20)
 - meten en meetkunde (20)
 - verbanden (15)

3. **Wereldoriëntatie** **(90)**
 - aardrijkskunde (30)
 - geschiedenis (30)
 - natuur & techniek (30)

De verdeling van de opgaven over de verschillende domeinen sluit aan bij de verdeling in de Toetswijzer en vormt een afspiegeling van de doorsnee onderwijsbesteding aan taal en rekenen zoals deze voor het reguliere basisonderwijs bekend is uit verschillende PPON-onderzoeken. Als we de PPON-balansen die over een reeks van jaren zijn verschenen als uitgangspunt zouden nemen, dan zouden 59% van de opgaven taalopgaven moeten zijn (overeenkomend met 7,25 uur onderwijsbesteding in groep 8) en 41% van de opgaven rekenopgaven (overeenkomend met 5 uur onderwijsbesteding in groep 8). Deze percentages komen vrijwel overeen met de percentages in de Centrale Eindtoets PO (61% en 39%).

Het onderdeel wereldoriëntatie is facultatief en kan afzonderlijk worden getoetst. Het betreft 90 opgaven die in gelijke verhoudingen zijn verdeeld over aardrijkskunde, geschiedenis en natuur & techniek. Deze verdeling komt overeen met de gelijke verdeling over deze vakken zoals deze in de PPON-balansen is terug te vinden (namelijk per vakgebied één uur per week in groep 8).

De opgaven zijn opgenomen in drie opgavenboekjes die op drie achtereenvolgende ochtenden worden afgenomen. Daarbij is gezorgd voor een optimale spreiding van taken. Binnen taken is ook steeds sprake van voldoende variatie in onderdelen. Waar dat technisch nodig is zijn opgaven van een onderdeel bij elkaar gehouden.

	Taak	Rubriek(en)	Aantal opgaven	Tijdschema	Afnameduur
Dag 1.			75+30		
		Inleiding		8.45 – 9.00	15 minuten
	Taak Taal 1	Woordenschat Schrijven	10 20	9.00 – 9.45	45 minuten
		Pauze		9.45 – 9.50	5 minuten
	Taak Rekenen 1	Rekenen 1	25	9.50 – 10.35	45 minuten
		Pauze		10.35 – 10.50	15 minuten
	Taak Taal 2	Lezen 1: - Begrijpend lezen - Samenvatten/Opzoeken	20 - 10 - 10	10.50 – 11.30	40 minuten
		Pauze		11.30 – 11.35	5 minuten

	Taak Wereldoriëntatie 1	Natuur	30	11.35 – 12.15	40 minuten
Dag 2.			70+30		
	Taak Rekenen 2	Rekenen 2	20	9.00 – 9.35	35 minuten
		Pauze		9.35 – 9.40	5 minuten
	Taak Taal 3	Taalverzorging 1: - Spelling niet-werkwoorden - Spelling werkwoorden Woordenschat	20 - 10 - 10 10	9.40 – 10.25	45 minuten
		Pauze		10.25 – 10.40	15 minuten
	Taak Rekenen 3	Rekenen 3	20	10.40 – 11.15	35 minuten
		Pauze		11.15 – 11.20	5 minuten
	Taak Wereldoriëntatie 2	Aardrijkskunde	30	11.20 – 12.00	40 minuten
Dag 3.			75+30		
	Taak Taal 4	Lezen 2: - Begrijpend lezen - Samenvatten/Opzoeken	25 - 15 - 10	9.00 – 9.45	45 minuten
		Pauze		9.45 – 9.50	5 minuten
	Taak Rekenen 4	Rekenen 4	20	9.50 – 10.25	35 minuten
		Pauze		10.25 – 10.40	15 minuten
	Taak Taal 5	Taalverzorging 2: - Interpunctie - Grammatica	30 - 15 - 15	10.40 – 11.20	40 minuten
		Pauze		11.20 – 11.25	5 minuten
	Taak Wereldoriëntatie 3	Geschiedenis	30	11.25 – 12.05	40 minuten

4.2 Inhoudelijke verantwoording

Hieronder worden de inhoudelijke kwaliteitseisen en indicatoren beschreven die de in de Centrale Eindtoets gehanteerd worden. De eerste bijlage bij deze verantwoording bevat de algemene inhoudelijke eisen die door de Toetswijzer PO worden voorgeschreven.

4.2.1 Taal

De Toetswijzer bij de Centrale Eindtoets PO bevat een gedetailleerde beschrijving van de getoetste taalonderdelen. Hieronder is deze informatie per (deel)vaardigheid samengevat.

Lezen

Leerlingen lezen in de Centrale Eindtoets informatieve teksten (waarbij de auteur vooral feitelijke informatie geeft), betogende teksten (waarbij de auteur niet alleen feitelijke informatie, maar ook zijn mening en opvattingen geeft) en instructieve teksten (waarin de lezer wordt uitgelegd hoe een bepaalde handeling te verrichten). Daarnaast lezen ze fictie. Qua inhoud, structuur, lay-out en stijl zijn de teksten die de leerlingen lezen divers.

Bij lezen onderscheiden we vijf deelvaardigheden:

1. Woordenschat
 Het gaat er hier om of de leerlingen in staat zijn de betekenis van woorden en woordgroepen uit de tekst te halen. De toets bevat o.a. opgaven die vragen naar het afleiden van de betekenis van een woord of woordgroep uit de context.
2. Begrijpen
 Hier gaat het om het verwerken van inhoudelijke informatie die een schrijver in de tekst vermeldt en om het kunnen aanwijzen van relaties tussen tekstdelen. De toets bevat onder andere opgaven die vragen naar het gebruik van signaalwoorden, oorzaak-gevolg en middel-doel, tijdsopvolgingen van

gebeurtenissen en/of specifieke, expliciet genoemde inhoudselementen in de tekst, zoals hoofdpersonen, thema, doel en publiek.

3. Interpretieren

Het gelijktijdig en in samenhang inzetten van tekst- en kennisgestuurde verwerking. Dat gebeurt bij het lezen altijd; de lezer kan zijn kennis over het onderwerp niet zo maar even uitschakelen. De lezer brengt datgene wat hij leest in verband met wat hij al weet van het onderwerp. De toets bevat o.a. opgaven waarbij stukjes informatie uit de tekst met elkaar worden vergeleken of verbonden. Voor fictieteksten kan het bijvoorbeeld gaan om het afleiden van het karakter van de hoofdpersoon, maar ook om het afleiden van de tijdsperiode, het thema of de toon van een verhaal. Of om het invullen van ontbrekende woorden en het anticiperen op het vervolg ('Hoe zal het verhaal aflopen?').

4. Samenvatten

Bij samenvatten gaat het erom dat er na het begrijpen en interpreteren van de informatie iets met die informatie gedaan moet worden, bijvoorbeeld hoofd- en bijzaken onderscheiden en de tekst verkorten of schematiseren. De toets bevat onder andere opgaven waarbij een tekstdeel moet worden voorzien van een passende titel of waarbij een gegeven schema moet worden aangevuld.

5. Opzoeken.

Om de veelheid van gedrukte en digitale informatiebronnen - waaronder tijdschriften, kranten, gidsen, OV-informatie, beeldbanken, websites en woordenboeken - te kunnen hanteren is het noodzakelijk dat lezers weten dat informatie alfabetisch geordend is of op een andere logische manier is ingedeeld, bijvoorbeeld op plaats of onderwerp. Ook het formuleren van zoekvragen, trefwoorden en synoniemen is hierbij van belang. De toets bevat onder andere opgaven waarbij het kiezen van een geschikte informatiebron centraal staat of die gericht zijn op het kiezen van de juiste ingang tot een informatiebron.

Woordenschat

Woordenschat wordt gedefinieerd als de kennis die mensen hebben over woordbetekenissen. Zowel de breedte als de diepte van de woordenschat spelen hierbij een rol. Bij de breedte van de woordenschat staat de omvang (kwantiteit) centraal. Bij de diepte van de woordenschat gaat het om de vraag: hoe goed worden woorden en hun betekenisaspecten in relatie tot andere woorden gekend (kwaliteit)?

De opgaven Woordenschat in de Centrale Eindtoets hebben op beide aspecten (kwalitatief en kwantitatief) betrekking en worden verdeeld in de rubrieken Betekenis en Betekenisrelaties.

1. Betekenis

In deze rubriek zijn de volgende typen opgaven ondergebracht:

- Opgaven waarin gevraagd wordt naar het geven van een synoniem van een gegeven woord.
- Opgaven die vragen naar definities of omschrijvingen van woorden of woordgroepen.
- Opgaven waarin het draait om het herkennen van het belangrijkste betekeniskenmerk van een woord of woordgroep.

2. Betekenisrelaties

In deze rubriek zijn de volgende typen opgaven ondergebracht:

- Opgaven waarin de tegengestelde betekenis van een woord of woordgroep centraal staat.
- Opgaven waarin gevraagd wordt naar betekenisassociaties.

- Opgaven waarin woorden voorkomen die op basis van een of meer gezamenlijke kenmerken onder één noemer of betekeniscluster geplaatst kunnen worden.
- Opgaven waarbij een reeks woorden op een vergrotende of verkleinende trap geplaatst moeten worden.

Schrijven

In de Centrale Eindtoets schrijven de leerlingen niet zelf teksten, maar bestaat dit onderdeel uit teksten met opgaven. De teksten zijn zogenaamd geschreven door leerlingen van de basisschool en voor het doel min of meer aangepast. De 'fouten' in de teksten zijn blijven staan; het zijn ook de fouten waar het in de opgaven om draait. In de opgaven zijn ruwweg drie soorten fouten te onderscheiden: fouten die betrekking hebben op de inhoud van de tekst, fouten die gemaakt worden tegen de gepastheid van de tekst en fouten die gemaakt worden tegen de correctheid van het taalgebruik in de tekst.

1. Inhoud

De rubriek Inhoud is gericht op de fouten die schrijvers maken bij het weergeven van informatie. Ze verstrekken dan onvoldoende of onduidelijke informatie, waardoor misverstanden bij de lezers kunnen ontstaan. De toets bevat onder andere opgaven waarin het beoordelen van de logica van de tekst centraal staat en opgaven waarin het gaat om begrijpelijk taalgebruik op woord-, zins- en/of tekstniveau.

2. Passendheid

In de rubriek Passendheid gaat het om het doel dat de schrijver met zijn tekst voor ogen heeft en de afstemming van de tekst op de lezer. De toets bevat onder andere opgaven waarbij de leerling beoordeelt of er sprake is van passend woordgebruik en een consistente schrijfstijl.

3. Correct taalgebruik

De rubriek Correct taalgebruik is gericht op fouten tegen regels die te maken hebben met conventies over de vorm van de taal. De toets bevat onder andere opgaven over de keuze van woorden en zinnen en het correct gebruik van uitdrukkingen.

Taalverzorging

Taalverzorging valt uiteen in twee onderdelen: spelling en het gebruik van interpunctie.

Spelling

Bij spelling gaat het om de schrijfwijze van woorden. Er is een relatie tussen letters (grafemen) en spraakklanken (fonemen). Soms is de relatie een-op-een, maar meestal moet je regels gebruiken om tot de correcte schrijfwijze te komen ('held' eindigt op een 'd' omdat het meervoud 'helden' is) of moet je gewoon weten dat dat een woord op een bepaalde manier wordt gespeld ('rauwe' groente en geen 'rouwe' groente). De opgaven spelling zijn te verdelen in twee rubrieken.

Spelling van niet-werkwoorden

In de Centrale Eindtoets kunnen de volgende spellingcategorieën aan de orde komen bij het spellen van niet-werkwoorden:

- Woorden met open en gesloten lettergreep (notaris, bemanning)
- Woorden op -ig(e) en -lijk(e) (akelig, koninklijk)
- Woorden waarin /s/ geschreven wordt als c (narcis, ceremonie)
- Woorden waarin /k/ geschreven wordt als c (cacao, accordeon)
- Woorden waarin /sj/ geschreven wordt als ch (chef, machine)

- Woorden beginnend met 's of eindigend op 's ('s ochtends, komma 's)
- Woorden op /tie/ waar t klinkt als ts (traditie)
- Woorden op -heid of -teit (veiligheid, majesteit)
- Woorden op -ig en -lijk (akelig, koninklijk)
- Woorden waarin /ie/ geschreven wordt als i (diploma)
- Gebruik van hoofdletters (Marjolein, Nederland, Pasen)
- Woorden met eind -b (krab, web)
- Woorden op -iaal, -ieel, -ueel, -eal (speciaal, actueel, commercieel)
- Woorden met -y- (typisch)
- Samenstellingen met tussen -e(n)- en tussen -s- (blokkendoos, dorpsplein)
- Woorden met een trema (zeeën, ruïne)
- Franse leenwoorden (bureau, journaal)
- Engelse leenwoorden (team, training, keeper)
- Woorden met th die klinkt als t (theater, apotheek)
- Woorden op -sch(e) (alfabetisch)
- Woorden waarin /ks/geschreven wordt als x (examen)
- Stoffelijk bijvoeglijke naamwoorden met -e(n) (wollen, leren, houten)
- Verkleinwoorden met -aatje, -eetje, -ootje, -uutje en met de uitgang -nkje (paraplutje, kettinkje)
- Woorden met assimilatieverschijnselen (ontdekking, enigszins)
- Gebruik van het koppelteken in samenstellingen (e-mail, zonne-energie)
- Restwoorden (museum, erwt)

Spelling van werkwoorden

In de Centrale Eindtoets kunnen de volgende spellingcategorieën aan de orde komen bij het spellen van werkwoorden:

- Onvoltooid tegenwoordige tijd: wel of geen -t achter een stam die eindigt op -d (ik verbind), een -t achter de stam van een zwak werkwoord dat in de onvoltooid verleden tijd de uitgang -de(n) krijgt (jij tekent) en bij inversie van persoonsvorm en onderwerp wel of geen -t achter een stam die eindigt op een -d (word jij).
- Onvoltooid verleden tijd: verdubbeling -d of -t bij een zwak werkwoord met een stam die eindigt op -d of -t (ik raadde, wij landden), geen -t bij een sterk werkwoord dat in 2e en 3e persoon eindigt op -d (jij werd, hij zond) en de uitgang -sde(n) of -fde(n) bij zwakke werkwoorden waarvan de stam eindigt op een -z of -v (ik bonsde, hij beefde).
- Voltooid deelwoord: keuze voor een eind-d of een eind-t bij een stam die niet eindigt op een -d of een -t (geblust, gebonsd) en dubbelvormen van het voltooid deelwoord / van de onvoltooid tegenwoordige tijd (het is verbrand, het verbrandt).
- Bijvoeglijk gebruik van het voltooid deelwoord: wel of geen -n aan het eind (de gekookte eieren, het gebraden vlees) en dubbelvormen die bijvoeglijk worden gebruikt / in de onvoltooid verleden tijd staan (de vergrote foto, jij vergrootte de foto).
- Infinitief: dubbelvormen van de infinitief en de onvoltooid verleden tijd van een zwak werkwoord met een stam die eindigt op een -d of een -t (praten/praatte, wedden/wedde).

Interpunctie

Interpunctie heeft betrekking op de leestekens die binnen en tussen zinnen worden gebruikt. Leestekens dienen ervoor het lezen te vergemakkelijken. Tevens komt het gebruik van hoofdletters hier aan bod.

De Centrale Eindtoets bevat:

- Opgaven waarmee via het plaatsen van een hoofdletter wordt nagegaan of leerlingen de grenzen van een zin kunnen bepalen.
- Opgaven die nagaan of de leerlingen weten waar er een punt, een uitroepeteken of een vraagteken geplaatst moet/kan worden.
- Opgaven waarbij de plaats van de komma moet worden bepaald bij een opsomming of tussen twee persoonsvormen.
- Opgaven die nagaan of de leerlingen weten dat er bij de directe rede aanhalingstekens geplaatst moeten worden.
- Opgaven over de plaatsing van een dubbele punt bij een opsomming en bij de directe rede.

Grammatica

Grammatica is een breed begrip; we verstaan er in het kader van de Centrale Eindtoets alleen de analyse van zinnen in zinsdelen (redkundige ontleding) en het benoemen van woordsoorten (taalkundige ontleding) onder. Het gaat om vier soorten opgaven:

- Opgaven over de zinsvorm, waarin hoofdzin en bijzin; vragende, bedrijvende en lijdende vorm worden onderscheiden.
- Opgaven over zinsontleding, waarbij onderwerp, persoonsvorm, (werkwoordelijk) gezegde en lijdend voorwerp bevraagd worden.
- Opgaven over het benoemen van werkwoorden, waarbij het werkwoord, de infinitief, het voltooid deelwoord en tijd en getal van het werkwoord aan bod komen.
- Opgaven over woordbenoeming, waarbij zelfstandig naamwoorden, bijvoeglijke naamwoorden, voornaamwoorden, lidwoorden, telwoorden, voegwoorden en voorzetsels aan de orde zijn.

Bovenstaande inhoudelijke beschrijving van de gewenste (typen) opgaven Taal wordt door de toetsdeskundigen van Cito als richtlijn gebruikt bij de constructie van opgaven voor de Centrale Eindtoets. De toets wordt formeel vastgesteld (gefiatteerd) als de vaststellingscommissie van het College voor Toetsen en Examens (met dezelfde beschrijving in de hand) heeft geconstateerd dat de qua inhoud en opgavetype gewenste opgaven ook daadwerkelijk zijn gerealiseerd. De leden van de vaststellingscommissie hebben geen rol bij de constructie. Zij beoordelen dus niet hun 'eigen' opgaven. Vanzelfsprekend wordt bij de vaststelling ook de psychometrische kwaliteit van de opgaven (zoals gebleken uit de proeftoets) betrokken. Het gebeurt regelmatig dat opgaven die inhoudelijk aan de eisen voldoen, worden afgewezen vanwege een slecht proeftoetsresultaat (bijvoorbeeld omdat de opgave te moeilijk is).

4.2.2 Rekenen

De Toetswijzer bij de Centrale Eindtoets PO bevat een gedetailleerde beschrijving van de getoetste rekenonderdelen, onderverdeeld in de vier domeinen die ook in het referentiekader worden gehanteerd:

- Domein Getallen
- Domein Verhoudingen
- Domein Meten en Meetkunde
- Domein Verbanden.

Voorafgaand aan de inhoudelijke beschrijving zijn nog enkele opmerkingen van belang:

Gebruik van kladpapier

In de Centrale Eindtoets mogen leerlingen bij elke opgave kladpapier gebruiken. Dit is conform het referentiekader.

Opgaven met en zonder context

In de Centrale Eindtoets B (Basis) zijn wat meer opgaven zonder context, ook wel bekend als 'kale opgaven' opgenomen dan in de Centrale Eindtoets N (Niveau). In de B-toets (bestemd voor leerlingen die naar het vmbo-gt, de havo of het vwo gaan) neemt het formeel rekenen een belangrijker plaats in dan in de eindtoets N (bestemd voor leerlingen die naar het vmbo-kb of -bb gaan).

Niet te toetsen vaardigheden

In de Centrale Eindtoets kunnen niet alle doelen getoetst worden die in het referentiekader Rekenen zijn geformuleerd. Dit hangt samen met het feit dat de Centrale Eindtoets op dit moment alleen meerkeuzevragen bevat. Hiermee zijn doelen als 'kunnen uitleggen waarom...' en 'inzien dat...' en vaardigheden als 'hanteren van meetinstrumenten' en 'tekenen van diagrammen en grafieken' niet goed te meten.

Vlot uit het hoofd rekenen

Het is belangrijk dat leerlingen bepaalde basisvaardigheden, zoals het kennen van de (deel)tafels en het rekenen tot 100, goed en op tempo beheersen. De leerling heeft ze nodig om berekeningen uit te voeren zoals in de Centrale Eindtoets zijn opgenomen. Dit onderdeel is niet apart opgenomen.

Rekenen met de rekenmachine

Hoewel 'het kunnen werken met de rekenmachine en deze verstandig inzetten' behoort tot de doelen uit het referentiekader, wordt deze vaardigheid (nog) niet getoetst in de Centrale Eindtoets.

Getallen

Dit domein heeft betrekking op getallen, getalsrelaties en het uitvoeren van de elementaire bewerkingen (optellen, aftrekken, vermenigvuldigen en delen en combinaties hiervan) met hele getallen, kommagetallen en breuken.

Hieronder worden eerst de doelen voor Getalbegrip beschreven (1) en vervolgens die voor Bewerkingen (2).

1. Getalbegrip

- Hele getallen, kommagetallen, breuken en gemengde getallen uitspreken en noteren (0,45 is 'vijfenveertig honderdsten'; 'honderdtwintigduizend schrijf je: 120.000). Hieronder vallen ook spreektaalaanduidingen als 'een miljard' en 'anderhalf'.
- Wiskundetaal (formele sommentaal) gebruiken; vertalen van situaties naar wiskundetaal (3 T-shirts van 9,95 per stuk kosten samen...: wiskundetaal: $3 \times 9,95 = 29,85$).
- Verder tellen en terugtellen met hele getallen en kommagetallen.
- Vergelijken en ordenen van hele getallen, kommagetallen en breuken (Welk getal is het grootste: 0,5 0,29 0,099? Zet in volgorde van klein naar groot: $\frac{2}{3}$, $\frac{5}{6}$, $\frac{3}{8}$).
- Globaal en precies plaatsen van hele getallen, kommagetallen en breuken op een getallenlijn.
- Afronden van hele getallen, kommagetallen (12,498 afgerond op een heel getal is 12) en breuken ($\frac{24}{25}$ is ongeveer 1).
- Splitsen en samenvoegen van hele getallen en kommagetallen en kennen van de betekenis van cijfers en hun plaats in getallen ($5300 = 3 \times 100 + 5 \times \dots$; $0,06 + 70 + 0,001 = \dots$. Hoeveel is de 5 waard in het getal 4,561?).

- Relaties weten tussen breuken en kommagetallen ($0,03 = 3/100$; $4/5 = 0,8$).
- Doorzien van de structuur van het tientalig stelsel (Met hoeveel moet je 0,001 vermenigvuldigen om 1 te krijgen?).

2. Bewerkingen

Optellen en aftrekken met hele getallen en kommagetallen

- Doelmatig optellen en aftrekken gebruikmakend van eigenschappen van getallen en van bewerkingen, waaronder ook het 'rekenen met nullen': optellen en aftrekken met getallen met één/twee cijfers, gevolgd door een aantal nullen ($4000 + 60.000$; $180.000 - 2000$).
- Standaardprocedures gebruiken bij optellen en aftrekken van grotere hele getallen en kommagetallen met meer cijfers.
- Globaal (benaderend) optellen en aftrekken met grotere hele getallen en kommagetallen ($49,95 + 128,95 + 32,35$ is ongeveer $50 + 120 + 30$).

Vermenigvuldigen en delen met hele getallen en kommagetallen

- Doelmatig vermenigvuldigen en delen gebruikmakend van eigenschappen van getallen en van bewerkingen, waaronder ook 'rekenen met nullen': vermenigvuldigen en delen met getallen van één/twee cijfers gevolgd door een aantal nullen (60×400 ; $3200 : 40$). Kommagetallen vermenigvuldigen met en delen door 10, 100, 1000 ($1,8 \times 100$).
- Standaardprocedures gebruiken bij vermenigvuldigen en delen van grotere hele getallen en kommagetallen; bij delen met rest in contexten de rest juist interpreteren of verwerken (349 kinderen vervoeren in bussen. Per bus kunnen 45 kinderen mee: $659 : 45 = 14$, rest 21, dus 15 bussen nodig).
- Globaal (benaderend) vermenigvuldigen en delen met grotere hele getallen en kommagetallen ($49 \times 198,97$ is ongeveer 50×200).

Combinaties van bewerkingen met hele getallen en kommagetallen

- Berekeningen uitvoeren met combinaties van bewerkingen.
- Berekenen van gemiddelde.
- De regels kennen voor de volgorde waarin rekenbewerkingen moeten worden uitgevoerd.

Bewerkingen uitvoeren met breuken

- Vereenvoudigen en gelijknamig maken van breuken en breuken als gemengd getal schrijven ($6/9 = 2/3$; $3/8 = \dots/16$; $24/5 = 4 \frac{4}{5}$).
- Optellen en aftrekken van gelijknamige en ongelijknamige breuken en gemengde getallen in opgaven met en zonder context ($3/7 + 5/7$; $4/5 - 2/7$; $2 \frac{2}{9} + 3 \frac{2}{3}$).
- Deel nemen van een geheel getal, heel getal vermenigvuldigen met een breuk en omgekeerd, zowel in contextopgaven als in opgaven zonder context ($2/5$ van 700; $300 \times 5/6$).
- Breuk vermenigvuldigen met een breuk, of deel nemen van een breuk ($3/4 \times 5/8$; $1/2$ deel van $1/3$ deel).
- Heel getal delen door een breuk/gemengd getal en breuk/gemengd getal delen door een breuk (10 kg aardappels verpakken in zakken van $2 \frac{1}{2}$ kg; $1 \frac{1}{2}$ liter in bakjes van $1/4$ liter; $3/4 : 2/5$; $1 \frac{1}{2} : 1/8$).

Verhoudingen

In het basisonderwijs gaat het er bij dit domein om dat leerlingen leren structuur en samenhang van verhoudingen op hoofdlijnen te doorzien en dat zij er in praktische situaties mee kunnen rekenen. Verhoudingen kunnen beschreven worden in verhoudingstaal ('één op de tien Nederlanders'; 'het aantal fietsers is twee keer zo groot als het aantal automobilisten'), maar ook in breukentaal ('driekwart van de inwoners is ouder dan 25 jaar, dat is drie op de vier') en met procenten ('70 procent van de mensen is voor de aanleg van een rondweg'). Binnen verhoudingen gaat het dus soms ook om het rekenen met breuken en procenten.

3. Verhoudingen

- Verhoudingen herkennen, benoemen, schrijven en gebruiken als 'zoveel op/per/van de zoveel', als deel van een geheel, als 1:100 (staat tot), als breuk of als percentage; en deze verschillende benamingen met elkaar in verband brengen.
- Oplossen van eenvoudige verhoudingsproblemen met hele getallen, kommagetallen en breuken, bijvoorbeeld bij gebruik van recepten, snelheid, prijs per stuk/kg/liter, mengen, afstanden, vergelijken van groepen met een kenmerk, vergroten en verkleinen.
- Het begrip 'schaal' kennen en rekenen met schaallijnen en schaalnotaties, bijvoorbeeld bij een plattegrond of bij modelbouw.
- Verhoudingen met elkaar vergelijken en bepalen of ze gelijk zijn aan elkaar (1 op de 3, is dat meer of minder dan de helft).
- Bij verdelingen de ontbrekende percentages berekenen op basis van de kennis dat het geheel 100% is (cirkeldiagram; samenstelling van stof: 80% katoen, de rest is viscose).
- Rekenen met percentages via het deel nemen van, via rekenen met breuken, via verhoudingen, via vermenigvuldigen met een bijbehorend kommagetal of via 1%.
- In een context met eenvoudige getallen berekenen hoeveel procent het deel is (18 van de 90 is 20%), of de toename of afname bedraagt (hoeveel procent winst/korting/verlies/toename).
- Betekenis geven aan percentages boven 100%, en hiermee rekenen in opgaven.
- Verhoudingen en procenten in elkaar omzetten. Procenten en kommagetallen in elkaar omzetten. Breuken en procenten in elkaar omzetten. En daar waar mogelijk verhoudingen en breuken in elkaar omzetten.

Metten en Meetkunde

Bij dit domein gaat het vooral om het meten van allerlei verschijnselen in de werkelijkheid: lengte, omtrek, oppervlakte, inhoud, gewicht, tijd, snelheid en geld, met verschillende instrumenten. Ook het omzetten van maateenheden en de opbouw en decimale structuur van het metrieke stelsel vallen hieronder.

Hieronder worden eerst de doelen voor de verschillende deeldomeinen van Metten beschreven (4) en vervolgens de doelen voor Meetkunde (5).

4. Metten

Lengte en omtrek

- Meetinstrumenten hanteren en aflezen: liniaal, meetlint, rolmaat, kilometerteller.
- Orde van grootte van lengte en omtrek in een situatie inschatten op basis van referentiematen (een deur/bed is ongeveer 2 meter hoog/lang; een mier is

wellicht 5 mm, maar geen 50 mm, 5 cm of 5 meter); betekenis geven aan lengtematen en kiezen van de juiste lengtemaat bij een situatie (de afstand tussen steden niet in meters of centimeters maar in kilometers).

- Omtrek berekenen of benaderen van rechthoekige en grillige figuren. Gebruiken van de formule 'omtrek is twee keer de lengte en twee keer de breedte' bij het berekenen van omtrek van rechthoeken.
- Samenhang tussen standaardmaten kennen en herleidingen uitvoeren (bijvoorbeeld 2,3 km = 2300 meter; 1,65 meter is 1 meter en 65 centimeter). Vergelijken en ordenen van lengtematen (afstand van 2,5 km is langer dan van 2400 meter). Het betreft: km, hm, dam, m, dm, cm, mm.
- Oplossen van toepassingsproblemen waarbij herleidingen en berekeningen met lengtematen uitgevoerd moeten worden, inclusief het gebruiken van samenstellingen als km/u; prijs/km.

Oppervlakte

- Orde van grootte van oppervlakte in een situatie inschatten op basis van referentiematen (de oppervlakte van een deur is ongeveer 2 m², van een slaapkamer zo'n 8 m² tot 20 m²; een hectare is ongeveer twee voetbalvelden); betekenis geven aan oppervlaktematen en kiezen van de juiste oppervlaktemaat bij een situatie (de oppervlakte van een tuin druk je niet uit in cm², van een stad niet in m²).
- Oppervlakte bepalen met natuurlijke of gegeven maten (bijvoorbeeld het aantal tegels dat op een vloer komt te liggen) en oppervlakte berekenen van rechthoeken en rechthoekige figuren en van grillige figuren benaderen. Gebruiken van de formule 'lengte keer breedte is oppervlakte' bij het berekenen van oppervlaktes van rechthoeken.
- Samenhang tussen standaardmaten kennen en herleidingen uitvoeren (2 m² = 20.000 cm²; 1 hectare is 100 are, is 100 m bij 100 m). Het betreft: km², hm², dam², m², dm², cm², mm², hectare en are. Vergelijken en ordenen van oppervlaktematen.
- Oplossen van toepassingsproblemen waarbij herleidingen en berekeningen met oppervlaktematen uitgevoerd moeten worden; hieronder valt ook het gebruiken van samenstellingen als prijs/m².

Inhoud

- Meetinstrumenten hanteren en aflezen: maatbeker, schaalverdeling.
- Inhoud berekenen van rechthoekige figuren (balken) en hierbij de formule 'inhoud is lengte keer breedte keer hoogte' gebruiken. De inhoud berekenen via het tellen van een natuurlijke maat (pakjes in een doos).
- Samenhang tussen standaardmaten kennen en herleidingen uitvoeren (3 dm³ = 3000 cm³; 75 cl is 0,75 l). Het betreft: liter, dl, cl en ml, m³, dm³, cm³, mm³. Vergelijken en ordenen van inhoudsmaten (1,5 l is meer dan 500 ml; 5 l is evenveel als 5 dm³).
- Oplossen van toepassingsproblemen waarbij herleidingen en berekeningen met inhoudsmaten uitgevoerd moeten worden. Hieronder valt ook het rekenen met samenstellingen als prijs/liter.

Gewicht

- Meetinstrumenten hanteren en aflezen: balans, analoge en digitale weegschaal.
- Betekenis geven aan gewichtsmaten en kiezen van de juiste maat bij een situatie (een pak suiker is meestal een kg, gewicht van een pilletje druk je uit in milligram, gewicht van een groter persoon in kg en niet in gram, uitdrukking

in tonnen wordt gebruikt bij echt zware en grote dingen (containers, lading van vrachtwagen)).

- Samenhang tussen standaardmaten kennen en herleidingen uitvoeren (2,3 kg = 2300 gram; 1 ton is 1000 kg). Het betreft: mg, cg, dg, g, dag, hg, kg, ton. Vergelijken en ordenen van gewichts-aanduidingen (2,5 kg is meer dan 2000 gram).
- Oplossen van toepassingsproblemen waarbij herleidingen en berekeningen met gewichtsmaten uitgevoerd moeten worden. Hieronder vallen ook samenstellingen als prijs/kg.

Tijd en snelheid

- Aflezen van tijden, zowel analoog als digitaal en analoge en digitale tijden koppelen.
- Rekenen met tijdsaanduidingen (tijdsverschillen; tijdsduur).
- Betekenis kennen van kalenderaanduidingen en datumaanduidingen (dag, week, maand, kwartaal, jaar, schrikkeljaar, decennium, eeuw; 27-11-1999); omgaan met de kalender en rekenen met kalenderaanduidingen (volgorde van de maanden van het jaar, dagen per maand, structuur van de kalender; weeknummers).
- Samenhang tussen tijdsaanduidingen en kalenderaanduidingen kennen en herleidingen uitvoeren (anderhalf uur is 90 minuten; in welk kwartaal valt week 15?). Samenstellingen gebruiken als snelheid: km/u; m/s.

Geld

- Gepast samenstellen van bedragen met munten/biljetten en totale waarde van munten/biljetten bepalen.
- Omwisselen van munten/biljetten in andere munten/biljetten.
- Rekenen met geldbedragen in euro's (betalen/wisselgeld/bijbetalen).
- Oplossen van toepassingsproblemen waarin met euro's en andere valuta (wisselkoersen) gerekend wordt; samenstellingen gebruiken als prijs/u; prijs/m; prijs/l.

5. Meetkunde

- De namen en eigenschappen kennen van meetkundige figuren (driehoek, vierkant, rechthoek, vierhoek, vijfhoek, zeshoek, cirkel, kubus, balk, bol).
- Relaties zien tussen 2D en 3D objecten als foto's, plattegronden, legenda's, bouwplaten.
- Aanzichten van ruimtelijke objecten herkennen en interpreteren.
- Interpreteren van plattegronden en bouwtekeningen.
- Mentaal innemen van standpunten (in de meetkundige betekenis van standpunt).
- Routes en plaatsen beschrijven en lezen op een kaart (ook met behulp van een rooster).
- Aanduidingen op een windroos kennen (N, NO, O, ZO, Z, ZW, W, NW).
- Symmetrie herkennen, spiegelen in 2D en 3D, spiegelassen in 2D-figuren vinden.
- Meetkundige patronen herkennen en voortzetten.

Verbanden

Binnen dit domein gaat het voor het basisonderwijs om het omgaan met tabellen, diagrammen en grafieken, legenda's en assenstelsels. Tabellen, diagrammen en grafieken worden frequent gebruikt om getalsmatige (kwantitatieve) gegevens op een

compacte en overzichtelijke manier weer te geven. Het aflezen, interpreteren en combineren van de verschillende representatievormen zijn belangrijke vaardigheden voor kinderen om te leren. Ook gaat het om het uitvoeren van berekeningen met deze verschillende gegevens.

6. Tabellen, diagrammen en grafieken

- Lezen en interpreteren van gegevens uit tabellen. Met deze gegevens rekenen en op basis van de gegevens trends ontdekken en voorspellingen doen.
- Lezen en interpreteren van gegevens uit diagrammen zoals cirkeldiagram, staafdiagram (horizontaal, verticaal) en beelddiagram. Met deze gegevens rekenen en op basis van de gegevens trends ontdekken en voorspellingen doen.
- Lezen en interpreteren van gegevens uit lijngrafieken, ook met meer lijnen in één grafiek. Met deze gegevens rekenen en op basis van de gegevens trends ontdekken en voorspellingen doen.
- Verschillende informatiebronnen met elkaar in verband brengen (tabellen, diagrammen en grafieken) en hieruit gegevens lezen, vergelijken en interpreteren. Met deze gegevens rekenen en op basis van de gegevens trends ontdekken en voorspellingen doen.

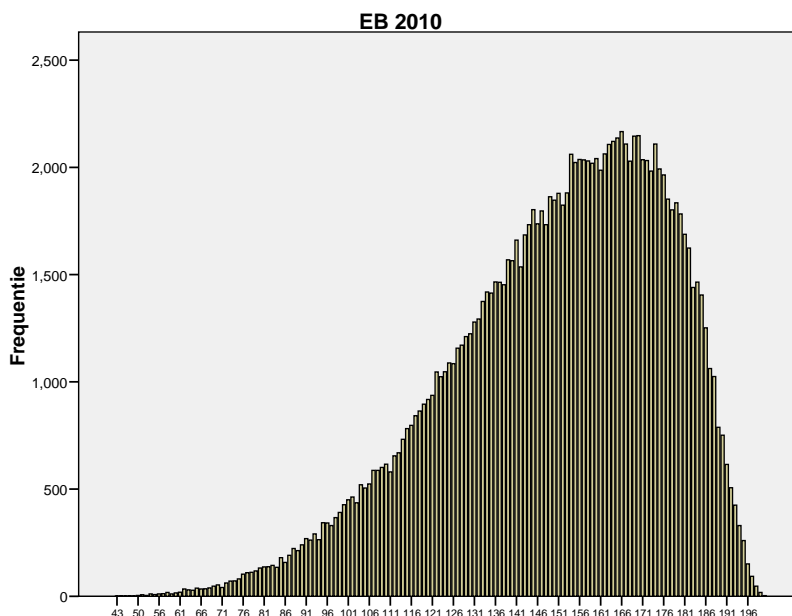
Bovenstaande inhoudelijke beschrijving van de gewenste (typen) opgaven Rekenen wordt door de toetsdeskundigen van Cito als richtlijn gebruikt bij de constructie van opgaven voor de Centrale Eindtoets. De toets wordt formeel vastgesteld (gefiatteerd) als de vaststellingscommissie van het College voor Toetsen en Examen (met dezelfde beschrijving in de hand) heeft geconstateerd dat de qua inhoud en opgavetype gewenste opgaven ook daadwerkelijk zijn gerealiseerd. De leden van de vaststellingscommissie hebben geen rol bij de constructie. Zij beoordelen dus niet hun 'eigen' opgaven. Vanzelfsprekend wordt bij de vaststelling ook de psychometrische kwaliteit van de opgaven (zoals gebleken uit de proeftoets) betrokken. Het gebeurt regelmatig dat opgaven die inhoudelijk aan de eisen voldoen, worden afgewezen vanwege een slecht proeftoetsresultaat (bijv. de opgave is te moeilijk).

4.3 Kenmerken

De toets is te beschrijven in de vorm van specifieke statistische descriptoren die aan bepaalde eisen dienen te voldoen. Dat geldt zowel voor beschrijving op itemniveau als voor beschrijving op het niveau van de gehele toets of deelttoetsen. Ook al is de toets gebaseerd op de item respons theorie, we geven hier kwaliteitsrichtlijnen in termen van parameters die zijn ontleend aan de klassieke testtheorie omdat deze wat eenvoudiger te beschrijven en voor niet-ingewijden gemakkelijker te interpreteren zijn. Voor de opgaven gelden de moeilijkheidsgraad (p -waarde: percentage correct beantwoord) en de item-totaalcorrelatie (R_{it}) als de belangrijkste kenmerken. Wat de moeilijkheid van de opgaven betreft: voor de opgavenselectie geldt het uitgangspunt dat de p -waarden bij voorkeur tussen .40 - .90 moeten liggen en dat de opgaven in de Centrale Eindtoets gemiddeld een p -waarde hebben van ongeveer .70 of iets hoger. Dat geldt voor elke deelttoets afzonderlijk. Per deelttoets mogen ten hoogste enkele opgaven buiten het aangegeven bereik liggen (bijvoorbeeld wanneer bij de itemselectie overwegingen van inhoudelijke representativiteit de doorslag geven). Voor R_{it} (de correlatie tussen de itemscore en de totaalscore op de betreffende deelttoets) geldt dat deze minimaal .20 bedraagt om voldoende te kunnen worden genoemd en minimaal .30 voor een beoordeling als goed (vergelijk de criteria in het COTAN Beoordelingssysteem; Evers et al., 2010). Voor de Centrale Eindtoets als geheel geldt dat ten hoogste enkele items (verdeeld over verschillende deelttoetsen) in de toets

mogen voorkomen met waarden lager dan .20. In die gevallen moeten inhoudelijke overwegingen de doorslag hebben gegeven. Het verdient aanbeveling om ook de item-resttotaal waarden te rapporteren (waarbij de betreffende itemscore niet in de toetsscore wordt meegeteld). Deze vormt een betere indicator voor de interne consistentie van een item, maar het COTAN Beoordelingssysteem kent daarvoor geen kwaliteitscriteria. De gemiddelde R_{it} -waarde per deelttoets dient uit te komen tussen .35 en .45 (afhankelijk van de aard van de vaardigheid; niet bij alle deelvaardigheden is een hoge interne consistentie even gemakkelijk te bereiken). Dat is weliswaar een hoge streefwaarde, maar in het licht van de gehanteerde constructie- en itemselectieprocedures moet dit haalbaar zijn. Uiteindelijk is het belangrijkste criterium dat de betrouwbaarheid in termen van homogeniteit van de toets of deelttoets aan de eisen voldoet (zie verder het hoofdstuk over betrouwbaarheid).

Voor de deelttoetsen geldt dat de gemiddelde score (in overeenstemming met de gemiddelde p-waarde) op ongeveer 70% van de maximaal te behalen toetsscore ligt (bij ongewogen itemscores). Gegeven de gekozen en gerealiseerde moeilijkheidsgraad zal de scoreverdeling voor de deelttoetsen eentoppig zijn en min of meer linksscheef. De totaalscore wordt via lineaire transformatie omgezet in een standaardscore. Dit houdt in dat de ruwe score wordt vermenigvuldigd met een getal (A) en dat er daarna een vast getal (B) bij wordt opgeteld. De waarden van A en B worden jaarlijks bepaald door middel van een IRT-analyse die de gegevens levert voor een daarop volgend equivaleringsonderzoek. Die equivalering levert vervolgens de A en B coëfficiënten (zie verder hoofdstuk 2). De resulterende standaardscore heeft het karakteristieke gemiddelde van omstreeks 535. De verdeling heeft de linksscheve vorm zoals deze hieronder is weergegeven (ontleend aan van Boxtel et al., 2011):



4.4 Procedures rond afname en rapportage

In het Toetsbesluit PO zijn een aantal proceseisen opgenomen waar elke eindtoets, dus ook de Centrale Eindtoets, aan moet voldoen. In de bijlage vindt u een puntsgewijze opsomming van deze eisen.

Voor de Centrale Eindtoets zijn vastomlijnde procedures beschreven in de handleiding. Deze handleiding wordt in gedrukte vorm aan alle deelnemende scholen gezonden samen met het toetsmateriaal (opgavenboekjes, antwoordbladen, etc.). Daarnaast is de handleiding voor iedereen vrij toegankelijk en beschikbaar op de website www.centraleeindtoetspo.nl

Kort samengevat voldoet de Centrale Eindtoets op de volgende wijze aan de proceseisen uit het Toetsbesluit PO:

- **Rapportage**
Elke leerling die deelneemt aan de Centrale Eindtoets krijgt een leerlingrapport waarop het resultaat op de toets vermeld is (standaardscore en scores op de onderdelen), het niveau waarop de toets is afgelegd (B of N), een advies voor een passend brugklastype (bijv. havo of havo/vwo) en een indicatie van de kansen op succes in het geadviseerde brugklastype (poppetjesgrafiek). Daarnaast krijgt elke leerling een rapport waarop vermeld is welk referentieniveau behaald is voor rekenen (alle domeinen) en voor de taaldomeinen lezen en taalverzorging (inclusief grammaticale begrippen). Het leerlingrapport referentieniveaus bevat voor deze drie gebieden ook profielanalyses. In blokjesgrafieken wordt getoond hoe de leerling presteerde op de afzonderlijke deelgebieden van rekenen, lezen en taalverzorging.
- **Geschiktheid voor alle leerlingen**
Voor leerlingen met specifieke ondersteuningsbehoeften zijn bij de Centrale Eindtoets verschillende aangepaste versies ontwikkeld. Naast de brailleversie voor blinde leerlingen zijn er vergrote versies voor slechtziende of dyslectische leerlingen, een zwart-witversie voor kleurenblinde leerlingen (alleen van toepassing bij het onderdeel Wereldoriëntatie) en cd's met een gesproken versie, een daisy-versie, een Kurzweil-versie of een pdf naar spraaksynthese, alle bedoeld voor dyslectische of slechtziende leerlingen. Voor leerlingen met concentratieproblemen of stoornissen in het autistisch spectrum kan ook de digitale afnamevorm van de Centrale Eindtoets erg geschikt zijn, omdat daarbij slechts één opgave tegelijk in beeld is. Ook kan bij de digitale toets de afname desgewenst over meer dan drie dagen worden gespreid. Voor leerlingen die ondanks deze aanpassingen problemen voorzien bij de afname, is maatwerk mogelijk als de school contact opneemt met het College voor Toetsen en Examens. Deze mogelijkheid wordt expliciet genoemd in de handleiding bij de Centrale Eindtoets, en in de brochure 'Leerlingen met speciale ondersteuningsbehoeften', die beschikbaar is op www.centraleeindtoetspo.nl
- **Afnamemoment**
De Centrale Eindtoets kent een papieren afname en een digitale afname. Beide afnames vallen tussen 15 april en 15 mei. De papieren afname vindt plaats op drie achtereenvolgende ochtenden. In 2015 betreft het de ochtenden van 21, 22 en 23 april. De digitale afname vindt in 2015 plaats op 15, 16 en 17 april en op 28, 29 en 30 april. Na 30 april start het tweede afnamemoment, voor leerlingen die (bijvoorbeeld door ziekte) niet mee konden doen op de hierboven genoemde dagen. Het tweede afnamemoment (de 'inhaaltoets') loopt in 2015 door tot en met vrijdag 29 mei. Na die dag is de afname van de Centrale Eindtoets niet meer mogelijk.
- **Wijze van afnemen**
De handleiding bij de Centrale Eindtoets bevat uitgebreide richtlijnen voor de voorbereiding en uitvoering van de afname, zowel op papier als digitaal. In het Toetsreglement Centrale Eindtoets PO (te downloaden via

www.centraleeindtoetpo.nl) wordt de juiste wijze van afnemen formeel beschreven. Daarbij wordt geregeld:

- de wijze waarop de directeur de leerlingen aanmeldt voor de eindtoets
 - welke hulpmiddelen de leerlingen kunnen gebruiken
 - de wijze waarop de eindtoets kan worden afgelegd door leerlingen voor wie een afwijkende wijze van toetsing noodzakelijk is
 - de wijze waarop de toetsopgaven aan de directeur ter beschikking worden gesteld
 - de wijze waarop de geheimhouding van de toetsopgaven wordt geregeld en de wijze en het moment waarop de toetsopgaven openbaar worden gemaakt
 - de wijze waarop door de directeur toezicht wordt gehouden op leerlingen die de eindtoets afleggen.
- Constructie en verantwoording
Uitgebreide informatie over de constructie en samenstelling van de Centrale Eindtoets is te vinden eerder in dit hoofdstuk. De hiernavolgende hoofdstukken bevatten een verantwoording van de gemaakte keuzes ten aanzien van de psychometrische aspecten van de toets.

5 Normering

De functie van de Centrale Eindtoets is het geven van onafhankelijke informatie over wat een leerling in acht jaar basisonderwijs heeft geleerd, ten behoeve van de keuze van het brugklatype in een school voor voortgezet onderwijs dat het best bij deze leerling past. Zo bezien vormen in principe alle leerlingen in groep 8 van het basisonderwijs in Nederland de doelgroep van het instrument. In de regel neemt een ruime meerderheid van de doelgroep deel aan de Centrale Eindtoets. In 2010 namen bijvoorbeeld ruim 137.000 leerlingen uit het reguliere basisonderwijs deel aan de Cito Eindtoets Basisonderwijs, naar schatting was dit toen een kleine 80% van de doelpopulatie. Bij de eerste afname van de Centrale Eindtoets in 2015 werden aanzienlijk meer (namelijk omstreeks 165.000) leerlingen ingeschreven. Dit aantal is overigens niet een-op-een met het aantal van 2010 te vergelijken, omdat het inclusief de leerlingen uit het Speciaal (Basis-)Onderwijs en leerlingen van Nederlandse scholen in het buitenland is.

Bij het normeren van een test wordt normaliter een representatieve steekproef van voldoende omvang getrokken uit de populatie van leerlingen die nu (of in de toekomst) de doelgroep van de test vormen. Daarbij heeft elk lid van de populatie idealiter dezelfde kans om in die steekproef terecht te komen. Bij de Centrale Eindtoets wordt een dergelijke steekproef niet getrokken. Men zou dit kunnen doen, maar het hoeft niet, eenvoudigweg omdat de normen worden vastgesteld nadat de toets bij alle leerlingen is afgenomen en voor alle leerlingen de scores bekend zijn. Bij het omzetten van ruwe scores in standaardscores, dan wel het bepalen van de bijbehorende percentielscores op basis van de cumulatieve frequenties wordt een leerling dus vergeleken met de totale groep leerlingen bij wie de toets in dat schooljaar wordt afgenomen. Iets soortgelijks gebeurt bij het koppelen van de score van een leerling aan het advies over het best passend brugklatype, een vorm van criteriumgerichte normering. Door middel van het Toelatings- en doorstroomonderzoek worden de toetsscores van de deelnemende leerlingen gekoppeld aan gegevens (met betrekking tot brugklatype) over de instroom in het eerste leerjaar van het voortgezet onderwijs, in het schooljaar volgend op de toetsafname.

De Centrale Eindtoets wordt jaarlijks opnieuw geconstrueerd en geëquivaaleerd. Ook de normering vindt steeds jaarlijks plaats. Omdat de normering geschiedt op basis van de totale groep deelnemers is de normeringssteekproef zeer groot. De normeringssteekproef valt als het ware min of meer samen met de populatie. Hoe groot de groep deelnemers aan de Centrale Eindtoets echter ook is, toch is het nodig om ten aanzien van een (beperkt) aantal variabelen na te gaan in hoeverre deze groep representatief is voor de populatie als geheel. De scholen die meedoen aan de Centrale Eindtoets hebben zich immers zelf geselecteerd en het is niet uitgesloten dat de factoren die hierbij een rol spelen ervoor zorgen dat de groep deelnemers aan de Centrale Eindtoets op een systematisch andere wijze is samengesteld dan de groep die ervoor heeft gekozen om niet deel te nemen, respectievelijk anders dan de totale doelpopulatie.

Het is daarom interessant en belangrijk om inzicht te hebben in de verdelingskenmerken van de groep scholen die deelneemt aan de Centrale Eindtoets in vergelijking met die van de totale doelpopulatie. Men kan dit nagaan op het niveau van de scholen en op het niveau van de leerlingen. Hierover handelt de volgende paragraaf.

5.1 Kenmerken

Kenmerken die normaliter relevant worden geacht om de relatie tussen normeringssteekproef en populatie in kaart te brengen worden besproken in het COTAN Beoordelingssysteem (Evers et al., 2010). De normeringssteekproef zou in

ieder geval moeten worden beschreven in termen van leeftijd, sekse, etniciteit en regio.

Op schoolniveau zijn de kenmerken *regio*, urbanisatiegraad of *mate van verstedelijking*, *schoolgrootte* en *percentage achterstandsleerlingen* bekend en relevant.

De verdeling naar *regio* gaat uit van vier regio's die overeenkomen met de volgende indeling naar provincies (CBS-indeling naar landsdeel):

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

De verdeling naar urbanisatiegraad of *mate van verstedelijking* is terug te voeren op de bij het CBS gebruikelijke indeling naar vijf niveaus, namelijk zeer sterk, sterk, matig, weinig en niet verstedelijkt.

Wat betreft *schoolgrootte* kan onderscheid worden gemaakt tussen grote scholen (≥ 200 leerlingen) en kleine scholen (< 200 leerlingen).

Het *percentage achterstandsleerlingen* is gebaseerd op het aantal kinderen per school met een afwijkend leerlinggewicht (formatiegewicht van 0.3 of 1.2. DUO levert deze gegevens per school aan).

Op leerlingniveau zijn de variabelen sekse, leeftijd en etnische herkomst van belang. De verdeling naar *sekse* levert geen probleem op.

Wat betreft *leeftijd* gaat het hier natuurlijk om een betrekkelijk homogene groep 11- à 12-jarige leerlingen die op een bepaald moment in groep 8 collectief de Centrale Eindtoets maakt. Daarbij is er sprake van de gebruikelijke spreiding van leeftijden over een schooljaar, met dien verstande dat – naast de grote groep reguliere leerlingen – sommige leerlingen voorlijk genoemd kunnen worden (bijvoorbeeld omdat ze ooit een klas hebben overgeslagen) en andere vertraagd (bijvoorbeeld omdat ze een of meer keer hebben gedoubleerd). Voor deze verdeling naar leeftijd zijn geen exacte populatiegegevens beschikbaar.

Wat betreft *etnische herkomst*: er zijn hierover geen populatiegegevens beschikbaar. Hetzelfde geldt voor de variabele *leerlinggewicht* (op het niveau van de individuele leerling) en de *taal die thuis gesproken wordt*. In het verleden konden hierover gegevens worden verzameld via het antwoordblad; bij de Centrale Eindtoets is dit niet geoorloofd. Dit betekent dat de analyses zich moeten beperken tot de achtergrondvariabele *percentage achterstandsleerlingen* op basis van leerlinggewicht (op schoolniveau). Op grond van peilingsonderzoek (PPON) is bekend dat dit kenmerk een vrij aardige benadering vormt van de achtergrondvariabele etnische herkomst.

Wil de normering deugen, dan zal de groep deelnemers aan de Centrale Eindtoets niet anders moeten zijn samengesteld dan de totale doelgroep-populatie. Om te checken of dit het geval is, worden de verdelingen vergeleken zoals in onderstaande tabel (voorbeeldgegevens hebben betrekking op de Cito eindtoets Basisonderwijs (EB) 2010 en zijn ontleend aan van Boxtel et al., 2011). Er wordt een soortgelijk beeld verwacht: de subgroep van scholen die de Centrale Eindtoets afnemen bij hun leerlingen vormt een nauwkeurige afspiegeling van de totale populatie van basisscholen in Nederland. Dat geldt voor alle besproken achtergrondkenmerken. Afwijkingen zullen worden getoetst, waarbij de effectgroottes in termen van Phi doorslaggevend zijn (deze moeten $< .10$ zijn).

De analyses op schoolniveau worden aangevuld met analyses op leerlingniveau voor de kenmerken sekse en leeftijd. Afwijkingen van de (geschatte) populatieverdeling worden

getoetst; ook hier zijn effectgroottes (Phi) doorslaggevend. De verwachting is dat ook ten aanzien van deze leerlingkenmerken sprake is van een nauwkeurige afspiegeling van de populatie. Dit zal sterker het geval zijn naarmate een groter percentage van de doelgroep deelneemt aan de Centrale Eindtoets.

Verdeling van basisscholen naar regio, mate van verstedelijking, schoolgrootte en percentage achterstandsl leerlingen voor de Nederlandse populatie en voor de groep deelnemers aan de Centrale Eindtoets (voorbeeld)

	Regio		Schoolgrootte		
	Populatie (%)	EB 2010 (%)	Populatie (%)	EB 2010 (%)	
noord	15.9	15.2	klein	49.5	49.2
oost	24.2	24.7	groot	50.5	50.8
west	41.1	39.9			
zuid	18.8	20.3	totaal	100	100
totaal	100	100			

	Mate van verstedelijking		Percentage achterstandsl leerlingen		
	Populatie (%)	EB 2010 (%)	Populatie (%)	EB 2010 (%)	
zeer sterk	11.7	12.6	0 – 10%	49.8	48.1
sterk	20.9	19.6	11 – 25%	31.1	31.7
matig	20.3	19.9	26 – 50%	12.0	12.3
weinig	26.7	27.7	51 – 100%	7.1	7.9
geen	20.4	20.2			
totaal	100	100	totaal	100	100

5.2 Kalibratie

5.2.1 De stappen in de kalibratie

Bij de kalibratie wordt iteratief een vast aantal stappen doorlopen:

1. Met behulp van het programma OPCAT worden de discriminatie-indices in OPLM ingesteld en worden indien noodzakelijk de antwoordcategorieën in de data gehercodeerd.
2. Vervolgens worden de itemparameters geschat (met behulp van de CML-methode).
3. Met behulp van M-toetsen wordt gecontroleerd of de discriminatie-indices goed zijn ingesteld.
4. Een volgende controle betreft de overschrijdingskansen van de S-toetsen en een grafische modelcontrole door middel van het programma OPDRAW (grafische inspectie van de itemkarakteristieke curven).
5. Vervolgens vindt een globale modelcontrole plaats in de vorm van een R1c-toets en de verdeling van de overschrijdingskansen van de S-toetsen.
6. Daarna wordt nadrukkelijk getoetst op vormen van itembias, in ieder geval naar sekse (analyses naar etnische herkomst kunnen niet worden uitgevoerd omdat de betreffende gegevens niet kunnen worden verzameld).

De stappen 1 tot en met 6 worden een aantal malen doorlopen tot het resultaat bevredigend is. Afhankelijk van de uitkomsten kunnen items worden verwijderd. Ook

inhoudelijke overwegingen spelen een rol in dit beslissingsproces. De opgaven vormen na de kalibratie een gekalibreerde opgavenbank, waarbij de opgaven per onderscheiden vaardigheidsdimensie een beroep doen op hetzelfde complex aan vaardigheden (dezelfde 'latente vaardigheid').

5.2.2 *Evaluatie van de kalibratieprocedure*

De procedures die in de kalibratie worden doorlopen en de in het kader daarvan uitgevoerde toetsen, lenen zich tevens voor een evaluatie van de kwaliteit van de kalibratie. De kalibratie wordt per onderscheiden deelvaardigheid (deeltoets) uitgevoerd, maar omdat de verslaglegging op die manier tot een onoverzichtelijke hoeveelheid resultaten zou leiden, wordt geëvalueerd op het niveau van de totaalscores voor in ieder geval Taal, Rekenen en Wereldoriëntatie. Wanneer de kalibratie-resultaten op dit niveau goed zijn, mag worden aangenomen dat een meer nauwkeurige kalibratie en evaluatie op detailniveau alleen maar betere resultaten laat zien.

Er wordt op de volgende manieren gerapporteerd (geëvalueerd):

- Grafische weergave van S-toetsen. Per hoofdonderdeel wordt een best passende, een slechtst passende en een representatieve (doorsnee) afbeelding gerapporteerd. De representatieve afbeelding moet een correct beeld vertonen (dat wil zeggen, er wordt beantwoord aan de eis dat de geobserveerde proportie binnen het 95%- betrouwbaarheidsinterval van de geschatte proporties ligt). Bij de slechtst passende opgaven zal die niet voor alle scoregroepen het geval zijn; dit mag slechts in uitzonderingsgevallen voorkomen.
- Steekproefsgewijs (dat wil zeggen voor tien steekproeven uit de kalibratiebestanden met elk een omvang van ongeveer 500) worden de rechteroverschrijdingskansen van alle S-toetsen gerapporteerd. Deze dienen gelijk verdeeld te zijn over het (0,1) interval. Significante waarden mogen voorkomen, maar dienen uitzonderingen te vormen.
- Voor dezelfde steekproeven worden overall R1c-toetsingen gerapporteerd. De R1c-waarden zijn bij voorkeur niet significant, maar in ieder geval niet groter dan anderhalf maal het aantal vrijheidsgraden.
- Constante 'c', een poging om de nauwkeurigheid van de itemparameterschattingen te beoordelen op basis van een constante die weergeeft hoe de relatie is tussen de standaardfout van de moeilijkheidsparameter van een item en de standaarddeviatie van de vaardigheidsverdeling van de kalibratiepopulatie (zie het COTAN Beoordelingssysteem; Evers et al., 2010). Het beoordelingssysteem geeft ook richtlijnen voor het beoordelen van de grootte van deze 'c'. Deze dient te worden beoordeeld als goed als de waarde lager is dan of gelijk aan .20. Waarden tussen .30 en .40 kunnen nog als voldoende worden beschouwd. Resultaten uit het verleden geven aan dat de lat voor de Centrale Eindtoets gerust aanzienlijk hoger kan worden gelegd, met gemiddelde waarden tussen .03 en .10, waarbij geen enkele waarde als onvoldoende beoordeeld mag worden (<.20).

5.2.3 *Adviescategorieën*

Op basis van de vastgestelde cumulatieve verdeling worden ruwe scores ('aantal goed') omgezet in percentielscores voor Taal, Rekenen en Wereldoriëntatie (indien van toepassing) en Totaal. Zowel ruwe scores als percentielscores worden vermeld op het zogenoemde leerlingrapport. Daarnaast wordt de totaalscore (zonder de scores te normaliseren, zoals dat in veel andere instrumenten wél gebeurt) omgezet in een standardscore met een bereik van 501 tot en met 550, een gemiddelde van (omstreeks) 535 en een standaarddeviatie van om en nabij de 9 (de werkelijke waarden kunnen van jaar tot jaar licht verschillen). De percentielscores gelden als een (adequate) vorm van relatieve normering.

De resulterende standaardscores zijn op zichzelf tamelijk betekenisloos (ook al heeft het jarenlang omgaan met deze scores in het onderwijs – in het gebruik van de Cito Eindtoets Basisonderwijs – voor een duidelijk omschreven betekenis-toekenning gezorgd). Hun werkelijke betekenis ontleenen de getallen aan de adviescategorieën die eraan verbonden worden met betrekking tot de brugklastypen die voor de betreffende leerling als het meest geschikt worden beschouwd. Deze adviescategorieën en de grenzen waarmee deze worden afgebakend zijn op te vatten als een vorm van criteriumgeoriënteerde normering.

In de zogenoemde 'poppetjesgrafiek' op het leerlingrapport wordt per brugklastype aangegeven hoeveel procent van de leerlingen met eenzelfde score in het betreffende brugklastype is geplaatst. Verder worden aan de leerkracht op grafische wijze ook aanvullende gegevens verstrekt over het percentage kinderen dat met deze score in het betreffende brugklastype niet werd bevorderd naar het tweede leerjaar of werd overgeplaatst naar een lager brugklastype. Deze gegevens die ouders, leerkrachten en kinderen steun moeten bieden bij het maken van een keuze voor vervolgonderwijs, zijn ontleend aan de zogenoemde Toelatings- en doorstroomonderzoeken (T&DO) die Cito met regelmaat in samenwerking met het CBS organiseert (zie verder onder criteriumvaliditeit).

Op basis van het meest recente toelatings- en doorstroomonderzoek met de Cito Eindtoets Basisonderwijs kunnen de volgende intervallen worden verwacht:

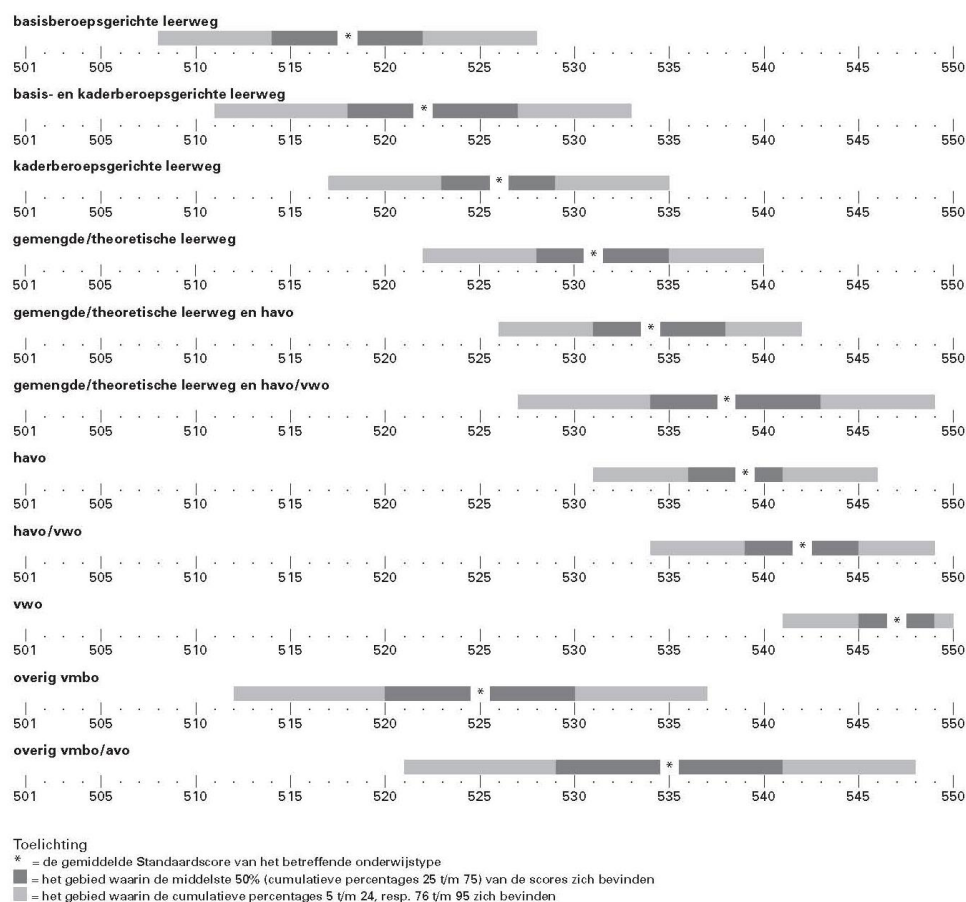
Interval dat het uitgangspunt vormt voor de interpretatie van de standaardscore per brugklastype

Brugklastype	Interval
basisberoepsgerichte leerweg	501-520
basis- en kaderberoepsgerichte leerweg	519-525
kaderberoepsgerichte leerweg	523-528
gemengde / theoretische leerweg	529-533
gemengde / theoretische leerweg en havo	533-536
havo	537-540
havo / vwo	540-544
vwo	545-550

Ten slotte wordt grafisch weergegeven hoe de prestatieverdeling voor de verschillende onderscheiden brugklastypen er bij de instroom in leerjaar 1 uitzag. Per brugklastype wordt de gemiddelde score aangegeven en wordt door middel van balkjes zichtbaar gemaakt waar de grenzen liggen van de volgende drie standaardscoregebieden (in termen van percentages):

- 5% tot en met 24%
- 25% tot en met 75%
- 76% tot en met 95%

Hieronder is door middel van een voorbeeld aangegeven hoe zo'n weergave er uit zou kunnen zien (figuur ontleend aan van Boxtel et al., 2011).



Overzicht van gemiddelde en spreiding van standaardscores per brugklatype (voorbeeld)

5.3 Beoordelingsnormen en de bijbehorende scores Centrale Eindtoets PO

Na afloop van de centrale eindtoets wordt voor iedere leerling een leerlingrapport opgemaakt met daarin de standaardscore en het advies voor het brugklatype. Wanneer een leerling minder dan de helft van het aantal opgaven per onderdeel gemaakt heeft, is er geen standaardscore mogelijk.

De standaardscore is berekend op basis van het totaal aantal goede antwoorden op de verplichte domeinen taal en rekenen. Door het werken met standaardscores is het CvTE in staat de resultaten van jaar tot jaar te vergelijken, ondanks (geringe) verschillen in moeilijkheid tussen de diverse jaargangen van de centrale eindtoets. Het domein wereldoriëntatie is facultatief in de centrale eindtoets en telt niet mee in de berekening van de standaardscore. De standaardscore is een getal op een schaal van 501 tot en met 550.

De twee verplichte domeinen hebben samen 220 opgaven. Het is de score op die 220 opgaven (die waarden kan aannemen van 0 tot en met 220) die de basis vormt voor de berekening van de standaardscore. Ieder goed beantwoorde opgave levert een score van 1 op.

De score op de 220 opgaven wordt aangeduid als 'ruwe score'.

De omzetting van ruwe score in standaardscore gebeurt in de volgende drie stappen:

1. Bereken de 'niet-afgeronde' standaardscore, verder symbolisch voorgesteld als $StSc^*$ met de volgende formule: $StSc^* = A \times \text{ruwe score} + B$, waarin A en B getallen zijn die door de equivaleringsprocedure worden bepaald. Deze getallen worden na afloop van de toetsafname vastgesteld. Voor de Centrale Eindtoets 2015 zal de omzetting van ruwe score in standaardscore in de periode van 24 april tot 1 mei 2015 plaatsvinden.
2. De niet-afgeronde standaardcores worden afgerond tot het dichtstbij gelegen gehele getal. Dit is de standaardscore, met uitzondering van de volgende twee gevallen.
3. Indien het afgerond getal uit de vorige stap:
 - a. groter is dan 550, wordt de standaardscore 550;
 - b. kleiner is dan 501, wordt de standaardscore 501.

5.4 Rapportage op referentieniveaus

Een tweede soort normering, die in 2015 voor het eerst bij de Centrale Eindtoets zal worden toegepast, heeft te maken met de rapportage op referentieniveaus. Voor het onderdeel rekenen (alle domeinen) en de taaldomeinen lezen en taalverzorging wordt voor elke leerling een referentieniveau berekend. Leerlingen kunnen in 2015 op de Centrale Eindtoets drie mogelijke 'scores' behalen: <1F (onder 1F), 1F, en 2F. 'F' staat voor fundamenteel niveau. Vanaf 2016 zal voor rekenen ook op 1S worden gerapporteerd ('S' staat voor streefniveau, en heeft bij rekenen vooral betrekking op de beheersing van meer wiskundige vaardigheden). Omdat nog niet eerder op referentieniveaus is gerapporteerd, blijft de beschrijving van de procedure vooralsnog noodgedwongen vrij algemeen. De precieze werkwijze moet immers nog gevonden worden.

De toetsing van referentieniveaus vindt plaats in verschillende onderwijstypen, met verschillende toetsen (en examens). De niveaus die leerlingen (al dan niet) behalen, moeten op elke soort toets en in elke school hetzelfde betekenen. Om dit te bereiken, wordt voor de normering gebruik gemaakt van referentiesets. Dit zijn verzamelingen opgaven die zijn samengesteld met het doel om onderlinge vergelijkingen tussen schooltypen en toetsen mogelijk te maken. Op elke referentieset is een prestatiestandaard vastgesteld door een panel van vakdeskundigen. Dit panel heeft alle opgaven bestudeerd en een uitspraak gedaan over het aantal opgaven dat een leerling ten minste goed moet maken om een bepaald referentieniveau te halen. Bijvoorbeeld: als een leerling 56 van de 80 opgaven uit de referentieset goed maakt, heeft hij of zij niveau 1F behaald.

De referentieset zelf is geen toets. De set wordt gebruikt om een schatting te maken van het aantal opgaven dat een leerling goed moet maken in een bepaalde toets, om een bepaald referentieniveau te halen. De set dient dus als ijkmateriaal. Om die functie te kunnen vervullen, zijn alle opgaven uit de referentieset afgenomen bij leerlingen in elk onderwijstype waarvoor dat referentieniveau relevant is. De referentiesets rekenen en lezen zijn afgenomen in het basisonderwijs, in het voortgezet onderwijs en in het middelbaar beroepsonderwijs. De referentieopgaven taalverzorging zijn alleen in het basisonderwijs afgenomen. Daarom wordt de set taalverzorging 'ankerset' genoemd, en geen referentieset. Mocht later blijken dat men ook in het voortgezet onderwijs en/of in het middelbaar beroepsonderwijs referentieniveaus op taalverzorging wil rapporteren, dan zijn alsnog afnames nodig in deze onderwijstypen. Momenteel is dat echter niet voorzien.

Om via een bepaalde toets, bijvoorbeeld de Centrale Eindtoets, te kunnen rapporteren op een referentieniveau, is het noodzakelijk dat een aantal opgaven uit de referentieset aan de leerlingen is voorgelegd samen met een aantal opgaven uit de toets. Dit is voor de Centrale Eindtoets gebeurd in de proeftoetsingen van 2014 (rekenen en lezen) en 2015 (taalverzorging). Na afloop van de gezamenlijke afname van toetsopgaven en referentiesetopgaven is een statistische vergelijking gemaakt tussen de prestaties op beide soorten opgaven. Als resultaat van deze stap kunnen de prestatiestandaarden van de referentiesets worden overgebracht naar de uiteindelijke toets, hetgeen concreet betekent dat de 56 goed op de referentieset (uit het eerdere voorbeeld) wordt vertaald naar bijvoorbeeld 34 goed op het toetsonderdeel. Deze vertaalslag wordt voor de Centrale Eindtoets in april 2015 voor het eerst gemaakt.

6 Betrouwbaarheid en meetnauwkeurigheid

6.1 Betrouwbaarheid

Voor het schatten van de betrouwbaarheid van de Centrale Eindtoets en de onderdelen daarvan wordt gebruik gemaakt van methoden uit zowel de klassieke testtheorie als de itemresponstheorie. Betrouwbaarheden worden gerapporteerd met betrekking tot de toetsonderdelen die ook daadwerkelijk voorkomen op het leerlingrapport. Dat zijn – naast de totaalscore – de scores op de onderdelen Taal en Rekenen. Ook voor het facultatieve onderdeel Wereldoriëntatie worden gegevens verstrekt.

Gegevens met betrekking tot de interne consistentie worden gerapporteerd in termen van *Cronbachs alfa*. Deze coëfficiënt vormt echter een onderschatting van de betrouwbaarheid. Waar de software dit toelaat zal daarom ook de *Greater lower Bound (GLB)* of *lambda* worden berekend. Als kwaliteitscriteria gelden waarden van minimaal .95 voor de totaalscore en .90 voor de genoemde onderdelen.

Naast *alfa*, *GLB* en/of *lambda*, wordt ook Macc berekend, een betrouwbaarheidsmaat die is ontleend aan de IRT (voor een toelichting op de berekening van deze coëfficiënt, zie van Boxtel et al., 2011). Macc kan op dezelfde wijze worden geïnterpreteerd, dus gelden ook hier criteria van minimaal .95 voor de totaalscore en minimaal .90 voor de scores op de onderdelen.

Ten slotte is ook de test-hertestbetrouwbaarheid van de toets van belang. Test-hertest onderzoek is echter lastig te organiseren. De afnamecontext en het karakter van de Centrale Eindtoets lenen zich hier niet goed voor. Het feit dat alle items echter OPLM-gekalibreerd zijn, maakt het mogelijk een hertest te simuleren. Daarbij wordt een dubbele afname gesimuleerd voor een groep van 150.000 leerlingen. Enerzijds wordt de vaardigheidsverdeling van alle leerlingen in de betreffende versie van de Centrale Eindtoets als uitgangspunt genomen, anderzijds alle itemparameters van deze toets. Steeds wordt een bepaalde vaardigheidsscore aselekt uit de verdeling genomen en worden twee bij deze vaardigheidsscore horende toetsafnames gesimuleerd. Uiteindelijk wordt de correlatie tussen deze 150.000 dubbele (virtuele) afnames berekend. Men kan deze simulatie beschouwen als een test-hertestonderzoek onder ideale condities. De tweede toetsafname is immers volledig onafhankelijk van de eerste en wordt niet beïnvloed door de kennis die de leerling mogelijk verworven heeft via de eerste toetsafname. Daarnaast is er geen sprake van invloed van een test-hertest-interval: beide afnames worden gesimuleerd alsof zij op hetzelfde moment plaats zouden vinden. De uitkomsten van deze procedure dienen vrijwel exact overeen te komen met de eerder aangeduide coëfficiënten.

Voor alle toetsscores worden 95%-betrouwbaarheidsintervallen berekend.

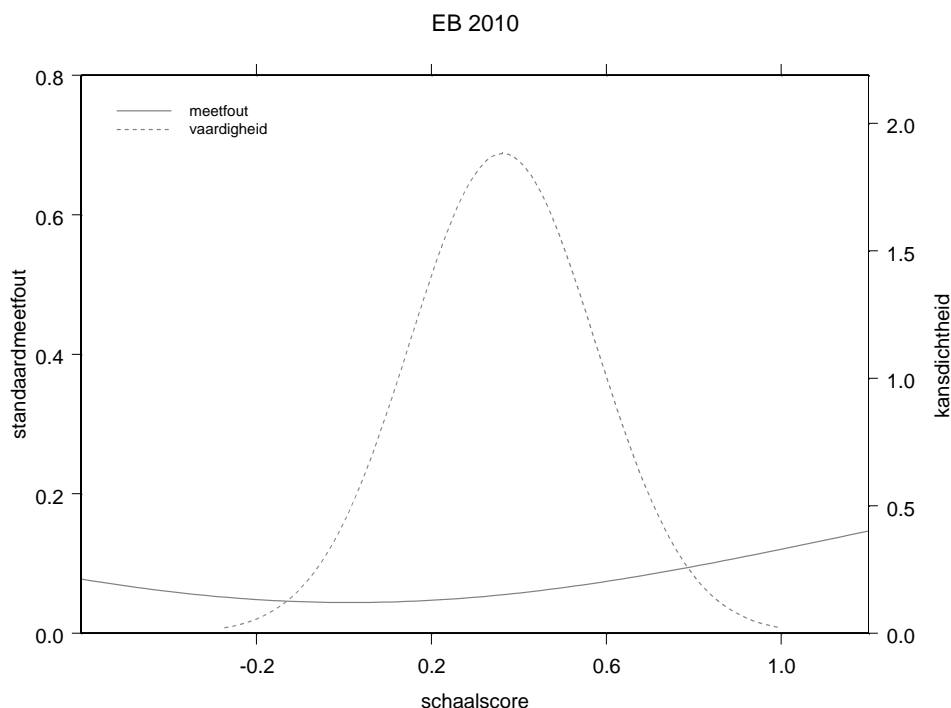
6.2 Nauwkeurigheid

De hier vermelde betrouwbaarheidscoëfficiënten en betrouwbaarheidsintervallen hebben uitsluitend betrekking op de globale (overall) meetnauwkeurigheid van de toetsen. Omdat de toetsen zijn geconstrueerd op basis van IRT is het mogelijk om ook gegevens te verstrekken over de *lokale* meetnauwkeurigheid van de verschillende onderdelen en de totaalscore. De meetfout van een toetsscore is immers niet bij elke toetsscore even groot. Deze hangt onder meer af van de moeilijkheidsgraad van de toets (omstreeks .70, zie eerder) en het aantal opgaven in de toets.

Ter evaluatie van de lokale meetfout zullen figuren worden geconstrueerd waarin naast kansdichtheidsfuncties (zeg maar de vaardigheidsverdeling) de grootte van de meetfout is afgebeeld.

Een voorbeeld van zo'n grafische weergave van de (lokale) meetfout is hier afgebeeld (ontleend aan van Boxtel et al., 2011).

Uit de figuren dient te blijken dat de meetfout kleiner is in de lagere en gemiddelde vaardigheidsregionen dan in de hogere vaardigheidsregionen. Dit zou overeenkomen met de bedoelingen van de toetsconstructeurs. Omdat in de lagere en gemiddelde vaardigheidsregionen tussen meer adviescategorieën moet worden gedifferentieerd dan in de hogere vaardigheidsregionen, is gekozen voor relatief gemakkelijke toetsen met een linksscheve verdeling. Bij dit type toetsen is ook de 'toetservaring' voor de leerling het prettigste. Het ligt voor de hand dat om die reden het differentiërend vermogen en de nauwkeurigheid in de lagere en gemiddelde regionen optimaal is.



Grootte van de meetfout voor de totaalscore (inclusief de kansdichtheidsfunctie voor deze score) (voorbeeld ontleend aan de Eindtoets Basisonderwijs 2010)

Het advies met betrekking tot het brugklatype dat voor een leerling als het meest geschikt wordt aangemerkt wordt gebaseerd op de totaalscore (standaardscore). Daarom is het voor deze score interessant om na te gaan wat het effect van de hierboven beschreven betrouwbaarheid en nauwkeurigheid is op de verstrekte adviezen. Daarom wordt een tabel opgesteld waarin wordt weergegeven welke percentages leerlingen op basis van hun werkelijke vaardigheidsscore een advies hebben gekregen dat overeenkomt met het advies op basis van de geschatte vaardigheidsscore. Uit zo'n tabel is af te lezen wat het effect van de (lokale) meetfout is op de nauwkeurigheid van de uitgebrachte adviezen. Gestreefd wordt naar percentages overeenstemming tussen 60% en 85% (ontleend aan van Boxtel et al., 2011). Voor zo goed als alle leerlingen moet daarnaast gelden dat het advies er op basis van de onnauwkeurigheid van de toetsscore nooit meer dan één adviescategorie naast mag zitten (i.e. de cellen op de diagonaal en de daaraan direct grenzende cellen).

7 Validiteit

Ten aanzien van de begripsvaliditeit onderscheidt het COTAN Beoordelingssysteem een aantal typen informatie die een bijdrage kunnen leveren aan de onderbouwing ervan. Daarbij wordt aan de soortgenootvaliditeit een belangrijke en aan de inhoudsvaliditeit een betrekkelijk bescheiden rol toegekend. In het RCEC Beoordelingssysteem staat juist de inhoudsvaliditeit centraal onder de invalshoek van de representativiteit van de toetsinhoud en wordt aan andere informatie nauwelijks aandacht besteed. Voor de Centrale Eindtoets is primair gekozen voor inhoudsvaliditeit, maar worden daarnaast ook andere bronnen van validiteit gebruikt en gerapporteerd.

7.1 Inhoudsvaliditeit

De inhoudsvaliditeit van een toets heeft betrekking op de vraag in hoeverre de opgaven in een toets een welomschreven en afgebakend geheel representeren van mogelijk in de toets op te nemen opgaven.

De opgaven in de Centrale Eindtoets sluiten nauw aan bij het doel en de inhoud van het onderwijs op de basisschool. Dat moet ook, want zij dienen een goede afspiegeling te vormen van wat leerlingen in het basisonderwijs in acht jaar tijd geleerd hebben. Daarbij moeten het opgaven zijn waarin leerstof behandeld wordt die elke basisschoolleerling zou kunnen beheersen, uitgaande van de communale onderwijsdoelen en –inhouden. Voor een optimaal voorspellend vermogen is dit noodzakelijk.

Hieronder vatten we nog eens samen hoe het CvTE ervoor zorgt dat de toetsinhoud representatief is voor de communale onderwijsdoelen en –inhouden.

- De onderdelen van de Centrale Eindtoets zijn geheel in lijn met de Toetswijzer, die gebaseerd is op de kerndoelen die voor het basisonderwijs geformuleerd zijn.
- De toets heeft een vaste structuur met taken Taal en Rekenen die door alle deelnemende leerlingen worden gemaakt, en een facultatief onderdeel Wereldoriëntatie. Binnen deze hoofdgebieden zijn weer rubrieken onderscheidbaar. Bij Taal wordt onderscheid gemaakt tussen Lezen (begrijpend lezen, opzoeken, samenvatten), Taalverzorging (spelling werkwoorden, spelling niet-werkwoorden, interpunctie) en Grammaticale begrippen, Woordenschat en Schrijven. De onderdelen van Rekenen (Getallen en bewerkingen; Verhoudingen, breuken en procenten; Meten en meetkunde, Verbanden) correleren onderling zeer hoog en worden dan ook als één homogene schaal opgevat.
- Bij Wereldoriëntatie worden drie onderdelen onderscheiden: Aardrijkskunde, Geschiedenis en Natuur & techniek.
- Elk jaar verstrekt de Vaststellingscommissie van het CvTE een constructieopdracht aan Stichting Cito voor een nieuwe jaargang van de Centrale Eindtoets. De constructieopdracht is een directe afgeleide van de Toetswijzer en bevat minutieuze voorschriften over de toetsinhoud.
- De verdeling van de aantallen opgaven over de onderdelen vormt een afspiegeling van de onderwijsbesteding in het reguliere basisonderwijs met betrekking tot de onderscheiden vakgebieden. Daarnaast laten de intercorrelaties tussen de toetsonderdelen een patroon zien dat van jaar tot jaar praktisch constant is en wordt gehouden. Adequate equivaleringsprocedures vormen de basis van deze constante toetsopbouw.
- De leden van de Vaststellingscommissie en de toetsdeskundigen van Cito analyseren regelmatig de lesmethoden die frequent in het onderwijs gebruikt

worden en controleren aan de hand van de toetsresultaten of de opgaven nog voldoende aansluiten bij de inhoud van het onderwijs.

- Een onderdeel van de jaarlijkse proeftoets is het verzoek om commentaar van leerkrachten op gehanteerde onderwijsdoelstellingen en opgaven. Ook anderszins is de Centrale Eindtoets een 'glazen huis': iedereen kan (na afname) kennis nemen van de toetsinhoud en deze desgewenst becommentariëren.
- De constructie van opgaven vindt plaats in commissies waarvan de leden ervaring hebben in het basisonderwijs.
- Opgaven worden pas in de toets opgenomen nadat de Vaststellingscommissie heeft geconstateerd dat ze voldoen aan de constructieopdracht.

Op basis van de beschreven procedures en bevindingen kan geconcludeerd worden dat de inhoudsvaliditeit van de Centrale Eindtoets uitstekend en vrijwel onomstreden is.

7.2 **Begripsvaliditeit**

Anders dan bij veel andere toetsen en examens kan de Centrale Eindtoets ook worden geëvalueerd vanuit het perspectief van de begripsvaliditeit. De belangrijkste reden hiervoor is dat de toets in overeenstemming met een IRT-model wordt geconstrueerd. Daarbij wordt aannemelijk gemaakt dat alle opgaven van een (deel)toets dezelfde onderliggende latente vaardigheid (i.e. een construct of begrip) meten. Inhouds- en constructvaliditeit gaan bij de Centrale Eindtoets dus hand in hand.

7.2.1 *Passing van het meetmodel; psychometrische kwaliteit van de opgaven*

Om ervoor te zorgen dat alle opgaven van een deelttoets dezelfde onderliggende vaardigheid meten wordt per deelttoets gekalibreerd waarbij diverse statistische toetsen worden ingezet in een iteratief proces. Daarnaast speelt visuele inspectie van de itemkarakteristieke curven een belangrijke rol in dit proces. Sommige toetsen (S-toetsen, R1c-toetsen) zijn ook geschikt om de uiteindelijke kwaliteit van de kalibratie te documenteren, evenals de visuele weergave van de S-toetsen (zie hiervoor paragraaf 5.2 over kalibratie). Deze verantwoording geschiedt wat globaler (op het niveau van de hoofdonderdelen van de toets) dan in de kalibratieprocedure zelf (per deelttoets). Dit impliceert dat de kwaliteit van de kalibratie in werkelijkheid nog hoger is dan in de verantwoording wordt gedocumenteerd. De kalibratie zorgt er uiteindelijk voor dat alle deelttoetsen kunnen worden opgevat als operationalisaties van unidimensionale latente vaardigheden, dit als noodzakelijke voorwaarde voor de validiteit. Dat de toetsen ook werkelijk meten wat zij beogen te meten, daarvoor zorgen de procedures die onder inhoudsvaliditeit zijn samengevat.

De genoemde procedures zorgen er tevens voor dat de kenmerken van de items in termen van moeilijkheidsgraden en itemresttotaalcorrelaties (in klassiek testtheoretische termen) optimaal zijn. De kwaliteitscriteria daarvoor zijn opgesomd in paragraaf 3.3.

7.2.2 *Structuur*

De ervaring heeft geleerd dat de onderscheiden toetsonderdelen op een karakteristieke wijze onderling samenhangen. Dit is zichtbaar te maken aan de hand van zowel de latente als de geobserveerde intercorrelaties. Kenmerkend voor leerjaar 8 is dat de onderdelen van Rekenen onderling zeer hoge correlaties laten zien. Rekenen is daar dus een sterk homogeen construct. Dat geldt in veel mindere mate voor de onderdelen van Taal, wat voornamelijk het gevolg is van het feit dat zowel niet-semantische onderdelen (zoals Spellingvaardigheid) als semantische onderdelen (bijvoorbeeld Begrijpend lezen en Woordenschat) kunnen worden onderscheiden. De semantische

onderdelen van Taal hangen onderling vrij sterk samen, maar veel minder sterk met niet-semantiche onderdelen zoals spellingvaardigheid.

Het optreden van dit karakteristieke, goed verklaarbare patroon van intercorrelaties kan men opvatten als een aanwijzing voor de begripsvaliditeit van de toets.

7.2.3 *Equivalentie met eerdere toetsversies*

Eerder zijn de procedures beschreven waarmee geborgd wordt dat de Centrale Eindtoets jaar in jaar uit dezelfde inhoud, moeilijkheidsgraad en structuur kent, met andere woorden equivalent zijn. Opeenvolgende Centrale Eindtoetsen voor verschillende leerjaren zijn daarmee te beschouwen als equivalente, parallelle versies van elkaar. Een hoge mate van equivalentie zoals deze onder meer in af te leiden uit de interne structuur duidt op begripsvaliditeit. Equivalentie is op deze wijze immers te beschouwen als een bepaalde vorm van soortgenootvaliditeit (in termen van gelijkheid tussen qua concrete inhoud volledig verschillende toetsen die operationalisaties vormen van hetzelfde construct). Het kan constructie-technisch gesproken opgevat worden als een prestatie van formaat dat een dergelijke equivalentie tot stand wordt gebracht. In het toetsconstructieproces kan in principe immers van alles mis gaan dat afbreuk doet aan die equivalentie. Als er bijvoorbeeld van jaar tot jaar steeds aanmerkelijke verschillen zouden worden vastgesteld in de intercorrelatiematrix of de gemiddelde moeilijkheidsgraad, zou dit sterk afbreuk hebben gedaan aan de validiteit van de Centrale Eindtoets. Het feit dat dit niet gebeurt wijst erop dat de toegepaste constructie- en equivalentieprocedures in orde zijn en dat er geen sprake is van toevallige of systematische fouten in die procedures die de validiteit zouden kunnen bedreigen.

7.2.4 *Soortgenootvaliditeit; convergente en discriminante validiteit*

Het is erg lastig om onderzoek te doen naar de soortgenootvaliditeit en convergente versus divergente validiteit van toetsen die een belangrijke formele functie hebben in het maatschappelijk verkeer. Over het algemeen wordt de Centrale Eindtoets opgevat als een toets waar veel van afhangt (en die dus 'high stake' is). Het is niet eenvoudig om een soortgenootonderzoek uit te voeren met hetzelfde 'high stake'-karakter en al helemaal als het gaat om een toets zoals de Centrale Eindtoets die meerdere dagdelen in beslag neemt. Gegeven de gehanteerde inhoudelijke voorwaarden (waar overigens ook alternatieve eindtoetsen aan gehouden zijn) lijkt empirisch soortgenootonderzoek bovendien overbodig. Eventueel kunnen kleine steekproeven worden gevormd om de uitkomsten van de Centrale Eindtoets te koppelen aan een tweede score op een andere eindtoets die als het ware bij wijze van 'second opinion' wordt afgenomen, maar dit zal niet gemakkelijk (maar mogelijk in samenwerking met andere toetsaanbieders) te realiseren zijn.

In het verleden werden scores van de Cito Eindtoets Basisonderwijs met het oog op de convergente en divergente validiteit wel gekoppeld aan de resultaten van eerder afgenomen toetsen, zoals de Entreetoets groep 7 en de toetsen van het Cito Leerlingvolgsysteem primair en speciaal onderwijs. Gegevens van de Centrale Eindtoets mogen echter niet zonder meer aan andere gegevens (van Cito, dan wel andere toetsaanbieders) worden gekoppeld. Wanneer dit wenselijk wordt geacht (bijvoorbeeld door het CvTE of door de Expertgroep Toetsen PO) zal hier alsnog de mogelijkheid voor gecreëerd moeten worden (bijvoorbeeld door Cito dataretour te benutten of scholen aan te schrijven om informatie op te vragen).

7.2.5 *Verschillen tussen relevante subgroepen; samenhang met achtergrondkenmerken*

Wanneer sprake is van daadwerkelijke samenhangen tussen leerprestaties en andere variabelen, of wanneer specifieke subgroepen (bijvoorbeeld naar sekse of etnische

herkomst) daadwerkelijk verschillen laten zien in leerprestaties, daar waar men dit op grond van theoretische overwegingen en/of eerder onderzoek verwacht, dan vormen deze samenhangen en verschillen impliciet bewijsmateriaal voor de validiteit van de betreffende instrumenten. De volgende kenmerken kunnen worden meegenomen:

- Verschillen naar *sekse*: op sommige onderdelen van de Centrale Eindtoets scoren meisjes traditioneel wat hoger, op andere jongens. Deze verschillen zijn deels op basis van de literatuur te verwachten.
- Verschillen naar *leeftijd*: vertraagde leerlingen scoren als groep lager dan reguliere leerlingen, voorlijke leerlingen juist hoger.
- Verschillen naar *urbanisatiegraad*: leerlingen uit de vier grote steden (G4) scoren *als groep* in de regel wat lager. Daarvoor zijn goede verklaringen aan te voeren.
- Verschillen naar *leerlinggewicht* (op schoolniveau, in termen van % *achterstandsleerlingen*): scholen met een hoger percentage achterstandsleerlingen scoren lager dan scholen met een lager percentage achterstandsleerlingen.
- Leerlingen die worden *aangemeld voor de niveautoets* scoren duidelijk lager dan leerlingen die daarvoor niet worden aangemeld.

Andere gegevens over samenhangen en verschillen die mogelijk interessant zijn (en vanuit het oogpunt van *fairness* misschien ook wel onontbeerlijk) betreffen die ten aanzien van *etnische herkomst* en *thuis taal*. Op het moment worden hierover geen gegevens verzameld op het antwoordblad omdat de Wbp (Wet bescherming persoonsgegevens) dit niet toelaat. Als onderzoek hiernaar noodzakelijk wordt geacht, zal dit alsnog moeten worden uitgevoerd, bijvoorbeeld door aanvullende gegevens hierover gericht op te vragen bij scholen.

7.2.6 *DIF-analyses en invariantie van structuur*

Toetsopgaven mogen leerlingen uit bepaalde subgroepen niet onbedoeld en oneigenlijk bevoordelen of benadelen. Men noemt dit Differential Item Functioning (DIF).

Onderzoek hiernaar kan gecombineerd worden met het kalibratie-onderzoek, waarbij de passing van het item voor verschillende subgroepen afzonderlijk wordt onderzocht. Ook meer klassieke methoden zoals die van Mantel-Hänszel (MH) kunnen in dit kader worden benut. Hierbij wordt gewerkt met steekproeven uit de kalibratiebestanden met een optimale statistische power (N omstreeks 500).

Per deoltoets mogen niet meer dan enkele opgaven DIF laten zien.

Op dit moment is het mogelijk om DIF-analyses uit te voeren naar sekse en gebeurt dit ook. DIF-analyses naar thuis taal of etnische herkomst, zoals deze bij de Cito Eindtoets Basisonderwijs gebruikelijk waren, kunnen niet worden uitgevoerd, omdat er geen systematische gegevens hierover bij de toetsafname kunnen worden verzameld. Mogelijk kan dit in de toekomst gebeuren door deze gegevens bij de proeftoetsingen op te vragen. Hetzelfde geldt voor de variabele leerlinggewicht (op het niveau van individuele leerlingen).

Het karakteristieke patroon van intercorrelaties waarvan eerder sprake was, dient zich voor verschillende subgroepen op dezelfde manier voor te doen. Voor de variabele sekse wordt dit onderzocht. Voor de variabelen etnische herkomst, thuis taal en leerlinggewicht (op individueel leerlingniveau) kan dit alleen als hierover gegevens mogen worden vastgelegd.

7.3 **Criteriumvaliditeit**

De Centrale Eindtoets is mede bedoeld om behulpzaam te zijn ten aanzien van het advies over de meest geschikte vorm van vervolgonderwijs na het reguliere

basisonderwijs. Dit impliceert dat aan de toets voorspellende waarde moet kunnen worden toegekend en dus dat criteriumvaliditeit nadrukkelijk aan de orde is.

7.3.1 *Relatie met het doorstroomadvies van de leerkracht*

Op het antwoordblad van de Centrale Eindtoets kan door de school (leerkracht groep 8) worden aangegeven wat diens mening is over de meest geschikte vorm van vervolgonderwijs. Het doorstroomadvies van de leerkracht vormt een vorm van operationalisatie van het niveau van vervolgonderwijs dat de leerling aan kan als criterium, maar valt daarmee niet samen. De leerkracht baseert zich op diens kennis van de leerling, zijn leerprestaties gedurende de periode dat de leerling basisonderwijs volgde en de wijze waarop de leerling zijn cognitieve mogelijkheden benutte in het verleden (inclusief de prestaties van de leerling op toetsen van een leerlingvolgsysteem). Daarmee kan dit advies ook worden opgevat als operationalisatie van diens leerprestaties in het verleden. Hoe dan ook, gericht op het heden en verleden (begripsvaliditeit), dan wel op de toekomst (criteriumvaliditeit), de samenhang met de totaalscore op de Centrale Eindtoets dient sterk te zijn (.80 of hoger bij toepassing van een CatReg analyse). De samenhang is ook weer te geven in de vorm van een kruistabel advies leerkracht versus advies op basis van Centrale Eindtoets en de mate van overeenstemming, zoals hieronder bij wijze van illustratie is weergegeven voor de Eindtoets Basisonderwijs 2010 (ontleend aan van Boxtel et al, 2011):

Overeenkomst van doorstroomadvies leerkracht en Cito-advies op basis van de Eindtoets Basisonderwijs 2010

	?	bb	bb/kb	kb	gt	gt/h	gt/h/v	havo	h/v	vwo	totaal
501-523 (bb)	4195	3552	2295	3071	1706	234	21	35	7	4	15120
524-529 (kb)	5237	800	1372	3429	5281	1342	81	281	52	4	17879
530-536 (gt)	9532	205	597	2262	9489	6153	460	3075	928	122	32823
537-544 (havo)	13144	23	62	305	3344	6098	828	9671	8529	3197	45301
545-550 (vwo)	7486	1	3	6	100	320	135	1711	5384	11032	26178
Totaal	39594	4581	4329	9073	20020	14147	1525	14773	14900	14359	137301

De cijfers dienen omstreeks 70% overeenstemming tussen beide adviezen te laten zien. Daarnaast dient in ten minste 95% van de gevallen het advies op basis van de standscore op de Centrale Eindtoets niet meer dan één adviescategorie af te liggen van het leerkrachtadvies.

7.3.2 *De score op de Centrale Eindtoets en de feitelijke plaatsing en doorstroom*

Ten aanzien van de voorspellende waarde van de Centrale Eindtoets geldt uiteindelijk dat 'the proof is in the pudding'. In dit opzicht is informatie over de feitelijke plaatsing in het voortgezet onderwijs en de eerste jaren die daarop volgen (doorstroom) van belang. Lange termijn studiesucces, zoals behalen van diploma's, aantal jaren waarin diploma wordt behaald, aantal malen doubleren is in vele opzichten interessant en relevant, maar geldt nadrukkelijk niet als criterium. De meetpretentie van de Centrale Eindtoets is daarvoor te bescheiden; deze reikt niet verder dan de onderbouw (plaatsing in het eerste jaar) en wil slechts ondersteuning bieden bij het tot stand komen van het advies. Bij het huidige afnamemoment is het feitelijke schooladvies bovendien al tot stand gekomen en is het behaalde resultaat op de Centrale Eindtoets een *second opinion* waarbij een hoger schooltype mogelijk is, niet een lager schooltype.

Ten behoeve van de onderbouwing van de criteriumvaliditeit wordt in samenwerking met het CBS nagegaan in welk school- en brugklatype de leerlingen feitelijk geplaatst worden. Bovendien wordt geregistreerd waar de leerlingen bij uitstroom naar het tweede en derde leerjaar terechtkomen. Tevens worden doublures geregistreerd. Net zoals ten aanzien van het doorstroomadvies is de sterkte van de relatie tussen de score op de Centrale Eindtoets en de plaatsing in het vervolgonderwijs van belang (te analyseren met behulp van van CatReg, vanwege het ordinale karakter van de data). Op basis van resultaten in het verleden mag verwacht worden dat deze samenhang nog wat sterker is dan de relatie met het doorstroomadvies, namelijk omstreeks .85. Van de totale groep leerlingen in het doorstroomonderzoek bij de meest recente verantwoording van de Cito Eindtoets Basisonderwijs (van Boxtel et al., 2011) had 80% een plaatsing gekregen die in overeenstemming was met het geadviseerde brugklatype. Bij de analyse van de doorstroom naar het tweede en derde leerjaar worden discrepanties tussen de score op de Centrale Eindtoets enerzijds en het leerkrachtadvies, respectievelijk de feitelijke plaatsing anderzijds betrokken. Verwacht wordt dat afstroom (naar een lager schooltype) vaker voorkomt bij een hogere plaatsing dan die welke de Centrale Eindtoets adviseert, en opstroom (naar een hoger schooltype) vaker bij een lagere plaatsing dan die welke de Centrale Eindtoets adviseert. Over enkele jaren zal blijken of deze verwachting terecht is. Desgewenst kan langere termijn informatie worden verzameld in samenwerking met CvTE.

8 Toezicht

8.1 Beoordelingskader andere eindtoetsen

De Expertgroep toetsen PO adviseert over het toelaten van eindtoetsen naast de Centrale Eindtoets. Dit gebeurt op basis van het *Beoordelingskader andere eindtoetsen*. Dit beoordelingskader beschrijft de aspecten waarop andere eindtoetsen worden beoordeeld, de wijze waarop toetsen worden beoordeeld en de wijze waarop de Expertgroep toetsen PO tot een advies komt. De inhoud van het beoordelingskader is gebaseerd op de kwaliteitseisen in het Toetsbesluit PO (2014) en het document Toetswijzer Eindtoets PO algemeen deel (CvTE, 2014) en bestaat uit twee inhoudelijke onderdelen: psychometrie en onderwijskunde & organisatie.

Op basis van een set beslisregels stelt de Expertgroep toetsen PO een definitief advies op. Dit advies gaat naar de Minister van Onderwijs, Cultuur en Wetenschappen die, op basis van het ontvangen advies, een definitieve beslissing neemt over het al dan niet toelaten van de betreffende eindtoets.

Als een eindtoets is toegelaten voor het primair onderwijs, volgt een jaarlijkse korte beoordelingsprocedure om de kwaliteit te beoordelen. Jaarlijks moeten toetsopgaven ververs worden (in de eerste jaren minimaal 50%). Daarom wordt ook jaarlijks opnieuw de validiteit, betrouwbaarheid en normering beoordeeld. Ook hiervoor gebruikt de Expertgroep het *Beoordelingskader andere eindtoetsen*. De Expertgroep toetsen PO stelt, voorafgaand aan de feitelijke advisering betreffende de toelating van toetsen, het beoordelingskader vast. De Expertgroep kan gemotiveerd afwijken van het beoordelingskader wanneer daar gegronde redenen voor zijn. Bijvoorbeeld wanneer een toetsaanbieder niet redelijkerwijs in staat is gebleken te kunnen voldoen aan de gestelde eisen, of wanneer de toetsaanbieder een goede inhoudelijke onderbouwing kan geven waarom aan bepaalde eisen niet is voldaan.

De Expertgroep neemt de Centrale Eindtoets niet mee in haar beoordelingsprocedures. De Centrale Eindtoets hanteert een eigen kwaliteitsnormering die parallel loopt met het beoordelingskader.

8.2 Extern toezicht door COTAN

De Commissie Testaangelegenheden Nederland (COTAN) beoordeelt psychodiagnostische instrumenten die in Nederland worden uitgegeven en bij de COTAN ter beoordeling worden aangeboden. De COTAN documentatie is ontwikkeld voor professionals die zich bezighouden met psychodiagnostiek.

De COTAN documentatie wordt in opdracht van het Nederlands Instituut voor Psychologen (NIP) namens de COTAN uitgegeven door Boom test uitgevers, Amsterdam. De COTAN Documentatie is de online opvolger en combinatie van de tweedelige Documentatie van tests en testresearch in Nederland, de aanvullingen op deze uitgave en het COTAN Testboek voor het onderwijs.

Zoals de Centrale Eindtoets niet beoordeeld wordt door de Expertgroep, geldt dit ook voor de COTAN-beoordeling. Ook hier hanteert het CvTE de COTAN-normen in zijn eigen kwaliteitsprocedures.

8.3 Extern toezicht door de onderwijsinspectie

De rol van de onderwijsinspectie ten aanzien van de Centrale Eindtoets is tweeledig. Enerzijds houdt zij toezicht op de afname en deelname. Zij richt zich niet op de inhoud,

maar op een correcte uitvoering en heeft daartoe een toezichtlijn om zo nodig te kunnen optreden bij overtreding van de afname- en deelnamevoorschriften. Anderzijds maakt zij gebruik van de toetsresultaten in haar analyse, vergelijking en beoordeling van scholen. Zij heeft geen inhoudelijk mening of inbreng bij de totstandkoming van de Centrale Eindtoets.

Een zorg van de onderwijsinspectie geldt de vergelijkbaarheid van de gegevens die verschillende aanbieders van eindtoetsen aanleveren. Deze zorg heeft niet alleen betrekking op de typen informatie die de verschillende eindtoetsen opleveren, maar ook de equivalering van inhoud en moeilijkheidsgraad over tijd en de normering van de toetsen.

8.4 Extern toezicht door audits

Een audit is een externe waarborging op het gehele proces rond de afname en constructie van de Centrale Eindtoets PO. Gelet op de discussie over de beveiliging van de examenketen in het voortgezet en middelbaar onderwijs initieert het CvTE incidenteel een audit op de toetsketen voor het primair onderwijs. Deze wordt uitgevoerd door een per uitvoering te benoemen externe instantie. Een audit richt zich niet op de kwaliteit van de inhoudelijke samenstelling en resultaten van de Centrale Eindtoets. Hiervoor zijn vaststellingscommissies ingericht. Wel kan de audit zich richten op de kwaliteitsprocedures, waaronder de psychometrische, zoals deze in dit document zijn beschreven. Zijn deze procedures adequaat, zijn ze *up to date*, zijn er betere denkbaar?

Omdat er een verschil in omgevingen en processen zit tussen de papieren en digitale afname, wordt de audit opgesplitst naar een papieren afname met een aanvulling voor de digitale afname.

Papieren afname

Bij de papieren afname worden de in het schema genoemde processtappen onderkend. Deze worden alle afzonderlijk in de audit meegenomen.



De aanpak van een audit wordt volgens een vast aantal stappen uitgevoerd:

1. Uitwerken onderzoeksopdracht en maken afspraken over uitvoering en rapportage.
2. Desk-research op basis van beschikbare documenten die betrekking hebben op de staat van de beveiliging van de toetsketen, alsmede eerdere onderzoeken en audits rond Questify en het ontwerp van het netwerkredesign.
3. Interviews, waarin wordt onderzocht welke problemen er verbonden zijn aan de beveiliging van de toetsketen en de problemen die verbonden zijn aan het aantal gelijktijdige digitale toetsafnames.
4. Technisch onderzoek naar zowel het gedrag van de applicatie, als naar de eisen die gesteld zijn aan de technische infrastructuur waarop de applicatie draait.
5. Toetsing van de netwerkinfrastructuur.
6. Opstellen rapportage en review concept versie.
7. Definitieve rapportage, oplevering aan de opdrachtgever.

Digitale afname

Voor de audit van de digitale afname is de volgende onderzoeksvraag geformuleerd: zijn er voldoende waarborgen om een toenemend aantal gelijktijdige afnamen aan te kunnen. Indien dit niet het geval is, welke aanvullende maatregelen moeten dan alsnog worden genomen.

De digitale afname kent een specifieke additionele audit van de software omgeving die gebruikt wordt voor het samenstellen en uitvoeren van de Digitale Centrale Eindtoets Primair Onderwijs. Deze is eind 2012 beoordeeld door de Software Improvement Group (SIG) door middel van een Software Risk Assessment. SIG is een gecertificeerd keuringsinstituut voor software. Met een Software Risk Assessment beoordeelt het SIG de software hierbij op de volgende aandachtsgebieden:

- Analyseerbaarheid
- Aanpasbaarheid
- Testbaarheid
- Modulariteit
- Herbruikbaarheid

9 Referenties

College voor Examens (2012). *Toetswijzer bij de Centrale Eindtoets PO taal en rekenen. Inhoudsverantwoording van de Centrale Eindtoets taal en rekenen*. Utrecht: College voor Examens.

De Staatssecretaris van Onderwijs, Cultuur en Wetenschap, J. M. van Bijsterveldt-Vliegenthart, De Minister van Onderwijs, Cultuur en Wetenschap, A. Rouvoet, De Minister van Landbouw, Natuur en Voedselkwaliteit, G. Verburg. *Besluit referentieniveaus Nederlandse taal en rekenen*. Den Haag: Staatsblad 2010 265.

De Staatssecretaris van Onderwijs, Cultuur en Wetenschap, S. Dekker (2014) *Toetsbesluit PO*. Den Haag: Staatsblad 2014 000

De Staatssecretaris van Onderwijs, Cultuur en Wetenschap, S. Dekker (2014) *Wet Centrale Eindtoets en leerling- en onderwijsvolgsysteem primair onderwijs*. Den Haag:

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: Nederlands Instituut van Psychologen.

Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008). *Over de drempels met taal. De niveaus voor de taalvaardigheid. Onderdeel van de eindrapportage van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen*. Enschede: Expertgroep Doorlopende Leerlijnen Taal en Rekenen.

Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2009). *Referentiekader taal en rekenen. De referentieniveaus*. Enschede: SLO.

NEN (2009) *Kwaliteitsmanagementsystemen – Eisen (ISO 9001:2008, IDT)*.

OCW (2006). *Kerndoelen Primair Onderwijs*. Den Haag: OCW.

RCEC (2011). *RCEC Beoordelingssysteem voor de kwaliteit van toetsen en examens*. Enschede: RCEC.

van Boxtel, Engelen & de Wijs (2011) *Wetenschappelijke verantwoording van de Cito Eindtoets Basisonderwijs*. Arnhem: Cito.

Verhelst, N.D., Verstralen, H.H.F.M., & Eggen, T.J.H.M. (1991). *Finding starting values for the item parameters and suitable discrimination indices in the one-parameter logistic model*. Measurement and Research Department Reports 91-10, Arnhem: Cito Instituut voor Toetsontwikkeling.

10 Bijlage Inhoudelijke eisen en indicatoren Centrale Eindtoets

10.1 Inleiding

Hier worden de inhoudelijke kwaliteitseisen en indicatoren beschreven die de in de Centrale Eindtoets gehanteerd worden. Deze bestaan uit een algemeen deel dat geldt voor alle eindtoetsen en een specifiek deel dat alleen geldt voor de Centrale Eindtoets.

In beide zijn eisen opgenomen die aandacht hebben voor het proces van ontwerp, constructie, uitvoering en evaluatie en inhoudelijke eisen waar de toets aan dient te voldoen.

Het uitgangspunt hierbij is dat het CvTE zijn eigen eisen en beoordelingskader opstelt, zoals dit ook gebeurt bij de examens voor het voortgezet onderwijs (VO). Dit beoordelingskader is opgesteld aan de hand van elementen uit de eerder genoemde relevante kwaliteitssystemen. Het resultaat hiervan is een kwaliteitssysteem dat garandeert dat de Centrale Eindtoets voldoet aan alle toetsingscriteria zoals die voor de andere eindtoetsen gelden. Een centrale rol in dit kwaliteitssysteem wordt gespeeld door de Toetswijzer. In de *Toetswijzer Eindtoets PO algemeen deel* staan de minimale inhoudelijke kwaliteitseisen die voor alle eindtoetsen gelden. Daarnaast geeft deze ruimte voor inhoudelijke keuzes. De inhoudelijke keuzes als ook de verdere eisen staan beschreven in de *Toetswijzer Eindtoets PO specifiek deel*, welke gericht is op de Centrale Eindtoets.

Eisen eindtoets PO

- 1 Een eindtoets PO toetst leerlingen op inhouden van de Kerndoelen taal en rekenen.
- 2 Een eindtoets PO toetst leerlingen op inhouden van en meet hun beheersingsniveau ten aanzien van het Referentiekader taal, tenminste voor de domeinen Lezen en, uit het domein Begrippenlijst en Taalverzorging, het subdomein Taalverzorging.
- 3 Een eindtoets PO toetst leerlingen op de inhouden van en meet hun beheersingsniveau ten aanzien van het Referentiekader rekenen, voor de domeinen Getallen, Verhoudingen, Meten en meetkunde en Verbanden.
- 4 De inhoud van een eindtoets PO maakt niveaubepaling mogelijk ten aanzien van de referentieniveaus Lezen en, uit het domein Begrippenlijst en Taalverzorging, het subdomein Taalverzorging, en Rekenen.
- 5 Een eindtoets PO gaat vergezeld van een handleiding met een beschrijving van keuzes die zijn gemaakt ten aanzien van de kwaliteitseisen 2, 3 en 4.

10.1.1 *Kwaliteitseis 1, Samenhang Taal en Rekenen*

Als een eindtoets voldoet aan de eisen 2 en 3, wordt deze geacht ook aan eis 1 te voldoen.

10.1.2 *Kwaliteitseis 2, Onderdeel Taal*

Het Referentiekader onderscheidt voor taal vier domeinen: Mondelinge taalvaardigheid, Lezen, Schrijven, en Begrippenlijst en Taalverzorging. In de eindtoetsen PO onderdeel Nederlandse taal (vanaf nu: taal) is in principe ruimte om alle domeinen te meten, met daarbij de kanttekening dat in ieder geval het domein Lezen en het subdomein Taalverzorging uit het domein *Begrippenlijst en Taalverzorging* getoetst moeten worden conform de eisen genoemd in het Toetsbesluit PO (2014). Deze Algemene Toetswijzer beschrijft de minimaal inhoudelijke kwaliteitseisen voor deze wettelijk verplichte (onderdelen van) domeinen, maar beschrijft daarnaast – als richtlijn – ook de kwaliteitseisen voor de domeinen Mondelinge taalvaardigheid, Schrijven en het

subdomein Begrippenlijst van het domein *Begrippenlijst en Taalverzorging*, die ook in de eindtoets aangeboden mogen worden.

Het is niet wettelijk verplicht om de domeinen Mondelinge taalvaardigheid, Schrijven en Begrippenlijst uit het domein Begrippenlijst en Taalverzorging te toetsen. Toetsaanbieders zijn vrij om vanuit deze Algemene Toetswijzer PO aan de andere domeinen invulling te geven in hun eigen handleidingen en deze domeinen op te nemen in hun eindtoets. Door ook andere domeinen in de eindtoets op te nemen kan meer recht gedaan worden aan de onderwijsinhouden voor taal en aan het evalueren van de prestaties van leerlingen aan het einde van het basisonderwijs (zie ook de paragraaf hieronder). Uiteraard blijft het de verantwoordelijkheid van de school om ervoor te zorgen dat de onderdelen die niet door een eindtoets getoetst worden, door kinderen beheerst worden.

Inhoudsbeschrijving in de handleiding voor het onderdeel Taal

Voor scholen is het belangrijk zich te realiseren welke domeinen in een eindtoets getoetst worden. In de eindtoetsen gaat het niet om dé taalvaardigheid van de leerling maar om een beperkt deel van de taalvaardigheid. Immers, het is niet verplicht om alle domeinen te toetsen en binnen de domeinen kunnen keuzes gemaakt worden in wat wel of niet getoetst wordt. Het is derhalve nodig om op andere wijzen dan via de eindtoets een beeld te verwerven van de ontwikkeling van de taalvaardigheid van leerlingen. Dit kan bijvoorbeeld aan de hand van methode gebonden toetsen, observaties, portfolio's en besprekingen van de uitvoering van taaltaken met leerlingen.

In de handleiding bij de eindtoets moet de toetsaanbieder een heldere beschrijving geven van de gemaakte keuzes. Deze handleiding bevat bijvoorbeeld een beschrijving waarom en hoe inhouden uit het domein Mondelinge taalvaardigheid getoetst worden. Indien inhouden uit het domein Mondelinge taalvaardigheid getoetst worden, geeft de handleiding een beschrijving van de gemaakte keuzes ten aanzien van de toetsing van deze inhouden (direct of indirect, keuzes in taken). Hetzelfde geldt voor de gemaakte keuzes ten aanzien van de toetsing van inhouden uit het domein Schrijfvaardigheid (direct of indirect, keuzes in genres of taken). Zo kan inzichtelijk gemaakt worden welke aspecten van de taalvaardigheid de eindtoets in beeld brengt.

Rapportage van de niveaubepaling voor het onderdeel Taal

In het Toetsbesluit PO (artikel 3) is vastgelegd dat de eindtoetsen meten welk eindniveau de leerling heeft behaald ten opzichte van de referentieniveaus (zie kwaliteitseis 4). Elke eindtoets dient minimaal te voldoen aan de inhoudelijke kwaliteitseisen voor het domein Lezen en het subdomein Taalverzorging uit het domein Begrippenlijst en Taalverzorging. Voor het domein Lezen zijn referentiesets van opgaven beschikbaar en voor het subdomein Taalverzorging komen ankersets van opgaven. Met de daarbij behorende normering kan de (wettelijke) referentiecesuur worden bepaald. Een indicatie van het behaalde referentieniveau kan in het leerlingrapport dus alleen voor het domein Lezen en het subdomein Taalverzorging worden gegeven. Voor de andere (sub)domeinen dienen eindtoetsaanbieders zelf een cesuur te kunnen verantwoorden.

Minimaal inhoudelijke kwaliteitseisen

Heel eenvoudig gezegd is lezen het betekenis geven aan een tekst. Lezen is een actief proces dat de lezer op heel veel verschillende manieren kan sturen, en dat hij in allerlei situaties kan inzetten (zie bijvoorbeeld Gibson, E.J. & Levin, H., 1975). Natuurlijk is er meer dan één manier van lezen: we kunnen lezen om een (studie)tekst te onthouden of te begrijpen, om ons te informeren, om instructies op te volgen, et cetera. Maar we kunnen ook lezen om ons te vermaken, onze geest te slijpen of te genieten van de stijl, dan doelen wij meer op het lezen van fictie en literatuur. Maar in alle vormen van lezen construeren lezers een voorstelling van de tekst. In deze tekstrepresentatie leggen zij zowel relaties tussen tekstdelen, als ook relaties met hun algemene kennis. Dit kan zijn: relaties met algemene kennis, onderwerp specifieke kennis, kennis over het teksttype, tekststructuur, literair historische kennis, kennis van literatuurtheorie et cetera. Het belang van het leggen van dergelijke relaties kan moeilijk onderschat worden voor zowel het begrip van de tekst, de waardering van de tekst, als voor het leesproces. Immers, diverse leesstrategieën kapitaliseren erop dat de lezer zijn kennis – in de brede zin van het woord – en gevoelens inzet om de tekst te begrijpen. Deze cognitieve representatie is de basis voor begrip en interpretatie van de tekst en aanknopingspunt voor evaluatie, reflectie en verwerking.

Begrippenlijst en Taalverzorging

Onder taalverzorging verstaan we de activiteiten die ondernomen moeten worden om correct te schrijven. Hieronder wordt ten minste verstaan het correct spellen van woorden en het gebruiken van de juiste leestekens. In principe zou onder taalverzorging ook begrepen kunnen worden het maken van een goede vormgeving (alineëring en lay-out). Maar dit deel is opgenomen bij het domein schrijfvaardigheid, omdat het in de praktijk van het onderwijs deel uitmaakt van de schrijfdidactiek.

Schrijven

Het lezen en analyseren van bronnen, het genereren van ideeën, het structureren van ideeën, het stellen van retorische doelen, het formuleren van tekst, het lezen van tekst die al geschreven is, het evalueren en verbeteren van die tekst, en niet in het minst, het coördineren van die activiteiten in de tijd.

Schrijven vervult verschillende functies: het is een communicatieve handeling (informeren, activeren, instrueren, argumenteren, overtuigen etc.) maar ook een conceptualiserende handeling: schrijven kan het leerproces ondersteunen (aantekeningen maken, op ideeën komen, samenhang inzien etc.). In die functie speelt schrijven een grote rol in het leren van leerlingen (zie bijvoorbeeld Grabowski, 2007). Kenmerkend voor goed schrijfonderwijs is niet zozeer het aantal teksten dat leerlingen moeten schrijven (frequentie van opdrachten), maar meer de didactische context waarin die teksten geschreven worden (Hillocks, 1986; Graham & Perrin, 2007). Daarbij dient aandacht te zijn voor het ontdekken van verschillende aanpakken en het vormgeven aan de inhoud (Breetvelt, Van den Bergh & Rijlaarsdam, 1994).

Mondelinge Taalvaardigheid

Mondelinge taalvaardigheid wordt in het onderwijs onderscheiden in drie subdomeinen: Gesprekken voeren, Luisteren en Spreken. In het subdomein Gesprekken voeren staat het interactieve karakter centraal: situaties waarin de leerling met een of meer

gesprekspartners direct communiceert en afwisselend de rol van spreker en luisteraar vervult (dialogische en polylogische situaties). Daarnaast staan de subdomeinen Luisteren en Spreken, waarin de leerling meer in een eenrichtingssituatie geplaatst is, met geen of een beperkte mogelijkheid tot directe interactie. De leerling heeft hoofdzakelijk één rol, die van luisteraar, of die van spreker. In werkelijkheid zijn de grenzen niet zo hard: na een mondelinge (monologische) presentatie, krijgen luisteraars immers de gelegenheid om vragen te stellen die de spreker moet kunnen begrijpen (in de rol van luisteraar) alvorens ze te beantwoorden (in de rol van spreker).

Begrippenlijst en Taalverzorging

Om te spreken over taal en taalverschijnselen is een beperkt aantal begrippen noodzakelijk. De meeste daarvan zijn aan het einde van het basisonderwijs wel aan de orde geweest (1F). Kennis van deze begrippen bevordert het gesprek binnen en buiten het taalonderwijs over taal en taalverschijnselen: het gaat erom dat docenten (en leerlingen) bepaalde verschijnselen kunnen benoemen in contextrijke taalsituaties. Dat wil zeggen dat docenten deze termen moeten kunnen gebruiken in hun onderwijs in de vaardigheidsdomeinen

Specifieke eisen Taal

Lezen, subdomein zakelijke teksten

Taken

- De leerlingen lezen en maken opdrachten bij informatieve teksten.
- De leerlingen lezen en maken opdrachten bij instructieve teksten.
- De leerlingen lezen en maken opdrachten bij betogende teksten.

Tekstkenmerken

De teksten die de leerlingen lezen voldoen aan de tekstkenmerken omschreven in het Referentiekader.

Kenmerken van de taakuitvoering

- In de toets wordt als kenmerk van de taakuitvoering techniek en woordenschat getoetst
- In de toets wordt als kenmerk van de taakuitvoering begrijpen getoetst.
- In de toets wordt als kenmerk van de taakuitvoering interpreteren getoetst.
- In de toets wordt als kenmerk van de taakuitvoering samenvatten getoetst.
- In de toets wordt als kenmerk van de taakuitvoering opzoeken getoetst.

Lezen, subdomein fictionele, narratieve en literaire teksten

Taken

- De leerlingen lezen en maken opdrachten bij fictionele, narratieve en literaire teksten.

Kenmerken van de taakuitvoering

- In de toets wordt als kenmerk van de taakuitvoering begrijpen getoetst.
- In de toets wordt als kenmerk van de taakuitvoering interpreteren getoetst.

Taalverzorging

- De leerlingen worden getoetst op verschillende categorieën van spellingsproblemen en – regels.
- De leerlingen worden getoetst op regels voor lettergreepgrenzen.
- De leerlingen worden getoetst op regels voor woordgrenzen.
- De leerlingen worden getoetst op morfologische spelling.
- De leerlingen worden getoetst op regels voor werkwoordspelling.
- De leerlingen worden getoetst op overige regels.
- De leerlingen worden getoetst op leestekens.
- De toets heeft een evenwichtige verdeling tussen opgaven rond niet-werkwoordspelling, werkwoordspelling en interpunctie.

Schrijven, via indirecte meting

Taken

- De leerlingen reviseren of beoordelen tekstsoorten uit de schrijftakencategorie Correspondentie.
- De leerlingen reviseren of beoordelen tekstsoorten uit de schrijftakencategorie Formulieren, berichten, advertenties, aantekeningen.
- De leerlingen reviseren of beoordelen tekstsoorten uit de schrijftakencategorie Verslagen, werkstukken, samenvattingen, artikelen.

Kenmerken van de taakuitvoering

- In de toets maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect samenhang.
- In de toets maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect afstemming op doel.
- In de toets maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect afstemming op publiek.
- In de toets maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect woordgebruik en woordenschat.

Begrippenlijst

Taken

- De leerlingen worden getoetst met opgaven die betrekking hebben op de begrippenlijst.

10.1.3 *Kwaliteitseis 3, Onderdeel Rekenen*

Een eindtoets PO toetst leerlingen op de inhoud van en meet hun beheersingsniveau ten aanzien van het Referentiekader rekenen, alle domeinen Getallen, Verhoudingen, Meten en meetkunde, en Verbanden.

Voor het einde van de basisschool zijn twee referentieniveaus van belang. Voor rekenen zijn dat niveau 1F en 1S. Het verschil tussen beide niveaus zit vooral in de complexiteit van contexten, bewerkingen en getallen. Op niveau 1S ligt de nadruk meer op formeel rekenen, op inzicht in het decimale systeem en in bewerkingen, op rekenen met grotere getallen en op complexere contexten waarin bijvoorbeeld ook gegevens gecombineerd moeten worden. Bij 1F gaat het meer om elementair rekenen, voornamelijk in eenvoudiger contexten met niet te grote getallen of met veel voorkomende getallen (breuken, kommagetallen) en maten. De nadruk ligt meer op het kunnen rekenen in dagelijkse situaties. Maar ook op het niveau van 1F moeten leerlingen tot op een bepaalde hoogte formeel kunnen rekenen.

Specifieke eisen rekenen

Rekenen, verdeling toetsopgaven over domeinen

De eindtoets bevat opgaven uit alle domeinen uit het referentiekader rekenen. Het betreft de domeinen Getallen (G), Verhoudingen (Vh), Meten en meetkunde (M/mk) en Verbanden (Vb).

De opgaven dienen in de toets als volgt over de domeinen te zijn verdeeld:

- Ten minste 30 procent en maximaal 40 procent van alle toetsopgaven is afkomstig uit het domein Getallen.
- Ten minste 20 procent en maximaal 30 procent van alle toetsopgaven is afkomstig uit het domein Verhoudingen.
- Ten minste 20 procent en maximaal 30 procent van alle toetsopgaven is afkomstig uit het domein Meten & meetkunde.
- Ten minste 15 procent en maximaal 20 procent van alle toetsopgaven is afkomstig uit het domein Verbanden.

Rekenen, verdeling toetsopgaven over onderdelen

Elk domein is opgebouwd uit drie onderdelen:

- A. Notatie, taal en betekenis, waarbij het gaat om de uitspraak, schrijfwijze en betekenis van getallen, symbolen en relaties en om het gebruik van wiskundetaal.
- B. Met elkaar in verband brengen, waarbij het gaat om het verband tussen begrippen, notaties, getallen en dagelijks spraakgebruik.
- C. Gebruiken, waarbij het er om gaat rekenkundige vaardigheden in te zetten bij het oplossen van problemen.

De toets bevat opgaven uit alle onderdelen.

De relatieve verdeling van opgaven over de onderscheiden onderdelen is als volgt5

- Ten minste 20 procent van de gehele toets is afkomstig uit onderdeel B (met elkaar in verband brengen).
- Ten minste 30 procent van de gehele toets is afkomstig uit onderdeel C (gebruiken).
- Elke opgave is onder één (primair) onderdeel geplaatst, met uitzondering van onderdeel A.

Rekenen, opgaven met en zonder context

In een eindtoets PO rekenen moeten zowel opgaven met context als zonder context (zogenoemde 'kale opgaven') worden opgenomen. De toetswijzercommissie – rekenen heeft de volgende definitie van contextopgaven en van contextloze opgaven geformuleerd:

'Een contextopgave is een opgave waarin benoemde getallen voorkomen. Dit zijn getallen die gekoppeld zijn aan aantallen (bijvoorbeeld aantallen appels) of aan hoeveelheden die uitgedrukt zijn in een eenheid (bijvoorbeeld van geld, tijd, lengte). Wordt aan getallen een illustratie gekoppeld dan is het ook een contextopgave. Een contextloze opgave is een opgave met onbenoemde getallen' (CvE, 2014).

Kwaliteitseisen

- De toets bestaat voor het domein Getallen uit ten minste 30 procent contextopgaven.
- De toets bestaat voor het domein Getallen uit ten minste 20 procent contextloze opgaven.
- De toets bestaat voor het domein Verhoudingen uit ten minste 30 procent contextopgaven.
- De toets bestaat voor het domein Verhoudingen uit ten minste 20 procent contextloze opgaven.

Rekenen, gebruik kladpapier

Leerlingen moeten bepaalde basisvaardigheden, zoals het kennen van de (deel)tafels en het rekenen tot 100 goed beheersen. Indien in een eindtoets PO rekenopgaven zijn opgenomen waarbij het nodig is dat deze vaardigheden worden gebruikt, hoeft dit niet als apart te toetsen onderdeel te worden opgenomen (CvE, 2014, p.20).

- De leerlingen mogen bij ten minste 80 procent van de toetsopgaven kladpapier gebruiken.

Rekenen, toetsopgaven op niveau 1S en 1F (S/F)

Een eindtoets PO moet voor rekenen zowel 1F als 1S opgaven bevatten. Ongeacht de toetsvorm, moet de toets het mogelijk maken om voor elke leerling uitspraken te doen ten aanzien van de beheersing van het 1F niveau en het 1S niveau. Dat betekent dat de opgaven van een eindtoets breed zijn samengesteld qua moeilijkheid; ten minste de inhouden van 1F en 1S moeten in een eindtoets zijn vertegenwoordigd

- De opgaven van een toets zijn breed samengesteld qua moeilijkheid.
- Ten minste de inhouden van 1F en 1S zijn in de toets vertegenwoordigd.

10.1.4 Kwaliteitseis 4, Niveaubepaling ten aanzien van de referentieniveaus

Een eindtoets PO meet het beheersingsniveau dat de leerling heeft behaald aan het einde van het PO ten aanzien van de referentieniveaus. Het grootste deel van de leerlingen zal een beheersingsniveau hebben dat valt binnen het bereik van de referentieniveaus 1F en 2F (taal) en 1F en 1S (rekenen).

10.1.5 Kwaliteitseis 5, Handleiding

De handleiding van een eindtoets beschrijft de inhoud van de betreffende eindtoets en geeft inzicht in de gemaakte keuzes. Daarmee biedt de handleiding tevens informatie voor andere belanghebbenden, zoals scholen. Een handleiding is vergelijkbaar met de specifieke toetswijzer van de Centrale Eindtoets PO en omvat ten minste:

- een beschrijving van de in de toets opgenomen domeinen;
- een beschrijving per domein van de in de toets opgenomen verschillende inhoudsonderdelen waarbij de relatie wordt aangegeven tussen de inhouden van het referentiekader en de inhouden van de toets en waarbij inzichtelijk wordt gemaakt of de toets voldoet aan de minimaal wettelijke kwaliteitseisen;
- een beschrijving die inzicht geeft in de wijze waarop aan kwaliteitseis 4 wordt voldaan;
- een beschrijving die inzicht geeft in de gemaakte keuzes ten aanzien van domeinen en inhoudsonderdelen;
- een beschrijving van de manier van toetsen: de vorm waarin opgaven of taken worden aangeboden.

De Expertgroep toetsen PO beoordeelt niet de handleiding op zich, maar gebruikt de handleiding bij de beoordeling van de vraag of de betreffende eindtoets voldoet aan de kwaliteitseisen 1 tot en met 4. Voor de Centrale Eindtoets wordt gebruik gemaakt van de Toetswijzer Eindtoets PO, specifiek deel. Daarnaast is er een praktische handleiding voor scholen die de Centrale Eindtoets afnemen.

11 Bijlage Procedurele eisen en indicatoren Centrale Eindtoets

11.1 Inleiding

Naast de inhoudelijke eisen is er in het Toetsbesluit PO een aantal proceseisen opgenomen voor een Eindtoets PO.

11.1.1 *Rapportage*

- De toetsaanbieder stelt voor iedere leerling die de eindtoets aflegt een leerlingrapport op.
- Het leerlingrapport bevat het resultaat van de eindtoets.
- Het leerlingrapport bevat het niveau waarop de toets is afgelegd.
- Het leerlingrapport geeft een eenduidig advies omtrent het te volgen vervolgonderwijs (maximaal twee schoolsoorten of leerwegen).
- Het leerlingrapport geeft een indicatie van de beheersing van de referentieniveaus Nederlandse taal en rekenen.

11.1.2 *Leerling*

De eindtoets is geschikt voor alle leerlingen, ook voor leerlingen met een specifieke ondersteuningsbehoefte (met uitzondering van meervoudig gehandicapte leerlingen en leerlingen die korter dan 4 jaar in Nederland zijn en daardoor de Nederlandse taal onvoldoende beheersen).

11.1.3 *Afname*

- De eindtoets wordt jaarlijks afgenomen in het tijdvak tussen 15 april en 15 mei.
- De eindtoets heeft een mogelijkheid voor een tweede afnamemoment.
- Het bij de eindtoets behorende toetsreglement regelt de wijze van afnemen, zoals beschreven in het Toetsbesluit, artikel 7.

11.1.4 *Constructie en verantwoording*

- De eindtoets bestaat jaarlijks uit nieuwe opgaven voor taal en rekenen.
- De toets is voorzien van een handleiding met daarin een verantwoording van de gemaakte keuzes ten aanzien van de toetsinhoud, de afnamevorm, en de psychometrische aspecten van de toets.

Colofon

Redactie
Bijdragen

College voor Toetsen en Examens
Herman van Boxtel,
Anja de Wijs,
Margit van Aalst,
Ronald Engelen

College voor Toetsen en Examens, Utrecht
April 2015

