



College voor Toetsen en Examens

RAPPORTAGE

REFERENTIESETS NEDERLANDSE TAAL (LEZEN) EN
REKENEN
VERANTWOORDING PROJECT

September 2014

Inhoud

Voorwoord	5
Samenvatting	6
1 Inleiding	7
2 Wat is een referentieset?	8
2.1 Illustratie referentieset	8
2.2 Verschil in moeilijkheidsgraad van opgaven	9
2.3 Vergelijkbare cesuren voor toetsen en examens	10
2.4 Wat is een referentieset <i>niet</i> ?	11
3 Welke referentiesets	13
3.1 Niet-openbare referentiesets taal en rekenen	13
3.2 Openbare referentiesets taal en rekenen	13
3.2.1 IJkingsfunctie openbare referentiesets	13
3.2.2 Verantwoordingsfunctie openbare referentiesets	13
3.3 Project ankersets taalverzorging 1F en 2F	14
4 Toetsmatrijs referentiesets	15
4.1 Toelichting toetsmatrijs referentiesets taal	15
4.2 Toelichting toetsmatrijs referentiesets rekenen	17
5 Inhoudsvalidering referentiesets	21
5.1 Opdracht voor experts	21
5.2 Samenstelling expertpanels	21
5.3 Proces inhoudelijke validering door experts	21
5.4 Adviezen expertpanels	22
6 Afnames	24
6.1 Afnames in welke onderwijssectoren?	24
6.2 Verzamelde afnamegegevens	25
7 Standaardbepaling referentiesets	28
7.1 Procedure standaardbepaling referentiesets	28
7.1.1 Expertpanels standaardbepaling	28
7.1.2 Procedure standaardbepaling rekenen	28
7.1.3 Procedure taal	29
7.2 Uitkomsten standaardbepaling	29
7.2.1 Uitkomsten standaardbepaling referentiesets rekenen	29
7.2.2 Uitkomsten standaardbepaling referentiesets taal	31
8 Implementatie referentiesets	33
8.1 Overbrengen referentiecesuur naar een centrale toets of examen	33
8.2 Overbrengen referentiecesuur naar een commercieel product	33
8.3 Publicatie openbare referentiesets	34

Literatuur	35
BIJLAGE I Expertpanels inhoudsvalidering referentiesets	36
BIJLAGE II Methodes voor standaardbepaling	38
BIJLAGE III Expertpanels standaardbepaling referentiesets	39

Voorwoord

Voor u ligt de rapportage van het College voor Toetsen en Examens (het CvTE) over het project Referentiesets taal en rekenen.

Deze rapportage heeft als doel een breed publiek te informeren over de manier waarop de referentiesets taal en rekenen tot stand gekomen zijn. De rapportage beschrijft de belangrijkste fasen in de ontwikkeling van de referentiesets taal en rekenen. Bovendien wordt u geïnformeerd over de concrete producten die dit project heeft opgeleverd.

Deze rapportage bevat géén overzicht van de opgaven uit de referentiesets taal (lezen) en rekenen. De opgaven en bijbehorende data van deze referentiesets zijn te vinden via www.referentiesets.nl.

De voorliggende rapportage is in samenwerking met Stichting Cito tot stand gekomen. Dit project en het bijbehorende onderzoek is oorspronkelijk beschreven in Wools & Béguin (2013, 2014) en Béguin & Wools (2014).

Samenvatting

In 2011 is het College voor Toetsen en Examens (het CvTE), in opdracht van het ministerie van OCW, gestart met het project *Referentiesets taal (lezen) en rekenen*. Stichting Cito werd aangewezen als uitvoerder van het project.

Een referentieset bestaat uit een set van opgaven die het inhoudelijk beschreven kennisdomein van het referentieniveau zo goed mogelijk representeert. De opgaven hebben hun oorsprong in verschillende onderwijssectoren waarop het betreffende referentieniveau van toepassing is. Alle opgaven zijn afgenomen in de onderwijssectoren die voor de referentieset relevant zijn. Er is een prestatiestandaard (= referentiecesuur) vastgesteld voor de referentieset.

Het project referentiesets had als doel het ontwikkelen van landelijke referentiecesuren voor de referentieniveaus Nederlandse taal (lezen) en rekenen. Sommige referentieniveaus worden in meerdere onderwijssectoren centraal getoetst of geëxamineerd. Voor deze referentieniveaus zijn sectoroverstijgende prestatiestandaarden nodig. Door middel van een referentieset is zo'n prestatiestandaard ontwikkeld.

Er zijn in totaal zes referentiesets ontwikkeld: rekenen 1F, 2F, 3F; taal (begrijpend lezen) 1F, 2F, 3F. Verder zijn ankersets rekenen 1S opgeleverd voor het referentieniveau rekenen 1S. Deze ankersets zijn alleen van toepassing op het primair onderwijs.

Voor elk referentieniveau zijn niet-openbare en openbare referentiesets ontwikkeld. De niet-openbare referentiesets hebben als doel het overbrengen van de referentiecesuur op centrale toetsen en examens Nederlandse taal en rekenen, die zodoende aan een vergelijkbare cesuur gerelateerd kunnen worden.

De openbare referentiesets zijn bedoeld voor commerciële toetsaanbieders en uitgeverijen. Zij kunnen de landelijke referentiecesuur op hun eigen toetsen taal en rekenen overbrengen met behulp van de referentiesets. Bovendien kan het CvTE, met behulp van de openbare referentiesets, verantwoorden hoe de normering van centrale toetsen en examens, die rapporteren op referentieniveaus, tot stand is gekomen.

Bij de ontwikkeling van de referentiesets hebben docenten en experts uit het veld een belangrijke rol gespeeld. Een panel van experts heeft de door Cito ontwikkelde opgaven inhoudelijk gevalideerd. Een uitgebreider panel van experts heeft deelgenomen aan de standaardbepaling van de referentiesets, waarmee de referentiecesuren bepaald konden worden.

In het najaar van 2013 heeft het College de referentiesets taal en rekenen en bijbehorende referentiecesuren vastgesteld. Daarmee is het hoofddoel van het project referentiesets behaald, namelijk het opleveren van landelijke referentiecesuren voor de referentieniveaus taal en rekenen.

1 Inleiding

In 2010 is de *Wet referentieniveaus Nederlandse taal en rekenen* van kracht geworden. Deze wet geldt voor alle onderwijssectoren en draagt daarmee zorg voor inhoudelijke samenhang tussen deze sectoren. In de wet is vastgesteld dat de referentieniveaus verplicht zijn van primair onderwijs (PO) tot uitstroom naar de arbeidsmarkt en instroom naar het hoger onderwijs (HO).

Het referentiekader taal en rekenen bestaat uit fundamentele niveaus en streefniveaus. Het fundamentele niveau (F-niveau) is de basis die zo veel mogelijk leerlingen moeten beheersen. Het streefniveau (S-niveau) is voor leerlingen die meer aankunnen. Er zijn drie F-niveaus, namelijk 1F, 2F 3F (en 4F taal); er zijn drie S-niveaus rekenen, 1S, 2S, 3S. Niveau 2F in taal en rekenen is nodig om in de samenleving te kunnen functioneren.

Om te kunnen bepalen of leerlingen en studenten het gevraagde referentieniveau beheersen, zijn toetsen nodig. Daarom wordt, onder verantwoordelijkheid van het CvTE, een doorlopende toetslijn geïmplementeerd met verplichte centrale toetsen en examens Nederlandse taal en rekenen.

Het is belangrijk dat in die toetsen en examens de eisen van de referentieniveaus voor alle leerlingen gelijk zijn, ongeacht het schooltype dat zij volgen. Op die manier kan de doorlopende leerlijn tussen de schooltypen worden gerealiseerd. Om dit mogelijk te maken heeft het ministerie van Onderwijs, Cultuur en Wetenschap (OCW) medio 2011 opdracht gegeven een vergelijkingsonderzoek uit te voeren onder de noemer *project referentiesets taal en rekenen*. Het CvTE is gedelegeerd opdrachtgever van het project, Stichting Cito (vanaf nu: Cito) uitvoerder.

In de navolgende hoofdstukken wordt uitgelegd wat het project referentiesets inhoudt, op welke wijze het project is uitgevoerd. Ook wordt ingegaan op de implementatie van de referentiesets in de doorlopende leerlijn taal en rekenen.

2 Wat is een referentieset?

Een referentieset bestaat uit een set van opgaven die samen het inhoudelijk beschreven kennisdomein van het referentieniveau zo goed mogelijk representeren. De opgaven hebben hun oorsprong in verschillende onderwijssectoren waarop het betreffende referentieniveau van toepassing is. Alle opgaven zijn afgenomen in de onderwijssectoren die voor de set relevant zijn. Er is een prestatie standaard (= referentiecesuur) vastgesteld voor de referentieset.

2.1 Illustratie referentieset

De betekenis van een referentieset kan het best weergegeven worden aan de hand van een illustratie. In figuur 1 wordt schematisch een referentieset afgebeeld. Elk balkje stelt een opgave voor. Roze balkjes zijn opgaven uit het primair onderwijs (PO), blauwe balkjes zijn opgaven uit het voortgezet onderwijs (VO) en groene balkjes zijn opgaven uit het middelbaar beroepsonderwijs (mbo).

Ter illustratie is voor elke sector een voorbeeldopgave opgenomen uit de openbare referentiesets. De blauwe lat geeft de 'prestatie standaard' weer. Dit is een door experts vastgelegde (inhoudelijke) grens die aangeeft hoeveel opgaven een leerling goed moet maken om een referentieniveau te behalen.

Figuur 1: referentieset rekenen 2F



Wanneer de opgaven in oplopende moeilijkheid worden weergegeven, moeten gemiddeld de opgaven onder de streep goed gemaakt worden en geeft beheersing van de opgaven boven de streep aan dat een leerling een hoger niveau beheerst. Dit geldt gemiddeld over alle leerlingen. Het kan immers voorkomen dat een leerling een paar opgaven onder de streep fout maakt, maar dat compenseert via opgaven boven de streep die hij goed maakt.

2.2 **Verskil in moeilijkheidsgraad van opgaven**

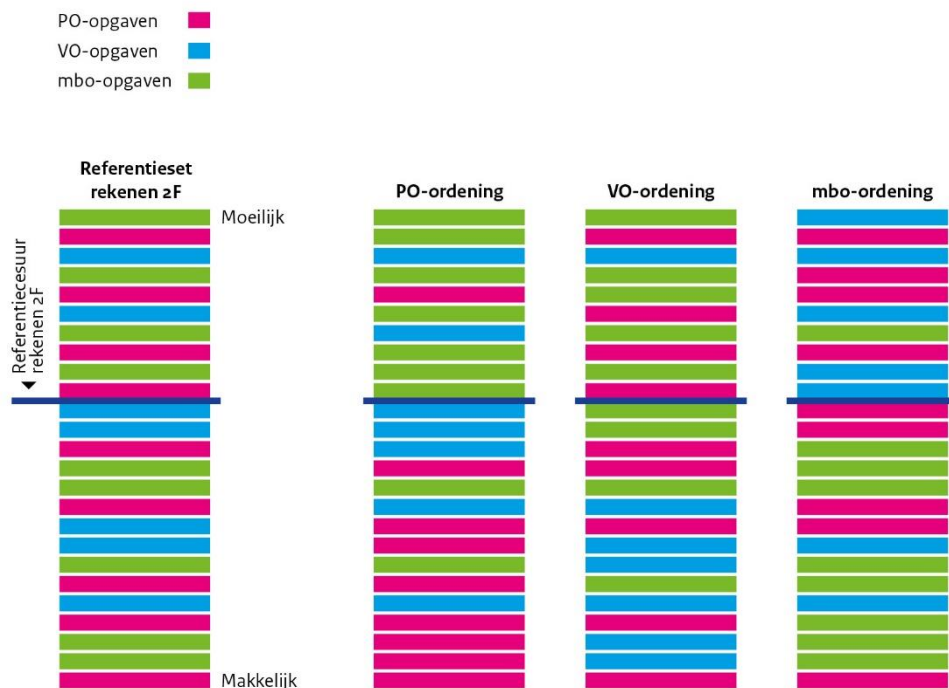
Een probleem dat kan optreden, is dat leerlingen uit de ene onderwijssector een vraag moeilijker vinden dan leerlingen uit een andere onderwijssector. Dit hoeft niet te komen doordat deze leerlingen minder vaardig zijn, maar doordat zij bijvoorbeeld niet gewend zijn met een bepaald type vraag om te gaan.

Zo kan het voorkomen dat opgaven, die gesorteerd worden op moeilijkheid, voor de ene onderwijssector anders geordend worden dan voor de andere. Deze verschillende ordening kan er voor zorgen dat een toets bepaalde groepen onterecht bevoordeeld of benadeeld. Denk bijvoorbeeld aan een rekentoets die veel opgaven bevat die voor kandidaten uit het mbo makkelijk zijn, terwijl leerlingen uit vmbo-GT juist moeite hebben met deze specifieke opgaven. Het zou dan kunnen voorkomen dat een kandidaat uit het mbo een hogere score haalt dan een leerling uit vmbo-GT die, als we zouden kijken naar een representatieve set met opgaven, eigenlijk even goed kan rekenen.

In Figuur 2 is het verschil in ordening van moeilijkheid gevisualiseerd. Wederom stelt elk balkje een opgave voor, de kleur van het balkje laat zien uit welke sector de opgave afkomstig is. Ook hier geeft de blauwe lat een prestatie standaard weer. De figuur laat zien dat de ordening van makkelijk naar moeilijk verschilt per onderwijssector (fictief voorbeeld). Naast de verschillende ordeningen wordt weergegeven dat, ongeacht de onderwijssector, de prestatie standaard voor dit referentieniveau altijd gelijk blijft.

De ordening in moeilijkheid wordt bepaald door de opgaven uit de referentiesets bij leerlingen uit alle betreffende onderwijssectoren af te nemen.

Figuur 2: Verschil in ordening van moeilijkheid



2.3

Vergelijkbare cesuren voor toetsen en examens

Wanneer de ordening van de opgaven in moeilijkheidsgraad voor elke onderwijssector bekend is, is het mogelijk om de prestatiestandaard om te zetten in een cesuur op een toets voor een specifieke doelgroep. Dit betekent dat de toets ontwikkeld is voor een specifieke doelgroep, bijvoorbeeld mbo-kandidaten, maar dat de cesuur afgeleid is van de referentieset (die bestaat uit opgaven voor alle sectoren).

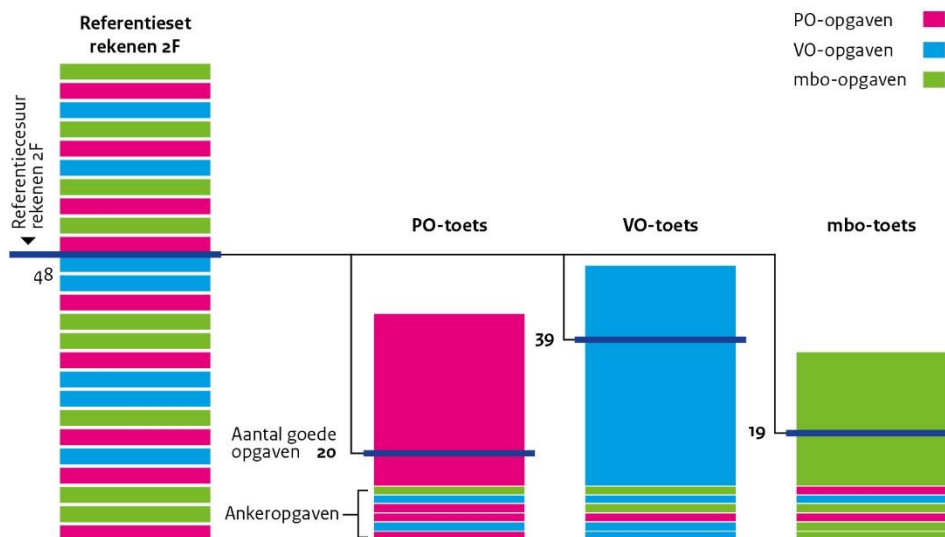
Dit is weergegeven in Figuur 3 (fictief voorbeeld). Elke score op een toets correspondeert met een aantal 'goed' op de referentieset. De score op de toets die correspondeert met de cesuur op de referentieset is de cesuur op de toets. Op deze manier wordt er voor gezorgd dat de cesuren tussen de toetsen vergelijkbaar zijn. Het is dan bijvoorbeeld mogelijk om in een relatief moeilijke toets voor mbo toch dezelfde prestatiestandaard te hanteren als voor een relatief makkelijke toets voor PO.

Het overbrengen van de referentiecensuur op de referentieset naar een specifieke toets is mogelijk door middel van opname van een anker van referentiesetopgaven in de toets. Via de ankeropgaven kan de referentiecensuur overgebracht worden. Het aantal opgaven in het anker is afhankelijk van de betreffende toets (gedacht kan worden aan ongeveer 15 à 20 opgaven). Het maakt daarbij in principe niet uit welke opgaven geselecteerd worden als anker. Het ligt echter het meest voor de hand die opgaven uit de referentieset te selecteren die het beste in de betreffende toets passen.

In de figuur stellen de blokken met één kleur specifieke voorbeeldtoetsen voor, die gericht zijn op één onderwijssector. De blauwe latten en bijbehorende getallen laten de

cesuren van de toetsen zien; deze cesuren zijn gekoppeld aan de prestatiestandaard (blauwe lat) zoals vastgelegd in de referentieset.

Figuur 3: één 2F cesuur voor alle toetsen



Fictief voorbeeld (behorend bij figuur 3):

Olaf (16 jaar) legt de rekentoets in het vmbo-KB af. Hij beantwoordt 39 van de 50 opgaven en behaalt daarmee referentieniveau rekenen 2F.

Thijmen (11 jaar) doet mee aan een verplichte eindtoets basisonderwijs. Voor het onderdeel rekenen heeft Thijmen 20 van de 40 opgaven goed beantwoord. Ook Thijmen behaalt met deze prestatie het niveau 2F.

Thijmen en Olaf hebben niet dezelfde opgaven gemaakt, toch stellen we vast dat ze even goed kunnen rekenen. Dit kan door de koppeling met de referentieset: als Thijmen en Olaf allebei alle opgaven van de referentieset 2F zouden maken, hebben zij (naar verwachting) beiden 48 opgaven goed. Iemand die graag wil weten wat het betekent dat Olaf en Thijmen op niveau 2F kunnen rekenen kan in de referentieset zien welke opgaven zij in principe zouden kunnen oplossen (ongeacht hun leeftijd).

2.4 Wat is een referentieset *niet*?

Een referentieset is niet bedoeld als voorbeeldmateriaal ter voorbereiding op centrale toetsen en examens taal en rekenen. Voorbeeldtoetsen en -examens zijn hiervoor beter geschikt dan een referentieset. De referentieset laat namelijk niet zien welke opgaven er in de specifieke toetsen komen. Zij laat daarnaast ook niet zien hoe de leerstof zich verhoudt tot een toets. De referentiesets zijn dus alleen geschikt om het gehele leerstofdomein (voor alle onderwijssectoren samen) te concretiseren en is er op gericht de prestatiestandaard vast te stellen.

Invoegen link voorbeeldexamens:

- Sector PO: www.centrale-eindtoetspo.nl (oefentoets beschikbaar vanaf het najaar 2014)

- Sector VO: www.examenblad.nl
- Sector mbo: www.examenbladmbo.nl

3 Welke referentiesets

Sommige referentieniveaus worden in meerdere onderwijssectoren centraal getoetst of geëxamineerd (bijvoorbeeld taal 2F). Voor deze referentieniveaus zijn sectoroverstijgende prestatiestandaarden nodig. Door middel van een referentieset kan zo'n prestatiestandaard ontwikkeld worden (zie hoofdstuk 2). Er zijn in totaal zes referentiesets opgeleverd: rekenen 1F, 2F, 3F; taal (begrijpend lezen) 1F, 2F, 3F. Sommige referentieniveaus komen enkel in één onderwijstype naar voren, bijvoorbeeld niveau taal 4F in het vwo. Het is daarom niet noodzakelijk een sectoroverstijgende prestatiestandaard te ontwikkelen.

Niveau rekenen 1S is hierop een uitzondering. Dit niveau is het wettelijk vastgelegde niveau in het PO voor leerlingen die meer aankunnen. Om tegemoet te komen aan de visie van experts om het niveau rekenen 1S te operationaliseren in een verzameling opgaven met bijbehorende prestatiestandaard, zijn ankersets rekenen 1S ontwikkeld. Het niveau 1S is alleen van toepassing op het PO. Er vindt geen vergelijking tussen schooltypen plaats. Daarom wordt gebruikgemaakt van de term 'ankersets' in plaats van 'referentiesets' (sectoroverstijgend). Deze ankersets bevatten rekenopgaven 1S en zijn alleen in het PO afgenomen.

3.1 Niet-openbare referentiesets taal en rekenen

Het hoofddoel van het project referentiesets is het opleveren van referentiesets die een landelijke referentiecesuur (= prestatiestandaard) leveren voor de referentieniveaus, die vergelijkbaar is over verschillende schooltypen heen. De opgaven uit deze referentiesets gaan mee als anker in de centrale toetsen en examens taal en rekenen, zodat deze aan een vergelijkbare cesuur gerelateerd kunnen worden. Vanwege het geheime karakter van centrale toetsen en examens worden deze referentiesets niet openbaar gemaakt.

3.2 Openbare referentiesets taal en rekenen

Naast de niet-openbare referentiesets worden openbare referentiesets opgeleverd. Deze hebben twee functies.

3.2.1 *Ijkingsfunctie openbare referentiesets*

Ten eerste gaat het om de ijkingsfunctie. De openbare referentiesets zijn voorzien van een vergelijkbare referentiecesuur als de niet-openbare referentiesets. Commerciële toetsenmakers, uitgeverijen en onderzoekers kunnen middels een ankeronderzoek met de referentiesets de referentiecesuur overbrengen op hun eigen toetsen/producten taal en rekenen (zie: *Headstart ankeronderzoek referentiesets*, paragraaf 8.2).

3.2.2 *Verantwoordingsfunctie openbare referentiesets*

Ten tweede hebben de openbare referentiesets een verantwoordingsfunctie. Met behulp van de openbare referentiesets kan het CvTE aan het onderwijsveld duidelijk maken hoe de normering van centrale toetsen en examens, die rapporteren op referentieniveaus, tot stand is gekomen.

3.3 Project ankersets taalverzorging 1F en 2F

Het ministerie van OCW heeft, in aanvulling op het project referentiesets en naar analogie van het project ankersets rekenen 1S, het CvTE opdracht gegeven om het project ankersets taalverzorging uit te voeren. In dit project worden sets met opgaven ontwikkeld die passen bij het referentieniveau 1F en 2F taalverzorging. Op de sets van opgaven komt een prestatiestandaard te liggen. De ankersets taalverzorging worden alleen afgenomen in het PO.

Het project is eind 2013 van start gegaan. De ankersets taalverzorging zullen, naar verwachting, medio 2015 ontwikkeld zijn. Net als bij het project referentiesets zal ook voor de ankersets taalverzorging een Headstart georganiseerd worden voor toetsenmakers en uitgeverijen. Tijdens de *Headstart ankeronderzoek ankersets taalverzorging* kan deze doelgroep de landelijke referentiecesuur overbrengen op hun eigen toetsen taalverzorging. Toetsaanbieders van alternatieve eindtoetsen kunnen zo vanaf 2016 rapporteren op het referentieniveau taalverzorging.

4 Toetsmatrijs referentiesets

Aan de hand van een toetsmatrijs heeft Cito opgaven voor de referentiesets geselecteerd en ontwikkeld. Een toetsmatrijs geeft weer hoeveel en welk type opgaven een referentieset moet bevatten. De toetsmatrijs van elke referentieset is gebaseerd op de domeinindelingen van het betreffende referentieniveau. Het is van belang dat de geselecteerde opgaven het inhoudsdomein zo goed mogelijk dekken. De opgaven zijn daarom voorgelegd aan een expertpanel (zie voor meer informatie hoofdstuk 5 'inhoudelijke validering')

De samenstelling van de referentiesets moet passend moeten zijn voor de hoofdfunctie: het overbrengen van de referentiecesuur op verschillende toetsen. De toetsmatrijzen zijn hier ook op gebaseerd. De toetsmatrijzen van de openbare referentiesets (OS) laten een diversiteit aan opgaven en vraagvormen zien. Een OS dient immers bruikbaar te zijn voor andere toetsaanbieders ten behoeve van hun ankeronderzoek (zie paragraaf 2.2). De opgaven uit de niet-openbare referentiesets (NOS) worden gebruikt als anker in de centrale toetsen en examens. Daarom bevatten de NOS enkel opgaven die passend zijn voor de centrale toetsen en examens die het betreffende referentieniveau meten.

4.1 Toelichting toetsmatrijs referentiesets taal

Tabel 1 geeft een beschrijving van de samenstelling van de referentiesets taal 1F, 2F en 3F. De referentiesets taal betreffen enkel het domein 'begrijpend lezen'. De tabellen geven weer uit hoeveel opgaven en teksten de niet-openbare en openbare referentiesets bestaan, welke tekstsoorten hierin voorkomen, en welke verdeling gehanteerd wordt tussen de verschillende onderwijssectoren. De referentiesets bestaan uit de volgende tekstsoorten:

- Fictionele teksten (1F en 2F, ten behoeve van het PO)
 - Boek
 - Fabel
 - Gedicht
 - Historisch
 - Mop
 - Parabelachtig
 - Sprookje
 - Realistische fictie
- Zakelijke teksten
 - Betogend
 - Informatief
 - Instructief

De referentiesets taal bevatten hoofdzakelijk meerkeuzevragen.

Aanvullende kenmerken van de referentiesets taal:

- Er is sprake van een oververtegenwoordiging van informatieve teksten in referentieset taal 1F.
- In referentiesets taal 1F en 2F ontbreken evaluatieve opgaven.
- Een operationalisering van de vaardigheid samenvatten is niet opgenomen in referentieset taal 2F.
- Referentieset taal 3F bevat geen instructieve tekst. Het blijkt erg lastig een instructieve tekst te ontwikkelen op niveau 3F.

De expertpanels, verantwoordelijk voor de inhoudelijke validering van de referentiesets, geven aan het belangrijk te vinden dat deze inhoudelijke kenmerken bij de referentiesets benoemd worden in de communicatie over de referentiesets (zie voor meer toelichting hoofdstuk 5).

Tabel 1. Toetsmatrijs referentiesets taal 1F, 2F en 3F

1F	Nederlandse taal (Lezen)		
Aantal teksten	NOS	OS	Totaal
Fictioneel	4	6	10
Boek		3	3
Fabel	1		1
Gedicht		2	2
Historisch		1	1
Mop	1		1
Parabelachtig	1		1
Sprookje	1		1
Zakelijk	10	14	24
Betogend		2	2
Informatief	8	10	18
Instructief	2	2	4
Totaal	14	20	34
Aantal opgaven	NOS	OS	Totaal
Totaal	50	85	135

2F	Nederlandse taal (Lezen)		
Aantal teksten	NOS	OS	Totaal
Fictioneel	3	3	6
Fictie, historisch	1		1

Fictieeel, grap		1	1
Realistische fictie	2	2	4
Zakelijk	9	18	27
Betogend	1	7	8
Informatief	7	9	16
Instructief	1	2	3
Totaal	12	21	33
Aantal opgaven	NOS	OS	Totaal
Totaal	69	89	158

3F	Nederlandse taal (Lezen)		
Aantal teksten	NOS	OS	Totaal
Zakelijk	7	13	20
Betogend	4	9	13
Informatief	2	3	5
Instructief	1	1	2
Totaal	7	13	20
Aantal opgaven	NOS	OS	Totaal
Totaal	66	98	164

4.2 Toelichting toetsmatrijs referentiesets rekenen

Tabel 2 is een beschrijving van de samenstelling van de referentiesets rekenen. De tabel geeft weer uit hoeveel opgaven de NOS en OS bestaan en hoe de verdeling over domeinen en schooltypen eruit ziet.

De volgende domeinen komen voor in de referentiesets rekenen:

- getallen
- meten en meetkunde
- verbanden

- verhoudingen

In de referentiesets komen open vragen en meerkeuzevragen voor. Verder zijn er opgaven met rekenmachine en opgaven zonder rekenmachine.

Aanvullend kenmerk bij de referentieset rekenen 3F:

- Het expertpanel rekenen 3F geeft aan dat het enkel opnemen van niveau 3F opgaven in de referentiesets rekenen 3F ertoe heeft geleid dat het domein meten en meetkunde wat ondervertegenwoordigd is in de referentieset rekenen 3F. Het expertpanel raadt daarom aan toetsenmakers aan om, ten behoeve van het overbrengen van de referentiecesuur, een mix van opgaven uit de 2F en 3F referentieset samen te stellen om zo het zo tot een goed mogelijk inhoudelijke dekking van het domein te komen.

Tabel 2. Toetsmatrijs referentiesets rekenen 1F, 2F en 3F, ankersets rekenen 1S

1F	Rekenen		
	NOS	OS	Totaal
Domein: getallen	20	32	52
Domein: meten	15	24	39
Domein: verbanden	5	8	13
Domein: verhoudingen	10	16	26
Totaal	50	80	130
	NOS	OS	Totaal
Vraagtype: meerkeuze	33	48	81
Vraagtype: open	17	32	49
Totaal	50	80	130

1S	Rekenen		
	NOS	OS	Totaal
Domein: getallen	16	16	32
Domein: meten	12	12	24
Domein: verbanden	4	4	8

Domein: verhoudingen	8	8	16
Totaal	40	40	80
	NOS	OS	Totaal
Vraagtype: meerkeuze	37	23	60
Vraagtype: open	3	17	20
Totaal	40	40	80

2F	Rekenen		
	NOS	OS	Totaal
Domein: getallen	15	24	39
Domein: meten	10	16	26
Domein: verbanden	10	16	26
Domein: verhoudingen	15	24	39
Totaal	50	80	130
	NOS	OS	Totaal
Vraagtype: meerkeuze	19	34	53
Vraagtype: open	31	46	77
Totaal	50	80	130
	NOS	OS	Totaal
Hulpmiddel: rekenmachine	46	44	90
Hulpmiddel: zonder rekenmachine	4	36	40
Totaal	50	80	130

3F	Rekenen		
	NOS	OS	Totaal
Domein: getallen	12	21	33
Domein: meten	11	18	29
Domein: verbanden	11	15	26
Domein: verhoudingen	16	26	42
Totaal	50	80	130
	NOS	OS	Totaal
Vraagtype: meerkeuze	1	31	32
Vraagtype: open	49	49	98
Totaal	50	80	130
	NOS	OS	Totaal
Hulpmiddel: rekenmachine	44	53	97
Hulpmiddel: zonder rekenmachine	6	27	33
Totaal	50	80	130

5 Inhoudvalidering referentiesets

5.1 Opdracht voor experts

Het is van belang dat de referentiesets het inhoudsdomen van het referentieniveau goed dekken. De opgaven zijn daarom voorgelegd aan expertpanels. De expertpanels hadden als taak het inhoudelijk valideren van de door Cito geselecteerde opgaven voor de openbare en niet-openbare referentiesets. Experts kregen de taak voorgelegd de opgaven uit de referentiesets te beoordelen op de volgende punten:

- Zijn de opgaven inhoudelijk passend bij het te meten kennisdomein?
- Zijn de opgaven geschikt voor het referentieniveau waarvoor de opgaven geselecteerd zijn?
- Is de volledige set met opgaven bruikbaar voor het doel van de referentiesets, namelijk een ijkinstrument voor verschillende toetsen en examens voor het overbrengen van de referentiecesuur?

5.2 Samenstelling expertpanels

Voor alle referentiesets is een apart panel van experts samengesteld. Een expertpanel heeft zowel de openbare als niet-openbare referentieset van een bepaald vak en niveau gevalideerd (bijvoorbeeld: expertpanel taal 2F valideert zowel de openbare als niet-openbare referentiesets taal 2F).

De ontwikkeling van ankersets rekenen 1S is op een later moment aan het project toegevoegd. Het expertpanel rekenen 1F is vanuit inhoudelijk oogpunt gevraagd ook de validering van de ankersets rekenen 1S op zich te nemen.

Het CvTE was verantwoordelijk voor de samenstelling van de expertpanels. Bij de samenstelling is rekening gehouden met de verschillende visies en stromingen uit het onderwijsveld, om te zorgen voor draagvlak voor en kwaliteit van de referentiesets. De experts zijn benaderd op basis van hun expertise en kennis van de referentieniveaus.

De zes expertpanels bestaan uit experts die geselecteerd zijn op persoonlijke titel en afkomstig zijn van onder andere:

- SLO
- leerplancentra
- onderwijsadviesbureaus
- wetenschappelijke instituten
- CvTE Vaststellingscommissies taal en rekenen VO en mbo
- CvTE Toetswijzercommissies PO
- Het desbetreffende scholenveld

Elk panel bestond uit 5 tot 8 experts. Het panel is telkens aangevuld met één toetsdeskundige van Cito. Bijlage I biedt een overzicht van de expertpanels.

5.3 Proces inhoudelijke validering door experts

Op een aantal beoordelingsmomenten hebben experts een oordeel gegeven over de samenstelling van de referentiesets. Onderstaand wordt het proces van inhoudelijke validering beschreven. Alle expertpanels hebben dezelfde stappen doorlopen.¹

¹ Op een later moment tijdens het valideringsproces van de referentiesets is besloten ankersets rekenen 1S te ontwikkelen. Vanaf de tweede valideringsronde heeft het expertpanel rekenen 1F ook de validering van de ankersets rekenen 1S op zich genomen.

Eerste valideringsronde

Ter voorbereiding op de eerste valideringsbijeenkomst hebben experts de opgaven beoordeeld op hun geschiktheid voor opname in de referentieset. Tijdens een valideringsbijeenkomst heeft het panel gediscussieerd over de samenstelling van de referentiesets. Uit deze eerste valideringsbijeenkomst bleek dat een deel van de opgaven geschikt werd bevonden.

Daarnaast werd een deel van de opgaven gezien als niet of minder bruikbaar. Zo gaven experts aan dat de referentiesets een grotere variatie aan opgaven moesten bevatten. Soms ging het om opgaven die bijvoorbeeld beter geschikt zijn voor een specifiek schooltype. Soms ging het om meer innovatieve vormen van bevraging of om meer open opgaven en minder meerkeuze-opgaven.

Aanvullende constructieopdracht

Naar aanleiding van de eerste valideringsbijeenkomst is een aanvullend constructieplan opgesteld. Dit constructieplan is voorgelegd aan de expertpanels. Aan de hand van dit plan heeft aanvullende itemconstructie plaatsgevonden.

Tweede valideringsronde

De opgaven die voortkwamen uit de aanvullende constructie zijn voorgelegd aan de experts. Wederom konden experts de opgaven beoordelen op hun geschiktheid voor opname in de referentieset. De expertpanels hebben vervolgens een advies geformuleerd richting het CvTE over de samenstelling en bruikbaarheid van de referentiesets, gelet op het beoogde doel: opname van een anker in specifieke toetsen en examens om de referentiesesuur over te brengen.

5.4 Adviezen expertpanels

Elk expertpanel heeft het CvTE geadviseerd de referentiesets vast te stellen voor het overbrengen van de referentiesesuur op verschillende toetsen en examens.² De expertpanels wijzen zonder uitzondering op het belang van een zorgvuldige communicatie over de functie van de referentiesets. In de communicatie moet volgens de expertpanels benadrukt worden dat de referentiesets ingezet worden om eenduidige cesuren vast te stellen. De referentiesets zelf zijn niet geschikt als voorbeeldtoets in enige sector, hiervoor zijn voorbeeldexamens en -toetsen uit die onderwijssectoren beter geschikt (zie ook §2.4).

Andere adviezen van de expertpanels:

- Het expertpanel rekenen 1F/1S wijst op het belang van een goede communicatie over de functie van het niveau 1F en 1S. Het expertpanel is van mening dat duidelijk moet worden dat het doel is dat leerlingen het 1S-niveau halen en dat het 1F-niveau eerder als vangnet fungeert dan als beoogd eindniveau.
- Ook het expertpanel taal 1F wijst op het belang van een goede communicatie over de functie van het niveau 1F en het niveau 2F. Toetsen in het basisonderwijs zullen volgens het expertpanel vaak beide niveaus in eenzelfde toets meten.
- Op advies van het expertpanel rekenen 2F zijn afspraken gemaakt over de oplevering van de openbare referentiesets. Allereerst is afgesproken dat de papieren opgaven van de referentiesets in zwart/wit worden afgenomen, maar dat de

² Dit advies is geformuleerd met het oog op de syllabi en toetswijzers voor centrale toetsen en examens Rekenen en Nederlandse Taal die ten tijde van de opgavenconstructie voor de referentiesets beschikbaar waren. Indien de inhoud van deze toetswijzers of syllabi drastisch wijzigt, kan het noodzakelijk zijn de consequenties van deze wijzigingen voor het overbrengen van de standaard te evalueren.

afbeeldingen ook in kleur worden opgeleverd. Op deze manier kunnen toetsaanbieders de opgaven ook digitaliseren en in de toekomst in kleur afnemen. Ten tweede is bepaald dat in de referentiesets niet zal worden weergegeven welk schooltype initieel aan een opgave was verbonden. Het is namelijk niet altijd eenduidig uit welk schooltype een opgave afkomstig is.

- Expertpanel rekenen 3F benadrukt het belang van een goede communicatie over de relatie met het 3S niveau.

Verder willen de expertpanels een aantal inhoudelijke kenmerken van de referentiesets nadrukkelijk benoemen. Deze kenmerken zijn te vinden in hoofdstuk 4.1 en 4.2 van dit rapport.

6 Afnames

Om te bepalen hoe leerlingen scoren op de opgaven uit de referentiesets (bepalen van de moeilijkheidsgraad van de opgaven) zijn de opgaven afgenomen in alle relevante onderwijssectoren. Deze afnamegegevens zijn op twee manieren verzameld: via losse afnames en via afnames van centrale pilottoetsen en –examens (zogenaamde zaaiopgaven).

6.1 Afnames in welke onderwijssectoren?

Tabellen 3 en 4 geven een overzicht van de onderwijssectoren waarin de referentiesets afgenomen zijn. Daarnaast wordt ook weergegeven wat het fundamentele niveau (F) voor deze leerlingen is, wat het streefniveau (S) is en wat het lager gelegen niveau (L) is. Het streef- dan wel lager gelegen niveau meenemen in de afname geeft zicht op de doorlopende leerlijn 1F->2F->3F.³ Hiermee kon ook gecontroleerd worden of de prestatiecriteria empirisch ook in oplopende volgorde gezet werden.

Tabel 3 laat bijvoorbeeld zien dat in het primair onderwijs zowel de 1F- als de 2F-referentieset is afgenomen. In het vmbo-BB-KB is zowel de 1F referentieset (lagergelegen niveau) als de 2F referentieset (fundamentele niveau) afgenomen.

Een '-' betekent dat de referentieset niet in de betreffende onderwijssector is afgenomen. Zo zijn in het mbo alleen de referentiesets van de fundamentele niveaus afgenomen. In het vwo is alleen referentieset 3F afgenomen (bij taal is 3F het lager gelegen niveau, bij rekenen het fundamentele niveau).

Tabel 3. Afnames referentiesets taal in onderwijssectoren

Onderwijssector	1F	2F	3F
primair onderwijs	F	S	-
vmbo-BB	L	F	-
vmbo-KB	L	F	-
vmbo-GT	-	F	S
Havo	-	L	F
Vwo	-	-	L
mbo-2	-	F	-
mbo-3	-	F	-
mbo-4	-	-	F

³ Hiermee kon vastgesteld worden dat de prestatiecriteria voor 1F, 2F en 3F qua moeilijkheid in oplopende volgorde te zetten zijn.

Tabel 4. Afnames referentiesets rekenen in onderwijssectoren

Onderwijssector	1F	1S	2F	3F
primair onderwijs	F	S	S	-
vmbo-BB	L	-	F	-
vmbo-KB	L	-	F	-
vmbo-GT	-	-	F	S
Havo	-	-	L	F
Vwo	-	-	-	F
mbo-2	-	-	F	-
mbo-3	-	-	F	-
mbo-4	-	-	-	F

6.2 Verzamelde afnamegegevens

In het voorjaar van 2013 vonden de losse afnames van de referentiesets plaats in het PO, VO en mbo. Onderstaande tabellen 5 en 6 geven weer hoeveel waarnemingen per opgave gemiddeld verzameld zijn in de drie sectoren. De gewenste aantallen waarnemingen per opgave zijn gebaseerd op de COTAN-richtlijnen voor normgroepen. Volgens deze richtlijnen zijn bij toetsen voor belangrijke beslissingen minimaal 300 waarnemingen per item noodzakelijk om voor het criterium 'normen' een voldoende beoordeling te krijgen. Er zijn minimaal 400 waarnemingen nodig voor het oordeel 'goed'.⁴

⁴ Deze aantallen zijn gebaseerd op aparte normgroepen; wanneer de richtlijnen van continue normering worden aangehouden is het aantal 150. Dit is bijvoorbeeld mogelijk wanneer binnen een schooltype voldoende data verzameld zijn om te compenseren voor een leerweg met te weinig data (zoals binnen het VO).

Tabel 5: gemiddeld aantal waarnemingen per item per schooltype – referentiesets rekenen

		PO	vmbo BB/KB	vmbo GT	havo	vwo	mbo 2/3	mbo 4
1F	OS	488	388					
	NOS	504	473					
1S	OS	503						
	NOS	474						
2F	OS	480	577	931	409		344	
	NOS	488	RTVO*	RTVO*	370		COE**	
3F	OS			696	491	441		315
	NOS			555	RTVO*	RTVO*		COE**

***/**De niet-openbare sets rekenen 2F en 3F zijn waar mogelijk meegenomen in de pilot van de Rekentoets VO en de Centraal Ontwikkelde Examens (COE's) in het mbo. Deze opgaven zijn daarom niet apart in de afnames van de referentiesets opgenomen.**

Tabel 6: gemiddeld aantal waarnemingen per opgave per schooltype – referentiesets taal

		PO	Vmbo- BB/KB	Vmbo-GT	Havo	Vwo	Mbo 2/3	Mbo 4
1F	OS	671	475					
	NOS	722	438					
2F	OS	666	521	398	274 ⁵		414	
	NOS	679	425	380	260		211*	
3F	OS			369	348	375		331
	NOS			353	279 ⁶	281 ⁷		165*

***De niet-openbare set 2F en 3F zijn waar mogelijk meegenomen in de pilot van de Centraal Ontwikkelde Examens (COE's) in het mbo. De opgaven die daarin niet konden worden meegenomen zijn wel in afgenomen in de losse afnames van de referentiesets.**

⁵ Voor de niet-openbare referentiesets taal 2F en 3F is het minimale aantal waarnemingen van 300 niet gehaald voor havo en vwo. Dit kan echter gecompenseerd worden door de overige data uit het VO mee te wegen (continue normering).

⁶ Voor de niet-openbare referentiesets taal 2F en 3F is het minimale aantal waarnemingen van 300 niet gehaald voor havo en vwo. Dit kan echter gecompenseerd worden door de overige data uit het VO mee te wegen (continue normering).

⁷ Voor de niet-openbare referentiesets taal 2F en 3F is het minimale aantal waarnemingen van 300 niet gehaald voor havo en vwo. Dit kan echter gecompenseerd worden door de overige data uit het VO mee te wegen (continue normering).

Uit de tabellen 5 en 6 blijkt dat voor de niet-openbare en openbare referentiesets taal en rekenen in het merendeel van de betreffende schooltypen voldoende data verzameld zijn.

Bij de afnames in het voorjaar van 2013 was het aantal waarnemingen voor de niet-openbare set taal 2F en 3F echter onvoldoende. Dit vormde een probleem voor het overbrengen van referentiecesuur op de COE's taal. Ook waren in het mbo onvoldoende data verzameld voor de openbare referentiesets taal 2F en 3F. Daardoor zou het voor commerciële toetsaanbieders en uitgeverijen moeilijker zijn de referentiecesuur van de referentieniveaus over te brengen via een ankeronderzoek wanneer zij gebruikmaken van een mbo-populatie.

Om de ontbrekende data van de referentiesets taal te verzamelen is in het mbo een reparatieafname georganiseerd. Deze reparatieafname heeft ook in het mbo geleid tot voldoende waarnemingen voor de referentiesets taal. Tabel 6 geeft het aantal waarnemingen weer inclusief deze reparatieafname.

7 Standaardbepaling referentiesets

Naast de ontwikkeling van de referentiesets taal (lezen) en rekenen moet een referentiecesuur worden vastgelegd. Dit betekent dat moet worden bepaald hoeveel opgaven de leerling uit een set referentieopgaven goed moet hebben om het betreffende referentieniveau te halen. De procedure waarmee dat is gebeurd is een standaardbepaling.

Voor elke referentieset is een standaardbepalingsprocedure uitgevoerd. In dit hoofdstuk is een verslag te vinden van deze procedure. Eerst wordt de gevolgde procedure uiteengezet. Daarna volgt een overzicht van de uitkomsten van de standaardbepalingen.

7.1 Procedure standaardbepaling referentiesets

7.1.1 *Expertpanels standaardbepaling*

Ten behoeve van een breed draagvlak uit het veld voor de referentiecesuur is voor elke standaardbepaling een panel van experts samengesteld dat een advies heeft uitgebracht over de te hanteren prestatiestandaard. Deze expertpanels bestaan uit een afvaardiging van stakeholders uit het veld (docenten, wetenschappelijke instituten, onderwijsadvies, SLO). Elk panel bevat 14 tot 21 experts. De expertpanels die verantwoordelijk waren voor de inhoudsvalidering, maken ook deel uit van de expertpanels van de standaardbepaling.

Zie bijlage II voor een overzicht van de indeling van experts in de panels voor de standaardbepaling.

7.1.2 *Procedure standaardbepaling rekenen*

Gedurende de standaardbepalingsprocedure van de referentiesets rekenen is gebruikgemaakt van de Extended Angoff methode.⁸ Voor deze procedure zijn de panelleden in drie groepen verdeeld. Bij de groepsindeling is rekening gehouden met de achtergrond van panelleden. Elke groep heeft een gedeelte van het materiaal beoordeeld.

Na een individuele beoordelingsronde (ronde 1) hebben de panelleden in kleine groepen de opgaven besproken. Tijdens deze discussieronde was het mogelijk om het oordeel dat in de individuele ronde gegeven was aan te passen (ronde 2). De discussieronde is zodanig georganiseerd dat de opgaven in kleine groepen van wisselende samenstelling zijn besproken. Aangezien slechts een gedeelte van de opgaven door een panellid beoordeeld is, is de grensscore op het beoordeelde gedeelte geëxtrapoleerd naar de volledige referentieset, door middel van een omzettingstabel (gebaseerd op een vaardigheidsschaal).⁹

Naast de Extended Angoff procedure is ook een Bookmarkprocedure uitgevoerd.¹⁰ Voor deze procedure is op basis van de verzamelde data de moeilijkheid van opgaven bepaald.¹¹ De opgaven zijn vervolgens geordend van makkelijk naar moeilijk.

⁸ Zie bijlage II voor uitleg over deze methode.

⁹ De grensscore is de grens tussen een onvoldoende en een voldoende.

¹⁰ Zie bijlage II voor uitleg over deze methode.

¹¹ Moeilijkheid is hier uitgedrukt in p-waarde. Hierbij is een gewogen p-waarde gehanteerd waarbij alle beschikbare data van alle schooltypen meegenomen is.

Panelleden konden aangeven tot en met welke opgave een grenskandidaat een kans van 67% of minder had om de opgave goed te maken.

De uitkomsten van beide procedures zijn in een plenaire discussie voorgelegd aan de panelleden. Hierbij werd weergegeven wat het percentage leerlingen zou zijn dat het referentieniveau zou halen. Daarnaast is deze voorlopige cesuur vergeleken met reeds vastgestelde cesuren, zoals de tot dan toe gehanteerde rekenen 2F en 3F standaarden voor de COE's en de Rekentoets VO en de voorlopige 1F standaard zoals tot dan toe gehanteerd. Vervolgens hebben alle experts op schrift een definitief oordeel gegeven. Het gemiddelde van deze oordelen geldt vervolgens als uiteindelijk advies aan het CvTE.

7.1.3 *Procedure taal*

Gedurende de standaardbepalingsprocedure is gebruikgemaakt van de Extended Angoff methode.¹² Voor deze procedure zijn de panelleden in vier groepen verdeeld. Elke groep heeft een gedeelte van het materiaal beoordeeld.

Na een individuele beoordelingsronde (ronde 1) hebben de panelleden in kleine groepen de opgaven besproken. Tijdens deze discussieronde was het mogelijk om het oordeel dat in de individuele ronde gegeven was aan te passen (ronde 2).

Na de Extended Angoff procedure is nagegaan wat de gemiddelde cesuur van alle panelleden op de referentieset zou zijn. Voor deze cesuur is bekeken wat het percentage leerlingen zou zijn dat het referentieniveau zou halen. Daarnaast is deze voorlopige cesuur vergeleken met reeds vastgestelde cesuren, zoals de tot dan toe gehanteerde 1F-cesuur zoals vastgesteld in PPON en de tot dan toe gehanteerde cesuren Taal 2F en 3F van de COE's mbo. Vervolgens hebben alle experts op schrift een definitief oordeel gegeven. Het gemiddelde van deze oordelen geldt vervolgens als uiteindelijk advies aan het CvTE.

7.2 **Uitkomsten standaardbepaling**

7.2.1 *Uitkomsten standaardbepaling referentiesets rekenen*

Onderstaande tabellen 7 tot en met 10 geven per referentieset rekenen weer wat de uitkomsten zijn van de Extended Angoff procedure, de Bookmarkprocedure en het definitieve oordeel dat door experts is geformuleerd. Het definitieve oordeel is tot stand gekomen door na afloop van de twee inhoudelijke procedures ook de haalbaarheid van standaarden in de populatie mee te wegen (zie paragraaf 7.1). Experts hebben op basis van de empirische gegevens en hun eerder geformuleerde inhoudelijke standaard een uiteindelijk advies gedefinieerd. Dit advies is over alle experts gemiddeld en wordt in de tabel weergegeven onder 'Definitief advies'.

In de tabellen worden tevens de laagste en hoogste score uit het panel weergegeven (minimum en maximum).

Onder elke tabel is het maximaal te behalen punten weergegeven. De openbare en de niet-openbare referentieset zijn samengenomen bij de standaardbepaling. Het

¹² Zie bijlage II voor uitleg over deze methode

definitieve advies van het expertpanel betreft daarom een advies over de cesuur op niet-openbare en de openbare referentiesets samengenomen.

Tabel 7: resultaten standaardbepalingsprocedure referentiesets rekenen 1F (14 panelleden)

	Extended Angoff procedure	Bookmarkmethode	Definitief advies
Gemiddelde cesuur	86	82	84
Minimum cesuur	82	64	78
Maximum cesuur	93	96	90
Standaarddeviatie	3,8	10,0	4,2

*maximaal te behalen punten: 130

Het uiteindelijke advies van het expertpanel voor de referentiecesuur rekenen 1F op de referentiesets rekenen 1F is: 83/84.

Tabel 8: resultaten standaardbepalingsprocedure ankeresets rekenen 1S (16 panelleden)

	Extended Angoff procedure	Bookmarkmethode	Definitief advies
Gemiddelde cesuur	59	65	59
Minimum cesuur	53	56	56
Maximum cesuur	64	74	61
Standaarddeviatie	2,9	4,9	1,6

*maximaal te behalen punten: 80

Het uiteindelijke advies van het panel voor de referentiecesuur op de ankeresets rekenen 1S is: 58/59.

Tabel 9: resultaten standaardbepalingsprocedure referentiesets rekenen 2F (15 panelleden)

	Extended Angoff procedure	Bookmarkmethode	Definitief advies
Gemiddelde cesuur	76	74	75
Minimum cesuur	59	60	65
Maximum cesuur	97	99	91
Standaarddeviatie	10,6	9,9	6,7

*maximaal te behalen punten: 130

Het uiteindelijke advies van het panel voor de referentiecesuur op de referentiesets rekenen 2F is: 74/75.

Tabel 10: resultaten standaardbepalingsprocedure referentiesets rekenen 3F (17 panelleden)

	Extended Angoff procedure	Bookmarkmethode	Definitief advies
Gemiddelde cesuur	80	75	74
Minimum cesuur	65	63	66
Maximum cesuur	99	93	86
Standaarddeviatie	8,9	9,1	4,8

*maximaal te behalen punten: 130

Het uiteindelijke advies van het panel voor de referentiecesuur op de referentieset rekenen 3F is: 73/74.

7.2.2

Uitkomsten standaardbepaling referentiesets taal

In onderstaande tabellen 11,12 en 13 is telkens een cesuur na de 'inhoudelijke procedure' en een 'Definitief advies' weergegeven. De eerste cesuur betreft het gemiddelde van experts na afloop van de inhoudelijke ronde. Na deze ronde zijn experts geconfronteerd met empirische gegevens en konden zij de impact van de door hen voorgestelde cesuur bekijken. Hierover is vervolgens een plenaire discussie gevoerd. Na de plenaire discussie konden experts hun oordeel wijzigen in een definitief advies. Dit advies is over alle experts gemiddeld en wordt in de tabel weergegeven onder 'Definitief advies'.

In de tabellen worden tevens de laagste en hoogste score uit het panel weergegeven (minimum en maximum).

Onder elke tabel is het maximaal te behalen punten weergegeven. De openbare en de niet-openbare referentieset zijn samengenomen bij de standaardbepaling. Het definitieve advies van het expertpanel betreft daarom een advies over de cesuur op niet-openbare en de openbare referentiesets samengenomen.

Tabel 11: Resultaten procedure standaardbepaling referentiesets taal 1F (19 panelleden)

	Extended Angoff procedure	Definitief advies
Gemiddelde cesuur	72	71
Minimum cesuur	62	62
Maximum cesuur	82	80
Standaarddeviatie	6,7	6,4

*maximaal te behalen punten: 121

Het uiteindelijke advies van het panel voor de referentiecesuur op de referentieset taal 1F: 70/71.

Tabel 12: Resultaten procedure standaardbepaling referentiesets taal 2F (21 panelleden)

	Extended Angoff procedure	Definitief advies
Gemiddelde cesuur	91	87
Minimum cesuur	82	82
Maximum cesuur	101	93
Standaarddeviatie	5,6	3,3

*maximaal te behalen punten: 144

Het uiteindelijke advies van het panel voor de referentiecesuur op de referentieset taal 2F is: 86/87.

Tabel 13: Resultaten standaardbepaling referentiesets taal 3F (20 panelleden)

	Extended Angoff procedure	Definitief advies
Gemiddelde cesuur	94	83
Minimum cesuur	75	72
Maximum cesuur	106	93
Standaarddeviatie	10,4	6,0

*maximaal te behalen punten: 170

Het uiteindelijke advies van het panel voor de referentiecesuur op de referentieset taal 3F is: uiteindelijke advies van het panel voor de referentiecesuur op de referentieset taal 3F is: 82/83.

8 Implementatie referentiesets

In de voorgaande hoofdstukken is uiteengezet dat het CvTE expertpanels heeft ingezet voor de inhoudsvalidering en de standaardbepaling van de referentiesets. Deze expertpanels hebben aan het CvTE adviezen gegeven over de samenstelling en bruikbaarheid van de referentiesets van opgaven (inhoudelijke validering) en de hoogte van de cesuur (standaardbepaling). Het College heeft deze adviezen overgenomen bij het vaststellen van de referentiesets en de referentiecesuren. Daarmee is het hoofddoel van het project referentiesets behaald, namelijk het opleveren van landelijke referentiecesuren voor de referentieniveaus taal en rekenen.

Vervolgens kunnen de referentiecesuren overgebracht worden naar afzonderlijke toetsen en examens taal en rekenen. Daarbij wordt ofwel gebruik gemaakt van de niet-openbare referentiesets (overheid) ofwel van de openbare referentiesets (commerciële toetsontwikkelaars). Onderstaand is dit nader toegelicht.

8.1 Overbrengen referentiecesuur naar een centrale toets of examen

De standaard op de referentiesets kan door middel van de niet-openbare referentiesets overgebracht worden op centrale toetsen en examens. De opgaven uit deze referentiesets dienen daarvoor eenmalig als anker opgenomen te worden in een afname van een dergelijke toets of examen. Doordat de opgaven uit de referentiesets samen met de reguliere opgaven worden afgenomen is het mogelijk om alle opgaven op één vaardigheidsschaal te plaatsen (door middel van IRT). De standaard op de referentieset wordt omgezet in een cesuur op de vaardigheidsschaal. Deze standaard kan vervolgens gebruikt worden voor de normering van toetsen en examens door middel van de reguliere procedure 'normering met een vaardigheidsschaal'.¹³

In dit rapport wordt niet nader ingegaan op de implementatie van de referentiecesuur in elke afzonderlijke centrale toets of centraal examen. In sectorspecifieke regelingen van een centrale toets of examen is beschreven hoe de normering tot stand is gekomen en op welke wijze dit gerelateerd is aan de in het project referentiesets opgeleverde referentiecesuren.¹⁴

8.2 Overbrengen referentiecesuur naar een commercieel product

Tijdens de *Headstart ankeronderzoek referentiesets taal en rekenen* (juli 2013 – juli 2014) hebben ontwikkelaars van referentieniveaugerichte toetsen taal en rekenen de mogelijkheid gekregen om de referentiecesuur over te brengen op hun eigen toetsen. Dit kan bijvoorbeeld door opgaven uit de openbare referentiesets eenmalig als anker op te nemen in een proefafname van een commerciële toets. Doordat een anker van de opgaven uit de referentiesets samen met de reguliere toetsopgaven worden afgenomen kan de toetsontwikkelaar alle opgaven op één vaardigheidsschaal plaatsen en de cesuur overbrengen.

Door middel van de *Headstart* heeft het CvTE toetsontwikkelaars de mogelijkheid geboden de landelijke referentiecesuur over te brengen op hun eigen toets. De toetsontwikkelaar is zelf verantwoordelijk voor het uitvoeren van een ankeronderzoek en de wijze waarop de referentiecesuur op de eigen toets overgebracht wordt.

¹³ Zie voor meer informatie over 'normeren met een vaardigheidsschaal' de volgende link: http://toetswijzer.kennisnet.nl/html/normering_vaardigheidsschaal/default.shtm.

¹⁴ Zie voor VO-regelingen: <http://wetten.overheid.nl/BWBR0035058>
Zie voor mbo-regelingen: <http://wetten.overheid.nl/BWBR0035241>; <https://www.examenbladmbo.nl/onderwerp/wet-en-regelgeving-aangaande-de/2014-2015>

Op het moment van publiceren van dit rapport is de regeling in het PO nog niet opgeleverd.

In paragraaf 2.2 van dit rapport is gesproken over de ijkingsfunctie van de openbare referentiesets. De ijkingsfunctie is tijdens de Headstart benut.

8.3 Publicatie openbare referentiesets

Tijdens de Headstart hebben toetsontwikkelaars de openbare referentiesets onder geheimhouding kunnen aanvragen bij het CvTE. Op dat moment waren de openbare referentiesets dus nog niet 'openbaar' gemaakt door het CvTE.

In het najaar van 2014 heeft het CvTE deze referentiesets wel openbaar gemaakt, met als doel verantwoording afleggen over de manier waarop de landelijke referentiecesuren zijn bepaald. Op die manier is ook beantwoord aan de verantwoordingsfunctie van de openbare referentiesets, waarover in paragraaf 2.2 van dit rapport is gesproken.

Literatuur

Béguin, A.A. & Wools, S. (2014). Vertical comparison using reference sets. In R.E. Millsap, L.A. van der Ark, D.M. Bolt, & W.-C. Wang (Eds.), *New developments in quantitative psychology: Presentations from the 78th annual psychometric society meeting*. New York: Springer.

Wools, S & Béguin, A.A. (2013). *Handleiding referentieset*. Arnhem: Cito.

Wools, S & Béguin, A.A. (2014). Het toetsen van referentieniveaus. *Examens*, 11, 5-8.

BIJLAGE I Expertpanels inhoudvalidering referentiesets

Expertpanel referentiesets rekenen 1F/ankersets rekenen 1S

Panellid	Organisatie
Jan van de Craats	UvA
Tijn Bloemendaal	HCO
Ceciel Borghouts	Freudenthal Instituut
Anneke Noteboom	SLO
Jan van Kessel	Bureau ICE

Expertpanel referentiesets rekenen 2F

Panellid	Organisatie
Irma van der Laan	CvTE Vaststellingscommissie mbo Rekenen 2F
Rudy Weemaes	CvTE Vaststellingscommissie mbo Rekenen 2F
Pieter van der Zwaard	CvTE Vaststellingscommissie vmbo Rekenen 2F
Jan van de Craats	UvA
Kees Hoogland	APS
Truus Dekker	SLO
Vincent Jonker	Freudenthal Instituut

Expertpanel referentiesets rekenen 3F

Panellid	Organisatie
Rudy Weemaes	CvTE Vaststellingscommissie mbo Rekenen 3F
Frans Schoenmakers	CvTE Vaststellingscommissie mbo Rekenen 3F
Jaap Vedder	CvTE Vaststellingscommissie havo/vwo Rekenen 3F
Kees Hoogland	APS
Metha Kamminga	-
Victor Schmidt	SLO
Henk van der Kooij	Freudenthal Instituut

Expertpanel referentiesets taal 1F

Panellid	Organisatie
Saskia van Berkel	CvTE Toetswijzercommissie PO
Gerrit Staphorsius	-
Corine Ballering	CPS
Bart van der Leeuw	SLO
Henk Blok	Kohnstamm Instituut
Heleen Strating	Expertisecentrum Nederlands

Expertpanel referentiesets taal 2F

Panellid	Organisatie
Klaas Heemskerk	VLLT
Clary Ravesloot	SLO
Huib van den Bergh	UU
Willemijn Stockmann	CvTE Vaststellingscommissie COE Nederlands 2F
Marielle Paulus	CvTE Vaksectie Examen NL vmbo
Inge van Meelis	ITTA

Expertpanel referentiesets taal 3F

Panellid	Organisatie
Piet Litjens	CvTE Vaststellingscommissie COE Nederlands 3F; ITTA
Wim van Dijk	CvTE Vaksectie Examen NL havo/vwo
Clary Ravesloot	SLO
Theo Witte	RUG
Eva-Maria Ternité	CINOP
Wilma van der Westen	VLLT

BIJLAGE II Methodes voor standaardbepaling

Standaardbepaling rekenen

Bij de standaardbepaling referentiesets rekenen is een combinatie gemaakt van een Angoff procedure en een bookmarkprocedure.

Met de term 'beoordelaars' in onderstaande tekst worden bij de standaardbepaling referentiesets de experts uit bijlage III bedoeld.

- **Angoff procedure**
In de Angoff procedure geven beoordelaars aan welk percentage van de grenskandidaten de dichotome opgaven (meerkeuzeopgaven en enkelvoudig open vragen) goed maakt.
- **Bookmark procedure**
In de bookmarkprocedure krijgen de beoordelaars de opgaven voorgelegd op volgorde van moeilijkheid. De beoordelaars moeten aangeven vanaf welke opgave de grenskandidaat een kans heeft van minder dan 67% om de opgaven goed te maken. (Vergelijkbaar kan gesteld worden dat de beoordelaars moeten aangeven vanaf welke opgaven minder dan 67% van de grenskandidaten de opgave goed maakt).

Standaardbepaling taal

Bij de standaardbepaling referentiesets taal is een Extended Angoff procedure toegepast.

- **Extended Angoff**
In de Extended Angoff procedure geven beoordelaars aan hoeveel punten grenskandidaten halen op polytome opgaven of op een combinatie van opgaven. In dit specifieke geval worden alle opgaven bij één tekst samengenomen en geven beoordelaars aan hoeveel punten een grenskandidaat op al deze opgaven samen zal halen.

BIJLAGE III Expertpanels standaardbepaling referentiesets

Expertpanel referentiesets rekenen 1F

Panellid	Organisatie
Ada André	CBS Bloemenhof
Anke Arts	Vrije School Zutphen
Ineke Bette	Marnix College
Ceciel Borghouts	Freudenthal Instituut
Tijn Bloemendaal	HCO
Job Burema	Corderius
Jan van de Craats	UvA
Mieke van Groenestijn	Lector Gecijferdheid, HU
Peter Jakobs	Koningin Julianaschool Bameveld
Jan Janssen	Cito
Jan van Kessel	Bureau ICE
Michèle Kraal-Hartemink	Basisschool Ichthus
Anneke Noteboom	SLO
Rianne Timmermans-Linnenkoper	O.B.S. De Klimop

Expertpanel ankerets rekenen 1S

Panellid	Organisatie
Ada André	CBS Bloemenhof
Ineke Bette	Marnix College
Ceciel Borghouts	Freudenthal Instituut
Job Burema	Corderius
Jan van de Craats	UvA
Wouter Fondse	Christelijk Lyceum Veenendaal
Wim Goossens	Basisschool St. Willibrordus
Just Heinis	De Hoge Berg
Peter Jakobs	Kon. Julianaschool- Barneveld
Jan Janssen	Cito
Jan van Kessel	Bureau ICE
Michèle Kraal-Hartemink	Basisschool Ichthus
Bert Kraai	Vrije School Zutphen
Anneke Noteboom	SLO
Rianne Timmermans-Linnenkoper	O.B.S. De Klimop
Marjo de Widt	CL Delft

Expertpanel referentiesets rekenen 2F

Panellid	Organisatie
Ans van der Ark	Maris College
Martin van Bentheim	Clusius College
Jan van de Craats	UvA
Truus Dekker	SLO
Jan van Dijk	Landstede
Marcel Engelen	Martinusschool
Jolanda Gerritse	Koningin Julianaschool
Vincent Jonker	Freudenthal Instituut
Gert de Kleuver	Ichthus College
Irma van der Laan	Vaststellingscommissie mbo Rekenen 2F
Maurice van Leeuwen	Graafschap College
Guus Swerts	Leeuwenborgh Opleidingen
Rudy Weemaes	Vaststellingscommissie mbo Rekenen 2F
Rob Wennink	College de Brink
Pieter van der Zwaard	Vaststellingscommissie Rekenen VO 2F

Expertpanel referentiesets rekenen 3F

Panellid	Organisatie
Gery Gorter	Instituut Archimedes van Hogeschool Utrecht
Teun Hommersom	Hoornbeek
Ronald Keijzer	Hogeschool iPabo
Wouter van der Knaap	CL Veenendaal
Jan Los	Kaj Munkcollege
Roelf Meijer	OSG Erasmus
Gerrit Ros	Aloysiuscollege
Victor Schmidt	SLO
Frans Schoenmakers	Vaststellingscommissie mbo Rekenen 3F
Jantien Stam	Instituut Archimedes van Hogeschool Utrecht
Jaap Vedder	Vaststellingscommissie havo/vwo Rekenen 3F
Fokko de Visser	Trinitas College
Fred Walthuis	ROC Amsterdam
Joris Westerbeek	Graafschap college
Rudy Weemaes	Vaststellingscommissie mbo Rekenen 3F
Harco Weemink	Cito
Bert Zwaneveld	OU

Expertpanel referentiesets taal 1F

Panellid	Organisatie
Annemiek Ammerlaan	Docente primair onderwijs
Corine Ballering	CPS
Saskia van Berkel	Cito
Ineke Bette	Marnix College
Henk Blok	Kohnstamm Instituut
Marjolein Bos	CSG Liudger Drachten
Mariska van Dijk	CBS De Oranjerie
Hiske Feenstra	Cito
Sandra Hellemons	Nutsbasisschool Teteringen
Roos Hilkes	NBS De Hoogakker
Pleuni Hooft van Huijsduijnen	IJburg College
Rianne Timmermans-Linnenkoper	O.B.S. De Klimop
John de Jong	Sjaloomschool Wierden
Corina Korthout	Docente primair onderwijs
Bart van der Leeuw	SLO
Arthur Pruijssers	CBS de Rank
Gerrit Staphorsius	-
Heleen Strating	Expertisecentrum Nederlands
Betty Verdouw	Augustinuscollege Diamantlaan afd. OPDC
Daphne Zijlstra	BS De Driemaster Alkmaar (SAKS)

Expertpanel referentiesets taal 2F

Panellid	Organisatie
Jolanda de Baaij	ROC Nijmegen
Huib van den Bergh	UU
Florien Bulens-Volders	Edison College
Cecile Doeser	mbo Prinsentuin College Breda
Jos Elbersen	Basisschool Willibrordus
Hiske Feenstra	Cito
Klaas Heemskerk	VLLT
Sonja Heynsdijk	Beroepscollege Parkstad Limburg
Stephan Iske	Zadkine
Bregje Kaars Sijpesteijn	ITTA
Nanke Kraehe-Vriezen	ROC Aventus
Roelof Liemburg	O.B.S. It Skriuwboerd
Dolf Lok	Pieter Zandt Scholengemeenschap
Patty Meuling	SG Spieringshoek Schiedam
Koos Monshouwer	CBS Bloemenhof
Marielle Paulus	Vaststellingscommissie NL vmbo
Clary Ravesloot	SLO
Willemijn Stockmann	Vaststellingscommissie COE 2F
Herman Stuijver	Vereinigde Scholen Alberdingk Thijm
Leanne Timmermans	ROC Tilburg

Expertpanel referentiesets taal 3F

Panellid	Organisatie
Corina Breukink-Prosper	Hogeschool Utrecht
Wim van Dijk	Vaststellingscommissie NL havo/vwo
Jean Luc van Engelen	Jacob Roelandslyceum Boxtel
Kooske Franken	Albeda College
Hans Goosen	FLOT
Françoise Huijssoon	ROC Mondriaan
Marleen Janssen-Boesten	ROC Gilde opleidingen
Marit de Jong	Hogeschool van Amsterdam
Alex van de Kerkhof	Cito
Piet Litjens	Vaststellingscommissie COE Nederlands 3F; ITTA
Aat Nieuwenhuijse	Driestar College Gouda
Han Niesing	ROC Aventus
Clary Ravesloot	SLO
Eva-Maria Ternité	CINOP
Polly den Tenter	Mondriaan College
Coen van 't Veer	SG Spieringshoek
Jessica de Weerd	ROC Aventus
Wilma van der Westen	VLLT
Vincent Westerwoudt	Fontys Lerarenopleiding
Klaas van der Wolf	CSG Reggesteyn Nijverdal

