



College voor Toetsen en Examens

RAPPORTAGE ANKERSETS KIJK- EN LUISTERVAARDIGHEID EN STANDAARDBEPALING MONDELINGE TAALVAARDIGHEID 1F EN 2F IN HET PRIMAIR ONDERWIJS

VERANTWOORDING PROJECT

December 2019

Inhoud

Voorwoord	3
Samenvatting	4
1 Inleiding	6
2 Doel en uitleg ankersets kijk- en luistervaardigheid	7
2.1 Wat is een ankerzet kijk- en luistervaardigheid?	7
2.2 Verschil ankerzet en referentieset	7
2.3 Wat is een ankerzet niet?	7
3 Inhoudelijke validering ankersets kijk- en luistervaardigheid	8
3.1 Opdracht voor experts	8
3.2 Samenstelling expertpanel	8
3.3 Proces inhoudelijke validering door experts	8
3.4 Afnames en aantal waarnemingen	11
4 Samenstelling ankersets kijk- en luistervaardigheid	12
4.1 Samenstelling ankersets kijk- en luistervaardigheid 1F	13
4.2 Samenstelling ankersets kijk- en luistervaardigheid 2F	14
5 Standaardebepaling ankersets	15
5.1 Procedure standaardbepaling	15
5.2 Uitkomsten standaardbepaling	17
6 Implementatie ankersets	19
6.1 Overbrengen referentiecases naar een commercieel product	19
6.2 Publicatie openbare ankersets kijk- en luistervaardigheid	19
7 Standaardebepalingen peilingsonderzoeken mondelinge taalvaardigheid	20
7.1 Peilingsonderzoek mondelinge taalvaardigheid in het reguliere po	20
7.2 Peilingsonderzoek mondelinge taalvaardigheid in het so en sbo	25
8 Literatuur	30
9 Bijlage - overzicht bronnen	31

Voorwoord

Voor u ligt de rapportage van het College voor Toetsen en Examens (het CvTE) over het project ankersets kijk- en luistervaardigheid 1F en 2F.¹

Deze rapportage heeft als doel een breed publiek te informeren over de manier waarop de ankersets kijk- en luistervaardigheid tot stand gekomen zijn. De rapportage beschrijft de belangrijkste fasen in de ontwikkeling van deze ankersets. Bovendien geeft het informatie over de concrete producten die dit project heeft opgeleverd.

Op de bronnen en bijbehorende opgaven uit de ankersets berusten auteursrechten. Deze rapportage bevat daarom géén bronnen of opgaven uit de ankersets kijk- en luistervaardigheid 1F en 2F. Onder voorwaarden zijn deze bronnen en opgaven wel opvraagbaar voor geïnteresseerden via www.referentiesets.nl.

De voorliggende rapportage is in samenwerking met Stichting Cito tot stand gekomen. Het project ankersets kijk- en luistervaardigheid 1F en 2F ligt in het verlengde van het project referentiesets Nederlandse taal (lezen) en rekenen en het project ankersets taalverzorging. Het project referentiesets en het bijbehorende onderzoek zijn oorspronkelijk beschreven in Wools & Béguin (2013, 2014) en Béguin & Wools (2014).

¹ Oorspronkelijk werd er gesproken over de ankersets luistervaardigheid, conform het referentiekader taal waarin ook gesproken wordt over 'luistervaardigheid'. Later is er echter overgegaan op 'ankersets kijk- en luistervaardigheid', mede op advies van het expertpanel. In het onderwijs is er veeleer sprake van kijk- en luistervaardigheid in plaats van enkel luisteren naar audio. In deze verantwoording wordt om die reden ook gesproken over ankersets kijk- en luistervaardigheid.

Samenvatting

In deze rapportage wordt beschreven hoe de ankersets kijk- en luistervaardigheid en de standaard mondelinge taalvaardigheid 1F en 2F tot stand gekomen zijn.

Achtergrond

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. Het is belangrijk dat in toetsen en examens de eisen van de referentieniveaus voor alle leerlingen gelijk zijn, ongeacht het schooltype dat zij volgen of de toets die zij maken. Dat kan door middel van ankersets of referentiesets en bijbehorende (landelijk vastgestelde) referentiecesuren. In oktober 2014 zijn daarom referentiesets Nederlandse taal (lezen) en rekenen opgeleverd en in september 2016 de ankersets taalverzorging. Het project ankersets kijk- en luistervaardigheid 1F en 2F en standaardbepaling mondelinge taalvaardigheid 1F en 2F liggen in het verlengde hiervan.

Doel en opbrengsten project ankersets kijk- en luistervaardigheid po

Een ankerzet bestaat uit een set van opgaven die het inhoudelijk beschreven kennisdomein van het referentieniveau zo goed mogelijk representeert. De opgaven hebben hun oorsprong in één onderwijstype, in dit geval het primair onderwijs (PO). Alle opgaven zijn ook afgenomen in het PO en er is een prestatie standaard (= referentiecesuur) vastgesteld voor de ankerzet. Het project ankersets kijk- en luistervaardigheid had als doel het ontwikkelen van landelijke referentiecesuren voor deze vaardigheid. Er zijn twee ankersets ontwikkeld: voor 1F en 2F. Deze ankersets zijn alleen van toepassing op het PO.

Om de inhoudelijke kwaliteit van de ankerzet te borgen is er een onafhankelijk expertpanel aangesteld. Het panel had als taak om via een zorgvuldige en intensieve procedure de inhoud van de ankerzet kijk- en luistervaardigheid te valideren. Vervolgens heeft een standaardbepalingsprocedure, in dit geval de Data Driven Direct Consensus methode (3DC-methode), met experts geleid tot een advies voor de prestatie standaard. Het referentiekader taal (onderdeel luisteren) vormde de basis voor de standaardbepaling. Het College voor Toetsen en Examens heeft het advies overgenomen, waarmee de standaard voor kijk- en luistervaardigheid 1F is gelegd op 22. Dat betekent dat het definitieve advies voor de cesuur 1F bij 22 van de 31 opgaven geplaatst wordt. Zowel bij niveau 1F als bij niveau 2F is er een vergelijking gemaakt tussen de adviezen van het expertpanel en de indicatieve cesuren die eerder geplaatst waren op de luistertoets die gebruikt is in de peilingsonderzoeken mondelinge taalvaardigheid. Bij 1F kwam het advies over de cesuur voor de ankerzet overeen met de indicatieve cesuur van de luistertoets uit het peilingsonderzoek. Bij 2F kwamen het advies en de indicatieve cesuur van de luistertoets niet volledig overeen. De standaard voor kijk- en luistervaardigheid 2F is daarom definitief op 25 van de 30 geplaatst: het gemiddelde tussen het voorlopige advies van het panel over de ankerzet enerzijds en de luistertoets uit het peilingsonderzoek anderzijds.

Standaardbepalingen peilingsonderzoeken luistervaardigheid en mondelinge taalvaardigheid regulier po

Ten behoeve van een peilingsonderzoek in 2017 van de Inspectie naar de beheersing van de mondelinge taalvaardigheid van leerlingen aan het einde van het basisonderwijs, zijn beheersingscesuren 1F en 2F voor de betreffende taken uit het peilingsonderzoek vastgesteld. Via dezelfde standaardbepalingsprocedure als bij de

ankersets luistervaardigheid is met een expertpanel tot een adviescesuur gekomen. Het College voor Toetsen en Examens heeft vervolgens de cesuur vastgesteld.

Voor luistervaardigheid 1F is het advies overgenomen en is de definitieve cesuur gesteld op 15 van de 30 opgaven. Bij luistervaardigheid 2F is na een vergelijking met de cesuren op de ankerzet kijk- en luistervaardigheid 2F, de cesuur gelegd op 23 van de 30 opgaven goed.

Bij de standaardbepaling spreekvaardigheid is de taakgerichte standaard 1F geplaatst op 8 van de 17 vragen goed en bij 2F op 12 van de 17 vragen goed. Bij een score van 12 punten of meer beheerst een leerling de taak dan op niveau 2F. Er wordt bewust gesproken over beheersing van de taak: aangezien er in het peilingsonderzoek slechts één spreektaak afgenomen kon worden. Generalisatie naar beheersing van spreekvaardigheid in het algemeen is daardoor niet mogelijk. Voor gesprekken is de taakgerichte standaard bij 1F op 5 van de 10 scorepunten geplaatst en voor 2F op 8 van de 10 scorepunten.

Standaardbepalingen peilingsonderzoeken luistervaardigheid en mondelinge taalvaardigheid speciaal onderwijs en speciaal basisonderwijs

Naast het peilingsonderzoek mondelinge taalvaardigheid in het reguliere basisonderwijs, heeft de Inspectie van het onderwijs in 2018 een onderzoek laten uitvoeren naar de beheersing van de mondelinge taalvaardigheid van leerlingen in het speciaal onderwijs en het speciaal basisonderwijs die uitstromen naar het voortgezet (speciaal) onderwijs. Hiervoor zijn andere taken gehanteerd dan in het peilingsonderzoek in het reguliere basisonderwijs, aangezien de daar gebruikte taken niet geschikt waren voor deze doelgroep. De procedure voor het bepalen van de cesuren op deze taken was vrijwel gelijk aan de procedure die beschreven is voor het reguliere basisonderwijs.

Ook hier gaat het om taakgerichte standaarden, waar bij Spreken de lat is gelegd op 11 (taak 1) en 12 (taak 2) van de 17 scorepunten voor 1F. Voor 2F is de standaard bij beide spreektaken geplaatst op 16 van de 17. De taakgerichte standaard 1F Gesprekken is geplaatst bij 8 van de 10 scorepunten en voor 2F Gesprekken bij 9 van de 10 scorepunten.

De taken uit het so/sbo blijken eenvoudiger en dat maakt dat leerlingen die net 1F behalen, voor deze taken een hogere score moeten halen dan voor de spreektaak uit het reguliere basisonderwijs. De so/sbo-taken lijken inhoudelijk minder geschikt om niveau 2F te kunnen meten. Dit verklaart vermoedelijk ook de hoge cesuren.

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 Nederlandse 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. Voor Nederlandse taal zijn er vier F-niveaus, namelijk 1F, 2F 3F en 4F. De niveaus geven een opklimmende moeilijkheidsgraad in basiskennis en -vaardigheden aan. Elk fundamenteel niveau omvat het voorgaande niveau. Bij Taal is het streefniveau voor leerlingen die meer aankunnen het eerstvolgende F-niveau. Niveau 2F in taal is gedefinieerd als het minimaal benodigde niveau om in de samenleving te kunnen functioneren.

Het is belangrijk dat in toetsen en examens de eisen van de referentieniveaus voor alle leerlingen gelijk zijn, ongeacht het schooltype dat zij volgen of de toets die zij maken. Dat kan door middel van ankersets of referentiesets² en bijbehorende (landelijk vastgestelde) referentiecesuren. In oktober 2014 zijn daarom referentiesets Nederlandse taal (lezen) en rekenen opgeleverd en in september 2016 de ankersets taalverzorging.

In 2016 heeft het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW) het College voor Toetsen en Examens (CvTE) en Stichting Cito (vanaf nu: Cito) de opdracht gegeven ankersets kijk- en luistervaardigheid 1F en 2F te ontwikkelen. Deze ankersets zijn bedoeld voor het primair onderwijs (PO). Het project ankersets kijk- en luistervaardigheid ligt qua opzet en uitvoering in lijn met het project referentiesets Nederlands taal (lezen) en rekenen en de ankersets taalverzorging.

In de navolgende hoofdstukken wordt uitgelegd wat het project ankersets kijk- en luistervaardigheid inhoudt en op welke wijze het project is uitgevoerd.

² Een referentieset bedient zowel het po als het vo. Een ankerset bedient slechts één onderwijstype, in dit geval het po. Zie ook paragraaf 2.2.

2 Doel en uitleg ankersets kijk- en luistervaardigheid

2.1 Wat is een ankerzet kijk- en luistervaardigheid?

Een ankerzet kijk- en luistervaardigheid bestaat uit een set van opgaven die het inhoudelijk beschreven kennisdomein van het referentieniveau luisteren zo goed mogelijk representeert. De opgaven hebben hun oorsprong in één onderwijstype, in dit geval het PO. Alle opgaven zijn ook afgenomen in het PO. Er is een prestatiestandaard (= referentiecesuur) vastgesteld voor de ankerzet.

In het project ankersets kijk- en luistervaardigheid is één type ankerzet ontwikkeld: een openbare ankerzet. Deze ankerzet heeft een ijkingsfunctie en een verantwoordingsfunctie.

Ijkingsfunctie

De openbare ankersets kijk- en luistervaardigheid zijn voorzien van prestatiestandaarden 1F en 2F. De ankersets zijn gebruikt in de peilingsonderzoeken in het reguliere po en het so/sbo (Inspectie van het onderwijs, maart 2019). Daarnaast zijn commerciële toetsontwikkelaars, uitgeverijen en onderzoekers tijdens een *headstart* in staat gesteld met deze ankersets de landelijke prestatiestandaard over te brengen op hun eigen producten kijk- en luistervaardigheid. Dit gebeurde door middel van ankeronderzoek.

Verantwoordingsfunctie

Met behulp van de openbare ankersets kan het CvTE aan het onderwijsveld duidelijk maken hoe het onderdeel kijk- en luistervaardigheid op niveau 1F en 2F geoperationaliseerd wordt en welke prestatiestandaarden hieraan verbonden zijn. In tegenstelling tot de referentiesets en de ankerzet taalverzorging zijn er geen niet-openbare ankersets ontwikkeld, aangezien de Centrale Eindtoets PO, die onder verantwoordelijkheid van de overheid valt, geen onderdeel (kijk- en) luistervaardigheid bevat.

2.2 Verschil ankerzet en referentieset

Het hoofddoel van het project referentiesets Nederlandse taal en rekenen is zorgen voor vergelijkbare prestatie-eisen binnen verschillende schooltypen waar eenzelfde referentieniveau van toepassing is. Gegeven dit doel is er voor de doorlopende toetslijn taal en rekenen vanuit de overheid geen aanleiding geweest om een referentieset kijk- en luistervaardigheid 1F en 2F voor zowel po als vo te maken.

Er was vanuit de peilingsonderzoeken in het po wel behoefte aan een ankerzet voorzien van gestandaardiseerde cesuren voor kijk- en luistervaardigheid. Daarnaast bestond bij verschillende toetsaanbieders de behoefte hun kijk- en luistertoetsen te ankeren aan een ankerzet. In het vo leefde die behoefte niet. Een ankerzet met referentiecesuur die zich beperkt tot enkel po, bleek daarom voldoende.

2.3 Wat is een ankerzet niet?

Een ankerzet is expliciet niet bedoeld als voorbeeldmateriaal ter voorbereiding op toetsen in het PO. Voorbeeldopgaven zijn hiervoor beter geschikt dan een ankerzet. De ankerzet laat namelijk niet zien welke typen opgaven er in een specifieke toets voorkomen. Zij laat daarnaast ook niet zien hoe de leerstof zich verhoudt tot een toets. De ankersets zijn alléén geschikt om het gehele leerstofdomein te concretiseren en de prestatiestandaard vast te stellen.

3 Inhoudelijke validering ankersets kijk- en luistervaardigheid

3.1 Opdracht voor experts

De ankerset kijk- en luistervaardigheid representeert het onderdeel luisteren van het referentiekader taal. Het is om die reden van belang dat de inhoud van de ankerset het referentiekader goed dekt. Om de inhoudelijke kwaliteit van de ankerset te borgen is er een onafhankelijk expertpanel aangesteld. Het panel had als taak om de door Cito ontwikkelde inhoud van de ankerset kijk- en luistervaardigheid te valideren.

3.2 Samenstelling expertpanel

Het CvTE was verantwoordelijk voor de samenstelling van het expertpanel. Bij de samenstelling is rekening gehouden met de verschillende visies en stromingen uit het onderwijsveld met het oog op het draagvlak voor en de kwaliteit van de ankersets. In onderstaand overzicht is weergegeven welke personen in het expertpanel zitting hebben genomen.

Panellid	Organisatie
Tjalling Brouwer	Impact Educatief
Resi Damhuis	Taaldenkgesprekken.nu
Marieke Melis / Daisy Mertens ¹	Leerkracht po De vuurvogel
Femke Scheltinga	ITTA
Hans Schepers	Oud-leerkracht/directeur PO
Heleen Strating	Expertisecentrum Nederlands

¹Marieke en Daisy wisselden elkaar af.

3.3 Proces inhoudelijke validering door experts

Via het voorbereiden en bijwonen van een aantal beoordelingsbijeenkomsten hebben de experts een oordeel gegeven over de samenstelling van de ankersets. Hieronder wordt het proces van inhoudelijke validering beschreven.

Bijeenkomst 1 en 2: introductie en toetsmatrijzen (13 december 2016 en 17 januari 2017)

Tijdens de eerste bijeenkomst heeft Cito aan de experts de toetsmatrijzen voor de ankerset voorgelegd. Een toetsmatrijs geeft weer hoeveel en welk type opgaven een ankerset moet bevatten. De toetsmatrijs van de ankerset is gebaseerd op de domeinindelingen van het betreffende referentieniveau luistervaardigheid. Het is van belang dat de geselecteerde opgaven het inhoudsdomein zo goed mogelijk representeren. Op basis van het commentaar van het expertpanel heeft Cito de toetsmatrijzen aangepast.

Bij de toetsmatrijs van de ankerset kijk- en luistervaardigheid zijn enkele opmerkingen te maken:

- Hoewel er in het referentiekader taal gesproken wordt over luistervaardigheid, is luistervaardigheid voor kinderen in groep 8 veelal kijk- en luistervaardigheid. Veel communicatie gaat gepaard met beeld. Om die reden zijn er in de ankerset op advies van het expertpanel zowel videofragmenten als

audiofragmenten opgenomen, waarbij de uiteindelijke verhouding ongeveer 2/3 deel videomateriaal en 1/3 deel audiomateriaal bedraagt.

- In de toetsmatrijsen is vastgelegd dat in zowel de ankerset 1F als de ankerset 2F de drie luistertaken die in het referentiekader luisteren voorkomen, moeten terugkomen:
 - luisteren naar instructies
 - luisteren als lid van live publiek
 - luisteren naar radio, tv en internet

Luisteren als lid van live publiek is in een toetssetting lastig realiseerbaar.

Op advies van het Expertpanel is het onderdeel wel opgenomen in de toetsmatrijs vanwege het streven om het referentiekader zo volledig mogelijk te dekken. De live situatie wordt daarbij nagebootst, bijvoorbeeld door kijk- en luisterfragmenten in de ankerset op te nemen waarbij een 'live publiek' in de studio aanwezig is.

- Op uitdrukkelijk advies van het Expertpanel zijn er bij zowel 1F als 2F fictionele kijk- en luisterfragmenten in de ankerset opgenomen.
- Het referentiekader onderscheidt vier kenmerken van taakuitvoering: begrijpen, interpreteren, evalueren en samenvatten. Op advies van het Expertpanel zijn er geen evaluerende items opgenomen in de ankerset. Evalueren is een belangrijke vaardigheid in het onderwijs, maar is moeilijk operationaliseerbaar in een (gesloten) toetsvorm. Samenvatten is wel opgenomen in de ankerset; het betreft hierbij echter opgaven waarbij leerlingen moeten aangeven wat de beste samenvatting van een kijk-/luisterfragment is. Leerlingen wordt niet gevraagd om actief een fragment samen te vatten.
- Het referentiekader geeft aan dat leerlingen op niveau 2F ook moeten kunnen luisteren naar langere teksten met een luisterduur tot ongeveer 20 minuten. Een dergelijke tekstlengte is uit praktische overwegingen in een ankerset niet haalbaar. In de ankerset worden fragmenten opgenomen met een maximum duur van 5 minuten.

Bijeenkomsten 3 en 4: validering ankersets (juni 2017 en 9 januari 2018)

Op basis van de toetsmatrijs is door Cito een selectie gemaakt van geschikte kijk- en luisterfragmenten voor de ankersets. Bij deze kijk- en luisterfragmenten zijn meerkeuzeopgaven ontwikkeld. In een expertbijeenkomst op 27 juni is het materiaal besproken met het expertpanel. In voorbereiding op deze bijeenkomsten hebben experts vooraf 39 kijk-/luisterfragmenten en de bijna 200 bijbehorende opgaven beoordeeld op hun geschiktheid voor opname in de ankersets. Aandachtspunten daarbij waren onder andere de geschiktheid van de opgaven voor het beoogde referentieniveau en de inhoudelijke uitwerking van de opgave.

Na de validering door de experts zijn de overgebleven 35 kijk-/luisterfragmenten met de bijbehorende 173 opgaven in het najaar van 2018 geproeftoetst bij ruim 1400 leerlingen van 63 basisscholen. In een tweede bijeenkomst op 9 januari 2018 is door het expertpanel op basis van de inhoud en de gegevens van de proeftoets een definitieve selectie gemaakt voor de ankersets. De uiteindelijke samenstelling van de ankersets is hieronder weergegeven.

Ankerzet 1F			
	kijk/luisterfragmenten	opgaven per fragment	totaal aantal opgaven reeks
Luisteren naar instructies	1F_hersenen	4	12
	1F_ridders	5	
	1F_voicemailbericht	3	
Luisteren als lid live publiek	1F_koningscobra	3	9
	1F_Vriens	3	
	1F_Joris	3	
Luisteren naar radio/tv/internet	1F_kom_in_actie	3	10
	1F_zebrastrepen	3	
	1F_hik	4	
Totaal aantal opgaven 1F			31

Ankerzet 2F			
	kijk/luisterfragmenten	opgaven per fragment	totaal aantal opgaven reeks
Luisteren naar instructies	2F_dynamo	4	11
	2F_papierscheppen	4	
	2F_yogaboom	3	
Luisteren als lid live publiek	2F_skatekeet	3	11
	2F_melkmeisje	4	
	2F_eekhoorn	4	
Luisteren naar radio/tv/internet	2F_bisjpalen	3	8
	2F_kindermanifest	2	
	2F_pindakaas	3	
Totaal aantal opgaven 2F			30

3.4 Afnames en aantal waarnemingen

Zoals in paragraaf 3.3. weergegeven is, zijn de opgaven uit de ankerzet afgenomen in groep 8 van het basisonderwijs om te bepalen hoe leerlingen scoren op de opgaven en zo moeilijkheidsgraad van de opgaven te bepalen. Dit onderzoek vond plaats in het najaar van 2017. De ankerzet is daarna in het voorjaar van 2018 nogmaals afgenomen bij 228 leerlingen tezamen met de luistertoets uit het peilingsonderzoek mondelinge taalvaardigheid einde basisonderwijs (zie paragraaf 8.1.1.).

Omdat de ankersets kijk- en luistervaardigheid alleen bedoeld zijn voor het ijken van metingen aan het eind van het primair onderwijs, heeft geen onderzoek plaatsgevonden in de sectoren van het vo of het mbo.

Aantal waarnemingen

De onderstaande tabel geeft weer hoeveel waarnemingen gemiddeld per opgave verzameld zijn. De gewenste aantallen waarnemingen 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'.

Gemiddeld aantal waarnemingen per item	proefafname najaar 2017
1F	348
2F	377

4 Samenstelling ankersets kijk- en luistervaardigheid

Via de ankerzet moet de referentiecesuur van het referentieniveau luistervaardigheid overgebracht kunnen worden op verschillende toetsen kijk- en luistervaardigheid. De samenstelling van de ankersets moet passend hiervoor zijn. De openbare ankersets laten een diversiteit aan opgaven en vraagvormen zien. Een openbare ankerzet dient immers bruikbaar te zijn voor andere toetsaanbieders ten behoeve van hun ankeronderzoek.

Hieronder is de definitieve samenstelling van de ankersets beschreven voor 1F en 2F. Deze samenstelling is tot stand gekomen na afloop van de proeftoets die in het najaar van 2017 heeft plaatsgevonden (zie hoofdstuk 4: Afnames).

4.1 Samenstelling ankersets kijk- en luistervaardigheid 1F

	Ankerzet 1F			
		begrijpen	interpreteren	samenvatten
Luisteren naar instructies				
	<i>video</i>			
	1F_hersenen	2	2	
	1F_ridders	3	2	
	<i>audio</i>			
	1F_voicemailbericht	3		
Luisteren als lid live publiek				
	<i>video</i>			
	1F_koningscobra	2		1
	1F_Vriens	1	2	
	<i>audio</i>			
	1F_Joris	1	2	
Luisteren naar radio, tv en internet				
	<i>video</i>			
	1F_kom_in_actie	2	1	
	1F_zebrastrepen	1	2	
	<i>audio</i>			
	1F_hik	2	1	1
	Totaal	17	12	2

4.2 Samenstelling ankersets kijk- en luistervaardigheid 2F

	Ankerset 2F			
		begrijpen	interpreteren	Samenvatten
Luisteren naar instructies				
	<i>video</i>			
	2F_dynamo	1	2	1
	2F_papierscheppen	2	1	1
	<i>audio</i>			
	2F_yogaboom	2		1
Luisteren als lid live publiek				
	<i>video</i>			
	2F_skatekeet	3		
	2F_melkmeisje	1	2	1
	<i>audio</i>			
	2F_eekhoorn	2	2	
Luisteren naar radio, tv en internet				
	<i>video</i>			
	2F_bisjpalen	1	2	
	2F_kindermanifest	1		1
	<i>audio</i>			
	2F_pindakaas	2	1	
	Totaal	15	10	5

5 Standaardbepaling ankersets

Voor de ankerzet kijk- en luistervaardigheid is een prestatie standaard bepaald door middel van een procedure met experts. Experts bogen zich over de vraag hoeveel opgaven van de ankerzet een leerling goed moet hebben om het referentieniveau te behalen. De procedure waarmee dat gebeurt is een standaardbepaling. Het referentiekader taal (onderdeel luisteren) vormde de basis voor de standaardbepaling.

Hierna wordt eerst de gevolgde procedure bij de standaardbepaling uiteengezet. Daarna volgen de uitkomsten van de standaardbepaling.

5.1 Procedure standaardbepaling

Standaardbepaling met experts

De opdracht die het panel van experts had, was het geven van een advies aan het CvTE over de referentieproeven voor de ankersets kijk- en luistervaardigheid. Het CvTE was verantwoordelijk voor de werving van het expertpanel. De leden zijn geselecteerd vanwege hun (les)ervaring in het PO en hun inhoudelijke expertise.

In totaal hebben 13 personen deelgenomen aan de standaardbepaling op 1 juni 2018. 4 van hen zijn taalexperts die ook zitting hebben in het expertpanel dat betrokken was bij de inhoudelijke validering van de ankerzet luisteren. De overige 9 personen zijn werkzaam als leerkracht of schooldirecteur in het basisonderwijs, van wie 7 personen werkzaam zijn het reguliere basisonderwijs en 2 personen in het so/sbo.

Onderstaand overzicht geeft weer welke experts hebben deelgenomen aan de standaardbepaling.

Naam	Organisatie
Hans Schepers	Oud-leerkracht/directeur PO
Tjalling Brouwer	Impact Educatief
Irene Neelis	OG Primus
Dorret de Leeuw van Weenen-Poel	Julianaschool
Janneke Tambach	Zaan Primair
Annemarie Heerink	Konot
Roelof Liemburg	OBS It Skriuwboerd
Rachid Belhocein	Michaelschool / De Baander
Loukili Hanane	Al Iman
Resi Damhuis	Taaldenkgesprekken.nu
Wieteke Klijnstra-Idema	CBS De Wendakker
Wim Goossens	Basisschool St. Willibrordus
Heleen Strating	Expertisecentrum Nederlands

Beschrijving procedure

De opzet van de standaardbepaling sluit aan bij de procedure die gehanteerd is bij eerdere standaardbepalingen voor de ankersets taalverzorging en bij de eerdere standaardbepaling voor het onderdeel luisteren uit het peilingsonderzoek mondelinge taalvaardigheid (zie paragraaf 8.1).

De eisen die het referentiekader taal aan het onderdeel luisteren stelt, vormden het uitgangspunt van de standaardbepalingen. Aan het begin van de bijeenkomst is daarom door de panelleden uitgebreid stilgestaan bij de beschrijvingen die het referentiekader geeft voor luistervaardigheid 1F en 2F. De panelleden bekeken deze beschrijvingen eerst individueel en noteerden voor zichzelf de hoofdpunten. Deze hoofdzaken werden vervolgens plenair besproken, waarbij ingegaan werd op karakteristieke onderdelen van kijk- en luistervaardigheid op niveau 1F en niveau 2F. Zo werd ervoor gezorgd dat de onderscheidende kenmerken van niveau 1F en 2F voor alle panelleden duidelijk waren en dat ze vanuit hetzelfde inhoudelijke vertrekpunt aan de standaardbepalingen konden beginnen.

Bij de standaardbepaling is gebruik gemaakt van de Data Driven Direct Consensus methode (3DC-methode). De 3DC-methode deelt een toets op in een aantal clusters, in dit geval 3 clusters variërend van 8 tot 12 opgaven per cluster. De clusters correspondeerden met de verschillende taken die het referentiekader en de ankerset onderscheiden: luisteren naar instructies, luisteren naar live publiek en luisteren naar radio, tv en internet. Aan de panelleden werd bij ieder cluster opgaven de vraag gesteld: 'Hoeveel opgaven van dit cluster zou een leerling goed moeten beantwoorden als zijn/haar vaardigheid zich precies op de grens wel of geen 1F (of 2F) bevindt?'

In de standaardbepaling is voor het referentieniveau 1F een andere verzameling opgaven gebruikt dan voor de standaardbepaling van 2F, omdat er een aparte ankerset ontwikkeld is voor 1F en voor 2F. Door middel van itemresponstheorie (IRT) kunnen deze grenzen op een en dezelfde vaardigheidsschaal vergeleken worden.³

De bepaling van de standaard 1F en de standaard 2F gebeurde telkens in twee rondes. In de eerste ronde bepaalde ieder panellid individueel de score van de zogenaamde 'grensleerling' per cluster (ronde 1, voorlopige standaard). Nadat alle panelleden hun oordeel gegeven hadden, werden de oordelen geanonimiseerd in een figuur weergegeven en aan de deelnemers getoond. Over de uitkomsten werd vervolgens gediscussieerd. Na deze discussieronde was het mogelijk om het oordeel aan te passen (ronde 2, definitieve standaard).

Het gemiddelde over alle experts van beoordelingsronde 2 gold vervolgens als voorlopig advies. Na de vergelijking van dit voorlopige advies met de uitkomsten van de standaardbepaling op de luistervaardigheidstoets van de peiling mondelinge taalvaardigheid (zie paragraaf 8.1.1) is er een definitief advies geformuleerd aan het CvTE. Het CvTE nam dit advies mee in de besluitvorming over de referentiecesuren.

³ Theorie binnen de testleer waarbij de kans dat een kandidaat met een bepaalde vaardigheid een bepaald item juist zal beantwoorden door een mathematisch model beschreven is. In tegenstelling tot het klassieke testmodel levert de itemresponstheorie toetsafhankelijke persoonsparameters en steekproefafhankelijke itemparameters.

5.2 Uitkomsten standaardbepaling

In onderstaande overzichten staat per referentieniveau weergegeven wat de uitkomsten zijn van de standaardbepaling. Het uiteindelijke advies is over alle experts gemiddeld en wordt in de tabel weergegeven achter 'voorlopig advies'. In de tabel worden tevens de laagste en hoogste score uit het panel weergegeven (minimum en maximum). Achter 'definitief advies' wordt het advies van Cito aan CvTE na vergelijking met de luistertoets uit het peilingsonderzoek vermeld.

Kijk- en luistervaardigheid 1F

Resultaten standaardbepaling 1F	Totale panel
Aantal items	31
Voorlopig advies cesuur	22
Minimum	17
Maximum	25
Definitief advies cesuur	22

Het advies van het panel voor de cesuur is: 22. Dat houdt in dat een leerling die 22 of meer opgaven van deze ankerset goed beantwoordt, het referentieniveau 1F heeft behaald. Leerlingen die minder dan 22 opgaven goed beantwoorden, hebben het referentieniveau 1F niet behaald.

In het peilingsonderzoek mondelinge taalvaardigheid is gebruikgemaakt van een enkele itemset van in totaal 30 opgaven waarmee zowel niveau 1F als niveau 2F gemeten wordt (zie paragraaf 8.1). De in 2017 gestelde indicatieve cesuur voor 1F bedroeg 15 van de 30 opgaven. Dat houdt in dat een leerling die 15 of meer opgaven goed beantwoord heeft, het referentieniveau 1F zou behalen. Met behulp van IRT is het mogelijk om de grens 1F van deze luistertoets te vergelijken met de voorlopige cesuur van de ankerset. Bij 1F zien we bij de vergelijking dat de standaarden van het peilingsonderzoek en de ankerset overeenkomen. Dat betekent dat het definitieve advies voor de cesuur 1F bij 22 van de 31 opgaven geplaatst wordt.

Kijk- en luistervaardigheid 2F

Resultaten standaardbepaling 2F	Totale panel
Aantal items	30
Voorlopig advies cesuur	24
Minimum	21
Maximum	27
Definitief advies cesuur	25

Het advies van het panel voor de cesuur is: 24. Dat houdt in dat een leerling die 24 of meer opgaven van deze ankerset goed beantwoordt, het referentieniveau 2F heeft behaald. Leerlingen die minder dan 24 opgaven goed beantwoorden, hebben het referentieniveau 2F niet behaald.

Net als bij niveau 1F is er een vergelijking gemaakt tussen de indicatieve cesuur 2F van de luistertoets uit het peilingsonderzoek en de ankerzet 2F. De destijds gestelde indicatieve cesuur voor 2F op de luistertoets van het peilingsonderzoek bedroeg 24 van de 30 opgaven. Dat houdt in dat een leerling die 24 of meer opgaven goed beantwoord heeft, het referentieniveau 2F zou behalen. Met behulp van IRT is het mogelijk om de grens 2F van deze luistertoets te vergelijken met de voorlopige cesuur van de ankerzet. Bij 2F kwamen deze cesuren niet volledig overeen.

Aangezien de standaardbepalingen van het peilingsonderzoek en de ankerzet niet kwalitatief van elkaar verschillen, wordt de informatie geoptimaliseerd door het gemiddelde van de twee te nemen. Dat betekent dat het definitieve advies voor de cesuur 2F van de ankerzet bij 25 van de 30 opgaven geplaatst wordt en het definitieve advies voor de cesuur 2F op de luistertoets van het peilingsonderzoek bij 23 van de 30 opgaven uit die luistertoets geplaatst wordt.

Definitieve adviezen cesuren Ankerzet Luisteren

De definitieve standaard 1F Luisteren is geplaatst bij **22 van de 31** van de 1F-set opgaven.

De definitieve standaard 2F Luisteren is geplaatst bij **25 van de 30** van de 2F-set opgaven.

6 Implementatie ankersets

In voorgaande hoofdstukken is uitgelegd dat het CvTE expertpanels heeft ingezet voor de inhoudsvalidering en de standaardbepaling van de ankersets. Deze expertpanels hebben aan het CvTE adviezen gegeven over de samenstelling en bruikbaarheid van de ankersets (inhoudelijke validering) en de hoogte van de referentiecesuur (standaardbepaling). Het College heeft deze adviezen overgenomen bij de vaststelling van de ankersets kijk- en luistervaardigheid. Daarmee is het hoofddoel van het project ankersets kijk- en luistervaardigheid behaald, namelijk het opleveren van door de overheid vastgestelde referentiecesuren voor kijk- en luistervaardigheid 1F en 2F in het primair onderwijs. Vervolgens kunnen de referentiecesuren kijk- en luistervaardigheid overgebracht worden naar verschillende toetsen.

6.1 Overbrengen referentiecesuren naar een commercieel product

Tijdens de headstart ankeronderzoek kijk- en luistervaardigheid (periode februari – december 2019) hebben ontwikkelaars van referentieniveaugerichte toetsen taal de mogelijkheid gekregen om de referentiecesuur van de openbare ankersets over te brengen op hun eigen toetsen kijk- en luistervaardigheid. Dit kan bijvoorbeeld door opgaven uit de ankersets eenmalig als anker op te nemen in een proefafname van een commerciële toets. Doordat een anker van opgaven uit de ankerzet samen met de reguliere toetsopgaven afgenomen wordt, kan de toetsontwikkelaar alle opgaven op één vaardigheidsschaal plaatsen en de referentiecesuur overbrengen.

In het document *Handleiding ankeronderzoek kijk- en luistervaardigheid* is meer informatie te vinden over de manier waarop een ankeronderzoek met de ankersets uitgevoerd kan worden.⁴ Ook is er een tool (Applicatie berekenen grensscore) beschikbaar voor toetsontwikkelaars waarmee zij kunnen bepalen waar de referentiecesuur ligt op een geselecteerd anker van opgaven uit de ankerzet.

6.2 Publicatie openbare ankersets kijk- en luistervaardigheid

Tijdens de headstart hebben toetsontwikkelaars de openbare ankersets Kijk- en Luistervaardigheid, onder geheimhouding, kunnen aanvragen bij het CvTE. Op dat moment waren de openbare ankersets dus nog niet (beperkt) toegankelijk gemaakt.

Vanaf 1 januari 2020 zijn de opgaven en de bijbehorende Kijk- en Luisterfragmenten van de ankersets opvraagbaar door contact op te nemen met CvTE.⁵

⁴ Meer informatie is te vinden via http://www.toetsspecials.nl/html/referentiesets/ankersets_kijkluister.shtm .

⁵ Zie voor gegevens en voorwaarden: www.referentiesets.nl .

7 Standaardbepalingen peilingsonderzoeken mondelinge taalvaardigheid

Naast de standaardbepaling voor de ankersets luistervaardigheid zijn er in het kader van dit project nog een vijftal standaardbepalingen uitgevoerd. Het betrof standaardbepalingen luistervaardigheid, spreekvaardigheid en gespreksvaardigheid in het kader van het peilingsonderzoek mondelinge taalvaardigheid einde basisonderwijs en standaardbepalingen spreekvaardigheid en gespreksvaardigheid voor het peilingonderzoek mondelinge taalvaardigheid einde so/sbo. Voor een uitgebreide verantwoording van deze standaardbepalingen verwijzen we naar eerdere rapportages (Linthorst e.a., 2017 en Hemker e.a., 2018). Hieronder worden de opzet en de resultaten in hoofdlijnen weergegeven.

7.1 Peilingsonderzoek mondelinge taalvaardigheid in het reguliere po

In 2017 heeft de Inspectie van het onderwijs een peilingsonderzoek laten uitvoeren naar de beheersing van de mondelinge taalvaardigheid van leerlingen aan het einde van het basisonderwijs.⁶ In dat onderzoek is er bij leerlingen onder andere een luistertoets, een spreektaak en een gesprekstaak afgenomen. Deze taken moesten voorzien worden van beheersingscesuren 1F en 2F. Om dat te bewerkstelligen zijn er op 29 mei, 2 juni en 9 juni 2017 standaardbepalingen uitgevoerd. De cesuren op de luistertoets betreffen indicatieve cesuren die na ankering met de ankerzet omgezet zijn naar definitieve cesuren. De cesuren spreekvaardigheid en gespreksvaardigheid betreffen taakgerichte cesuren. Leerlingen hebben in dit peilingsonderzoek slechts één spreektaak en één gesprekstaak uitgevoerd, omdat het praktisch niet haalbaar was om hen meer taken te laten uitvoeren. Hierdoor kunnen de standaarden niet representatief zijn voor de beheersing van de eisen die het Referentiekader stelt aan spreekvaardigheid of gespreksvaardigheid in het algemeen. Hier moet rekening mee gehouden worden bij de interpretatie van de gezette standaarden.

7.1.1 Standaardbepaling met experts

De standaardbepalingen zijn uitgevoerd door een panel van experts. De samenstelling van dit expertpanel is voor alle uitgevoerde standaardbepalingen voor de ankerzet luistervaardigheid (zie paragraaf 6.1), het peilingsonderzoek mondelinge taalvaardigheid einde basisonderwijs en het peilingsonderzoek mondelinge taalvaardigheid einde so/sbo zoveel mogelijk gelijk gehouden vanwege de inhoudelijke overlap tussen de onderdelen.

Het CvTE was verantwoordelijk voor de werving van het expertpanel. In onderstaand overzicht is de samenstelling van het expertpanel voor de standaardbepalingen van het peilingsonderzoek mondelinge taalvaardigheid weergegeven.

⁶ Peil.Mondelinge taalvaardigheid - Einde basisonderwijs (Inspectie van het Onderwijs, 2019)

naam	Organisatie	luisteren	spreken	gesprekken
Annemarie Heerink	Aloysiusschool	x	x	x
Dorret de Leeuw van Weenen-Poel	Julianaschool	x	x	x
Hans Schepers	Pabo Hogeschool Zeeland	x	x	x
Heleen Strating	Expertisecentrum Nederlands	x	x	
Inge van Meelis	ITTA Kennisinstituut voor Taalontwikkeling – Universiteit van Amsterdam		x	x
Irene Neelis	Onderwijsgroep Primus	x	x	x
Janneke Tambach	De Mei	x	x	x
Jos Boerema	SBO de Zwaaiikom	x	x	x
Hanane Loukili	Al Iman	x	x	x
Nanne Klompstra	Leeuwarder Schoolvereniging	x	x	x
Petra Popma	ITTA Kennisinstituut voor Taalontwikkeling – Universiteit van Amsterdam	x		
Rachid Belhocein	SBO Michaelschool	x	x	x
Resi Damhuis	Taaldenkgesprekken.nu		x	x
Roelof Liemburg	OBS It Skriuwboerd	x		x
Tjalling Brouwer	Impact Educatief	x	x	x
Wieteke Klijnstra-Idema	CBS De Wendakker	x	x	x
Wim Goossens	Kindcentrum De Bakelaar		x	x
		14	15	15

Het panel dat de standaarden heeft gezet, bestond in totaal uit achttien personen. Vier van hen zijn taalexperts die ook zitting hebben in het expertpanel voor de ankersets kijk- en luistervaardigheid. Elf personen zijn werkzaam als leerkracht of directeur in het reguliere basisonderwijs en twee personen zijn werkzaam als leerkracht in het sbo. De mix van leerkrachten uit het reguliere basisonderwijs en het so/sbo is gemaakt met het oog op de standaardbepalingen voor het peilingsonderzoek mondelinge taalvaardigheid so/sbo. In totaal hebben 14 personen uit het panel deelgenomen aan de standaardbepaling luistervaardigheid, 15 personen aan de standaardbepaling spreekvaardigheid en 15 personen aan de standaardbepaling gespreksvaardigheid.

7.1.2 *Beschrijving procedure* *Luistervaardigheid*

In het peilingsonderzoek is bij leerlingen uit groep 8 een luistertoets afgenomen. Deze luistertoets bestond uit 30 opgaven die de onderdelen *luisteren naar radio, tv en internet* en *luisteren naar instructies* uit het referentiekader taal maten. De toets bevatte opgaven die ontwikkeld zijn voor niveau 1F en 2F en is digitaal en individueel afgenomen. De toets is geschaald volgens een IRT-methode.

De procedure die gevolgd is bij de standaardbepaling is gelijk aan de procedure die gehanteerd is bij de standaardbepaling voor de ankerzet luistervaardigheid (zie paragraaf 6.1). De standaardbepaling startte met het grondig doornemen van de eisen die het Referentiekader Taal stelt aan luistervaardigheid, aangezien die eisen de basis vormen voor het bepalen van de grenzen 1F en 2F. Vervolgens is de daadwerkelijke standaardbepaling uitgevoerd op basis van de 3DC-methode. De toets is opgedeeld in 3 clusters die varieerden van 9 tot 11 opgaven per cluster. Tijdens de standaardbepaling werd aan beoordelaars gevraagd om per cluster opgaven aan te geven welke score leerlingen zouden moeten behalen als zij zich precies op de grens van het desbetreffende referentieniveau bevinden. De bepaling van de standaard 1F en 2F gebeurde in twee rondes: in de eerste ronde bepaalde ieder panellid individueel de score van de 'grensleerling', vervolgens volgde er een plenaire discussie waarna de panelleden in de twee ronde nog de mogelijkheid kregen om hun oordeel bij te stellen op basis van de discussie.

Spreekvaardigheid en gespreksvaardigheid

Voor de standaardbepalingen spreekvaardigheid en gespreksvaardigheid is een andere opzet gehanteerd dan voor luistervaardigheid, omdat de gehanteerde spreektaak en gesprekstaak niet op te delen is in afzonderlijk te interpreteren opgaven.

De spreektaak die in het kader van het peilingsonderzoek afgenomen is, betrof een monoloog in de vorm van een vlog. In de vlog moesten leerlingen verslag doen van een bijzondere gebeurtenis: in hun straat was de straatprijs van een bekende loterij gevallen en via de vlog informeerden de leerlingen hun volgers over wat ze hadden meegemaakt en wat ze daarvan vonden. De gesprekstaak betrof een groepsgesprek met drie leerlingen die ongeveer tien minuten duurde. In dit gesprek moesten de leerlingen plannen maken om geld in te zamelen voor een goed doel. Het doel van het gesprek was dat de leerlingen in gezamenlijk overleg tot een onderbouwd idee kwamen. Zowel de vlogs als de gesprekken zijn vastgelegd op video en de filmfragmenten zijn beoordeeld door getrainde beoordelaars. De oordelen zijn geschaald volgens een IRT-methode.

Omdat de spreek- en gesprekstaak niet in te delen zijn in afzonderlijke opgaven, is er bij de standaardbepaling gebruikgemaakt van een leerlinggerichte methode in plaats van een itemgerichte methode. Dit betekent dat de standaard niet bepaald is door de evaluatie van items maar door de evaluatie van (voorbeelden van) spreekproductie (video-opnamen van vlogs en gesprekken) van leerlingen. Voorafgaand aan de standaardbepalingen zijn leerlingwerken geselecteerd waarbij de gehanteerde beoordelingsschaal als uitgangspunt genomen is. Voor spreken liepen de mogelijke scores op van 0 tot 17 punten, voor gespreksvaardigheid van 0 tot 10 punten. Voor alle mogelijke scorepunten op de schaal zijn leerlingwerken geselecteerd, waarbij bij de selectie van leerlingwerken rekening gehouden is met de factoren beoordelaar (ervarenheid en strengheid van beoordelaar), beoordelingspatroon (manier waarop tot eendoordeel gekomen is), school en leerlingen (spreiding over scholen en geslacht).

Net als bij luistervaardigheid startte de standaardbepaling met het doornemen van de eisen die het Referentiekader Taal stelt voor spreekvaardigheid en gesprekvaardigheid 1F en 2F. Deze inhoudelijke eisen vormden het vertrekpunt voor de standaardbepalingen.

De standaardbepaling startte met een globale ronde, waarbij de experts filmpjes van leerlingenwerken kregen te zien die opliepen in beheersingsniveau. De deelnemers gaven individueel aan waar volgens hen de grens 1F of 2F lag. Op basis van de globale oordelen uit deze ronde, kregen de deelnemers in een tweede ronde filmpjes te zien van leerlingen met een score vlak onder, op en vlak boven de globale cesuur van de eerste ronde. Ze stelden opnieuw individueel de cesuur vast voor deze serie filmpjes. Deze ronde werd besproken en de experts kregen in een derde en laatste ronde de kans om hun oordeel nog bij te stellen op basis van deze discussie. Door te werken met een globale en een gedetailleerde ronde werd de standaard bepaald op basis van zo veel mogelijk leerlingprestaties, waarbij tegelijkertijd een zo nauwkeurig mogelijke standaard gesteld werd.

7.1.3

Uitkomsten standaardbepalingen

Uitkomsten standaardbepaling luistervaardigheid

Resultaten standaardbepaling 1F	Totale panel
Aantal items	30
Voorlopig advies cesuur	15
Minimum	11
Maximum	18
Definitief advies cesuur	15

Het advies van het panel voor de cesuur op de luistertoets 1F is: 15. Dat houdt in dat een leerling die 15 of meer opgaven van deze luistertoets goed beantwoordt, het referentieniveau 1F heeft behaald. Leerlingen die minder dan 15 opgaven goed beantwoorden, hebben het referentieniveau 1F niet behaald.

Op het moment dat de standaardbepaling op de luistertoets plaatsvond, was de ankerzet kijk- en luistervaardigheid nog niet beschikbaar. Deze ankerzet is in 2018 opgeleverd en voorzien van een standaard. Om de cesuren op de luistertoets en de ankerzet op elkaar te laten aansluiten, is er een aanvullend onderzoek verricht waarbij leerlingen de luistertoets in combinatie met (een deel van) de ankerzet gemaakt hebben. Met behulp van IRT konden de cesuren met elkaar vergeleken worden. Bij 1F zagen we dat de cesuren van de luistertoets en de ankerzet overeenkwamen, wat betekent dat de cesuur van de luistertoets voor 1F definitief gesteld is op 15 van de 30 opgaven goed.

Resultaten standaardbepaling 2F	Totale panel
Aantal items	30
Voorlopig advies cesuur	24
Minimum	22
Maximum	25
Definitief advies cesuur	23

Het advies van het panel voor de cesuur luisteren 2F is: 24. Dat houdt in dat een leerling die 24 of meer opgaven van deze luistertoets goed beantwoordt, het referentieniveau 2F heeft behaald. Leerlingen die minder dan 24 opgaven goed beantwoorden, hebben het referentieniveau 2F niet behaald.

Ook voor 2F zijn de voorlopige cesuren vergeleken met de cesuren van de ankerzet kijk- en luistervaardigheid. Bij 2F trad daarbij een klein verschil op. Aangezien beide standaardbepalingen niet kwalitatief van elkaar verschilden, is besloten om te middelen (zie paragraaf 6.2). Daardoor is de definitieve cesuur voor de luistertoets van het peilingsonderzoek gelegd op 23 van de 30 opgaven goed.

Definitieve adviezen cesuren Luistertoets Peilingsonderzoek regulier basisonderwijs

De definitieve standaard 1F Luisteren is geplaatst bij **15 van de 30** opgaven.

De definitieve standaard 2F Luisteren is geplaatst bij **23 van de 30** opgaven.

Uitkomsten standaardbepaling spreekvaardigheid regulier basisonderwijs

De standaardbepaling voor spreekvaardigheid 1F leverde na de derde ronde een gemiddelde grensscore op van 7,27 van de 17 mogelijke scorepunten, met een standaardfout⁷ van het gemiddelde van 0,23. De standaardbepaling voor spreken 2F leverde na de derde ronde een gemiddelde grensscore op van 11,73 van de 17 mogelijke scorepunten, met een standaardfout van het gemiddelde van 0,18. De standaardbepalingen leverden zodoende de volgende grenzen op bij deze taak:

De taakgerichte standaard 1F Spreken is geplaatst bij **8 van de 17** scorepunten.

De taakgerichte standaard 2F Spreken is geplaatst bij **12 van de 17** scorepunten.

Indien een leerling minder dan 8 scorepunten op de spreektaak behaald heeft, beheerst hij deze spreektaak niet op niveau 1F. Bij een score tussen 8 en 11 punten beheerst een leerling het referentieniveau 1F voor deze spreektaak wel, maar het referentieniveau 2F nog niet. Bij een score van 12 punten of meer beheerst een leerling de taak op niveau 2F. Er wordt bewust gesproken over beheersing van de taak: aangezien er in het peilingsonderzoek om eerder genoemde redenen slechts één

⁷ De standaardfout is een indicatie voor de onnauwkeurigheid van een meting.

spreektaak afgenomen kon worden en het Referentiekader veelomvattender is, is generalisatie naar beheersing van spreekvaardigheid in het algemeen niet mogelijk.

Uitkomsten standaardbepaling gespreksvaardigheid regulier basisonderwijs

De standaardbepaling voor gespreksvaardigheid 1F leverde na de derde ronde een gemiddelde grensscore op van 4,60 van de 10 mogelijke scorepunten, met een standaardfout van het gemiddelde van 0,16. De standaardbepaling voor gesprekken 2F leverde na de derde ronde een gemiddelde grensscore op van 7,87 van de 10 mogelijke scorepunten, met een standaardfout van het gemiddelde van 0,09.

De standaardbepalingen leverden zodoende de volgende grenzen op bij deze taak:

De taakgerichte standaard 1F Gesprekken is geplaatst bij **5 van de 10** scorepunten.

De taakgerichte standaard 2F Gesprekken is geplaatst bij **8 van de 10** scorepunten.

Indien een leerling minder dan 5 scorepunten op de gesprekstaak behaald heeft, beheerst hij deze gesprekstaak niet op niveau 1F. Bij een score tussen 5 en 7 punten beheerst een leerling het referentieniveau 1F voor deze gesprekstaak wel, maar het referentieniveau 2F nog niet. Bij een score van 8 punten of meer beheerst een leerling de taak op niveau 2F.

7.2 Peilingsonderzoek mondelinge taalvaardigheid in het so en sbo

Naast het peilingsonderzoek mondelinge taalvaardigheid in het reguliere basisonderwijs, heeft de Inspectie van het onderwijs in 2018 een onderzoek laten uitvoeren naar de beheersing van de mondelinge taalvaardigheid van leerlingen in het speciaal onderwijs en het speciaal basisonderwijs die uitstromen naar het voortgezet (speciaal) onderwijs.⁸ Aangezien de gehanteerde taken in het peilingsonderzoek in het reguliere basisonderwijs niet geschikt waren voor de specifieke doelgroep van het so/sbo, is er in dit peilingsonderzoek gebruikgemaakt van andere taken spreekvaardigheid en gespreksvaardigheid. Om deze taken te kunnen voorzien van taakgerichte cesuren 1F en 2F hebben er nieuwe standaardbepalingen plaatsgevonden. Voor luistervaardigheid is er in het peilingsonderzoek gebruikgemaakt van de ankerzet luistervaardigheid, waardoor er geen nieuwe standaardbepaling luistervaardigheid nodig was. De standaardbepalingen voor het so/sbo vonden plaats op 12 en 15 oktober 2018.

7.2.1 Standaardbepaling met experts

De standaardbepalingen zijn uitgevoerd door een panel van experts dat, vanwege de inhoudelijke overlap, qua samenstelling zo veel mogelijk aansloot bij de standaardbepalingen mondelinge taalvaardigheid in het reguliere basisonderwijs. Het CvTE was verantwoordelijk voor de werving van het expertpanel. In onderstaand overzicht is de samenstelling van het expertpanel weergegeven.

⁸ Peil. Mondelinge taalvaardigheid - Einde speciaal (basis)onderwijs (Inspectie van het Onderwijs, 2019)

naam	organisatie	spreken so/sbo	gesprekken so/sbo
Annemarie Heerink	Konot	X	X
Dorret de Leeuw van Weenen-Poel	Julianaschool	X	X
Hans Schepers	Oud-leerkracht/directeur PO	X	X
Heleen Strating	Expertisecentrum Nederlands	X	
Irene Neelis	OG Primus	X	X
Loukili Hanane	Al Iman	X	X
Nanne Klompstra	LSV / CBS De wegwijzer	X	X
Rachid Belhocein	Michaelschool / De Baander		X
Resi Damhuis	Taaldenkgesprekken.nu	X	X
Roelof Liemburg	OBS It Skriuwboerd	X	X
Tjalling Brouwer	Impact Educatief	X	X
Wieteke Klijnstra-Idema	CBS De Wendakker	X	X
Maud van Druenen	Expertisecentrum Nederlands		X
Jacinta Kuijl	De Meulenridders		X
Annie van der Beek	Van der Beek Taalontwikkeling	X	X

Het panel dat de standaarden heeft gezet, bestond in totaal uit 16 personen. 5 van hen zijn taalexperts, van wie er 4 ook zitting hebben in het expertpanel voor de ankersets kijk- en luistervaardigheid. 7 personen zijn werkzaam als leerkracht of directeur in het reguliere basisonderwijs en 3 personen zijn werkzaam als leerkracht in het so/sbo. Eén persoon is gepensioneerd leerkracht basisonderwijs. Mede om de gelijkheid van de standaarden te borgen, is er een grote overlap tussen de deelnemers aan deze standaardbepalingen en de standaardbepalingen voor het peilingsonderzoek mondelinge taalvaardigheid in het reguliere basisonderwijs: 12 van de 16 experts hebben ook deelgenomen aan die eerdere standaardbepalingen. Omdat één persoon de dag voorafgaand aan de standaardbepalingen aangaf toch niet aanwezig te kunnen zijn, zijn de standaardbepalingen uiteindelijk uitgevoerd door 15 personen.

7.2.2 *Beschrijving procedure*

In het peilingsonderzoek in het so/sbo zijn twee spreektaken afgenomen. De eerste spreektaak, die verder aangeduid zal worden als de taak 'Jeugdjournaal', is een

naverteltaak. Leerlingen bekeken een nieuwsfragment uit het Jeugdjournaal over een baby-otter en moesten vervolgens aan de toetsleider, die het fragment niet gezien had, zo precies mogelijk navertellen wat er in het filmpje te zien was.

De tweede spreektaak, de taak 'Dobbelsteen', betrof een taak waarbij leerlingen aan de hand van afbeeldingen op dobbelstenen ('Rory's Story Cubes') een verhaaltje moesten vertellen aan de toetsleider.

Naast de twee spreektaken is er ook een gesprekstaak afgenomen: in drie- of viertallen voerden leerlingen een samenwerkingsspel uit waarbij ze als vissers gezamenlijk alle vissen binnen moesten halen voordat de storm opstak. Bij dit spel is het belangrijk dat leerlingen goed met elkaar overleggen over de te volgen strategie.

Net als bij de standaardbepalingen voor spreek- en gespreksvaardigheid in het reguliere basisonderwijs is er bij deze standaardbepalingen een leerlinggerichte methode van standaardbepalen gehanteerd, waarbij de leerlingresponsen het uitgangspunt vormden. De procedure voor het bepalen van de cesuren is volledig gelijk aan de procedure die beschreven is voor het reguliere basisonderwijs. In paragraaf 8.1 is deze procedure beschreven. Er zijn twee kleine verschillen: omdat de beoordelingsmodellen voor de spreektaken en de gesprekstaken gelijk waren aan die voor het reguliere basisonderwijs, was er voorafgaand aan de standaardbepalingen al enige informatie beschikbaar over het bereik waarin de standaarden mogelijk verwacht zouden worden, waardoor het scorebereik van de filmpjes in de globale ronde enigszins ingeperkt kon worden. Daarnaast waren er in dit peilingsonderzoek op het moment van de standaardbepalingen dubbele beoordelingen van leerlingwerken beschikbaar. In de standaardbepalingen zijn daarom filmpjes gebruikt waarbij grote overeenstemming was tussen de eerste en tweede beoordelaar.

7.2.3 *Uitkomsten standaardbepaling spreekvaardigheid so/sbo*

Het gehanteerde beoordelingsmodel spreekvaardigheid kende een score-range van 0 tot 17 punten. De standaardbepaling voor Jeugdjournaal 1F leverde na de derde ronde een gemiddelde grensscore op van 11,21 van de 17 mogelijk scorepunten, met een standaardfout van het gemiddelde van 0,26. Bij de taak Dobbelsteen 1F betrof dit een grensscore van 11,86 van de 17 mogelijke scorepunten, met een standaardfout van het gemiddeld van 0,63. Voor 2F was de grensscore bij de taak Jeugdjournaal 15,5 van de 17 mogelijke scorepunten, met een standaardfout van 0. (Alle experts waren unaniem in hun oordeel over deze grens.) Bij de taak Dobbelsteen 2F lag de grensscore op 15,5 scorepunten met een standaardfout van 0,18.

De standaardbepalingen leverden zodoende de volgende grenzen op bij deze taken:

De taakgerichte standaard 1F Spreken Jeugdjournaal is geplaatst bij **11 van de 17⁹** scorepunten.

De taakgerichte standaard 1F Spreken Dobbelsteen is geplaatst bij **12 van de 17** scorepunten.

De taakgerichte standaard 2F Spreken Jeugdjournaal is geplaatst bij **16 van de 17** scorepunten.

De taakgerichte standaard 2F Spreken Dobbelsteen is geplaatst bij **16 van de 17** scorepunten.

Uit het bovenstaande overzicht is op te maken dat de cesuren hoger liggen dan bij de spreektaak uit het peilingsonderzoek in het reguliere basisonderwijs. Daar lag de grens 1F bij 8 van de 17 scorepunten en de grens 2F bij 12 van de 17 punten (zie paragraaf 7.1.4). Een mogelijke verklaring hiervoor zou kunnen zijn dat de taken uit het so/sbo makkelijker zijn dan de taak uit het reguliere basisonderwijs. Om de taken uit het so/sbo te kunnen vergelijken met de taak uit het reguliere basisonderwijs zijn alle taken in een aanvullend onderzoek afgenomen bij een groep van 178 leerlingen van 18 scholen uit het reguliere basisonderwijs. Op basis van dit aanvullende onderzoek kon geconcludeerd worden dat de cesuren nagenoeg vergelijkbaar zijn en dat aanpassing van cesuren niet nodig is. De taken uit het so/sbo zijn inderdaad eenvoudiger en dat maakt dat leerlingen die net 1F behalen, voor deze taken een hogere score moeten halen dan voor de spreektaak uit het reguliere basisonderwijs.

Uit het overzicht blijkt ook dat leerlingen om niveau 2F te kunnen behalen, vrijwel alle scorepunten op de taak moeten behalen (16 van 17). Tijdens de standaardbepalingen kwam naar voren dat de so/sbo-taken inhoudelijk iets minder geschikt lijken om niveau 2F te kunnen meten, omdat de eisen die het Referentiekader stelt met de taken niet volledig in beeld gebracht kunnen worden met de taken. Dit verklaart vermoedelijk ook de hoge cesuren.

⁹ Het consortium dat het Peilingsonderzoek so/sbo uitvoerde heeft in de het onderzoek deze scoreschaal vertaald naar een eigen scoreschaal (de ongewogen somscore van de scores op alle beoordelingselementen) met een range van 0-30. De standaarden op deze scoreschaal zijn:
1F Jeugdjournaal: 20 van de 30 scorepunten
1F Dobbelstenen: 21 van de 30 scorepunten
2F Jeugdjournaal en Dobbelstenen: 27 van de 30 scorepunten.

7.2.4

Uitkomsten standaardbepaling gespreksvaardigheid so/sbo

Het gehanteerde beoordelingsmodel gespreksvaardigheid kende een scorereange van 0 tot 10 punten. De standaardbepaling voor de gesprekstaak 1F leverde na de derde ronde een gemiddelde grensscore op van 7,57 van de 10 mogelijk scorepunten, met een standaardfout van het gemiddelde van 0,27. De standaardbepaling 2F leverde een gemiddelde grensscore op van 9 van de 10 mogelijke scorepunten, met een standaardfout van het gemiddelde van 0. (Alle experts waren bij 2F unaniem in hun oordeel.) De standaardbepalingen leverden zodoende de volgende grenzen op bij deze taak:

De taakgerichte standaard 1F Gesprekken is geplaatst bij **8 van de 10** scorepunten.¹⁰

De taakgerichte standaard 2F Gesprekken is geplaatst bij **9 van de 10** scorepunten.

Net als bij spreken liggen de cesuren bij de so/sbo-taak hoger dan bij de gesprekstaak uit het peilingsonderzoek in het reguliere basisonderwijs. Ook hier zijn beide taken in een aanvullend onderzoek afgenomen bij een groep van 174 leerlingen uit groep 8 van het reguliere basisonderwijs en ook daaruit kwam geen reden tot aanpassing naar voren. De gesprekstaak 2F uit het so/sbo is net als de spreektaken ook eenvoudiger dan de taak uit het peilingsonderzoek in het reguliere basisonderwijs en lijkt ook iets minder geschikt om de beheersing van niveau 2F in beeld te brengen.

¹⁰ Het consortium dat het Peilingsonderzoek so/sbo uitvoerde heeft in de het onderzoek deze scoreschaal vertaald naar een eigen scoreschaal (de ongewogen somscore van de scores op alle beoordelingselementen) met een range van 0-34. De standaarden op deze scoreschaal zijn:
1F Gesprekken: 28 van de 34 scorepunten
2F Gesprekken: 31 van de 34 scorepunten.

8 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.

Hemker, B, Heesters, K., Koops, J. & Linthorst, R. (2018). *Standaardebepalingen Project Referentieset Mondelinge Taalvaardigheid (RMT) 2018, voorlopig rapport na eerste standaardebepaling 2018*. Arnhem: Cito.

Hemker, B., Ingen, K. van, Linthorst, R. (2018). *Rapportage standaardebepalingen peilingsonderzoek mondelinge taalvaardigheid so/sbo*. Arnhem: Cito.

Inspectie van het onderwijs (2019). *Peil.Mondelinge taalvaardigheid - Einde basisonderwijs*. Utrecht.

Inspectie van het Onderwijs (2019). *Peil.Mondelinge taalvaardigheid - Einde speciaal (basis)onderwijs*. Utrecht.

Linthorst, R., Hemker, B., Koerhuis, I., Feskens, R. & Koops, J. (2017). *Rapportage standaardebepalingen peilingsonderzoeken mondelinge taalvaardigheid*. Arnhem: Cito.

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

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

9 Bijlage - overzicht bronnen

Fragment	Naam programma	Eigenaar / omroep	datum uitzending
1F_hersenen	Het Klokhuis	NTR	15-3-2016
1F_ridders	Youtube / Ridders geschiedenis	Ridders geschiedenis	19-06-2016
1F_voicemailbericht	n.v.t.	CvTE/Stichting Cito	n.v.t.
1F_koningscobra	Schooltv weekjournaal	VPRO	02-10-2009
1F_Jacques Vriens	040Nieuws	Studio 040	09-04-2015
1F_Joris en geheimzinnige toverdrank	Boekstarter	Tanneke Hartzuiker / Uitgeverij de Fontein	25-06-2014
1F_kom in actie	Schooltv	NTR - Schooltv	30-01-2013
1F_zebrastrepen	Clipphanger	NTR - Schooltv	29-04-2015
1F_hik	Evers staat op	Radio538	05-09-2016

Fragment	Naam programma	Eigenaar / omroep	datum uitzending
2F_dynamo	Willem Wever	KRO/NCRV	29-04-2016
2F_papierscheppen	CNME Maastricht	CNME Maastricht	19-04-2011
2F_yogaboom	n.v.t.	Squla	n.v.t.
2F_Skatekeet	Zaplive	KRO/NCRV	18-03-2017
2F_Melkmeisje	Schooltv	NTR - Schooltv	21-01-2017
2F_eekhoorn en de mier	Frank	BNN / 3FM	18-04-2016
2F_bisjpalen	n.v.t.	Tropenmuseum	18-08-2014
2F_kindermanifest	n.v.t.	Vereniging van Openbare Bibliotheken	30-05-2016
2F_pindakaas	Bureau Kijk in de Vegte	KRO/NCRV	10-04-2016

