

Rapportage Standaardbepalingen
Project Referentieset Mondelinge Taalvaardigheid (RMT)
2018

voorlopig rapport na eerste standaardbepaling 2018

Bas Hemker

Karin Heesters

Jesse Koops

Roelien Linthorst

Cito

juni 2018

Inleiding

Stichting Cito voert in 2018 in opdracht van het College voor Toetsen en Examens (CvTE) drie standaardbepalingen uit binnen het project Referentieset Mondelinge Taalvaardigheid (RMT). In de eerste plaats is er inmiddels een standaard gezet op de ontwikkelde Ankerset Luistervaardigheid. Deze ankerset weerspiegelt het Referentiekader Luisteren op de Referentieniveaus 1F en 2F aan het einde van het basisonderwijs. Deze standaardbepaling heeft plaatsgevonden op 1 juni jongstleden. In dit voorlopige rapport wordt alvast verslag gedaan van deze standaardbepaling, aangezien de gezette cesuren van belang zijn voor de oplevering van het Peilingsonderzoek Mondelinge Taalvaardigheid in het reguliere basisonderwijs.

In het najaar van 2018 vinden vervolgens nog standaardbepalingen plaats voor de onderdelen spreekvaardigheid en gespreksvaardigheid van het Peilingsonderzoek Mondelinge Taalvaardigheid in het speciaal onderwijs en speciaal basisonderwijs. Na afloop van die standaardbepalingen zal er een definitief verslag opgeleverd worden, waarin naast de standaardbepalingen van 2018 ook een verslag opgenomen is van de standaardbepalingen die in 2017 in het kader van dit project hebben plaatsgevonden.

In 2017 hebben er drie standaardbepalingen plaatsgevonden die alle drie betrekking hadden op het Peilingonderzoek Mondelinge Taalvaardigheid in het reguliere basisonderwijs. Ten eerste is er een indicatieve cesuur gezet op de luistertoets die in het Peilingsonderzoek ingezet is om het onderdeel luistervaardigheid te meten. Nu de standaardbepaling op de Ankerset Luistervaardigheid heeft plaatsgevonden, kan deze indicatieve cesuur op basis van de gezette standaarden op de Ankerset en een aanvullend onderzoek waarin de Luistertoets samen met de Ankerset is afgenomen, omgezet worden in een definitieve cesuur. Daarnaast zijn er in 2017 standaarden gezet op een spreektaak (monoloog) en een gesprekstaak (polyloog) uit het Peilingsonderzoek. Aangezien er voor de onderdelen spreekvaardigheid en gespreksvaardigheid geen ankerset ontwikkeld is, zullen deze indicatieve cesuren indicatief blijven en niet omgezet worden in definitieve cesuren.

In dit voorlopige rapport doen we, zoals gezegd, verslag van de standaardbepaling op de Ankerset Luistervaardigheid en de implicaties van de gezette cesuren. In paragraaf 1 geven we een korte beschrijving van de Ankerset Luistervaardigheid. In paragraaf 2 beschrijven we de samenstelling van het expertpanel dat de standaardbepaling heeft uitgevoerd. In paragraaf 3 en 4 bespreken we de opzet en de uitkomsten van de standaardbepaling en in paragraaf 5 ten slotte geven we weer wat de implicaties van de standaardsetting zijn voor de definitieve cesuren voor de Ankerset Luisteren en voor de Luistertoets uit het Peilingsonderzoek Mondelinge Taalvaardigheid in het reguliere basisonderwijs.

1. De Ankerset Luisteren

Stichting Cito heeft in 2016 van het CvTE de taak gekregen om een Ankerset Luisteren te ontwikkelen. Deze ankerset heeft veel gelijkenissen met de eerdere Referentiesets Lezen en Rekenen en de Ankerset Taalverzorging die in eerdere jaren ontwikkeld zijn. In tegenstelling tot deze Referentiesets is de Ankerset Luisteren echter enkel ontwikkeld voor groep 8 van het basisonderwijs en niet voor andere leerjaren en/of andere onderwijsniveaus en biedt het daarmee geen referentie over onderwijssectoren heen.

De Ankerset Luisteren is in feite een ankerset kijk- en luistervaardigheid. 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 Luisteren zowel videofragmenten als audiofragmenten opgenomen, waarbij de uiteindelijke verhouding ongeveer 2/3 deel videomateriaal en 1/3 deel audiomateriaal bedraagt.

De Ankerset Luisteren sluit nauw aan bij het onderdeel Luisteren uit het Referentiekader Taal. Het Referentiekader Taal onderscheidt drie luistertaken: luisteren naar instructies, luisteren naar radio, tv en internet en luisteren naar live publiek. Voor al deze onderdelen bevat de ankerset video- en audiomateriaal met bijbehorende opgaven die bedoeld zijn om de kijk- en luistervaardigheid van leerlingen in kaart te brengen. Er is een ankerset ontwikkeld voor niveau 1F en voor niveau 2F.

De ankerset is voorafgaand aan de proeftoetsing in september 2017 gevalideerd door een extern expertpanel dat bestond uit taalexperts en leerkrachten uit het basisonderwijs. Het expertpanel heeft beoordeeld of de opgaven goed aansloten bij de doelstellingen uit het Referentiekader en heeft ook beoordeeld of de kwaliteit van de opgaven in orde was. Opgaven die niet goed gekeurd waren, zijn niet meegegaan in de proeftoetsing. Na afloop van de proeftoetsing zijn de opgaven opnieuw voorgelegd aan het expertpanel, nu voorzien van psychometrische gegevens. Op basis van die bespreking is een selectie gemaakt van kijk-/luisterfragmenten en bijbehorende opgaven die enerzijds het Referentiekader goed representeren en anderzijds voldoende onderscheidend zijn om de vaardige en minder vaardige leerlingen op niveau 1F en niveau 2F te kunnen onderscheiden. In Tabel 1 is de samenstelling van de definitieve ankerset weergegeven¹.

Ankerset 1F			
	kijk/luisterfragmenten	opgaven per fragment	totaal aantal opgaven reeks
Luisteren naar instructies	<ul style="list-style-type: none"> • Fragment 1 • Fragment 2 • Fragment 3 	4 5 3	12
Luisteren naar live publiek	<ul style="list-style-type: none"> • Fragment 4 • Fragment 5 • Fragment 6 	3 3 3	9
Luisteren naar radio/tv/internet	<ul style="list-style-type: none"> • Fragment 7 • Fragment 8 • Fragment 9 	3 3 4	10
Totaal aantal opgaven 1F			31
Ankerset 2F			
Luisteren naar instructies	<ul style="list-style-type: none"> • Fragment 10 • Fragment 11 • Fragment 12 	4 4 3	11
Luisteren naar live publiek	<ul style="list-style-type: none"> • Fragment 13 • Fragment 14 • Fragment 15 	3 4 4	11
Luisteren naar radio/tv/internet	<ul style="list-style-type: none"> • Fragment 16 • Fragment 17 • Fragment 18 	3 2 3	8
Totaal aantal opgaven 2F			30

Tabel 1 Samenstelling Ankerset Luistervaardigheid

¹ De inhoud van de Ankerset Luistervaardigheid is in deze tabel nog niet weergegeven. Die zal openbaar gemaakt worden via www.referentiesets.nl na afloop van de headstart die in 2019 plaatsvindt.

2. Het expertpanel standaardbepalingen

Voor de standaardbepaling op de Ankerset Luistervaardigheid zijn experts uit het veld geworven. In totaal hebben 13 personen deelgenomen aan de standaardbepaling op de ankerset. 5 van hen zijn taalexperpts die ook zitting hebben in het expertpanel dat betrokken was bij de vaststelling van de definitieve Ankerset Luisteren. De overige 8 personen zijn werkzaam als leerkracht of schooldirecteur in het basisonderwijs, van wie 6 personen werkzaam zijn het reguliere basisonderwijs en 2 personen in het SO/SBO.

De deelnemende leerkrachten zijn afkomstig van stads- en plattelandsscholen met verschillende denominaties uit alle regio's van Nederland en hebben een leservaring die varieert van 5 tot 38 jaar. Alle 13 personen hebben vorig jaar ook deelgenomen aan de standaardsetting die toen uitgevoerd is op de luistertoets uit het Peilingsonderzoek in het reguliere basisonderwijs. Eén deelnemer van de standaardsetting van vorig jaar was helaas verhinderd. Het panel is bewust zo veel mogelijk gelijk gehouden aan het panel van vorig jaar, om ervoor te zorgen dat de standaardbepaling op de Ankerset Luistervaardigheid zo veel mogelijk gelijk zou verlopen aan de standaardbepaling op de luistertoets uit het Peilingsonderzoek. Om die reden is ook de opzet van de standaardbepaling zo veel mogelijk gelijk gehouden.

3. Procedure standaardbepaling 2018

De opzet van de standaardbepaling 2018 is zoals aangegeven gelijk gehouden aan de opzet van 2017. Net als toen vormden de eisen die het Referentiekader Taal stelt aan de mondelinge taalvaardigheid van leerlingen het uitgangspunt. Om die reden is er bij de start van de standaardbepaling luisteren uitgebreid stilgestaan bij de beschrijvingen die het Referentiekader geeft voor luistervaardigheid 1F en 2F. De panelleden bekeken deze beschrijvingen voorafgaand aan de daadwerkelijke standaardsetting individueel en noteerden voor zichzelf de hoofdpunten. Deze hoofdzaken werden vervolgens plenair besproken, waarbij uitgebreid stilgestaan werd bij 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.

Tijdens de daadwerkelijke standaardbepaling is aan de beoordelaars gevraagd een inhoudelijk oordeel te geven over de hoeveelheid goed gemaakte opgaven die nodig is om een bepaald referentieniveau te halen. De methode die hierbij gebruikt werd, is de 3DC-methode (Data Driven Direct Consensus methode, zie Keuning et al, 2017). 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 onderscheidt: luisteren naar instructies, luisteren naar live publiek en luisteren naar radio, tv en internet. Tijdens de standaardbepalingsprocedure 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. Aan de beoordelaars werd dus de vraag gesteld: "Hoeveel opgaven zou een leerling op dit cluster goed moeten maken als zijn/haar vaardigheid zich precies op de grens wel of geen 1F of wel of geen 2F bevindt?"

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. Nadat alle panelleden hun oordeel gegeven hadden, werden de oordelen in een figuur weergegeven en aan de deelnemers getoond. Over de uitkomsten werd vervolgens gediscussieerd. De discussie werd gestart door aan een aantal beoordelaars te vragen om hun beoordeling bij een cluster toe te

lichten. De andere deelnemers konden hierop reageren. Daarna startte de tweede ronde. In de tweede beoordelingsronde werd de beoordelaars opnieuw gevraagd om in de figuur per cluster de grensscore te markeren. Op basis van de gevoerde discussie kregen de beoordelaars hierdoor de gelegenheid om hun grensscore bij te stellen als ze nieuwe inzichten uit de discussie wilden integreren in hun oordeel.

In de standaardbepaling van 2018 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 item response theorie (IRT) kunnen deze grenzen op een en dezelfde vaardigheidsschaal vergeleken worden.

4. Uitkomsten standaardbepaling 2018

In Tabel 2 worden de uitkomsten van de standaardbepalingen van de 3 clusters weergegeven. Door de grensscores (de afgeronde gemiddeldes) van de 3 clusters op te tellen, krijgen we de standaarden voor de hele luistertoets.

	1F op de 1F set			2F op de 2F set		
	cluster 1	cluster 2	cluster 3	cluster 1	cluster 2	cluster 3
aantal items	12	9	10	11	11	8
gemiddelde	7,62	6,54	6,92	8,00	9,08	7,08
afgerond gemiddelde	8	7	7	8	9	7
mediaan	8	7	7	8	9	7
maximum	9	8	8	9	10	8
minimum	6	5	6	7	8	6

Tabel 2 Grensscores voor clusters luistervaardigheid (na de tweede ronde)

De standaardbepaling in 2018 leverde zodoende de volgende grenzen op voor luistervaardigheid:

De indicatieve standaard 1F Luisteren is geplaatst bij 22 van de 31 van de 1F set opgaven goed.

De indicatieve standaard 2F Luisteren is geplaatst bij 24 van de 30 van de 2F set opgaven goed.

Bij deze grenzen heeft een verwacht percentage van ongeveer 93% van de leerlingen minstens het 1F-niveau bij luisteren, en een verwacht percentage van ongeveer 49% van de leerlingen minstens het 2F-niveau. Deze cesuren zijn echter nog indicatief. In de volgende paragraaf zal op basis van de vergelijking tussen de Ankerset en de luistertoets uit het Peilingsonderzoek een voorstel gedaan worden voor de definitieve cesuren.

4. Vergelijking standaarden Ankerset en standaarden Luistertoets Peilingsonderzoek

De standaardbepaling die in 2017 plaatsvond op de Luistertoets die afgenomen is in het Peilingsonderzoek vond plaats op basis van een enkele itemset van in totaal 30 opgaven. Zowel het niveau 1F als het niveau 2F is op deze verzameling van in totaal 30 items bepaald. Hierbij is ook gebruikgemaakt van de 3DC-methode. De destijds gestelde indicatieve cesuren waren dat er voor het niveau 1F 15 opgaven op de schaal van 30 opgaven goed gemaakt moeten worden om niveau 1F te behalen, en voor het niveau 2F 24 opgaven. Met behulp van IRT is het mogelijk om deze grenzen van

2017 en de grenzen op de Ankerset uit 2018 met elkaar te vergelijken, en ook om de grenzen van 1F en 2F zoals die in 2018 bepaald zijn met elkaar te vergelijken. Om deze vergelijking te kunnen maken, is het materiaal uit de Ankerset in het voorjaar van 2018 samen met de Luistertoets afgenomen bij een groep van 228 leerlingen uit groep 8.

Met behulp van IRT kan de relatie tussen de test en de vaardigheid van de leerlingen afgebeeld worden. Het is hierdoor mogelijk elk van de grensscores eerst te vertalen naar de vaardigheid, en dan naar iedere verzameling opgaven die ook op die schaal is af te beelden.

Door middel van die techniek kunnen de resultaten van de twee jaren op de verschillende verzamelingen opgaven met elkaar vergeleken worden. De grenzen kunnen over de verschillende verzamelingen heen verschillen: de ene schaal is nu eenmaal moeilijker dan de andere. In Tabel 3 worden de cesuren die gezet zijn op zowel de luistertoets uit het Peilingsonderzoek en de Ankerset Luisteren met elkaar vergeleken en vertaald naar de standaarden die in 2017 en 2018 gezet zijn.

laagste score die het niveau oplevert (cesuurpunten)

Standaard bepaling gebruikte itemset	Niveau Resultaten uit	1F		2F	
		2017	2018	2017	2018
	2017 set Peilingsonderzoek	15	15	24	22
	1F set Ankerset	22	22	29	27
	2F set Ankerset	18	18	26	24

Tabel 3 Vertaling cesuren 2017 naar 2018 en andersom

De resultaten over de verschillende jaren horen niet van elkaar te verschillen. Bij de 1F is dat ook niet het geval. Bij de standaarden voor 2F zien we wel verschillen. Aangezien de standaardbepaling van 2017 en 2018 niet kwalitatief van elkaar verschillen, wordt de informatie geoptimaliseerd door het gemiddelde van de twee jaren te nemen. Dat houdt in dat de grens voor 2F ligt op 25 opgaven goed van de 30 op de Ankerset 2F-verzameling. De grens op de standaard 2F uit het Peilingsonderzoek wordt daarmee op dezelfde wijze gezet op 23 van de 30 opgaven goed.

5. Definitieve cesuren Ankerset Luisteren en Luistertoets Peilingsonderzoek

De standaardbepaling in 2018 en de vergelijking van de standaardsetting op de ankerset en de standaardsetting op de Luistertoets uit het Peilingsonderzoek leveren zodoende de volgende definitieve cesuren op voor de Ankerset Luistervaardigheid:

Definitieve cesuren Ankerset 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 goed.

Bij deze grenzen zal ongeveer 93% van de leerlingen het niveau 1F behaald hebben en ongeveer 31% het niveau 2F hebben.

Voor de Luistertoets uit het Peilingsonderzoek worden de definitieve standaarden op basis van de hierboven weergegeven procedure:

Definitieve cesuren Luistertoets Peilingsonderzoek reguliere 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.

Bij deze grenzen behaalt ongeveer 95% van de leerlingen uit het Peilingsonderzoek niveau 1F en ongeveer 39% van de leerlingen niveau 2F.

Literatuur

Keuning, J., Straat, H., and R.C.W. Feskens (2017). The Data-Driven Direct Consensus (3DC) Procedure: A New Approach to Standard Bepaling. In *Standard Bepaling – International State of Research and Practices in the Nordic Countries* (Eds: Sigrid Blömeke and Jan-Eric Gustafsson). Springer series Methodology of Educational Measurement and Assessment, pp 263-278).