

Rapportage Standaardbepalingen  
Peilingsonderzoek mondelinge taalvaardigheid so/sbo  
2018

Bas Hemker

Kirsten van Ingen

Roelien Linthorst

Stichting Cito

5 november 2018

## Inleiding

Stichting Cito heeft in opdracht van het College voor Toetsen en Examens (CvTE) twee standaardbepalingsbijeenvakommen uitgevoerd in het kader van het Peilingsonderzoek Mondelinge Taalvaardigheid einde speciaal (basis)onderwijs. Het betreft een standaardbepaling op twee spreektaken en een standaardbepaling op een gesprekstaak, allen voor niveau 1F en niveau 2F. In dit rapport doen we verslag van deze standaardbepalingen en de gezette cesuren.

De cesuren die op basis van de standaardsettingen opgeleverd worden, betreffen taakgerichte cesuren. Om begrijpelijke redenen is het niet mogelijk om in het Peilingsonderzoek so/sbo alle mogelijke spreek- en gespreksdoelen en taken uit het Referentiekader Taal terug te laten komen. Er is daarom een selectie gemaakt van twee spreektaken een één gesprekstaak. Omdat de gezette standaarden taakafhankelijk zijn, zijn de standaarden niet representatief voor de eisen die het Referentiekader stelt aan spreekvaardigheid en gespreksvaardigheid in zijn geheel. Omdat er geen Referentieset spreek- en gespreksvaardigheid beschikbaar is, zoals bij leesvaardigheid, luistervaardigheid en taalverzorging wel het geval is, kunnen de gezette standaarden daarnaast ook niet geankerd worden aan een Referentieset die een breder beeld van de domeinen zou geven. Daarmee moet rekening gehouden worden bij de interpretatie van de gezette standaarden.

De standaardbepalingen hebben plaatsgevonden op vrijdag 12 oktober en maandag 15 oktober 2018. Alle standaardbepalingen zijn uitgevoerd door hetzelfde expertpanel. In paragraaf 1 beschrijven we de samenstelling van dit expertpanel. Vervolgens beschrijven we in de volgende paragrafen de opzet en de uitkomsten van de standaardbepalingen en de implicaties van deze uitkomsten. Daarbij gaan we ook in op een aanvullend onderzoek dat uitgevoerd is in het kader van de vergelijkbaarheid van de gezette standaarden so/sbo met die van het Peilingsonderzoek in het reguliere basisonderwijs.

Nog twee opmerkingen vooraf:

- 1) Het Peilingsonderzoek Mondelinge Taalvaardigheid einde speciaal (basis)onderwijs is uitgevoerd door een consortium van onder andere ITTA en Cito BV De standaardsettingen zijn uitgevoerd door Stichting Cito op vergelijkbare wijze als de standaardsettingen voor het in het reguliere basisonderwijs. Dit Peilingsonderzoek in het reguliere basisonderwijs is uitgevoerd door een consortium van onder andere Bureau ICE en het Expertisecentrum Nederlands. In het Peilingsonderzoek in het reguliere basisonderwijs is gebruikgemaakt van een scoreschaal die in dit rapport aangeduid wordt met ICE-schaal. In het Peilingsonderzoek in het so/sbo is naast de ICE-schaal gebruikgemaakt van een tweede scoreschaal. Deze schaal wordt aangeduid met Cito.bv-schaal.
- 2) Stichting Cito heeft voor de standaardsettingen een databestand ontvangen van het consortium van het Peilingsonderzoek so/sbo met daarin de scores op de ICE-schaal en de Cito.bv-schaal van de leerlingen die dubbel beoordeeld zijn. Dit betreft een deel van de leerlingen die meegedaan heeft aan het Peilingsonderzoek. Omdat Stichting Cito niet de beschikking had over het complete databestand, kan niet gerapporteerd worden over de implicaties van de gezette standaarden op de totale leerlingpopulatie so/sbo, die overigens ook zeer divers is. Relevanter is echter de vergelijking van de gezette standaarden so/sbo met de standaarden uit het reguliere basisonderwijs. Die standaarden moeten dezelfde betekenis hebben. Het aanvullende onderzoek geeft hierover enige uitsluitel en daarover wordt in dit rapport wel gerapporteerd.

## 1. Het expertpanel standaardbepalingen

Het panel dat de standaarden op de taken uit het Peilingsonderzoek so/sbo gezet heeft, bestond in totaal uit zestien personen. Omdat één persoon de dag voorafgaand aan de standaardbepalingen aangaf toch niet aanwezig te kunnen zijn, zijn de standaardbepalingen uiteindelijk uitgevoerd door in totaal vijftien verschillende personen. In Tabel 1 wordt de achtergrond van de deelnemers en hun aanwezigheid weergegeven.

Tabel 1 Overzicht deelnemers standaardbepalingen

deelnemer	achtergrond	aanwezig 12 oktober	aanwezig 15 oktober	deelnemer standaardsetting regulier po voorjaar 2017
1	basisonderwijs	x	x	ja
2	taalexpert	x	x	ja
3	basisonderwijs	x	x	ja
4	basisonderwijs	x	x	ja
5	basisonderwijs	x	x	ja
6	(speciaal) basisonderwijs	x	x	ja
7	(speciaal) basisonderwijs	x	x	ja
8	taalexpert	x	x	ja
9	basisonderwijs	x	x	ja
10	taalexpert	x		ja
11	basisonderwijs	x	x	ja
12	taalexpert	x	x	nee
13	(speciaal) basisonderwijs		x	ja
14	taalexpert		x	nee
15	basisonderwijs		x	nee
	totaal	12 personen	14 personen	

Bij de samenstelling van het expertpanel is met verschillende factoren rekening gehouden. In de eerste plaats is er gezorgd voor een mix van taalexperts, leerkrachten uit het reguliere basisonderwijs en leerkrachten uit het so of sbo. Het uiteindelijke expertpanel bestond uit vijf taalexperts, zeven leerkrachten uit het reguliere basisonderwijs en drie leerkrachten die (ook) werkzaam zijn in het so of sbo. Daarnaast is er een gezorgd voor een grote overlap met het expertpanel dat in het voorjaar van 2017 deelgenomen heeft aan de standaardsettingen spreken en gesprekken voor het Peilingsonderzoek dat uitgevoerd is in het reguliere basisonderwijs. Omdat het belangrijk was dat de standaardsettingen zoveel mogelijk een gelijke opzet hadden als deze eerdere standaardsettingen (zie ook paragraaf 2), was het van belang dat er een grote overlap in deelnemers zou zijn. Twaalf van de vijftien leden van dit expertpanel hadden dan ook al deelgenomen aan de standaardsettingen in 2017. Tot slot is ervoor gezorgd dat de deelnemende leerkrachten afkomstig zijn uit verschillende regio's van Nederland, dat ze werkzaam zijn op zowel stads- als plattelandsscholen en op scholen met een verschillende denominatie en dat ze variëren in mate van leservaring. Op deze manier is gepoogd tot een representatieve groep leerkrachten te komen.

## **2. Aanvullend onderzoek in het reguliere basisonderwijs**

Het Peilingsonderzoek Mondelinge Taalvaardigheid einde speciaal (basis)onderwijs volgt op het Peilingsonderzoek Mondelinge Taalvaardigheid in het reguliere basisonderwijs. In beide peilingsonderzoeken worden de spreekvaardigheid, de gespreksvaardigheid en de luistervaardigheid van leerlingen in kaart gebracht. De prestatiestandaarden mondelinge taalvaardigheid 1F en 2F zijn voor het speciaal (basis)onderwijs hetzelfde als voor het reguliere basisonderwijs. Het was echter om inhoudelijke redenen niet mogelijk om in het Peilingsonderzoek so/sbo gebruik te maken van dezelfde spreek- en gesprekstaken als in het Peilingsonderzoek in het reguliere basisonderwijs. Dit is problematisch wanneer aan de uiteindelijke standaarden 1F en 2F wel dezelfde betekenis toegekend moet worden.

Om de standaardsettingen so/sbo zoveel mogelijk te laten aansluiten bij de standaardsettingen in het reguliere basisonderwijs zijn verschillende maatregelen getroffen. Zoals in paragraaf 1 al weergegeven is, is er veel overlap tussen de leden van de verschillende expertpanels. Ook de opzet van de standaardbepalingen so/sbo is zoveel mogelijk gelijk gehouden aan die van de standaardbepalingen in het reguliere basisonderwijs. Daarnaast is in beide Peilingsonderzoeken gebruikgemaakt van dezelfde beoordelingsmodellen voor spreekvaardigheid en gespreksvaardigheid. Desondanks was er ook behoefte aan een empirische validatie. Daarom is er een kleinschalig onderzoek uitgevoerd om meer te kunnen zeggen over de gezette standaarden in het so/sbo ten opzichte van de standaarden uit het Peilingsonderzoek in het reguliere basisonderwijs.

In dit aanvullende onderzoek zijn de spreek- en gesprekstaken die gebruikt zijn in het Peilingsonderzoek so/sbo voorgelegd aan representatieve groep leerlingen uit groep 8 van het reguliere basisonderwijs. Bij de opzet van dit onderzoek werd uitgegaan van deelname van 200 leerlingen van 25 scholen, waarbij telkens 8 leerlingen per school uit één groep (klas) deel zouden nemen aan het onderzoek. Gezien het krappe tijdsbestek waarin het onderzoek uitgevoerd moest worden en het gegeven dat tijdens het onderzoek nog enkele scholen afhaakten, is de gewenste verdeling niet helemaal behaald. Uiteindelijk hebben 18 scholen meegedaan, waarbij de spreektaken voorgelegd zijn aan 178 leerlingen en de gesprekstaken aan 175 leerlingen. Wel is de eis vastgehouden dat er maximaal 8 leerlingen per groep meededen aan het onderzoek. Doordat 4 scholen met 2 verschillende groepen meededen, zijn de leerlingen uiteindelijk afkomstig uit 22 verschillende groepen.

Het aanvullende onderzoek is uitgevoerd in september 2018. De afname van de taken gebeurde door dezelfde toetsleiders die ook betrokken waren bij het Peilingsonderzoek in het so/sbo. In de volgende paragrafen waarin de uitkomsten van de standaardsettingen besproken worden, zal ook ingegaan worden op de resultaten van dit aanvullende onderzoek.

### **3. Standaardbepalingen spreekvaardigheid**

#### **3.1 De spreektaak**

De standaardbepalingen spreekvaardigheid zijn uitgevoerd op de twee spreektaken die in het Peilingsonderzoek so/sbo afgenomen zijn. De eerste spreektaak, die hieronder aangeduid wordt als de taak 'Jeugdjournaal' is een navertel-taak. Leerlingen kregen een fragment te zien uit het Jeugdjournaal over een baby-otter<sup>1</sup>. In het fragment wordt verteld over een baby-otter die gevonden is in Alaska. Vervolgens werd de leerlingen gevraagd om aan de toetsleider, die het fragment niet gezien had, zo precies mogelijk te vertellen wat er in het filmpje te zien is. De leerlingen kregen daarbij een blad aangereikt waarop enkele woorden stonden die als geheugensteuntje konden dienen.

De tweede spreektaak, die hieronder aangeduid wordt als de taak 'Dobbelsteen', betrof een taak waarbij de leerlingen zelf een verhaaltje moesten vertellen aan de hand van vijf dobbelstenen met afbeeldingen. Bij deze spreektaak werd gebruikgemaakt van de 'Rory's Story Cubes': een setje van negen dobbelstenen met verschillende afbeeldingen. Leerlingen gooiden met deze negen dobbelstenen, selecteerden vervolgens vijf dobbelstenen en vertelden aan de hand van de afbeeldingen op deze dobbelstenen aan de toetsleider een verhaaltje waarbij ze begonnen met de woorden 'Er was eens..'. Leerlingen mochten meerdere malen (in ieder geval twee) met deze dobbelstenen gooien, waarbij het laatste verhaaltje beoordeeld werd.

Beide spreektaken zijn op video opgenomen. Een selectie van alle leerlingprestaties is dubbel beoordeeld. Voor de standaardbepalingen is gebruikgemaakt van deze dubbel beoordeelde leerlingwerken.

#### **3.2 Procedure standaardbepaling**

Voor de standaardbepalingen spreken is gebruikgemaakt van een studentgerichte methode in plaats van een itemgerichte methode. Dit is gedaan omdat bij deze spreektaken en het gebruikte beoordelingsmodel de leerlingrespons veel meer een eenheid vormt en niet goed op te delen is in afzonderlijk goed te interpreteren opgaven. Dit betekent dat de standaard niet bepaald werd door de evaluatie van items maar van de evaluatie van (voorbeelden van) leerlingwerken. Hiermee kon achterhaald worden welke vaardigheid net voldoende is om indicatief te zijn voor het niveau 1F en welke vaardigheid net noodzakelijk is om het niveau 2F te behalen.

Voor de standaardbepaling plaatsvond, zijn de leerlingwerken al geëvalueerd met behulp van een vastliggend beoordelingsmodel<sup>2</sup>. In de peiling van groep 8 van het reguliere basisonderwijs in 2017 heeft Bureau ICE, één van de uitvoerders van dit Peilingsonderzoek, een scoringsmodel ontwikkeld, met een mogelijke scorering van 0 tot en met 17. Deze score wordt in het vervolg aangeduid met de 'ICE-score'. Voor de peiling in het so/sbo is hetzelfde beoordelingsmodel gebruikt. Naast het ICE-scoringsmodel heeft consortiumpartner Cito BV echter ook een eigen scoremodel gehanteerd met

---

<sup>1</sup> Het fragment is te zien via: <https://jeugdjournaal.nl/artikel/2195522-baby-otter-gered-in-alaska.html>

<sup>2</sup> Het is eerde observatie-codering-model: een oordeel wordt hier niet zozeer gegeven, als wel dat een specifieke geobserveerde respons gecodeerd is. In de naamgeving wordt het wel een beoordelingsmodel genoemd. De beoordeling volgt echter pas nadat er een waardering op basis van de codering toegepast wordt.

een mogelijk scorebereik van 0 tot en met 30. Deze 'Cito.bv-score' is de ongewogen somscore van alle beoordelingsscores, waarbij in tegenstelling tot de 'ICE-score' geen scores worden samengevoegd.

Met behulp van dit scoringsmodel kan bepaald worden hoe de verschillende gecodeerde observaties van de leerlingresponsen tot een beoordeling van de prestatie van de leerling leiden. Dit proces is in een eerder stadium bepaald, waardoor de kwantificering van de prestaties voor alle leerlingen vastlagen voordat de standaardbepalingen plaatsvond. Deze scores vormen de basis voor de standaardbepaling. In het geval van de taken Jeugdjournaal en Dobbelsteen waren er door de twee scoringsmodellen twee scores per uitvoering beschikbaar: de ICE-score van 0 tot en met 17 en de Cito.bv-score van 0 tot en met 30. De score die gebruikt is als basis van de standaardbepaling was de ICE-score. De keuze is gemaakt omdat bij de eerdere standaardbepaling in 2017 ook deze score is gebruikt en we deze standaarden met elkaar wilden vergelijken. We zullen hieronder echter op wens van het consortium ook rapporteren over de cesuur op de Cito.bv-scoreschaal.

Tijdens de standaardbepalingen kregen de standaardbepalers voorbeelden van leerlingprestaties voorgelegd. Om deze voorbeelden te kunnen selecteren zijn de leerlingprestaties geclusterd zodat per (ICE-)scorepunt voorbeelden geselecteerd konden worden. Er zijn, met mogelijk scores van 0 tot en met 17, 18 mogelijke scores die behaald kunnen worden. Een verschil met de keuze in voorbeelden bij de eerdere peiling in het reguliere basisonderwijs was dat er voor de standaardbepaling in 2017 slechts één beoordeling per leerlingprestatie beschikbaar was. Bij de standaardbepaling bij so/sbo waren er bij een deel van de leerlingen twee oordelen beschikbaar, namelijk die van de testleider en een expert van Cito BV of ITTA. Bij de selectie van de voorbeelden hebben we ons gebaseerd op de dubbel beoordeelde leerlingwerken. De geselecteerde voorbeelden betreffen leerlingwerken waarbij de twee beoordelaars het in (zeer) grote mate met elkaar eens waren, niet alleen op de ICE-scoreschaal, maar ook op de Cito.bv-scoreschaal. Bij het merendeel van de voorbeelden verschilden de twee beoordelaars maximaal 1 ICE-scorepunt en maximaal 2-Cito.bv-scorepunten van elkaar. Door dit uitgangspunt te hanteren bij de selectie van de voorbeelden, waren we bij de keuze van voorbeelden redelijk zeker dat de verkregen score passend was bij de leerlingprestatie.

Omdat we ons beperkten tot de voorbeelden die dubbel beoordeeld waren, was het aantal mogelijke voorbeelden in vergelijking met de eerdere standaardbepaling in het reguliere basisonderwijs echter lager. Ook de mogelijke keuzes per score om als voorbeeld per score te nemen lag daardoor lager. Bij een aantal scores waren er geen observaties te vinden, met name aan de onderkant van de scoreschaal: geen enkele leerling had de score 0 of 1 gekregen, en bij de opdracht Dobbelsteen was de laagst gevonden score zelfs 5 (geen scores 0 tot en met 4) op de ICE-schaal. We gaan ervan uit dat de scores van de dubbel beoordeelde voorbeelden niet sterk van de enkel beoordeelde voorbeelden verschillen. Het peilingsverslag waarin de resultaten voor so/sbo beschreven worden zal daarover uitsluitsel moeten geven.

Bij de taak Jeugdjournaal is gebruik gemaakt van een twintigtal voorbeelden. Bij de taak Dobbelsteen is er gebruik gemaakt van 17 voorbeelden. In bijlage 1 is een overzicht weergegeven van de gebruikte voorbeelden en de bijbehorende ICE- en Cito.bv-score. De standaardbepaling bij de twee verschillende taken werd volgens eenzelfde procedure uitgevoerd. Deze procedure is gelijk aan de eerder uitgevoerde procedure voor het Peilingsonderzoek in het reguliere basisonderwijs. Voorafgaand aan de standaardbepalingen werd aan de deelnemers het Referentiekader Mondelinge

Taalvaardigheid en een beschrijving van de spreektaken toegestuurd met het verzoek deze thuis te bestuderen. De standaardbepaling zelf startte met een bespreking van de eisen die in het Referentiekader Taal gesteld worden aan spreekvaardigheid niveau 1F. De deelnemers werd gevraagd dit niveau te karakteriseren, waarbij de kernbegrippen die de deelnemers aangaven, genoteerd werden op scherm. Hierbij was het belangrijk dat alle deelnemers zich konden vinden in de genoemde kernbegrippen. Deze inhoudelijke beschrijving van het Referentiekader spreekvaardigheid vormde het uitgangspunt van de standaardbepaling<sup>3</sup>.

Vervolgens vond de daadwerkelijke standaardbepaling plaats in twee rondes. In een eerste *globale ronde* kregen de deelnemers een aantal filmpjes te zien van leerlingen die de spreektaak uitgevoerd hebben. Deze filmpjes –die in vertoonde prestaties vrijwel de gehele vaardigheidsschaal besloegen<sup>4</sup>– liepen op in beheersingsniveau (plaats op de vaardigheidsschaal spreken), van het laagste behaalde vaardigheidsniveau van de schaal tot het hoogst behaalde vaardigheidsniveau. De deelnemers bekeken de voorbeelden en gaven vervolgens eerst individueel aan bij welk filmpje volgens hen de grens ligt voor beheersing van 1F. Dat wil zeggen: het eerste filmpje waarbij de leerling laat zien dat hij aan de genoemde inhoudelijke eisen vanuit het Referentiekader voor 1F voldoet. Hierna volgde een plenaire bespreking waarbij de deelnemers uitwisselden bij welk filmpje ze de grens gesteld hadden en welke inhoudelijke overwegingen aan hun keuze ten grondslag lagen. De deelnemers konden op basis van deze discussie hun oordeel nog aanpassen. Er werd gevraagd hun oordeel door te geven aan degenen die de standaardbepaling begeleiden, waarmee de *globale cesuur* gesteld was.

Na het stellen van de globale cesuur uit de eerste ronde, volgde op de tweede dag van de standaardbepalingen een tweede ronde, *de gedetailleerde ronde*. In deze gedetailleerde ronde werd ingezoomd op de globale cesuur die in de eerste ronde gesteld is. Deelnemers kregen nu enkele filmpjes te zien van leerlingen die presteren 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 daarna de kans om hun standaard nog bij te stellen op basis van deze discussie. Daarna bepaalden ze hun *definitieve cesuur* voor de taak Jeugdjournaal of Dobbelstenen op 1F. Voor niveau 2F werd de hele bovengenoemde procedure herhaald.

Door te werken met een globale en een gedetailleerde ronde is de standaard bepaald op basis van zoveel mogelijk leerlingprestaties, waarbij tegelijkertijd een zo nauwkeurig mogelijke standaard gesteld wordt. Dit levert een zo goed mogelijke verdeling van benodigde tijd en nauwkeurigheid zonder de standaardbepalers te sturen op het niveau dat ze verwacht worden aan te geven. In bijlage 2 is een overzicht gegeven van het volledige programma van de standaardbepalingen.

De standaardbepalingen in het so/sbo zijn voorafgegaan door een standaardbepaling in het reguliere basisonderwijs en een aanvullend onderzoek waarbij de spreek- en gesprekstaken so/sbo afgenomen zijn in het reguliere basisonderwijs (zie paragraaf 2). Hierdoor hadden we voorafgaand aan de standaardbepalingen al enige informatie over het bereik waarin je het de uiteindelijke standaarden wellicht zou mogen verwachten. In de eerste plaats is het gehanteerde beoordelingsmodel gelijk aan dat van de standaardbepalingen in het reguliere basisonderwijs. Bij die eerdere standaardbepaling in

---

<sup>3</sup> In bijlage 3 is een overzicht gegeven van de kenmerken die de deelnemers op basis van het Referentiekader als uitgangspunt namen bij de bepaling van de standaarden.

<sup>4</sup> Met uitzondering van de laagste punten op de scoreschaal die bij geen van de dubbel beoordeelde leerlingen voorkwamen.

het reguliere PO is de grens voor 1F bepaald tussen 7 en 8. Voor 2F lag dit tussen 11 en 12 scorepunten op de ICE-schaal. Als het scoremodel onafhankelijk van de taken is, dan zouden de te verwachten grenzen in het so/sbo ook rond dit scorebereik moeten liggen. Een andere bron van informatie is de het verwachte aantal percentage leerlingen met vaardigheid 1F. In eerdere peilingsonderzoek in groep 8 was naar voren gekomen dat op basis van de gestelde cesuren 92% van de leerlingen 1F behaalt (en 62% 2F). Bij de standaardbepalingen so/sbo waren er door het aanvullende onderzoek ook (beperkte) gegevens beschikbaar over hoe leerlingen uit groep 8 presteren op de so/sbo -taken. De grens of 92% voor 1F wordt gehaald bij een score van 12. Bij 2F haalt 62% dat vaardigheidsniveau bij een score van 15. De grenzen op basis van deze relatieve percentages liggen hoger dan op basis van de aanname van een scoremodel onafhankelijk van de taak.

### **3.3 Uitkomsten standaardbepaling**

De eerste globale ronde voor de standaardbepaling spreekvaardigheid is uitgevoerd met oordelen van 12 experts. In de eerste ronde kregen deelnemer voorbeelden voorgelegd in het bereik tussen 7,5 en 14,5 scorepunten op de ICE-schaal (1F) en in het bereik tussen 11 en 16,5 scorepunten (2F). Na die eerste globale ronde kon het scorebereik waarin de grenswaarde zou moeten liggen voor 1F met ongeveer 40% verkleind worden. Dit was het geval voor beide scoreschalen en beide taken. Voor de grens van 2F kon het scorebereik zelfs met ongeveer 50% verminderd worden.

De tweede en laatste (gedetailleerde) ronde vond plaats op een (werk)dag na de globale ronde. Het aantal experts was op die dag 14, van wie 10 experts er de eerste dag ook bij waren geweest. Tijdens deze ronde kregen de deelnemers leerlingwerken voorgelegd van scorepunt 9 tot scorepunt 12,5 voor 1F (met kleine verschillen tussen de taken) en van scorepunt 13,5 tot 16,5 voor 2F.

De standaardbepaling bij 1F leverde na gedetailleerde ronde bij de taak Jeugdjournaal een gemiddelde grensscore op van 11,21 op de ICE-schaal en 19,71 op de Cito.bv-schaal. De experts kozen allen óf score 11 óf 12 op de ICE-schaal, en óf 19,5 óf 20 op de Cito.bv-schaal. De standaardafwijking was 0,26 op beide schalen. Bij de taak Dobbelsteen lagen de grensscores net iets hoger: 11,86 en 20,96 op de ICE- en de Cito.bv-schaal. De standaardafwijking was ook iets groter (0,63 en 1,13). Op één expert na kozen de experts allen óf score 12 óf 13 op de ICE-schaal en óf 21 óf 22 op de Cito.bv-schaal<sup>5</sup>.

Op basis van dit gemiddelde en deze standaardfout verwachten we met ongeveer 90% zekerheid dat het werkelijke gemiddelde tussen de 10,8 en 11,6 ligt voor de taak Jeugdjournaal en tussen 10,8 en 12,9 voor Dobbelsteen. Op de meer Cito.bv-schaal zijn die grenzen 19,3 en 20,1, en 19 en 22,8. Voor Jeugdjournaal is de grens daarmee bij 11 punten gelegd op de ICE-schaal en op 20 op de Cito.bv-schaal. Voor Dobbelsteen is dat 12 punten voor de ICE-schaal en 21 punten op de Cito.bv-schaal.

De standaardbepaling bij 2F leverde na de tweede ronde bij Jeugdjournaal een unaniem oordeel op. Alle experts vonden dat de grens bij 15,5 op de ICE-schaal en 27 op de Cito.bv-schaal moest liggen. Bij de taak Dobbelsteen vonden 12 van de experts dat de grens bij 15,5 (ICE) of 27 (Cito.bv) moest liggen en 2 beoordelaars vonden 16 (ICE) en 28 (Cito.bv). De gemiddelde grenzen lagen daardoor bij 15,57 en 27,14 met een standaardafwijking van 0,18 en 0,36 respectievelijk. De 90%'s-range ligt

---

<sup>5</sup> De expert die hiervan afweek koos een score van één punt lager.



hiermee op 15,3-15,9 en 26,5-27,7. De grenzen voor 2F komen hiermee voor beide taken op 16 punten (ICE-schaal) en 27 punten (Cito.bv-schaal).

De standaard is de laagste score die het beschreven niveau oplevert. De standaardbepaling levert zodoende de volgende grenzen op bij deze twee taken en op de twee scoreschalen:

<b>ICE-scoreschaal:</b>	
De standaard 1F Spreken bij Jeugdjournaal is geplaatst bij 11 van de 17 scorepunten.	
De standaard 1F Spreken bij Dobbelsteen is geplaatst bij 12 van de 17 scorepunten.	
De standaard 2F Spreken bij beide taken geplaatst bij 16 van de 17 scorepunten.	
<b>Cito.bv-scoreschaal:</b>	
De standaard 1F Spreken bij Jeugdjournaal is geplaatst bij 20 van de 30 scorepunten.	
De standaard 1F Spreken bij Dobbelsteen is geplaatst bij 21 van de 30 scorepunten.	
De standaard 2F Spreken bij beide taken geplaatst bij 27 van de 30 scorepunten	

Het is duidelijk dat de grenspunten hoger liggen dan bij de eerder bekeken taak in het Peilingsonderzoek in het regulier basisonderwijs. Daar lag de grens (ICE-schaal) op 8 punten (1F) en 12 punten (2F). Als de vaardigheid nodig om de grens te halen gelijk is, betekent dat het op de nieuwe taken dus gemakkelijker hoge scores te behalen. Dat is ook niet verwonderlijk aangezien de taken opgezet waren om geschikt te zijn binnen het so/sbo. De oude taak werd daarvoor onder andere te moeilijk geacht. Het geeft ook aan dat de Cito.bv-scoreschaal hier een toegevoegde waarde heeft. Zo is het scorebereik om het niveau 2F te behalen bij de twee nieuwe taken wel heel beperkt (16 of 17), en daarmee is ook de mogelijkheid onderscheid tussen de verschillende niveaus van vaardige leerlingen te maken beperkt.

Zoals eerder genoemd zijn de so/sbo -taken ook bij een beperkt aantal leerlingen in groep 8 van het reguliere basisonderwijs afgenomen om een vergelijking te kunnen maken tussen beide Peilingsonderzoeken. In Tabel 2 is weergegeven welk percentage van deze leerlingen uit groep 8 van het reguliere basisonderwijs niveau 1F en 2F zouden behalen, gezien de hierboven gezette cesuren. Daarbij moet opgemerkt worden dat de taken so/sbo slechts bij een beperkte groep leerlingen (N=178) uit het reguliere PO zijn afgenomen.

Tabel 2: percentage leerlingen (%Iln) in de steekproef dat per taak de referentieniveaus haalt

Niveau	ICE-scoreschaal					Cito.bv-scoreschaal			
	Jeugdjournaal		Dobbelsteen			Jeugdjournaal		Dobbelsteen	
	Cesuur	% Iln	Cesuur	% Iln		Cesuur	% Iln	Cesuur	% Iln
<b>1F(+)</b>	11	98%	12	97%	<b>1F(+)</b>	20	97%	21	96%
<b>2F(+)</b>	16	42%	16	47%	<b>2F(+)</b>	27	50%	27	57%

Bij het Peilingsonderzoek in het reguliere basisonderwijs lag het percentage 1F op 92% en het percentage 2F op 62%. Bij de 178 leerlingen die hier onderzocht zijn, liggen de waarden op de

Cito.bv-schaal daar dichterbij in de buurt dan op de ICE-schaal. Als de 178 geheel representatief zouden zijn zou dat betekenen dat de grens bij 1F hier nu iets lager gezet is, en bij 2F iets hoger. Het aanpassen van de cesuren<sup>6</sup> zou de percentages meer gelijk kunnen brengen met die eerder gevonden waren. Het aantal observaties waar deze gegevens mee verzameld zijn, is echter te beperkt om op basis hiervan de grenzen aan te passen. Bovendien gaat het om geringe afwijkingen. Daarbij moet nog opgemerkt worden dat tijdens de standaardbepalingen naar voren kwam dat de so/sbo -taken inhoudelijk gezien iets minder geschikt lijken om niveau 2F te meten, omdat de eisen die het referentiekader 2F stelt aan de spreekvaardigheid van leerlingen met de taak Jeugdjournaal en de taak Dobbelstenen niet volledig in beeld gebracht konden worden<sup>7</sup>.

### 3.4 Discussie standaardbepaling

Bij de standaarden voor spreken moet vooraf opgemerkt worden dat het hier gaat om taakgerichte cesuren. Het generaliseren naar al het spreken in het basisonderwijs wordt daarmee redelijk lastig.

Ook moet er rekening mee gehouden worden dat het oordeel een enkele verzameling van experts betreft, die deze producten in een standaardbepalingssessie moeten beoordelen. Het is natuurlijk mogelijk dat een andere verzameling experts op een ander oordeel uit zou komen. Daarbij is het wel zo dat de experts met zorg zijn geselecteerd, zoals in paragraaf 1 is weergegeven. Daarnaast is enige evaluatie van de oordelen van de experts ten opzichte van de beoordelaars wel mogelijk. De laatste vraag in het beoordelingsschema bij de spreektaken betrof een globale beoordeling waar de beoordelaars (in dezen de toetsleiders en de tweede onafhankelijke beoordelaar) de beoordeelde leerling op een driepuntschaal (matig-voldoende-goed) diende te beoordelen. Deze beoordelingen moesten plaatsvinden met ook de Referentieniveau is het achterhoofd.

Het is mogelijk te bepalen wat de samenhang is tussen deze inschatting ter plekke met de inschatting op basis van de toepassing van de cesuurscores voor de standaarden voor de leerlingen die meegedaan hebben aan het aanvullende onderzoek. In dat geval komt 75% tot 80% van de kandidaten in dezelfde categorie terecht (<1F/Matig; 1F/Voldoende; 2F/Goed) bij de twee taken en bij de twee typen scores. Geen enkele 'matig' beoordeelde leerling heeft 2F en geen enkele 'goed' beoordeelde leerling zit onder 1F. De gewogen kappa's (overeenstemmingsmaat) liggen tussen 0,61 en 0,71. De Cito.bv-scores leveren een betere overeenstemming op. Een belangrijk verschil met de standaardbepalingsprocedure is echter dat het niet is vastgesteld dat een voldoende het niveau 1F aanduidt, en goed het niveau 2F.

Op basis van de maximale overeenstemming zouden alternatieve cesuurscores bepaald kunnen worden. Deze cesuren verschillen maximaal 1 scorepunt met de huidige cesuren. Aanpassing van de cesuren is echter niet wenselijk omdat de standaardbepalingsprocedure veel nauwkeuriger is uitgevoerd, met ook echt de Referentieniveaus als ijkpunt. Een ander belangrijk punt is dat het aantal

---

<sup>6</sup>Een aanpassing hier zou zijn: zet voor beide taken en op beide schalen de cesuur van 1F een punt hoger, en de cesuur van 2F een punt lager. In dat geval zou op beide taken en met beide schalen het percentage leerlingen dat 1F haalt op ongeveer 94% liggen. Het percentage leerling dat dan 2F haalt zou verschillen over de taken: bij Jeugdjournaal zou dan 62% 2F halen (beide schalen) en bij Dobbelsteen zou dat 74% zijn (beide schalen).

<sup>7</sup>Het gaat dan met name om zaken als eigen mening kunnen geven en beargumenteren en zelf verfijndere structuur kunnen aanbrenge. Omdat de spreektaken erg kort waren en behoorlijk voorgestructureerd kwamen deze punten niet naar voren in de leerlingwerken.

leerlingen waar dit nu op gebaseerd is, beperkt is tot 178 leerlingen. Het is echter wel goed te weten dat de gezette standaarden ook om en nabij het 'optimale beslispunt' zitten voor de grenzen onvoldoende/voldoende en voldoende/goed.

Daarbij speelt ook mee dat de taken spreekvaardigheid inhoudelijk gezien wat minder geschikt lijken om niveau 2F te meten, zoals ook in paragraaf 3.3 genoemd is. Het scorebereik dat een kwalificatie 2F oplevert, is daarbij ook smal en ligt (zeer) hoog op de scoreschaal. De grenzen voor 2F moeten diensgevolge met enige voorzichtigheid bekeken worden, waarbij een vertaalslag naar het reguliere PO ook met enige voorzichtigheid gemaakt moet worden.

## **4. Standaardbepaling gesprekken**

### **4.1 De gesprekstaak**

De standaardbepaling Gespreksvaardigheid is gebaseerd op een gesprekstaak waarbij de leerlingen in drie- of viertallen een bestaand samenwerkingsspel uitvoeren. Leerlingen kruipen tijdens het spel in de rol van 'vissers' die vissen naar de haven moeten brengen met hun bootjes. De leerlingen moeten gezamenlijk alle vissen binnenhalen voordat de storm opsteekt.

Het spel wordt gespeeld op een speelbord met daarop een eiland en daaromheen een zee, ingedeeld in vakjes, waarin vissen liggen. Elke leerling heeft een eigen haven en een bootje in een eigen kleur waarmee de vissen vervoerd kunnen worden. Wanneer de leerling de beurt heeft, gooit hij met een dobbelsteen. Vervolgens bespreken de leerlingen welke speler het gegooide aantal ogen het beste kan gebruiken om bij een vis te komen en/of deze vis naar een haven op het eiland te brengen. Als alle vissen op het eiland zijn, hebben de leerlingen samen gewonnen. Op de dobbelsteen staat echter ook een wolkje. Wanneer een wolkje wordt gegooid, moet er naast het spel één van de 12 wolkjes omgedraaid worden. Wanneer alle 12 wolkjes zijn omgedraaid, stormt het zo hard dat de boten omwaaien en "het spel" heeft gewonnen.

### **4.2 Procedure standaardbepaling**

De procedure voor de standaardbepalingen bij gesprekken komt overeen met de procedure die gebruikt is bij spreken. Er is om dezelfde redenen als bij spreken gebruikgemaakt van een studentgerichte methode. Dit betekent dat de standaard niet bepaald werd door de evaluatie van items maar van de evaluatie van (voorbeelden van) leerlingprestaties. Hiermee kan achterhaald worden welke vaardigheid net voldoende is om indicatief te zijn voor het niveau 1F en welke vaardigheid net noodzakelijk is om het niveau 2F te behalen.

Voordat de standaardbepaling plaatsvond zijn - net als bij spreken - de leerlingresponsen geëvalueerd met behulp van een vastliggend beoordelingsmodel. Op dit beoordelingsmodel is een scoringsmodel toegepast waardoor de leerlingresponsen te ordenen zijn. De ordening op deze scores dient overeen te komen met de gespreksvaardigheid van de leerling. Ook hier kan er weer onderscheid gemaakt worden tussen de eerder gebruikte ICE-score en de Cito.bv-score. Net als bij de spreektaken waren de ICE-scores scores samengevoegd en vond er een hercodering plaats, terwijl de aangeleverde Cito.bv-scores de ongewogen somscores van alle oordelen betroffen. Het scorebereik van de Cito.bv-scores is groter dan van de ICE-scores. Net als bij spreekvaardigheid vormen beide type scores de basis voor de standaardbepaling.

Om de voorbeelden te selecteren zijn de leerlingprestaties geclusterd, zodat per scorepunt een voorbeeld geselecteerd konden worden. Bij gesprekken lopen deze ICE-scores van een minimumwaarde 0 tot en met een maximumwaarde 10. Bij de Cito.bv-scores lopen deze van een minimumwaarde 0 tot en met een maximumwaarde 34. Net als bij de gesprekstaken is de keuze van de voorbeelden bepaald op basis van de voorbeelden die door twee beoordelaars beoordeeld waren. Dat waren ook hier weer een testleider en een expert. Dit had tot voordeel de keuze van voorbeelden bepaald kon worden door hoe zeker we waren dat de verkregen score passend was bij de prestatie. Bij de gekozen voorbeelden waren de twee beoordelaars het in grote mate met elkaar eens. Hierbij is niet alleen naar de ICE-score gekeken maar ook naar de Cito.bv-score. Bij het

merendeel van de voorbeelden verschilden de twee beoordelaars maximaal 1 ICE-scorepunt en maximaal 2-Cito.bv-scorepunten van elkaar. Er zijn net als bij spreekvaardigheid voorbeelden geselecteerd van zowel jongens als meisjes, afkomstig van verschillende scholen. De voorbeelden die gebruikt zijn tijdens de standaardbepalingen zijn weergegeven in bijlage 1.

Het aantal beschikbare voorbeelden verschilt wel per scorepunt. Met name uit de onderste helft van de scoreschaal was het lastig voorbeelden te vinden, aangezien het merendeel van de leerlingen – ook in het so/sbo – zowel bij de ICE-score als bij de Cito.bv-score meer dan 50% van de maximumscore behaalt.

Net als bij spreken startte de standaardbepaling 1F bij gesprekken met een bespreking van de eisen die in het Referentiekader Taal gesteld worden aan niveau 1F. De beschrijvingen vormden het uitgangspunt van de standaardbepaling. In een eerste *globale ronde* van de standaardbepaling kregen de deelnemers een aantal filmpjes te zien van leerlingen die de gesprekstaak uitgevoerd hebben. Hierbij werd duidelijk aangegeven welke leerling van het getoonde drie- of viertal hier beoordeeld werd. Ook hier liepen de scores van de te beoordelen leerlingenresponsen op van laag naar hoog en dekte deze reeks vrijwel de gehele vaardigheidsschaal voor gesprekken. De deelnemers gaven aan bij welk filmpje volgens hen de grens ligt voor beheersing van 1F.

Ook de tweede ronde was gelijk aan die bij spreken: na het stellen van de globale cesuur op basis van de oordelen in de eerste ronde, volgde een tweede *gedetailleerdere ronde*. Daarin werd ingezoomd op de globale cesuur. De experts kregen nu enkele filmpjes te zien van leerlingen die presteren vlak onder, op en vlak boven de cesuur van de eerste ronde. Ze stelden opnieuw de cesuur vast voor deze serie filmpjes. Op basis van eerdere informatie kon de eerste ronde al relatief gericht zijn. Het scorebereik van de scores in de tweede ronde was dan ook al heel klein: op de ICE-schaal 2 punten zowel bij 1F als 2F, en op Cito.bv-schaal 8 punten bij 1F en 4 punten bij 2F.

De ervaringen bij de standaardbepaling bij spreken bevestigde de voordelen van deze methode. Bij gesprekken was het bepalen van de standaard nog iets arbeidsintensiever omdat de filmpjes langer waren, en de interactie tussen de leerlingen in het gesprek ook een rol speelde, waarmee een grotere concentratie van de standaardbepalers gevergd werd om alleen de beoordeelde leerling te evalueren.

### 4.3 Uitkomsten standaardbepaling

De standaardbepaling bij 1F leverde na de tweede ronde een gemiddelde grensscore op van 7,57 van de 10 mogelijke scorepunten, met een standaardafwijking van het gemiddelde van 0,27. Op de Cito.bv-scoreschaal was het gemiddelde 27,82 met een standaardafwijking van 1,19. Met een zekerheid van 90% ligt de standaard tussen de 7 en de 8 op de ICE-scoreschaal en tussen 25,9 en 29,8 op de Cito.bv-scoreschaal.

De standaardbepaling bij 2F leverde na de tweede ronde een gemiddelde grensscore op van 9 van de 10 mogelijke scorepunten op de ICE-scoreschaal, en 31 van de 34 mogelijke scorepunten op de Cito.bv-scoreschaal. In de eerste ronde kozen alle 12 experts voor een voorbeeld met score 9 op de ICE-schaal en 31 op de Cito-bv-schaal. In de tweede ronde kozen weer alle (nu 14) beoordelaars voor het voorbeeld met score 9 op de ICE-schaal en 31 op de Cito.bv-schaal. Dit was een ander voorbeeld dan in de eerste ronde werd vertoond. Voor 2F is zodoende evident dat de cesuur voor 2F moet liggen op 9 dan wel 31, afhankelijk van de gebruikte scoreschaal.

De standaardbepaling levert zodoende de volgende grenzen op bij deze gesprekstaak:

**ICE-scoreschaal**

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

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

**Cito.bv-scoreschaal**

De standaard 1F Gesprekken is geplaatst bij 28 van de 34 scorepunten.

De standaard 2F Gesprekken is geplaatst bij 31 van de 34 scorepunten.

Het is duidelijk dat het scorebereik van scores dat de kwalificatie 1F oplevert wel zeer klein is. Bij de ICE-schaal is dat zelfs maar 1 score, namelijk bij 8. Bij de Cito.bv-scoreschaal zijn er drie scores die de kwalificatie 1F opleveren: 28, 29 en 30.

Net als bij spreken liggen de cesuurgrenzen bij de so/sbo-taak een hoger dan bij de taak uit het Peilingsonderzoek in het reguliere basisonderwijs. Die grenzen lagen bij de oude taak bij 5 voor 1F en 8 voor 2F. De nieuwe taak lijkt daarmee (veel) makkelijker. Als het percentage leerlingen dat 1F dan wel 2F haalt niet enorm daalt bij deze grenzen bij de nieuwe taak, dan is dat een bevestiging dat de nieuwe taak inderdaad gemakkelijker is. Bij de oude taak haalde 87% van de leerlingen het niveau 1F of hoger. Bijna de helft van de leerlingen (49%) haalde bij de oude taak het niveau 2F. Dat houdt in dat 38% van de leerlingen 1F behaalde, maar niet 2F.

Bij de nieuwe taak hebben we weer beschikking over een kleine groep van 174 leerlingen uit groep 8 van het basisonderwijs. Bij het toepassen van de cesuren op de nieuwe taken (zie Tabel 3) is te zien dat de percentages eerder hoger dan lager liggen dan gevonden bij de oude taak. Er zijn met name relatief meer leerlingen in de steekproef die het niveau 2F halen. De beoordelaars waren bij deze 2F-cesuur echter wel heel eensgezind dus het verhogen van de cesuur is niet een optie. Op de ICE-schaal zou bij verhoging met 1 scorepunt alleen de maximumscore 10 de kwalificatie 2F opleveren, en dan nog zou met 53% een hoger percentage leerlingen 2F hebben dan bij de eerdere peiling. Ook bij gespreksvaardigheid moet echter opgemerkt worden dat tijdens de standaardbepaling door de experts werd aangegeven dat de gehanteerde taak geschikter lijkt om niveau 1F te meten dan niveau 2F. Niet alle doelen die in het Referentiekader beschreven worden bij gespreksvaardigheid 2F komen in het samenwerkingsspel naar voren. Net als bij spreekvaardigheid moet een vergelijking op niveau 2F tussen de taak uit het reguliere Peilingsonderzoek en die uit het Peilingsonderzoek so/sbo dan ook met de nodige voorzichtigheid gebeuren.

Tabel 3: *percentage leerlingen (%Iln) in de steekproef reguliere basisonderwijs dat per taak de referentieniveaus haalt op basis van de cesuren so/sbo*

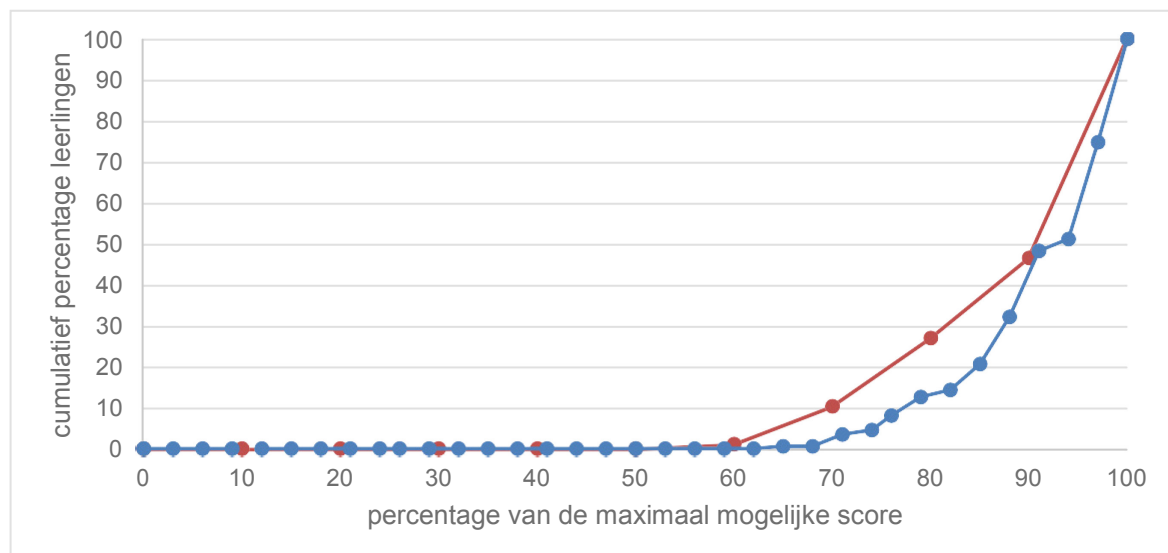
Gesprekken				
Niveau	ICE-schaal		Cito.bv-schaal	
	Cesuur	% Iln	Cesuur	% Iln
<b>1F(+)</b>	8	90%	28	87%
<b>2F(+)</b>	9	73%	31	68%

#### 4.4 Discussie standaardbepaling

Net als bij de gezette standaarden spreekvaardigheid moet er bij de interpretatie van de standaarden gespreksvaardigheid rekening mee gehouden worden dat het taakgerichte cesuren betreft. Ook geldt bij gespreksvaardigheid dat het oordeel een enkele verzameling van experts beoordelaars betreft. Door de zorgvuldige selectie van de standaardbepalers en een zorgvuldige opzet van de standaard sessies hebben we de invloed van ‘toeval’ echter zoveel mogelijk proberen uit te sluiten.

Bij spreekvaardigheid hebben we de gezette standaarden vergeleken met de oordelen matig-voldoende-goed die de beoordelaars (d.w.z. de toetsleiders en de onafhankelijke tweede beoordelaar) als laatste oordeel op het beoordelingsformulier invulden. Bij gespreksvaardigheid is een dergelijke vergelijking echter niet zinvol. Bij de gesprekstaking is duidelijk dat bij de leerlingen in groep 8 van het basisonderwijs (de leerlingen uit het ‘aanvullende onderzoek’) de scoreverdeling zeer scheef is. We kunnen beide scoreschalen in een figuur plaatsen door de scores te transformeren naar het percentage van de maximumscore op de schaal. Een score 0 levert bij beide schalen een percentage 0 op. Het percentage 100 wordt bij de ICE-schaal gehaald bij score 10, en bij de Cito.bv-scoreschaal bij 34. Het is duidelijk dat ongeveer 70% van de leerlingen uit het reguliere basisonderwijs 90% tot en met 100% van de maximale score haalt.

Figuur 1: *cumulatieve verdeling van de scores op de ICE- (rood) en Cito.bv-schaal (blauw)*



Als we de evaluatie door de toetsleider tijdens de toets bekijken, dan wordt de prestatie van 82% van de leerlingen als goed beoordeeld, en slechts 1% als matig. Met een dergelijke zeer scheve verdeling is het ook niet heel zinvol de vergelijking te maken tussen de globale beoordeling door de toetsleider

en het met de cesuur bepaalde niveau. De nieuwe taak voor gesprekken lijkt niet goed geschikt om het niveau 2F te bepalen. De taak is daar te gemakkelijk voor. Echter, als taak voor so/sbo lijkt deze zeker wel te voldoen aangezien de vaardigheid van de leerlingen lager ligt.



## Bijlage 1 Overzicht gebruikte leerlingwerken ('filmpjes') in de standaardbepalingen Spreken en Gesprekken

### Spreektaak 1 Jeugdjournaal

globale ronde				detailronde			
	jeugdjournaal 1F				jeugdjournaal 1F		
	leerlingnummer	ICE-score	Cito-score		leerlingnummer	ICE-score	Cito-score
A	LL0494	7,5	13,5	A	LL0311	9	16
B	LL1271	10	18	B	LL0796	10,5	18
C	LL1379	11	20	C	LL0321	11	19,5
D	LL0409	12,5	22	D	LL1282	11,5	20
E	LL0167	14	24	E	LL9046	12,5	22,5

globale ronde				detailronde			
	jeugdjournaal 2F				jeugdjournaal 2F		
	leerlingnummer	ICE-score	Cito-score		leerlingnummer	ICE-score	Cito-score
A	LL1379	11	20	A	LL0179	13,5	22,5
B	LL0409	12,5	22	B	LL0818	14,5	25
C	LL0167	14	24	C	LL1647	15	26
D	LL0131	16	28,5	D	LL0065	15,5	27
E	LL0670	16,5	29	E	LL0202	16	28

### Spreektaak 2 Dobbelstenen

globale ronde				detailronde			
	dobbelstenen 1F				dobbelstenen 1F		
	leerlingnummer	ICE-score	Cito-score		leerlingnummer	ICE-score	Cito-score
A	LL1427	9,5	15,5	A	LL1616	9,5	16
B	LL0237	11	19,5	B	LL0987	10,5	18
C	LL1503	12	21	C	LL0311	11,5	20,5
D	LL1270	14,5	25	D	LL1150	12,5	22
E				E			

globale ronde				detailronde			
	dobbelstenen 2F				dobbelstenen 2F		
	leerlingnummer	ICE-score	Cito-score		leerlingnummer	ICE	Cito
A	LL0237	11	19,5	A	LL0008	14,5	25,5
B	LL1503	12	21	B	LL0409	15	26
C	LL1270	14,5	25	C	LL0955	15,5	27
D	LL1264	16	28,5	D	LL1181	16	28
E				E	LL0913	16,5	29,5

Gesprekstaak Samenwerkingsspel

globale ronde				detailronde			
	gesprekken 1F				gesprekken 1F		
	leerlingnummer	ICE-score	Cito-score		leerlingnummer	ICE-score	Cito-score
A	LL0317	4,5	14,5	A	LL1264	6,5	22,5
B	LL0018	6	20	B	LL0777	7	24
C	LL0167	7	24,5	C	LL0270	7,5	25,5
D	LL1048	8	27,5	D	LL1091	7,5	28
E				E	LL0701	8,5	30,5

globale ronde				detailronde			
	gesprekken 2F				gesprekken 2F		
	leerlingnummer	ICE	Cito		leerlingnummer	ICE-score	Cito-score
A	LL0018	6	20	A	LL1091	7,5	28
B	LL0167	7	24,5	B	LL0701	8,5	30,5
C	LL1048	8	27,5	C	LL0503	9	31
D	LL1290	9	31	D	LL0387	9,5	32
E				E			

## Bijlage 2 Programma standaardbepalingen

Opzet programma vrijdag 12 oktober

10:00 uur	<b>Welkom</b>	Voorstelrondje (i.v.m. nieuwe deelnemers) Korte schets achtergrond standaardsettingen
10:15 uur	<b>Spreken</b> Taken + beoordelingsmodel	Introductie taken: - toelichting opdracht - toelichting beoordelingsmodel Bekijken van twee voorbeeldfilmpjes - toelichting werkwijze standaardsetting
11:00 uur	<b>Spreken 1F globaal</b>	Start 1F: wat vraagt referentiekader? 1F globale ronde jeugdjournaal Bespreking globale ronde Evt. aanpassing standaard globale ronde Globale ronde dobbelsteen Bespreking globale ronde Evt. aanpassing keuzes
	<b>Lunch</b>	
12:45 uur	<b>Spreken 2F globaal</b>	Start 2F: wat vraagt referentiekader? 2F globale ronde jeugdjournaal Bespreking globale ronde jeugdjournaal Evt. aanpassing standaard globale ronde 2F globale ronde dobbelstenen Bespreking globale ronde dobbelstenen Evt. aanpassing standaard globale ronde
13:45 uur	<b>Gesprekken</b> Taken + beoordelingsmodel	Introductie taak: - toelichting opdracht - toelichting beoordelingsmodel
14:15 uur	<b>Gesprekken 1F globaal</b>	Start 1F: wat vraagt referentiekader? 1F globale ronde samenwerkingsspel Bespreking globale ronde samenwerkingsspel Evt. aanpassing standaard globale ronde
	<b>Pauze</b>	
15:30 uur	<b>Gesprekken 2F globaal</b>	Start 2F: wat vraagt referentiekader? 2F globale ronde samenwerkingsspel Bespreking globale ronde samenwerkingsspel Evt. aanpassing standaard globale ronde

Opzet programma maandag 15 oktober

10:00 uur	<b>Welkom</b>	Voorstelrondje (i.v.m. nieuwe deelnemers) Korte terugblik bijeenkomst vrijdag introduceer werkwijze
10:15 uur	<b>Spreken 1F</b>	Korte terugblik punten die vrijdag aangestipt zijn
10:30 uur	<b>Spreken 1F detail</b>	Detailronde jeugdjournaal Bespreking detailronde Evt. aanpassing standaard detailronde Detailronde dobbelstenen (indien nodig) Bespreking detailronde Evt. aanpassing standaard detailronde
11:45 uur	<b>Spreken 2F detail</b>	Korte terugblik punten die vrijdag aangestipt zijn Detailronde jeugdjournaal Bespreking detailronde Evt. aanpassing standaard detailronde
12:30 uur	<b>Lunch</b>	
13:00 uur		Detailronde dobbelstenen Bespreking detailronde Evt. aanpassing standaard detailronde
13:45 uur	<b>Gesprekken 1F detail</b>	Korte terugblik punten die vrijdag aangestipt zijn Detailronde samenwerkingsspel Bespreking detailronde Evt. aanpassing standaard detailronde
15:15 uur	<b>Pauze</b>	
15:30 uur	<b>Gesprekken 2F detail</b>	Korte terugblik punten die vrijdag aangestipt zijn Detailronde samenwerkingsspel Bespreking detailronde Evt. aanpassing standaard detailronde

## Spoken – Wat vraagt referentiekader?

### Kenmerken spreekvaardigheid op niveau 1F en 2F:

1F:	2F:
<p>Eenvoudige bewoordingen begrijpelijk, trouw aan doel Structuur nog niet consequent Beschrijvend Dichtbij belevingswereld Weinig gevarieerd woordgebruik Haperingen in vloeiendheid mogen / zoeken naar woorden Houding/intonatie/mimiek ondersteunen, maar haperingen mogen</p>	<p>Vloeiender Helder: structuur / samenhang Meer gevarieerd taalgebruik Blikveld breder Samengestelde zinnen (voegwoorden/ verbindingswoorden) Formeel / informeel + register Eigen mening + beargumenteren Voorbeelden Kort mag nog steeds</p>



## Gesprekken – Wat vraagt referentiekader?

### Kenmerken gespreksvaardigheid op niveau 1F en 2F:

1F	2F
<p>Eenvoudig Vertrouwde onderwerpen Gesprek gaande kunnen houden / voorspelbaar / eenvoudige structuur Non-verbale ondersteuning van woorden Weinig variatie in woordgebruik Eenvoudige zinsbouw Mening met argumenten Kritisch luisteren en reactie kunnen geven Uitspraak duidelijk genoeg Zoeken naar woorden mag Gesprekken beginnen en beëindigen</p>	<p>Grotere geheel zien: bijdrage aan planning/gezamenlijk doel Meer eisen aan zelfregulatie (checken, doorvragen) Minder vertrouwde onderwerpen Reacties anderen beter in kunnen schatten en kunnen becommentariëren Meer variatie in woord- en taalgebruik Fouten mogen, maar soms wel kunnen herstellen Op de juiste manier aan woord komen of reactie uitstellen Persoonlijke mening en gevoelens Redelijke grammaticale beheersing</p>

