Mondelinge Taalvaardigheid in het s(b)o

Technisch rapport peiling 2018

Jos Keuning
Silvia Rietdijk
Sanneke Schouwstra
Femke Scheltinga

Instituut Taalonderzoek en Taalonderwijs Amsterdam / Cito BV
Amsterdam, 2019

Colofon

De Peiling Mondelinge Taalvaardigheid bij schoolverlaters in het sbo en so is uitgevoerd door een consortium bestaande uit Kennisinstituut voor Taalontwikkeling Amsterdam (ITTA) en Cito BV. Het project is uitgevoerd in opdracht van het Nationaal Regieorgaan Onderwijsonderzoek en de Inspectie van het Onderwijs. Bij vermelding van dit rapport in een publicatie dient gerefereerd te worden aan:

Keuning, J., Rietdijk, S., Schouwstra, S. & Scheltinga, F. (2019). Mondelinge Taalvaardigheid in het s(b)o: Technisch rapport peiling 2018. Amsterdam: Instituut Taalonderzoek en Taalonderwijs Amsterdam.

Projectmanagement
Femke Scheltinga, Jos Keuning

Werving
CED-groep

Logistieke ondersteuning
Cito Process Support, Cito Logistiek support, Cito Data support

Ontwerp toetsen en vragenlijsten
Judith Hollenberg, Femke Scheltinga, Linda Horsels, Beitske Bijleveld, Saskia van Berkel

Psychometrie en methodologie
Silvia Rietdijk, Sanneke Schouwstra, Jos Keuning

Analyse
Silvia Rietdijk, Sanneke Schouwstra, Jos Keuning

Eindredactie
Jos Keuning, Femke Scheltinga

© Instituut Taalonderzoek en Taalonderwijs Amsterdam / Cito BV (2019)

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van de uitgever worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Inhoud

1 Inleiding ... 5

2 Dataverzameling ... 6
2.1 Samenstelling onderzoeksgroep ... 6
2.1.1 Steekproeftrekking ... 6
2.1.2 Werving en respons ... 10
2.1.3 Representativiteit ... 11
2.1.4 Unit-non-respons ... 14
2.2 Materiaal en procedure ... 15
2.2.1 Instrumenten om zicht te krijgen op de mondelinge taalvaardigheid van leerlingen ... 16
2.2.2 Instrumenten om zicht te krijgen op de onderwijs- en ondersteuningsbehoeften van leerlingen ... 17
2.2.3 Instrumenten om het onderwijsleerproces op sbo/so scholen te duiden ... 18
2.2.4 Constructieproces en try-out ... 18
2.2.5 Werving en training toetsleiders ... 20
2.2.6 Organisatie van de afname en het interview ... 20
2.3 Analyseplan ... 21

3 Quickscan ... 23
3.1 Inhoud en afname ... 23
3.1.1 Leerlingkenmerken ... 23
3.1.2 Observatiepunten Luisteren, Spreken en Gesprekken ... 25
3.2 Constructie en beschrijving indices ... 26
3.2.1 Leerlinggegevens ... 27
3.2.2 Onderwijsloopbaan en -behoeften ... 28
3.2.3 Zorgprofiel ... 30
3.3 Missende waarden ... 32
3.4 Variabelen ... 33

4 Luisteren ... 35
4.1 Inhoud en afname ... 35
4.2 Missende waarden ... 36
4.3 Toets- en itemanalyse ... 36
4.4 Item respons theorie analyse ... 37
4.5 Meetnauwkeurigheid ... 39
4.6 Variabelen ... 40

5 Spreken ... 42
5.1 Inhoud en afname ... 42
5.2 Missende waarden ... 44
5.3 Toets- en itemanalyse ... 45
5.4 Beoordelaarsovereenstemming ... 48
5.5 Dimensionaliteit ... 49
5.6 Variabelen ... 52

6 Gesprekken ... 53
6.1 Inhoud en afname ... 53
6.2 Missende waarden ... 55
6.3 Toets- en itemanalyse ... 56
6.4 Beoordelaarsovereenstemming ... 58
6.5 Dimensionaliteit ... 59
6.6 Extra onderdeel: Luisteren in een praktijksituatie ... 60
6.7 Variabelen ... 61

7 Interview ... 63
7.1 Inhoud en afname ... 63
7.2 Onderwijsaanbod en methodegebruik ... 65
7.3 Doelgerichtheid en planmatigheid ... 66
7.4 Onderwijstijd ... 67
7.5 Monitoring en planning ... 67
7.6 Differentiatie ... 68
7.7 Teamontwikkeling ... 69
7.8 Variabelen ... 70

8 Resultaten ... 73
8.1 Beschrijving leerlingprestaties en onderwijsleerproces op scholen ... 73
8.1.1 Prestatiestandaarden voor Luisteren, Spreken en Gesprekken ... 73
8.1.2 Prestatieniveau leerlingen ... 75
8.1.3 Prestatieniveaus in relatie tot achtergrondkenmerken ... 75
8.1.4 Onderwijs op s(b)o scholen ... 80
8.2 Duiding en verklaring leerlingprestaties ... 86
8.2.1 Beheersing op itemniveau ... 86
8.2.2 Samenhang tussen (sub)onderdelen ... 91
8.3 Trends in luisteren ... 93
8.3.1 Equivalering PPON2007 en Peil.Onderwijs ... 93
8.3.2 Prestatievergelijking 2007 en 2018 ... 94
8.4 Voorspellers op schoolniveau ... 96
8.4.1 Imputatie van missende waarden ... 96
8.4.2 Modelspecificatie en procedure ... 98
8.4.3 Duiding uitkomsten multilevelanalyses ... 99

9 Samenvatting en conclusies ... 104

10 Databestand ... 107
10.1 Algemene omschrijving ... 107
10.2 Administratieve variabelen ... 107
10.3 Achtergrondvariabelen op schoolniveau ... 107
10.4 Achtergrondvariabelen op leerlingniveau ... 108
10.5 Variabelen voor de toetsondersdelen ... 110

11 Referenties ... 111

Bijlage 1 Quickscan ... 113
Bijlage 2 Beoordelingsmodel Spreken ... 117
Bijlage 3 Beoordelingsmodel Gesprekken ... 122
Bijlage 4 Interviewleidraad ... 132
Bijlage 5 Aanpassing aan (extra) onderwijsbehoeften ... 139
Bijlage 6 Instructie Gesprekstaak ... 142
Bijlage 7 Geheugensteuntje spreektaak (jeugdjournaal) ... 146

1 Inleiding

In 2017/2018 hebben Kennisinstituut voor Taalontwikkeling Amsterdam (ITTA) en Cito BV, in opdracht van het Nationaal Regierorgaan Onderwijsonderzoek (NRO) en de Inspectie van het Onderwijs, het Peilingsonderzoek Mondelinge Taalvaardigheid bij schoolverlaters in het s(b)o uitgevoerd. Bij de werving van scholen is samengewerkt met de CED-groep. Het peilingsonderzoek was bedoeld om antwoord te geven op de volgende onderzoeksvragen:

1. Op welk referentieniveau (<1F, 1F en 2F) functioneren schoolverlaters in het sbo en so bij de onderdelen Luisteren, Spreken en Gesprekken?
2. Hoe kunnen de prestaties van zwakke, gemiddelde en sterke leerlingen op referentieniveaus <1F, 1F en 2F worden getypeerd als het gaat om kenmerken van de taakuitvoering? En hoe kan de samenhang in uitkomsten op de onderdelen Luisteren, Spreken en Gesprekken worden verklaard?
3. In hoeverre kunnen prestatieverschillen tussen leerlingen verklaard worden vanuit leerlingkenmerken zoals onderwijsbehoeften, geslacht, thuistaal, thuissituatie, en zorgprofielen?
4. Hoe kunnen de deelnemende scholen worden getypeerd als het gaat om kenmerken van het onderwijsleerproces? En in hoeverre hangen verschillen in het onderwijsleerproces samen met de gemiddelde prestaties van leerlingen?
5. In hoeverre zijn de prestaties van schoolverlaters in het sbo in 2018 veranderd ten opzichte van 2007?
6. Welke verschillen zijn er tussen het onderwijsleerproces in het sbo en het so? En hoe zijn deze verschillen (kwalitatief) te duiden?

Om deze onderzoeksvragen te kunnen beantwoorden zijn verschillende instrumenten afgenomen bij leerlingen en (vertegenwoordigers van) scholen. Aan het peilingsonderzoek hebben ruim 1500 leerlingen van 68 verschillende sbo- en so-scholen deelgenomen.

In het eerste deel van dit rapport wordt een technische beschrijving gegeven van de gegevens die in het peilingsonderzoek in het voorjaar van 2018 verzameld zijn. Enerzijds gaat het om toetsgegevens op de verschillende domeinen van mondelinge taalvaardigheid: Luisteren, Spreken en Gesprekken. Anderzijds gaat het om achtergrondkenmerken van leerlingen en scholen die respectievelijk met behulp van een Quickscan en een interview verzameld zijn. In Hoofdstuk 2 wordt eerst de dataverzameling beschreven. Er wordt ingegaan op de constructie van het steekproefkader, de werving van de scholen, de gang van zaken bij de afname van de instrumenten, en de representativiteit van de onderzoeksgroep. Vervolgens wordt in Hoofdstukken 3 tot en met 7 uitgebreid ingegaan op de inhoud en de (psychometrische) eigenschappen van respectievelijk de Quickscan, de toetsen Luisteren, Spreken en Gesprekken, en het interview. Na de technische beschrijving worden in het tweede deel van dit rapport (Hoofdstuk 8) de onderzoeksvragen beantwoord. De leerlingprestaties worden beschreven en geduid in het perspectief van de referentieniveaus, het onderwijsleerproces op de scholen wordt in kaart gebracht, trends ten opzichte van eerdere peilingen worden geanalyseerd, en met behulp van multilevelanalyse wordt onderzoek gedaan naar voorspellers op leerling- en schoolniveau. Dit rapport sluit af met een samenvatting (Hoofdstuk 9) en een omschrijving van het databestand dat bij dit rapport hoort (Hoofdstuk 10).

2 Dataverzameling

2.1 Samenstelling onderzoeksgroep

Het belangrijkste doel van de peiling is om het prestatieniveau van schoolverlaters in het sbo en so cluster 4 in kaart te brengen. Daarvoor is het van belang dat dat de leerlingen die aan het onderzoek deelnemen representatief zijn voor deze populaties. Hieronder beschrijven we hoe geprobeerd is om de representativiteit van de onderzoeksgroep te borgen. We gaan in paragraaf 2.1.1 eerst in op de constructie van het steekproefkader en de wijze van steekproeftrekking. Vervolgens beschrijven we in paragraaf 2.1.2 de werving en presenteren we in paragraaf 2.1.3 de resultaten van de representativiteitsanalyse. We sluiten in paragraaf 2.1.4 af met een non-responsanalyse.

2.1.1 Steekproeftrekking

In de zomer van 2017 is de steekproef getrokken voor de Peiling Mondelinge Taalvaardigheid in het s(b)o. Het uitgangspunt was om een representatieve groep scholen (vestigingen) uit te nodigen voor de peiling, waarna in principe alle schoolverlaters op die scholen aan de peiling konden deelnemen. Ten behoeve van de representativiteit van de onderzoeksgroep is een tweevoudige stratificatie gebruikt, namelijk naar regio en naar het verwachte aantal schoolverlaters. De stratificatie naar regio was expliciet. Dat wil zeggen dat er aparte steekproeven getrokken zijn uit de landelijke groep van scholen in elk van de vier regio's (Noord, Oost, Zuid en West), zodat de verdeling van scholen over deze vier regio's in de steekproef zoveel mogelijk leek op de verdeling in de populatie. De stratificatie naar het verwachte aantal schoolverlaters was impliciet; de scholen binnen elke regio zijn op basis van het verwachte aantal schoolverlaters geordend van klein naar groot en vervolgens is vanaf een willekeurig beginpunt in de lijst elke zoveelste school in de steekproef opgenomen. Op deze manier is ervoor gezorgd dat de steekproef ook qua schoolgrootte zo representatief mogelijk zou moeten zijn voor de hele populatie van scholen. Bij het trekken van de steekproef zijn geen exclusiecriteria gehanteerd: alle scholen konden in de steekproef vallen. Op leerlingniveau was er wel sprake van exclusiecriteria. Leerlingen met een gediagnosticeerde taalontwikkelingsstoornis vielen buiten de onderzoeksgroep en bij de volgende leerlingen is alleen de Quickscan (zie Paragraaf 2.2. en Hoofdstuk 3) ingevuld:

- leerlingen met een beperkt gehoor;
- leerlingen met selectief mutisme;
- leerlingen met afasie;
- leerlingen die minder dan 2 jaar in Nederland verbleven.

Een complicerende factor bij de ontwikkeling van het steekproefkader en het trekken van de steekproef was dat er nauwelijks gegevens beschikbaar waren over het te verwachten aantal schoolverlaters per school. Op basis van de open onderwijsdata die door DUO *online* beschikbaar worden gesteld, is geprobeerd een geïnformeerde schatting te maken van het te verwachten aantal uitstromende leerlingen in de twee populaties (sbo en so cluster 4) en de beoogde steekproeven. Hieronder staat voor het sbo en so beschreven welke informatie we hebben gebruikt, en welke keuzes we hebben gemaakt bij het maken van het steekproefkader.

sbo. Voor de ontwikkeling van het steekproefkader voor het sbo zijn de volgende bestanden gekoppeld en vergeleken aan de hand van de unieke combinaties van BRIN-code vestigingsnummer:

Bestandsnaam	01.-leerlingen-po-soort-po,-cluster,-leeftijd-2016-2017.csv
Peildatum en status	1 oktober 2016, voorlopig
Omschrijving	Het aantal leerlingen in het primair onderwijs (po) per schoolvestiging, het soort primair onderwijs (bo, sbo, vso en so), cluster en leeftijd. Per schoolvestiging staan kenmerken van de vestiging weergegeven zoals postcode, plaats, gemeente, provincie, bevoegd gezag, denominatie.
Bron	https://duo.nl/open_onderwijsdata/databestanden/po/leerlingen-po/po-totaal/bo-gewicht-leeftijd.jsp

Bestandsnaam	04.-alle-vestigingen-speciaal-(basis)onderwijs.xlsx
Peildatum en status	1 juni 2017, definitief
Omschrijving	De adresgegevens van hoofd- en nevenvestigingen van scholen in het speciaal basisonderwijs, speciaal onderwijs en voortgezet speciaal onderwijs: het vestigings- en correspondentieadres, bijhorend bestuur, denominatie en regionale gebiedsindelingen (nodaal, RPA, WGR, COROP, onderwijsgebied, RMC regio).
Bron	https://duo.nl/open_onderwijsdata/databestanden/po/adressen/adressen-po-4.jsp
Bestandsnaam	03.-leerlingen-bo-sbo-swv-vestiging,-leerjaar,-geslacht-2016-2017.csv
Peildatum en status	1 oktober 2016, voorlopig
Omschrijving	Het aantal leerlingen in het basisonderwijs (bo) en speciaal basisonderwijs (sbo) per schoolvestiging, samenwerkingsverband (in de bestanden van 2014 en daarna), leerjaar en geslacht. Per schoolvestiging staan kenmerken van de vestiging weergegeven zoals postcode, plaats, gemeente, provincie, bevoegd gezag, denominatie.
Bron	https://duo.nl/open_onderwijsdata/databestanden/po/leerlingen-po/bo-sbo/bo-leerjaar.jsp

In het eerste bestand stonden 290 scholen, maar één daarvan is niet meegenomen in het steekproefkader omdat het om een dubbeling leek te gaan en er geen koppeling met de andere bestanden mogelijk was (BRIN-code 15LO). Dit betekent dat in het steekproefkader 289 sbo-scholen konden worden opgenomen. Van elke school was per leerjaar het aantal leerlingen zoals geregistreerd in oktober 2016 bekend. Direct bruikbare gegevens over het uitstromen van leerlingen in het sbo waren echter niet beschikbaar. In de onderwijsdata van DUO zijn weliswaar schooladviezen geregistreerd, die ook in het sbo verplicht gegeven moeten worden in leerjaar 8, maar als een leerling in groep 8 bleef zitten, kon dit op verschillende manieren in de administratie verwerkt zijn. Een school kon er bijvoorbeeld voor kiezen om “geen advies mogelijk” in te vullen of de leerling achteraf alsnog aan te merken als een leerjaar 7 leerling. Daarnaast was het mogelijk dat het advies ontbrak, terwijl de leerling weldegelijk uitstroomde, en andersom, er kon een advies geregistreerd zijn, terwijl de leerling in werkelijkheid niet uitstroomde. Om toch een indruk te krijgen van de omvang van de groep schoolverlaters hebben we eerst voor iedere schoolvestiging een schatting gemaakt van het totale aantal leerlingen in leerjaar 8. Op basis van deze schatting en globale informatie over het aantal leerlingen dat in groep 8 van het sbo blijft zitten, hebben we vervolgens een zo realistisch mogelijke schatting gemaakt van het aantal schoolverlaters in schooljaar 2017/2018[1]. Hierbij is als volgt geredeneerd:

a) Het aantal leerlingen in leerjaar 7 in 2016/2017 op een school is een goede indicatie voor leerjaar 8 in 2017/2018 op diezelfde school;
b) Globaal gezien blijft echter zo’n 35 procent van de leerlingen in het sbo in leerjaar 8 zitten. Het bij (a) genoemde aantal moet dus vermenigvuldigd worden met factor 1.35 om een goede schatting te krijgen van het **totale** aantal leerlingen in leerjaar 8 in 2017/2018;
c) Het verwachte aantal schoolverlaters in 2017-2018 is af te leiden uit (a) en (b) door voor elke schoolvestiging 65 procent (100-35) te nemen van de leerlingen die voor het eerst in leerjaar 8 zitten, en 100 procent te nemen van de leerlingen die leerjaar 8 voor de tweede keer doen.

In bovenstaande berekening is geen rekening gehouden met zittenblijvers in leerjaar 7, vervroegde uitstroom vanuit leerjaar 7, of met verschillen tussen scholen in hoeveel leerlingen uitstromen of blijven zitten. Dergelijke gegevens waren namelijk niet beschikbaar. De gemiddelde verwachte omvang van leerjaar 8 in 2017-2018 was op basis van deze redenatie 32.4 leerlingen en de gemiddelde verwachte uitstroom 24.2 leerlingen.

Uit de 289 sbo schoolvestigingen is een steekproef van 70 scholen getrokken. Daarbij hebben we eerst binnen elk expliciet stratum (dus binnen elke regio) de schoolvestigingen gesorteerd op basis van het verwachte aantal schoolverlaters in schooljaar 2017/2018. Vervolgens hebben we willekeurig een getal X gekozen tussen 1 en het aantal scholen in die regio, en is de steekproef gevormd door eerst X te kiezen, en daarna telkens 3 rijen over te slaan en de volgende school te kiezen totdat we het beoogde aantal scholen hadden. Met het oog op mogelijke uitval is iedere schoolvestiging in de steekproef gematcht aan een tweede en derde schoolvestiging uit dezelfde regio en een zo vergelijkbaar mogelijk verwacht leerlingaantal. Hiertoe zijn de schoolvestigingen, die in de gesorteerde lijst per regio direct boven of direct onder een school uit de steekproef staan, willekeurig toegewezen als “eerste reserve” en “tweede reserve”. Deze scholen konden uitgenodigd worden voor de peiling als de school uit de eerste steekproef weigerde om deel te nemen. Gegeven de 70 scholen in de steekproef en de schatting van het aantal schoolverlaters was de verwachting dat in het sbo in totaal een kleine 1700 leerlingen zouden deelnemen aan de peiling.

[1] In een informatievideo van de Academische Werkplaats Onderwijs (zie http://www.academischewerkplaatsonderwijs.nl/over/nieuws/animatie-over-zittenblijven-het-sbo/) wordt gesteld dat in het reguliere bo ongeveer 0.5 procent van de leerlingen blijft zitten in leerjaar 8 en in het sbo ongeveer 35 procent. Vaak gaat het om leerlingen die relatief laat zijn ingestroomd in het sbo en voor wie het voordelig lijkt een extra jaar te blijven. Omdat we geen nadere statistieken of informatie hadden op school- of leerlingniveau gebruikten we de landelijke indicatie van 35 procent in onze schatting van het verwachte aantal leerlingen in leerjaar 8 en van het verwachte aantal uitstromers in 2017/2018.

Dat aantal van 1700 lag aanzienlijk hoger dan het aantal van 700 waar in het onderzoeksvoorstel vanuit was gegaan. Dat komt doordat in het onderzoeksvoorstel een andere werkwijze was gevolgd; het aantal schoolverlaters was toen geschat op basis van het totale aantal 12- en 13-jarigen op de scholen. Tabel 2.1 geeft een gedetailleerde weergave van de aantallen scholen in het steekproefkader en in de getrokken steekproef in het sbo.

Tabel 2.1. *Aantallen (en percentages) sbo-scholen en verwachte aantallen schoolverlaters in 2018*

Telling	Regio	Kader	Steekproef
Schoolvestigingen	Noord	28 (9,7%)	7 (10,0%)
	Oost	67 (23,2%)	16 (22,9%)
	West	139 (48,1%)	34 (48,6%)
	Zuid	55 (19,0%)	13 (18,6%)
	Totaal	289	70
Verwacht aantal uitstromende leerlingen	Noord	783 (11,2%)	225 (13,4%)
	Oost	1709 (24,4%)	399 (23,8%)
	West	3161 (45,2%)	755 (45,5%)
	Zuid	1348 (19,3%)	300 (17,9%)
	Totaal	7001	1679

so cluster 4. Voor de ontwikkeling van het steekproefkader voor het so zijn de volgende bestanden gekoppeld en vergeleken aan de hand van de unieke combinaties van BRIN-code vestigingsnummer:

Bestandsnaam	01.-leerlingen-po-soort-po,-cluster,-leeftijd-2016-2017.csv
Peildatum en status	1 oktober 2016, voorlopig
Omschrijving	Het aantal leerlingen in het primair onderwijs (po) per schoolvestiging, het soort primair onderwijs (bo, sbo, vso en so), cluster en leeftijd. Per schoolvestiging staan kenmerken van de vestiging weergegeven zoals postcode, plaats, gemeente, provincie, bevoegd gezag, denominatie.
Bron	https://duo.nl/open_onderwijsdata/databestanden/po/leerlingen-po/po-totaal/bo-gewicht-leeftijd.jsp
Bestandsnaam	04.-alle-vestigingen-speciaal-(basis)onderwijs.xlsx
Peildatum en status	1 juni 2017, definitief
Omschrijving	De adresgegevens van hoofd- en nevenvestigingen van scholen in het speciaal basisonderwijs, speciaal onderwijs en voortgezet speciaal onderwijs: het vestigings- en correspondentieadres, bijhorend bestuur, denominatie en regionale gebiedsindelingen (nodaal, RPA, WGR, COROP, onderwijsgebied, RMC regio).
Bron	https://duo.nl/open_onderwijsdata/databestanden/po/adressen/adressen-po-4.jsp
Bestandsnaam	Scholen en leerlingen cluster 4 2014.xlsx
Omschrijving	Alle vestigingen met geregistreerde so leerlingen uit cluster 4 (met provincie en regio), waarbij de uitstroom bepaald is o.b.v. koppeling tussen de inschrijvingen 2014 en 2015. T.b.v. het verkrijgen van een indicatie van het aantal schoolverlaters is de eindgroep hierbij gedefinieerd als uitgestroomde leerlingen die op 31 juli 2014 ouder waren dan 10 jaar. Omdat er in het SO geen leerjaren of eindgroepen gedefinieerd zijn, leek dit de beste indicatie. Het betreft echter uitdrukkelijk slechts een benadering van het aantal schoolverlaters uit het so cluster 4.
Bron	Persoonlijke communicatie, Inspectie van het Onderwijs
Bestandsnaam	01.-leerlingen-po-soort-po,-cluster,-leeftijd-2013-2014.csv
Peildatum en status	1 oktober 2013, definitief
Omschrijving	Het aantal leerlingen in het primair onderwijs (po) per schoolvestiging, het soort primair onderwijs (bo, sbo, vso en so), cluster en leeftijd. Per schoolvestiging staan kenmerken van de vestiging weergegeven zoals postcode, plaats, gemeente, provincie, bevoegd gezag, denominatie.
Bron	https://duo.nl/open_onderwijsdata/databestanden/po/leerlingen-po/po-totaal/bo-gewicht-leeftijd.jsp

Het onderscheid tussen cluster 3 en 4 van het so werd in sommige open databestanden al niet meer gemaakt, omdat deze scholen tegenwoordig samen tot het cluster “overig” of “3/4” worden gerekend. In het eerste bestand leken er 298 scholen te zijn uit so cluster 3/4, maar na een koppeling met het tweede

databestand kwamen we uit op een kleiner aantal scholen specifiek binnen cluster 4. In dat bestand hebben namelijk 52 scholen de aanduiding “so cluster 4”, en 167 scholen werden gecategoriseerd als “(v)so cluster 4”, omdat ze (ook) voortgezet speciaal onderwijs verzorgen. Verder stonden er vijf scholen in dat bestand die werden ingedeeld onder cluster 3/4 en (v)so.

Na het koppelen van de gegevens zijn alle scholen verwijderd die voor de peiling niet relevant waren, zoals (exclusief) vso-scholen of cluster 3 scholen. Scholen met een gecombineerd onderwijsaanbod, die aparte rijen hadden met hun aantallen so- en vso-leerlingen, zijn ontdubbeld. Van de meeste scholen werd door de koppeling ook duidelijk of het (voornamelijk) cluster 3 of 4 scholen waren. Er bleven echter vier scholen over (00KM32, 02GD00, 23JT09, en 30EF00) die in beide databestanden stonden als cluster 3/4. Deze vestigingen hadden ook allemaal zowel so- als vso-leerlingen. Via de websites van de scholen is nagegaan of deze tot cluster 4 van het so gerekend kunnen worden en dus tot de beoogde populatie bij deze peiling. Dat bleek niet het geval te zijn en daarom zijn deze scholen niet meegenomen in het steekproefkader. Verder is één school verwijderd (02YN30) omdat deze in recentere bestanden niet meer voorkwam. Samenvattend bevatte het uiteindelijke steekproefkader 130 cluster 4 scholen. Per school was het aantal leerlingen van een bepaalde leeftijd zoals geregistreerd in oktober 2016 bekend. Direct bruikbare gegevens over het uitstromen van leerlingen waren echter ook in het so niet beschikbaar.

Het schatten van het aantal schoolverlaters in so cluster 4 was lastiger dan in het sbo, omdat er in het so geen leerjaren gedefinieerd zijn en er geen uitstroomgegevens geregistreerd worden. Leeftijd leek in eerste instantie de beste (en enige) indicator, maar deze zou weinig zekerheid geboden hebben, omdat er geen vaste leeftijd is waarop leerlingen het so moeten verlaten. Daarom is bij het schatten van het aantal schoolverlaters niet alleen naar leeftijd gekeken maar ook naar uitstroomgegevens van de Inspectie van het Onderwijs. Het ging om een schatting van het aantal leerlingen per schoolvestiging dat aan het einde van schooljaar 2013-2014 was uitgestroomd, onder de aanname dat een leerling die **niet** opnieuw was ingeschreven in schooljaar 2014-2015 én in juli 2014 minstens 11 jaar oud was waarschijnlijk het so had verlaten. Door het aantal geschatte uitstromers, volgens deze definitie, te combineren met het aantal leerlingen per schoolvestiging dat in oktober 2013 10 jaar of ouder was, kon een schatting gemaakt worden van de proportionele uitstroom. Circa 14 procent van de leerlingen die in oktober 2013 10 jaar of ouder waren, leek te zijn uitgestroomd in juli 2014. Omdat de uitstroominformatie niet voor alle schoolvestigingen beschikbaar was en bovendien erg gevoelig zou kunnen zijn voor jaarlijkse variaties op een school, is alleen deze globale schatting gebruikt van het aantal schoolverlaters binnen de genoemde leeftijdscategorie. Deze schatting is in een vervolgstap gecombineerd met het verwachte aantal leerlingen dat in oktober 2017 10 jaar of ouder zou zijn, gebaseerd op het aantal leerlingen van 9 jaar of ouder dat in het open databestand van oktober 2016 vermeld stond. Samenvattend is het aantal schoolverlaters per vestiging dus als volgt berekend:

a) Er is per school gekeken hoeveel leerlingen er in oktober 2016 geregistreerd stonden als 9, 10, 11 of 12 jaar oud.
b) Het geschatte aantal schoolverlaters op de school was dan gelijk aan 14 procent van dit aantal leerlingen. Dat waren immers de leerlingen die in oktober 2017 10 jaar of ouder zouden zijn.

Deze werkwijze leidde tot een gemiddelde verwachting van 7 schoolverlaters per vestiging.

Uit de 130 so-schoolvestigingen hebben we een steekproef van 31 scholen getrokken. De werkwijze bij de steekproeftrekking was hetzelfde als bij het sbo. Eerst zijn binnen elke regio de scholen geordend op basis van het verwachte aantal leerlingen dat in oktober 2017 10 jaar of ouder zou zijn. Vervolgens hebben we willekeurig een schoolvestiging gekozen, en daarna elke vierde vestiging uit de lijst in de steekproef opgenomen. Zo is de hele lijst doorlopen. Net als bij de steekproeftrekking voor het sbo, werden de scholen direct boven en onder de geselecteerde school de gematchte reserveschoolen, die willekeurig werden toegewezen als eerste of tweede reserve. Gegeven de 31 scholen in de steekproef en de schatting van het aantal schoolverlaters was de verwachting dat in het so in totaal ongeveer 230 leerlingen zouden deelnemen aan de peiling. Dat aantal van 230 lag lager dan het aantal van 300 waar in het onderzoeksvoorstel vanuit was gegaan. Dat komt, net als bij het sbo het geval was, door een verschil in werkwijze. In het onderzoeksvoorstel was gekeken naar het totale aantal 12- en 13-jarigen op de scholen, en hierboven naar een (laag geschat) uitstroompercentage van 10+jarige leerlingen. Tabel 2.2 geeft een gedetailleerde weergave van de aantallen scholen in het steekproefkader en in de getrokken steekproef in het so.

Tabel 2.2. *Aantallen (en percentages) so-scholen en verwachte aantallen schoolverlaters in 2018*

Telling	Regio	Kader	Steekproef
Schoolvestigingen	Noord	11 (8,5%)	2 (6,5%)
	Oost	28 (21,5%)	7 (22,6%)
	West	58 (44,6%)	14 (45,2%)
	Zuid	33 (25,4%)	8 (25,8%)
	Totaal	130	31
Verwacht aantal uitstromende leerlingen	Noord	72 (7,5%)	20 (8,7%)
	Oost	226 (23,6%)	50 (21,7%)
	West	404 (42,1%)	101 (43,9%)
	Zuid	257 (26,8%)	59 (25,7%)
	Totaal	959	230

2.1.2 Werving en respons

In september 2017 is gestart met het werven van de scholen. Er zijn verschillende maatregelen genomen om de peiling bij scholen onder de aandacht te brengen:

- Het onderzoek is aangekondigd op de websites van de consortiumpartners. De aankondiging omvatte een korte toelichting op het onderzoek en een filmpje (zie https://youtu.be/Mjq5f-4QzEM).

- Enkele landelijke organisaties hebben scholen geïnformeerd over de peiling. Het ging om het SBO Werkverband, LECSO, de Onderwijsinspectie en OCW.
- Er is een algemeen informatiemailadres aangemaakt: peilMTV@itta.uva.nl.
- Alle scholen in de steekproef zijn via e-mail geïnformeerd over het onderzoek.
- Naar aanleiding van de mailing is met alle scholen telefonisch contact opgenomen om te bespreken wat deelname in zou houden. Bij de telefoongesprekken is een belscript gevolgd, zodat alle scholen dezelfde informatie kregen.
- In een enkel geval is een school bezocht door een onderwijsadviseur van CED of een onderzoeker van ITTA of Cito.

Als een school had aangegeven deel te willen nemen aan de peiling werd aanvullende informatie verzameld en geregistreerd via de volgende vragenlijst:
https://www.survio.com/survey/d/gegevensdeelnemendescholenpeilingsonderzoek.

Medio februari 2018 is de balans opgemaakt en de werving van scholen stopgezet. Tabel 2.3 laat zien wat de deelnamebereidheid was van scholen in de eerste steekproef en hoeveel reservescholen uit de tweede of derde lijst zich hadden aangemeld voor deelname aan de peiling. In de tabel is onderscheid gemaakt tussen sbo- en so-scholen. Zoals we in Tabel 2.3 kunnen zien, lag het gerealiseerde aantal scholen in het sbo (68) en so (23) iets lager dan het beoogde aantal scholen (70 sbo; 30 so). Dat was niet ideaal, maar als we naar de (verwachte) leerlingaantallen keken – gebaseerd op de schattingen van paragraaf 2.1.1 of de schatting van de school zelf – zagen we dat de doelstellingen wel behaald waren. In het sbo streefden we ernaar om ongeveer 1600 á 1700 leerlingen (oorspronkelijk 700) bij de peiling te betrekken en dat leek met 1706 deelnemende leerlingen ruimschoots gelukt te zijn. In het so zaten we met 407 leerlingen ook aanzienlijk hoger dan het leerlingaantal waar we naar streefden (230, oorspronkelijk 300). De keuze om de werving stop te zetten leek dan ook gerechtvaardigd, en zelfs noodzakelijk, omdat het organisatorisch onhaalbaar zou zijn om meer leerlingen bij het onderzoek te betrekken.

Na afloop van de dataverzameling is in kaart gebracht hoeveel scholen en leerlingen feitelijk hadden deelgenomen. Vanwege uitval van scholen en onnauwkeurigheid in de schattingen van de leerlingaantallen bleek het beeld na afloop van de dataverzameling nogal af te wijken van het beeld dat we vooraf hadden.

Aan de rechterkant van Tabel 2.3 is te zien dat 15 van de 68 scholen in het sbo van deelname hebben afgezien, ondanks hun eerdere toezegging om wél aan het onderzoek mee te doen. Op de resterende sbo-scholen bleken in totaal 1253 schoolverlaters te zitten. In het so hebben 8 scholen hun toezegging om deel te nemen ingetrokken. Op de 15 resterende so-scholen zaten in totaal 282 schoolverlaters. Hoewel er in elk onderzoek sprake is van uitval, was de uitval in dit onderzoek onverwacht groot. De oorzaak is niet geheel duidelijk. Wel waren er opvallend veel personele wijzigingen op de scholen en bovendien bestreek de tijd tussen toezegging en feitelijke dataverzameling bij sommige scholen meerdere maanden. Mogelijk was deze periode te lang. Samenvattend laat Tabel 2.3 zien dat de geformuleerde ambities op leerlingniveau, ondanks de onvoorziene uitval, wel in voldoende mate behaald zijn. Op schoolniveau stelde het wervingsresultaat enigszins teleur.

Tabel 2.3. *Deelnamebereidheid in het sbo en so*

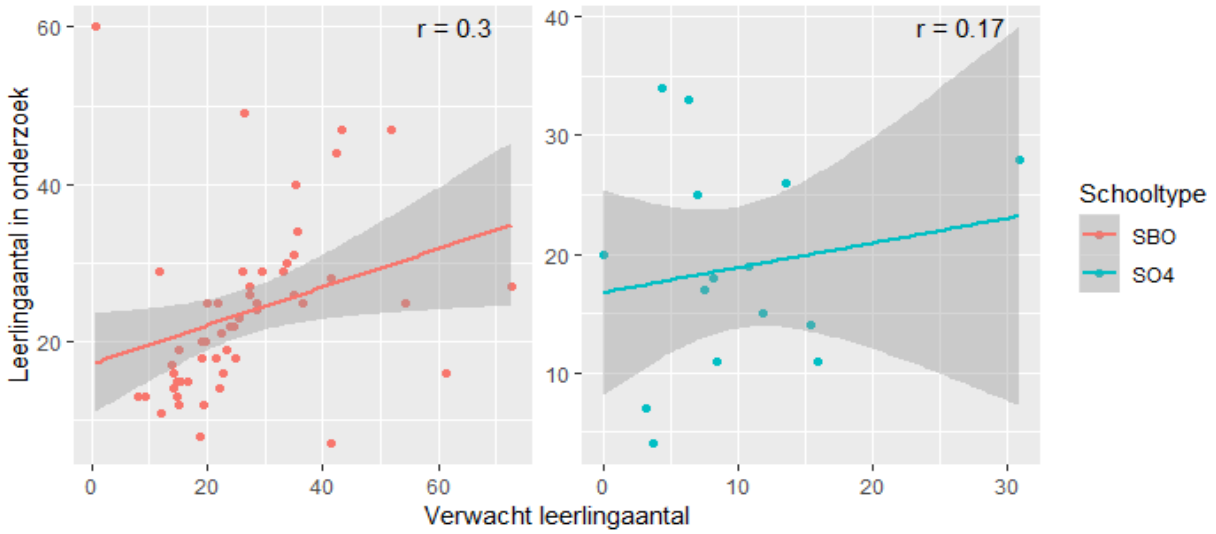
		Inschatting februari 2018		Feitelijke deelname	
Schooltype	Bron	*N* scholen	*N* leerlingen	*N* scholen	*N* leerlingen
sbo	Steekproef	36 (52.9%)	950 (55.7%)	28 (52.8%)	714 (57.0%)
	reserve 1	17 (25.0%)	395 (23.2%)	14 (26.4%)	275 (21.9%)
	reserve 2	15 (22.1%)	361 (21.1%)	11 (20.8%)	264 (21.1%)
	Totaal	68	1706	53	1253
so	Steekproef	11 (47.8%)	195 (47.9%)	5 (33.4%)	107 (37.9%)
	reserve 1	10 (43.5%)	168 (41.3%)	8 (53.3%)	138 (48.9%)
	reserve 2	2 (8.7%)	44 (10.8%)	2 (13.3%)	37 (13.2%)
	Totaal	23	407	15	282

2.1.3 Representativiteit

Voor de interpretatie van de onderzoeksresultaten is niet alleen de steekproefomvang van belang; het is ook belangrijk om te weten of de onderzoeksgroep een goede afspiegeling vormt van de populatie met betrekking tot een aantal kenmerken. Alleen als dit het geval is, zijn conclusies (optimaal) te generaliseren naar de populatie. Twee aspecten werden van primair belang geacht: representativiteit naar schoolgrootte en representativiteit naar achtergrondkenmerken zoals regio en sekse.

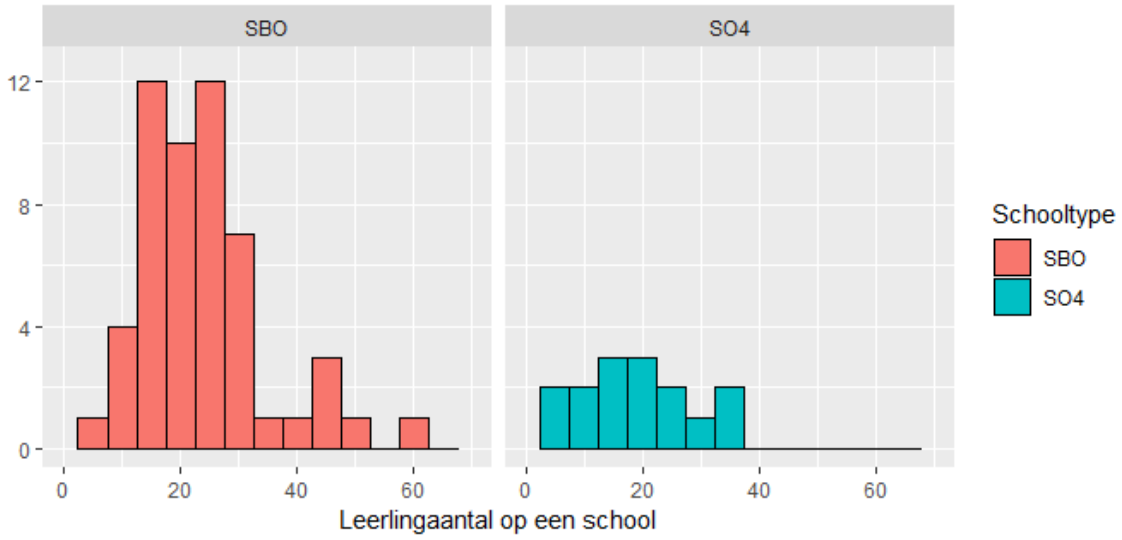
Schoolgrootte
Bij de steekproeftrekking is zo goed als mogelijk gezorgd voor een (impliciete) stratificatie naar de geschatte schoolgrootte, zodat de onderzoeksgroep representatief zou worden wat schoolgrootte betreft. Zoals in paragraaf 2.1.1. al duidelijk is geworden, was het moeilijk om zicht te krijgen op de schoolgrootte, omdat er weinig gegevens beschikbaar waren over de leerlingaantallen in, en specifiek de uitstroom uit, het speciaal (basis-)onderwijs. Er moesten verschillende databronnen gekoppeld worden om vervolgens op basis van de nodige assumpties te komen tot een schatting van het aantal uitstromende leerlingen op een school. Op basis van het definitieve deelnemersbestand hebben we in kaart gebracht in hoeverre het deelnemende leerlingaantal (de schoolverlaters) correspondeerde met het verwachte leerlingaantal. In Figuur 2.1 zijn de verwachte leerlingaantallen op de deelnemende scholen afgezet tegen de daadwerkelijke ingeschreven leerlingaantallen. Daarnaast is de regressielijn (en het 95% betrouwbaarheidsinterval van de voorspelling) weergegeven.

Als de afwijking voornamelijk zou bestaan uit een structurele onderschatting of overschatting van het daadwerkelijke aantal leerlingen, zou dit niet zoveel uitmaken, omdat onze interesse bij de representativiteit uitgaat naar de rangordening van de scholen (van klein naar groot). Figuur 2.1 laat echter zien dat de vooraf ingeschatte leerlingaantallen onverwacht zwak correleren met de daadwerkelijke leerlingaantallen; $r = 0.30$ voor het sbo en $r = 0.17$ voor het so. Hieruit volgt dat de rangordening van de scholen in het steekproefkader nogal onbetrouwbaar is en er in feite geen uitspraken gedaan kunnen worden over de representativiteit van de onderzoeksgroep in relatie tot schoolgrootte. Betrouwbare gegevens over het aantal schoolverlaters in de populatie ontbraken immers, en ook al zouden we het steekproefkader achteraf enigszins bijstellen op basis van de gegevens die feitelijk verzameld waren, blijft het problematisch om het **empirische** aantal schoolverlaters op de deelnemende scholen te vergelijken met het **geschatte** aantal schoolverlaters op de overige scholen.



Figuur 2.1. *Relatie tussen het verwachte en daadwerkelijke aantal schoolverlaters op de scholen*

Hoewel de representativiteit van de onderzoeksgroep niet goed te analyseren was in relatie tot schoolgrootte, was wel eenvoudig in kaart te brengen in hoeverre de deelnemende scholen varieerden in omvang. In Figuur 2.2 is de verdeling van de scholen weergegeven wat betreft het aantal uitstromende leerlingen dat deelnam aan het onderzoek. Bij de sbo-scholen varieerde het aantal schoolverlaters van 7 tot 60, en bij de so-scholen van 4 tot 34. Het gemiddelde aantal schoolverlaters dat op een school meedeed aan het onderzoek was in het sbo 23.6 en in het so 18.8. Samenvattend kan dus gesteld worden dat er zowel (erg) kleine als (erg) grote scholen aan het onderzoek hebben meegedaan, en dat niet goed vastgesteld kan worden in hoeverre de verhouding tussen kleine en grote scholen lijkt op de verhouding in de sbo en so populatie van scholen.



Figuur 2.2 *Verdeling van het aantal uitstromende leerlingen* op *de deelnemende scholen*

Achtergrondkenmerken

Het is ook belangrijk dat de onderzoeksgroep op leerlingniveau een goede afspiegeling vormt van de populatie met betrekking tot relevante achtergrondkenmerken. Er is gekeken naar de volgende achtergrondkenmerken: regio, urbanisatiegraad, sekse, herkomst en leerroute. Voor de populatiegegevens hebben we gebruikgemaakt van data van CBS, DUO en de Onderwijsinspectie. De verschillende achtergrondkenmerken zijn als volgt gedefinieerd:

- ***Regio***. Bij de definitie van de variabele *Regio* is uitgegaan van de CBS-indeling naar landsdeel. Dit betekent dat er vier regio's onderscheiden zijn. Regio *Noord* omvatte de provincies Groningen, Friesland en Drenthe; regio *Oost* de provincies Overijssel, Gelderland en Flevoland; regio *West* de

provincies Utrecht, Noord-Holland, Zuid-Holland en Zeeland en regio *Zuid* de provincies Noord-Brabant en Limburg.

- ***Urbanisatiegraad***. De gehanteerde verdeling naar *urbanisatiegraad* of *mate van verstedelijking* die we hebben gebruikt is terug te voeren op de bij het CBS gebruikelijke indeling naar vijf niveaus, namelijk *zeer sterk*, *sterk*, *matig*, *weinig* en *niet verstedelijkt*. Voor het onderzoek is de gebruikelijke vijfdeling echter gereduceerd tot een tweedeling in enerzijds *niet tot matig verstedelijkt* en anderzijds *sterk tot zeer sterk verstedelijkt*. Een dergelijke tweedeling blijkt in de praktijk goed te volstaan (zie bijvoorbeeld Van Boxtel & Hemker, 2009).
- ***Sekse***. In de analyses omtrent *sekse* is een tweedeling naar jongens en meisjes gehanteerd.
- ***Herkomst***. Bij de variabele *herkomst* is uitgegaan van een tweedeling naar Nederlands en Anders. Leerlingen zijn ingedeeld op basis van de volgende vraag uit de Quickscan: "Is de leerling in Nederland geboren?"
- ***Leerroute***. Bij de definitie van deze variabele is uitgegaan van de uitstroomprofielen die worden onderscheiden in *Passende Perspectieven*. Dat zijn vmbo-gl/tl, havo en vwo (leerroute 1), vbmo-bb/kb (leerroute 2) en vso arbeid-PrO (leerroute 3).

Door middel van kruistabellen, χ^2-toetsing en berekening van Φ_{Cramer} is voor elk van de genoemde achtergrondkenmerken nagegaan in hoeverre de onderzoeksgroep representatief was voor de populatie. De belangrijkste resultaten zijn samengevat in Tabel 2.4. Φ_{Cramer} is als volgt berekend:

$$\Phi_{\text{Cramer}} = \sqrt{\frac{\chi^2}{N \times df}}$$

Een waarde van .50/√*df* is aangemerkt als groot, een waarde van .30/√*df* als gemiddeld en een waarde van .10/√*df* als klein (zie Cohen, 1988).

Tabel 2.4. *Vergelijking tussen de populatie en de onderzoeksgroep naar achtergrondkenmerken*

		sbo		*so*	
Variabele		*Populatie*	*Onderzoeksgroep*	*Populatie*	*Onderzoeksgroep*
Regio	Noord	11.18%	20.59%	7.51%	17.02%
	Oost	24.41%	25.38%	23.57%	20.92%
	West	45.15%	40.86%	42.13%	50.00%
	Zuid	19.25%	13.17%	26.80%	12.06%
Urbanisatie-graad	Sterk / zeer sterk	64.14%	62.41%	55.47%	62.06%
	Niet tot matig	35.86%	37.59%	44.53%	37.94%
Sekse	Jongen	67.91%	62.89%	75.45%	87.54%
	Meisje	32.09%	37.11%	24.55%	12.46%
Herkomst	Nederland	86.00%	87.64%	71.67%	89.45%
	Ander land	14.00%	12.36%	28.33%	10.55%
Leerroute / uitstroomprofiel	vmbo-gl/tl, havo, vwo	6.36%	6.25%	36.93%	29.11%
	vmbo-bb/kb	49.18%	46.47%	42.95%	54.85%
	vso arbeid / PrO	44.46%	47.28%	20.12%	16.03%

We zien dat de verdeling van leerlingen over de vier regio's in de onderzoeksgroep iets afwijkt van de verdeling in de populatie. Zowel voor het sbo als so zijn de noordelijke provincies oververtegenwoordigd en de westelijke en zuidelijke iets ondervertegenwoordigd. Dit verschil is voor beide populaties significant: (sbo) χ^2 = 128.9, *df* = 3, *p* < .001, Φ_{Cramer} = 0.185; (so) χ^2 = 61.8, *df* = 3, *p* < .001, Φ_{Cramer} = 0.270. De urbanisatiegraad is voor de sbo-onderzoeksgroep goed vergelijkbaar met die in de populatie: χ^2 = 1.630, *df* = 1, *p* = .2, Φ_{Cramer} = 0.036. Voor het so is er desalniettemin sprake van een significante, maar weinig betekenisvolle, afwijking: χ^2 = 4.953, *df* = 1, *p* = .026, Φ_{Cramer} = 0.133. Bij de variabele sekse zien we dat jongens iets ondervertegenwoordigd zijn in de sbo-onderzoeksgroep (χ^2 = 14.495, *df* = 1, *p* < .001, Φ_{Cramer} = 0.108), en juist behoorlijk oververtegenwoordigd in de so-onderzoeksgroep (χ^2 = 22.191, *df* = 1, *p* < .001, Φ_{Cramer} = 0.281). Verder is de herkomst van de sbo-leerlingen goed representatief voor de populatie (χ^2 = 2.173, *df* = 1, *p* = .140, Φ_{Cramer} = 0.047), maar dit is minder het geval bij de so-onderzoeksgroep, waar leerlingen met een niet-Nederlandse afkomst ondervertegenwoordigd zijn (χ^2 = 36.906, *df* = 1, *p* < .001, Φ_{Cramer} = 0.395). Ten slotte zien we bij leerroute dat de sbo-leerlingen in het onderzoek een goede

afspiegeling vormen van de populatie (χ^2 = 3.270, *df* = 2, *p* = .195, Φ_{Cramer} = 0.040), maar de so-leerlingen minder (χ^2 = 13.705, *df* = 2, *p* = .001, Φ_{Cramer} = 0.170).

Samenvattend kan gesteld worden dat de sbo-onderzoeksgroep de verdeling in de populatie op de meeste van de onderzochte achtergrondkenmerken goed tot redelijk goed benadert. De effectgroottes van de afwijkingen zijn klein en bij de variabele met de meeste discrepantie (regio) ligt het niet direct voor de hand dat deze sterk samenhangt met de mondelinge taalvaardigheid van leerlingen. De so-onderzoeksgroep wijkt daarentegen op meerdere punten af van de populatie, niet alleen qua regio, maar ook qua sekse, herkomst en leerroute. Het is de vraag hoe bij de inhoudelijke analyses het beste omgegaan kan worden met de geconstateerde afwijkingen. In dit onderzoek is ervoor gekozen om de sbo-onderzoeksgroep als "voldoende" representatief te beschouwen. De verschillende leerroutes zijn evenwichtig vertegenwoordigd en bij de overige achtergrondkenmerken zijn er geen theoretische gronden om te corrigeren voor het gebrek aan representativiteit. In het so is ervoor gekozen om de populatieschattingen wél te corrigeren, en dan specifiek voor leerroute, omdat dat achtergrondkenmerk (theoretisch gezien) het meest verband houdt met de prestaties van de leerlingen. Er is niet óók voor herkomst en geslacht gecorrigeerd, omdat

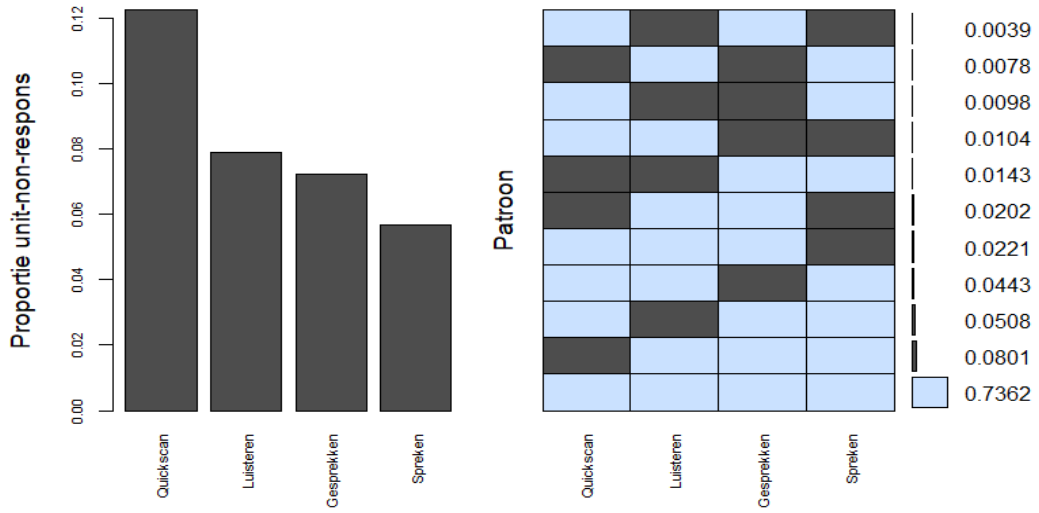
(a) er geen populatiegegevens beschikbaar zijn voor de combinatie leerroute-herkomst-geslacht;
(b) het aantal leerlingen in sommige cellen erg klein zal zijn de gewichten daardoor erg hoog worden;
(c) de variabele *herkomst* een minder eenduidige (of bruikbare) maat is, omdat deze in onderzoek op verschillende manieren wordt gedefinieerd.

De gewichten zijn in de databestanden te vinden onder variabelen **LUI_WEIGHT**, **SPR_WEIGHT** en **GSP_WEIGHT**.

2.1.4 Unit-non-respons

De representativiteitsanalyse in paragraaf 2.1.3 is gebaseerd op álle leerlingen in het onderzoeksbestand. Niet alle leerlingen hebben echter aan alle onderdelen van het onderzoek deelgenomen. In het geval bepaalde gegevens ontbreken en er sprake is van non-respons, is het handig om onderscheid te maken tussen unit-non-respons en item-non-respons. Bij unit-non-respons ontbreekt een volledig onderdeel (bijvoorbeeld de luistertoets). Bij item-non-respons missen alleen de scores op één of meerdere items of beoordelingsaspecten. In deze paragraaf wordt ingegaan op unit-non-respons en de mate waarin deze samenhangt met de in paragraaf 2.1.3 beschreven achtergrondkenmerken. Op item-non-respons gaan we hier niet in. Dit komt aan de orde in de hoofdstukken over de meetinstrumenten.

Hoewel unit-non-respons leidt tot een geringer aantal waarnemingen dan beoogd, hoeft dat geen probleem te vormen. Afhankelijk van de omvang van de unit-non-respons vallen immers alleen de onzekerheidsmarges van de schattingen bij een bepaald onderdeel wat groter uit. Als de unit-non-respons echter selectief is, kunnen de populatieschattingen vertekend worden. Eerst is in kaart gebracht in hoeverre er sprake is van unit-non-respons bij de vier onderdelen die op leerlingniveau zijn afgenomen. In Figuur 2.3 is aan de linkerkant, per onderdeel, weergegeven voor welk percentage van de leerlingen we **géén** gegevens hebben voor dat onderdeel. Aan de rechterkant van Figuur 2.3 is te zien hoe vaak bepaalde patronen van unit-non-respons voorkwamen (grijs gearceerd = ontbrekend). De Quickscan blijkt bij ongeveer 12 procent van de leerlingen te ontbreken. Bij de drie vaardigheidstoetsen is de proportie unit-non-respons kleiner (< 8 procent). Bovendien komt het zelden voor (bij minder dan 2.5 procent van de leerlingen) dat twee of meer vaardigheidstoetsen ontbreken.



Figuur 2.3. *Relatieve frequenties van het ontbreken van bepaalde onderdelen of combinaties daarvan*

In een vervolgstap is nagegaan in hoeverre de gesignaleerde unit-non-respons als “selectief” te kwalificeren is. Daartoe hebben we een vergelijking gemaakt tussen de leerlingen die een bepaald onderdeel wél hebben gemaakt en de leerlingen die dat onderdeel níet hebben gemaakt. De vergelijking is gemaakt in relatie tot regio, urbanisatiegraad, sekse, herkomst en leerroute. Het betreft hier, anders dan eerder, een vergelijking van groepen leerlingen **binnen** het deelnemersbestand en geen vergelijking met de landelijke populatie. In Tabel 2.5 is voor de onderdelen Luisteren, Spreken en Gesprekken te zien voor welk percentage leerlingen we beschikken over data. Daarnaast zijn de resultaten van de uitgevoerde χ^2-contingentietoetsen vermeld, samen met een indicatie van de effectgrootte (Φ_{Cramer}). De Quickscan is niet in deze analyse meegenomen, omdat dit onderdeel louter gebruikt wordt om prestaties van leerlingen te verklaren; een leerlingkenmerk is in dit onderzoek nooit de afhankelijke variabele.

Tabel 2.5. *Samenhang tussen achtergrondkenmerken en unit-non-respons in de onderzoeksgroep*

	% leerlingen met data				χ^2	*df*	*p*	Φ_{cramer}
Regio	Noord	Oost	West	Zuid				
Luisteren	94.8	93.6	90.5	90.5	7.259	3	0.064	0.040
Spreken	93.5	98.1	91.9	96.5	19.716	3	<.001	0.065
Gesprekken	95.4	94.2	90.5	93.5	9.444	3	0.024	0.045
Urbanisatiegraad	(zeer) sterk		Niet tot matig					
Luisteren	92.1		92.2		0.012	1	0.913	0.003
Spreken	93.4		95.8		3.983	1	0.046	0.051
Gesprekken	92.1		93.9		1.911	1	0.167	0.035
Sekse	Jongen		Meisje					
Luisteren	92.0		92.4		0.085	1	0.771	0.007
Spreken	92.8		92.6		0.030	1	0.863	0.004
Gesprekken	92.0		92.4		0.085	1	0.771	0.007
Herkomst	Nederland		Ander land					
Luisteren	92.4		97.2		4.610	1	0.032	0.062
Spreken	96.0		95.2		0.197	1	0.657	0.013
Gesprekken	92.6		91.7		0.131	1	0.718	0.010
Leerroute	1	2	3					
Luisteren	87.0	92.0	94.5		8.711	2	0.013	0.059
Spreken	96.9	96.6	95.1		2.040	2	0.361	0.029
Gesprekken	89.3	92.9	93.1		2.321	2	0.313	0.031

Bij het onderdeel Luisteren hangt de unit-non-respons samen met herkomst en leerroute. In beide gevallen geldt echter dat de effectgrootte (erg) klein is en de samenhang weinig betekenisvol is. Bij het onderdeel Spreken zijn deze twee variabelen niet significant gerelateerd aan unit-non-respons, maar de variabelen regio en urbanisatiegraad juist wel, al zijn de effectgroottes ook hier klein. Bij Gesprekken, ten slotte, hangt alleen regio significant samen met de unit-non-respons, maar wederom gaat het om een erg klein effect. Er is dus geen reden om aan te nemen dat de unit-non-respons in dit onderzoek selectief was. De generaliseerbaarheid van de uitkomsten is, met andere woorden, niet in het geding doordat niet alle leerlingen alle onderdelen gemaakt hebben.

2.2 Materiaal en procedure

Bij de peiling is gebruikgemaakt van verschillende meetinstrumenten. De meetinstrumenten waren bedoeld om (a) de mondelinge taalvaardigheid van leerlingen in kaart te brengen, (b) zicht te krijgen op de onderwijs- en ondersteuningsbehoeften van leerlingen, en (c) het onderwijsleerproces op sbo/so scholen te duiden. Op de verschillende meetinstrumenten wordt ingegaan in respectievelijk paragrafen 2.2.1, 2.2.2 en 2.2.3. In paragraaf 2.2.4 gaan we vervolgens in op de werving en training van de toetsleiders. Ten slotte beschrijven we in paragraaf 2.2.5 hoe de afname georganiseerd is op de scholen.

2.2.1 Instrumenten om zicht te krijgen op de mondelinge taalvaardigheid van leerlingen

De mondelinge taalvaardigheid van leerlingen is met behulp van verschillende meetinstrumenten in kaart gebracht. De instrumenten waren bedoeld voor het meten van de luistervaardigheid, de spreekvaardigheid en de gespreksvaardigheid van leerlingen.

- ***Luistervaardigheid.*** De luistervaardigheid van leerlingen is op twee verschillende manieren gemeten, namelijk in een klassikale pen-en-papier situatie en in een praktische spelsituatie. Bij de klassikale afname zijn via een digibord (met bijbehorend toetsboekje) vragen gesteld over kijk- en luisterfragmenten. In totaal zijn 52 items in een onvolledig design met drie toetsversies afgenomen. Het toetsboekje voor leerroute 1 bevatte 28 items, in het toetsboekje voor leerroute 2 zaten 24 items en in het toetsboekje voor leerroute 3 waren 21 items opgenomen. Bij de praktische afname speelden leerlingen een zogenaamd samenwerkingsspel (zie ook gespreksvaardigheid). Tijdens het spel kregen de leerlingen een instructie met diverse opdrachten waarmee ze konden laten zien of ze de (mondelinge) instructie hadden begrepen. Een toetsleider scoorde op een observatieformulier of dat het geval was.
- ***Spreekvaardigheid.*** De spreekvaardigheid van leerlingen is met behulp van twee taken onderzocht. Bij de eerste taak kregen leerlingen een jeugdjournaalfragment te zien en moesten zij aan de toetsleider (na)vertellen wat zij hadden gezien. Bij de tweede taak is gebruikgemaakt van het bestaande (educatieve) spel “Story Cubes”. Bij dit spel gooiden leerlingen enkele dobbelstenen met afbeeldingen op. Vervolgens vertelden zij aan de hand van vijf gegooide afbeeldingen een kort verhaal aan de toetsleider. Alle leerlingen in de steekproef maakten beide taken. Er was dus sprake van een volledig afnamedesign. De afname van de twee taken vond individueel plaats en de toetsleider scoorde de twee taken na afloop van de afname op basis van een speciaal ontwikkeld observatieformulier.
- ***Gespreksvaardigheid.*** De gespreksvaardigheid van leerlingen is in kaart gebracht met behulp van een bestaand (samenwerkings)spel uit Duitsland (het zogeheten *Fischierspiel*). Het doel van het spel was om samen alle vissen naar de haven te brengen voordat de storm begon. Het spel werd gespeeld op een speelbord met daarop een eiland en daaromheen een zee, ingedeeld in vakjes, waarin vissen lagen. Elke leerling had een eigen haven en een bootje in een eigen kleur waarmee de vissen vervoerd konden worden. Wanneer de leerling de beurt had, gooide hij met een dobbelsteen. Leerlingen bepaalden elke beurt samen wie dit aantal ogen het beste kon gebruiken om bij een vis te komen en/of deze vis naar een haven op het eiland te brengen. Zodra alle vissen op het eiland waren, hadden de leerlingen samen gewonnen. Op de dobbelsteen stond echter ook een wolkje. Wanneer een wolkje werd gegooid, moest er naast het spel één van de tien wolkjes omgedraaid worden. Wanneer alle tien wolkjes waren omgedraaid, stormde het zo hard dat de boten omwaaiden en “het spel” gewonnen had. De toetsleider vulde op basis van de gesprekken tijdens het spel een observatieformulier in. Het samenwerkingsspel werd in groepjes met 3 tot 4 leerlingen gespeeld.

Bij de ontwikkeling van de instrumenten vormden de leerroutes van *Passende perspectieven* (Langberg et al., 2012) het uitgangspunt. Er is rekening gehouden met de volgende leerlingkenmerken:

- ***Doeners.*** Kenmerkend voor de populatie in het so en sbo is dat relatief veel leerlingen doeners zijn. Deze leerlingen komen dikwijls beter tot hun recht wanneer zij praktische handelingen uitvoeren dan wanneer zij een opdracht met pen en papier of op de computer uitvoeren (van Boxtel et al., 2012). Om tegemoet te komen aan zowel leerlingen die in praktische situaties het beste uit de verf komen, als aan leerlingen die sterker zijn in pen-en-papier-situaties, zijn in de peiling verschillende toetsvormen gebruikt.
- ***Beperkte aandachtsspanne.*** Leerlingen in het sbo en so hebben vaak een beperkte aandachtsspanne. De toetsen mogen daarom niet te veel tijd (achter elkaar) kosten en moeten voldoende afwisseling bevatten. Om deze reden bestond het totale instrumentarium uit relatief korte toetsen en gevarieerde werkvormen.
- ***Toetsdruk.*** Voor alle leerlingen, maar voor leerlingen in het so en sbo in het bijzonder, geldt dat zij het beste presteren in een vertrouwde situatie zonder toetsdruk. In schoolse situaties en vrije situaties zal een leerling veel vaardigheden op het gebied van luisteren, spreken en gesprekken toepassen. Gegevens over de leerlingen kunnen in dat geval het gemakkelijkst verzameld worden met behulp van een observatielijst. De leerling wordt dan niet ‘lastiggevallen’ met een toets en kan (ongemerkt) in een vertrouwde setting laten zien wat hij kan. Waar mogelijk hebben we in de peiling daarom gewerkt met observaties.
- ***Faalervaringen.*** In de periode voordat leerlingen op het sbo of so komen, hebben zij soms al veel faalervaringen opgedaan. Voor alle leerlingen, maar zeker ook voor leerlingen in het sbo en so, is het prettig als een toets een succeservaring is. De toetsen moeten echter ook voldoende uitdagend zijn om de aandacht erbij te kunnen houden. De toetsen moeten, met andere woorden, passen bij het niveau van de leerling. Doordat de niveaus in het sbo en so sterk uiteenlopen, is in de peiling waar nodig adaptief getoetst.

- ***Toetsvorm.*** Door uiteenlopende problematieken zijn sommige toetsvormen voor leerlingen met specifieke onderwijsbehoeften minder geschikt. Daarom is bij elke toets per zorgprofiel aangegeven welke aanpassingen (eventueel) toegestaan waren om op een passende wijze de toets bij de leerling af te nemen.

Naast bovengenoemde leerlingkenmerken is bij de ontwikkeling van de meetinstrumenten rekening gehouden met het gegeven dat (a) de prestaties van leerlingen gerelateerd moesten worden aan de referentieniveaus en (b) de peiling bij het onderdeel Luisteren zicht moest geven op veranderingen in het prestatieniveau tussen 2007 en 2018. Om aan deze voorwaarden te voldoen, heeft CvTE voor de instrumenten Spreken en Gesprekken prestatiestandaarden voor referentieniveaus 1F en 2F ontwikkeld. Voor Luisteren is gebruikgemaakt van de landelijke Ankerset Luisteren: enkele items uit die Ankerset zijn in deze peiling meegenomen, zodat het mogelijk werd om de beschikbare prestatiestandaarden voor 1F en 2F over te brengen naar de luistertoetsen die in deze peiling zijn afgenomen. Om vergelijkingen met 2007 te kunnen maken, zijn in de luistertoetsen tevens items opgenomen uit PPON2007. Daarnaast is in de instrumenten gebruikgemaakt van materiaal uit het Cito Leerlingvolgsysteem voor het primair en speciaal onderwijs (LVS). Het ging om materiaal uit de LVS-toets begrijpend luisteren (zie Van Berkel et al., 2016) en de LVS-toetsen taalvaardigheid voor het zml en vso dagbesteding (zie Van Boxtel et al., 2012; Alberts et al., in voorbereiding). In Tabel 2.6 is nog eens samengevat met welke activiteiten, en met welke bronnen, de prestaties van leerlingen in kaart zijn gebracht.

Tabel 2.6. *Overzicht gebruikte instrumenten om leerprestaties in kaart te brengen*

Instrument	Type activiteit	Onderdeel	Aantal items/observatiepunten
Taak luistervaardigheid	Film- en audiofragmenten met bijbehorende meerkeuzevragen in een toetsboekje	Luisteren	Leerroute 1 – 28 Leerroute 2 – 24 Leerroute 3 – 21
Dobbelspel	Spel met bijbehorende observatie door toetsleider	Spreken	15
Navertellen filmpje	Verteltaak met bijbehorende observatie door toetsleider	Spreken	15
Samenwerkingsspel	Spel met bijbehorende observatie door toetsleider	Gesprekken Luisteren	17 4

Zoals we in Tabel 2.6 kunnen zien was het aantal items (of observatiepunten) bij de verschillende instrumenten niet erg groot. Dit is bewust gedaan. Bij metingen op systeemniveau is het namelijk optimaal om de opgaven per te meten vaardigheid zo gespreid mogelijk af te nemen bij zoveel mogelijk leerlingen (Lord & Novick, 1968). Met andere woorden: het is beter dat relatief veel leerlingen relatief weinig opgaven maken over een bepaalde vaardigheid dan dat relatief weinig leerlingen relatief veel opgaven over die vaardigheid krijgen. De invloed van meetonnauwkeurigheid op de stabiliteit van uitkomsten op systeemniveau is namelijk erg klein in vergelijking tot de invloed van andere foutenbronnen zoals de steekproef of een (item respons theorie) equivalering (Keuning & Hemker, 2014).

2.2.2 Instrumenten om zicht te krijgen op de onderwijs- en ondersteuningsbehoeften van leerlingen

De onderwijs- en ondersteuningsbehoeften van leerlingen zijn in kaart gebracht met behulp van een zogeheten Quickscan. De Quickscan bestond uit twee delen die ingevuld zijn door de leerkracht. In het eerste deel is op basis van een aantal vragen het volgende in kaart gebracht: sekse, geboortedatum, uitstroomprofiel, verblijfsduur sbo/so, stimulerende en belemmerende factoren, onderwijsbehoeften, thuistaalsituatie, herkomst, en aanwezigheid van stoornissen en beperkingen. In Hoofdstuk 3 is terug te vinden hoe de informatie precies is verzameld. Bij de ontwikkeling van de vragen is gestreefd naar een zo volledig mogelijke beeld zonder al te veel tijd van de leerkrachten te vragen. Bij het samenstellen is dus kritisch gekeken in hoeverre de leerkrachten het haalbaar vonden om de gevraagde gegevens voor hun schoolverlaters in te vullen.

Het tweede gedeelte van de Quickscan omvatte een korte observatielijst op basis van doelen uit Passende Perspectieven en de Cito observatielijst voor vso dagbesteding mondelinge taal (Alberts et al., in voorbereiding). De observatiepunten waren gericht op gedrag dat de leerkracht in dagelijkse situaties kan waarnemen. Concreet ging het om de volgende punten:

Luisteren

1. De inhoud van een gesproken tekst in eigen woorden weergeven
2. De hoofdpunten uit gesproken (nieuws)berichten halen
3. Een oordeel over een tekst(deel) of tv- of radioprogramma verwoorden

Spreken en gesprekken voeren

1. Verstaanbaarheid
2. Een verhaal vertellen
3. Bij het gespreksonderwerp blijven of afronden
4. Een gesprek beginnen
5. Een gesprek op gang houden
6. Gesprekspartner volgen

De vragen zijn beantwoord op basis van indicatoren, waarbij de omschrijving van indicator 1 qua inhoud paste bij leerroute 3 (vso arbeid / PrO). Indicator 2 paste bij leerroute 2 (vmbo-bb/kb) en indicator 3 paste bij leerroute 1 (vmbo-gl/tl, havo, vwo). Op deze manier verkregen we enerzijds informatie op doelniveau: welke doelen van luisteren, spreken en gesprekken beheerst de leerling in welke mate? Anderzijds verkregen we informatie over de leerroute die een leerling volgt.

2.2.3 Instrumenten om het onderwijsleerproces op sbo/so scholen te duiden

Om het onderwijsleerproces op sbo/so scholen te duiden is per school een gesprek gevoerd met vertegenwoordigers van die school. Het streven was om drie personen te spreken die zicht hebben op wat er schoolbreed wordt gedaan aan mondelinge taalvaardigheid. Om concreet zicht te krijgen op wat er in de klassen gebeurt, nam de leerkracht van de eindgroep (waarin de peiling plaatsvond) in elk geval deel aan het gesprek. Om de informatie van het gesprek gestructureerd te verwerken, vulde de “interviewer” na afloop van het gesprek een formulier in. Het formulier bevatte vragen over zes onderwerpen: (a) onderwijsaanbod en methodegebruik, (b) doelgerichtheid en planmatigheid, (c) onderwijstijd, (d) monitoring en planning, (e) differentiatie en (f) teamontwikkeling. Naast algemene vragen werden ook meer specifieke subvragen gecodeerd die duidelijk maken op welke wijze een school concrete invulling geeft aan deze aspecten van het onderwijs mondelinge taalvaardigheid. In Hoofdstuk 7 is terug te vinden hoe de informatie precies verwerkt is. Waar mogelijk is het gesprek met een voicerecorder opgenomen.

2.2.4 Constructieproces en try-out

Bij de constructie van de materialen om de leerprestaties en de onderwijs- en ondersteuningsbehoeften van leerlingen in kaart te brengen, zijn de volgende stappen doorlopen: (a) screenen bestaande materialen, (waar nodig) ontwikkeling nieuwe materialen, en samenstellen conceptinstrumenten, (b) try-out conceptinstrumenten en (c) vaststellen definitieve instrumenten.

Screening, ontwikkeling en samenstelling conceptinstrumenten

Voor Luisteren waren al veel materialen beschikbaar en uiteindelijk kon het instrument daar ook volledig op gebaseerd worden. Alle beschikbare materialen, te weten de items uit eerdere peilingsonderzoeken, de Ankerset en het cito-LVS, zijn eerst gescreend op geschiktheid voor het speciaal (basis-)onderwijs. Er is onder meer gekeken naar het niveau en de inhoud van de materialen. Bij de screening is gebruikgemaakt van de screeningslijst die Cito ook gebruikt bij de ontwikkeling van toetsen voor speciale leerlingen (Cito, 2010). Materialen waarbij de leerling zich moet verplaatsen in een ander vielen bijvoorbeeld af, omdat in de onderzoeksgroep relatief veel leerlingen met een autismespectrumstoornis zitten. Na de screening is een selectie gemaakt, waarbij expliciet is gelet op variëteit en tijdsduur.

Voor Spreken en Gesprekken was er veel minder materiaal voor handen. In vorige peilingsonderzoeken in het sbo/so werden de onderdelen Spreken en Gesprekken bijvoorbeeld nog niet expliciet onderscheiden zoals nu, na de invoering van het referentiekader Taal en Rekenen (Meijerink et al., 2009), wel gebruikelijk is. De materialen die wel beschikbaar waren, vroegen bovendien nogal veel inlevingsvermogen van de leerlingen doordat ze zich steeds in verschillende rollen moeten verplaatsen. Sommige andere taken sloten onvoldoende aan bij de belevingswereld van de leerlingen. De taak die bij de meest recente peiling in het reguliere onderwijs is gebruikt (zie Van Langen et al., 2017) was daardoor bijvoorbeeld ongeschikt; te weinig leerlingen zijn bekend met een vlog. Indien een dergelijke taak wel afgenomen wordt, kan dat zorgen voor een onveilig gevoel en een gebrek aan structuur, waardoor de leerlingen onvoldoende tot hun recht komen. Voor Spreken en Gesprekken zijn daarom nieuwe materialen ontwikkeld. Voor Gesprekken, en voor een deel ook voor Spreken, is gekozen voor een spelvorm. Dit haalt de toetscontext, die voor veel leerlingen een gevoel van onveiligheid oproept, weg en biedt de leerlingen zo de mogelijkheid zich veilig en betrokken te voelen. Dat zijn belangrijke voorwaarden voor leerlingen om het beste van zichzelf te kunnen laten zien.

Voor het eerste gedeelte van de Quickscan is gezocht naar bestaande en beproefde materialen waarmee de beoogde leerlingkenmerken in kaart gebracht konden worden. Uiteindelijk is gebruikgemaakt van de vragenlijsten zoals die bij COOL Speciaal (Ledoux, Roeleveld, van Langen & Smeets, 2012) en een NRO-onderzoek naar het plannen van leerroutes (Hollenberg, Keuning & Meijer, 2016) gebruikt zijn. Voor de

leerlingkenmerken waarvoor geen vragen beschikbaar waren, zijn nieuwe vragen geconstrueerd. Voor het tweede gedeelte van de Quickscan vormde *Passende Perspectieven* de basis. Concreet zijn uit de vragenlijst mondelinge taalvaardigheid van de toetsen VSO Dagbesteding Mondelinge taal de vragen geselecteerd die aansloten bij de doelen uit *Passende Perspectieven*. Deze zijn aangevuld met zelf geconstrueerde vragen en bijbehorende antwoordcategorieën. Elke antwoordcategorie correspondeerde met één van de drie leerroutes. Daarnaast is een vierde antwoordcategorie toegevoegd voor de leerlingen die lager scoorden dan de omschrijving die hoorde bij leerroute 3.

Try-out conceptinstrumenten
De conceptinstrumenten zijn voorgelegd aan enkele leerkrachten uit het sbo en so. Zij hebben feedback gegeven en op basis van hun feedback zijn enkele wijzingen doorgevoerd. Vervolgens zijn de instrumenten in oktober 2017 bij leerlingen uitgeprobeerd. Het voornaamste doel van de try-out was om vast te stellen of de materialen voldoende geschikt waren voor leerlingen met uiteenlopende onderwijsbehoeften. Tijdens de afname is bijvoorbeeld gekeken in hoeverre de leerlingen hun aandacht bij de taak konden houden, gemotiveerd waren en niet gefrustreerd raakten. Concreet bleek uit de try-out het volgende:

Luisteren
- De items waren grofweg in te delen in drie soorten: (a) items bij een filmfragment, (b) items bij een audiofragment, en (c) items waarbij leerlingen moesten aangeven welk plaatje het beste past bij een gesproken kort verhaal. Het luisteren naar audiofragmenten bleek de meeste concentratie te vergen. De audiofragmenten zijn in de definitieve versie daarom bewust als één-na-laatste fragment in de toets opgenomen. Zo startte en eindigde de toets altijd met een fragment waar leerlingen gemakkelijker de aandacht bij konden houden.
- De items waren afkomstig uit verschillende bronnen (Cito-LVS, Ankersets, eerdere peilingsonderzoeken) met elk hun eigen afnamecondities. Dit had tot gevolg dat de vragen in sommige gevallen werden voorgelezen, dat in andere gevallen zowel de vragen als de antwoordalternatieven werden voorgelezen, en dat er in weer andere gevallen helemaal niets werd voorgelezen. In de try-out is vastgehouden aan de oorspronkelijke afnamecondities, maar dit bleek zeer verwarrend te werken. Omdat de leerlingen een sterke voorkeur uitspraken voor het voorlezen van de vragen maar niet van de antwoordalternatieven, is dit in de definitieve versie doorgevoerd bij alle items.
- In de conceptversie stonden zowel de items als de bijbehorende instructies op de dvd. In de try-out bleek echter dat de leerlingen moeite hadden hun aandacht bij de instructie op de dvd te houden. Instructie van een toetsleider, die tussendoor verwees naar het toetsboekje, werkte beter. In de definitieve versie zijn op de dvd zijn daarom “rustmomenten” (scherm met een ‘i’) ingebouwd. Tijdens die momenten kon de toetsleider instructie geven.
- Elk fragment werd eerst als geheel afgespeeld, en daarna in delen herhaald. In de try-out bleek dat veel leerlingen direct na het beluisteren van het gehele fragment de vragen al beantwoordden. Als het fragment daarna nog in delen werd opgeknipt en herhaald, haakten ze soms af. Toch is ervoor gekozen om het opknippen van het fragment in delen te handhaven, omdat dit min of meer de standaard is en een deel van de leerlingen wel gebruikmaakte van de herhaling in delen.

Spreken
+ Alle leerlingen waren in staat om in meer of mindere mate te vertellen wat ze hadden gezien bij het jeugdjournaalfragment. Sommigen gebruikten hierbij het geheugensteuntje, anderen niet. De beoordelingspunten waren aan de hand van de opdracht voor vrijwel alle leerlingen goed te scoren. Aanpassingen waren niet nodig.
+ Leerlingen vonden het leuk om het (educatieve) spel Story Cubes te spelen. Doordat het geen standaard (toets)taak is, vonden ze het minder spannend.
+ Zoals verwacht mag worden, was er variatie in de hoeveelheid tekst die leerlingen bij de dobbelstenen vertelden, maar alle leerlingen kwamen tot een verhaal dat beoordeeld kon worden.
- Sommige dobbelstenen pasten beter bij de belevingswereld van leerlingen dan andere. Dit bleek van invloed op de moeilijkheidsgraad van de taak. In de definitieve versie mochten leerlingen daarom naar eigen inzicht vijf dobbelstenen kiezen; vier dobbelstenen hoefden niet in het verhaal meegenomen te worden.
- Toetsleiders deden het spel één keer voor. Dat ging niet direct goed, omdat de lengte van de verhalen van de toetsleiders uiteenliepen en de instructies daardoor nogal van elkaar gingen verschillen. In de definitieve instructie voor toetsleiders is hier daarom meer aandacht aan besteed. De richtlijn is geworden: geef een voorbeeld met één dobbelsteen per zin, dus in totaal vijf korte zinnen.

- Eén keer oefenen met Story Cubes was noodzakelijk, en voor de onzekere leerlingen soms zelfs een tweede keer. De definitieve afname-instructie is hierop aangepast: de eerste (oefen)ronde is niet meegenomen in de scoring en bij twijfel is de scoring gebaseerd op een toegevoegde derde ronde in plaats van op de “echte” tweede ronde.

Gesprekken
- + Alle leerlingen deden mee met het spel en de betrokkenheid was groot.
- + Ook leerlingen die in de klas nauwelijks spreken kwamen tijdens het spel tot spreken, soms tot verrassing van de leerkracht.
- + De afwisseling tussen instructie en korte opdrachtjes zoals even oefenen met gooien of een samenvatopdracht werkte goed.
- − De instructie duurde voor sommige leerlingen iets te lang. Vooraf vertellen dat het belangrijk is om goed naar de instructie te luisteren en aangeven dat leerlingen na de instructie direct aan de slag mogen bleek effectief. Dit is aan de definitieve instructie toegevoegd.
- − Sommige beoordelingspunten bleken onduidelijk. De toelichtingen bij de beoordelingspunten zijn daarom aangescherpt. De definitieve toelichtingen zijn te vinden in Bijlage 3.
- − De opdracht waarbij leerlingen op een werkblad moesten aangeven waar een boot naartoe moet varen werkte niet. De leerlingen hadden er moeite mee om de aandacht van het spel naar het papier te verplaatsen. Een vergelijkbare opdracht op het bord werkte wel goed. De opdracht is in de definitieve versie aangepast.

Quickscan
- + De Quickscan was duidelijk en in de regel kon de leerkracht de vragen binnen de gestelde 10 minuten per leerling invullen.
- − Het tweede deel van de Quickscan werd aanvankelijk ingevuld aan de hand van drie antwoordcategorieën. De zwakste leerlingen bleken dan echter niet goed te scoren. Om deze reden is in de definitieve versie een extra kolom (of vierde vakje) toegevoegd voor leerlingen die niet voldoen aan de doelen van leerroute 3.
- − Sommige formuleringen, zoals die over verstaanbaarheid en luidheid/volume, waren niet optimaal. Deze zijn in de definitieve versie aangescherpt.

Vaststelling definitieve instrumenten
De uitkomsten van de try-out zijn verwerkt in de definitieve versie van de instrumenten. Zoals hierboven al is aangegeven, zijn onder meer wijzigingen doorgevoerd in de instructies voor leerlingen en toetsleiders. Daarnaast zijn bepaalde opdrachten aangepast, formuleringen aangescherpt of afnamevolgordes gewijzigd. Aspecten die tijdens de try-out soms fout gingen, zijn bovendien benadrukt en behandeld in de training voor toetsleiders.

2.2.5 Werving en training toetsleiders

Toetsleiders zijn geworven bij de opleidingen (ortho)pedagogiek en (ontwikkelings)psychologie van verschillende universiteiten. Er is onder meer een oproep verspreid waarin de werkzaamheden kort werden toegelicht en de functie-eisen waaraan toetsleiders moesten voldoen om in aanmerking te komen, werden omschreven. Zo moesten de toetsleiders onder andere derde- of vierdejaars studenten zijn, ervaring hebben met de afname van (diagnostische) tests, en ervaring en/of affiniteit hebben met leerlingen in het (speciaal) basisonderwijs. Van elke toetsleider werd een verklaring omtrent gedrag geëist. Alle toetsleiders moesten verder verplicht deelnemen aan een trainingsdag die in totaal 6 uur in beslag nam. Er zijn drie trainingsdagen gehouden op verschillende locaties in Nederland. In de training werden de achtergrond en het doel van het onderzoek toegelicht. De nadruk in de training lag echter op het oefenen in de toetsafname en de scoring aan de hand van de beoordelingsschema’s. De beoordeling is, na het bestuderen van het referentiekader voor mondelinge taalvaardigheid, geoefend aan de hand van voorbeeldfilmpjes. Alle instructies zijn doorlopen en toetsleiders hebben elkaar de toetsen voor Spreken en Gesprekken voorgelegd. Met elke toetsleider is na de eerste toetsdag contact opgenomen om de afname te bespreken. Er is tevens een groepsapp opgezet, zodat toetsleiders ervaringen met elkaar konden uitwisselen.

2.2.6 Organisatie van de afname en het interview

Toetsleiders zijn aan scholen toegewezen en hebben vervolgens zelf toetsdagen met de scholen ingepland. De scholen ontvingen hierover eerst schriftelijke informatie. Quickscans en toetsboekjes zijn van tevoren naar de scholen gestuurd. De Quickscan werd per leerling door de leerkracht ingevuld voorafgaand aan de eerste toetsdag. In overleg met de contactpersoon op de school maakte de toetsleider een dagplanning voor elke toetsdag. Er werd begonnen met afname van de luistertoets, vervolgens werd het *Fischerspiel* afgenomen, en daarna werden de individuele spreektaken afgenomen (d.w.z., het jeugdjournaalfragment en het dobbelspel *Story Cubes*). Indien mogelijk werden de drie leerroutes van de luistertoets gelijktijdig in

verschillende ruimtes afgenomen door de toetsleider, leerkracht(en) en/of ander onderwijsprofessional(s). De afname van het samenwerkingsspel en de spreektaken vond plaats in een aparte, bij voorkeur rustige, ruimte.

De contactpersonen van de scholen zijn vanaf het voorjaar 2018 benaderd voor het inplannen van het interview. Het interview vond bij voorkeur op de school zelf plaats, of eventueel via Skype of een telefoon met speakerfunctie. Er werd gevraagd om te spreken met een goede vertegenwoordiging zoals leerkracht(en), taalcoördinatoren en/of intern begeleiders, vanuit de wens om een zo goed mogelijk beeld te krijgen van het onderwijsleerproces op het gebied van mondelinge taalvaardigheid. Tijdens een speciale bijeenkomst is de interviewleidraad doorgenomen met de interviewers. Daarbij is de onafhankelijke rol van de interviewer benadrukt, die met een open en niet-sturende houding een duidelijk beeld wil krijgen van het onderwijs. Er werden indien mogelijk audio-opnames gemaakt van de interviews.

2.3 Analyseplan

Zoals al eerder aangegeven wordt in dit onderzoek systematisch in kaart gebracht (a) hoe schoolverlaters in het sbo/so presteren bij Luisteren, Spreken en Gesprekken, (b) hoe deze prestaties zich verhouden tot voorgaande jaren, (c) hoe het onderwijsleerproces in het sbo en so gekarakteriseerd kan worden, en (d) hoe leerlingprestaties samenhangen met het onderwijsleerproces, taakkenmerken en leerlingkenmerken. Om deze doelstellingen te realiseren, zijn verschillende instrumenten afgenomen bij leerlingen en (vertegenwoordigers van) scholen. Daarna moesten allerlei statistische analyses plaatsvinden. Daarbij is onderscheid te maken tussen analyses die betrekking hebben op de samenstelling van de onderzoeksgroep (zie paragraaf 2.1.3 en 2.14), op de meetinstrumenten, en op de inhoudelijke onderzoeksvragen. Bij de meetinstrumenten gaat het om:

- De kwaliteit van de data en het instrument;
- De omgang met (incidenteel of structureel) missende waarden;
- De constructie van schalen en variabelen.

Deze onderdelen staan centraal in Hoofdstukken 3 tot en met 7 en worden hier niet toegelicht. Bij de inhoudelijke analyses is de volgende vierdeling te maken: (a) het beschrijven van de leerlingprestaties en het onderwijsleerproces op scholen, (b) het duiden en verklaren van de leerlingprestaties, (c) het onderzoeken van trends en (d) het voorspellen van leerlingprestaties op basis van variabelen op leerling- en schoolniveau.

Beschrijving leerlingprestaties en onderwijsleerproces op scholen
Eén van de doelen van dit onderzoek was om de prestaties van leerlingen te beschrijven in het licht van de referentieniveaus. Om dat te kunnen doen, was voor elke (deel)vaardigheid (Luisteren, Spreken en Gesprekken) een prestatiestandaard nodig. In paragraaf 8.1.1 wordt beschreven hoe de prestatiestandaarden bepaald zijn. Na bepaling van de prestatiestandaarden is in kaart gebracht hoeveel procent van de leerlingen beschikt over welk referentieniveau. Deze beschrijvende statistieken worden in paragraaf 8.1.2 apart gerapporteerd voor de leerlingen in het sbo en het so. De percentages leerlingen die de referentieniveaus behalen worden in beschrijvende analyses ook gerelateerd aan achtergrondkenmerken van de leerlingen, zoals bijvoorbeeld thuistaal, geslacht en onderwijsbehoeften (paragraaf 8.1.3). Ten slotte worden in paragraaf 8.1.4 enkele beschrijvende analyses gepresenteerd naar aanleiding van de interviews die hebben plaatsgevonden. Deze betreffen de kenmerken van het onderwijsleerproces op de so- en sbo-scholen.

Duiding en verklaring leerlingprestaties
Een tweede doel van dit onderzoek was om de prestaties van leerlingen te relateren aan kenmerken van de taakuitvoering. De samenhang tussen de verschillende toetsonderdelen wordt in paragraaf 8.2 in kaart gebracht en vervolgens wordt nagegaan hoe zwakke (<1F), gemiddelde (1F) en sterke leerlingen (2F) presteren op items of subonderdelen die verschillen qua taaktype of taakuitvoering. Hierbij gaat het onder meer om de volgende opdelingen:

- Feiten overbrengen (bij het filmfragment) versus creatief vertellen (bij het dobbelspel);
- Taalgebruik bij Spreken versus Taalgebruik bij Gesprekken. En daarnaast dezelfde vergelijking voor de subonderdelen Taalvorm en Taalinhoud.

Bovendien wordt een opdeling gemaakt naar (meer) theoretische/zakelijke taken enerzijds (de luistertoets en, in mindere mate, het navertellen van het filmfragment), en meer praktische/speelse taken anderzijds (het samenwerkingsspel en het dobbelspel). In deze vergelijking wordt ingestoken op de constatering van leerkrachten in het so en sbo dat leerlingen in praktische situaties ‘anders’ scoren dan wanneer zij een meer theoretische opdracht, bijvoorbeeld op papier, uitvoeren.

Trends
Een derde doel van dit onderzoek was om na te gaan hoe de prestaties van schoolverlaters in het sbo in 2018 zijn veranderd ten opzichte van eerdere peilingen. Idealiter zou de vergelijking gemaakt zijn voor alle drie de vaardigheden (Luisteren, Spreken en Gesprekken), maar dat is niet mogelijk. Het onderdeel Gesprekken maakte namelijk geen deel uit van eerdere peilingsonderzoeken en het onderdeel Spreken is in dit onderzoek dusdanig anders geoperationaliseerd dat een vergelijking met eerdere peilingen niet zinvol is. Voor het onderdeel Luisteren ligt dit anders: de vergelijking is vanwege een gelijke operationalisatie inhoudelijk zinvol en bovendien is er gezorgd voor overlappende items (zie ook paragraaf 2.2.1). In paragraaf 8.3.1 wordt eerst beschreven hoe de resultaten van PPON2007 gerelateerd zijn aan die van dit onderzoek. Daarna wordt in paragraaf 8.3.2 middels beschrijvende statistiek het verschil tussen 2007 en 2018 in kaart gebracht.

Voorspellers op leerling- en schoolniveau
Een vierde doel van dit onderzoek was het verkrijgen van inzicht in de samenhang tussen leerlingprestaties en kenmerken van de leerlingen en van het onderwijsleerproces op de scholen. De relatie tussen de leerlingprestaties en achtergrondkenmerken van de leerlingen wordt met beschrijvende analyses in kaart gebracht in paragraaf 8.1.3, maar de relatie met het onderwijsleerproces op de scholen kan het beste onderzocht worden door middel van multilevelanalyses waarin wordt gecorrigeerd voor relevante leerlingkenmerken. Hiervoor is gebruikgemaakt van de variabelen die met behulp van de Quickscan en de interviews verzameld zijn. Voor elke uitkomstvariabele is een multilevel regressie-analyse toegepast, ongeacht de geobserveerde intraklasse-correlatie (ICC). Hoe we het beste tot een eindmodel kunnen komen is bij dergelijke analyses een lastige vraag, zeker als we over veel variabelen beschikken die kunnen dienen als modelpredictoren. In dit onderzoek is ervoor gekozen om de volgende stappen te doorlopen (apart per uitkomstvariabele):

1. fit een random-intercept model en bekijk de varianties op leerling- en schoolniveau;
2. controleer voor effecten van geslacht, herkomst en thuistaal;
3. schat de effecten van onderwijsbehoeften, stoornissen, stimulerende factoren, en laag IQ;
4. schat het model opnieuw na verwijderen van niet-significante leerlingkenmerken;
5. onderzoek (in verschillende modellen) het effect van aspecten van het onderwijsleerproces;
6. fit een eindmodel met de significante (en meest relevante) leerling- en schoolvariabelen.

Kenmerken voor deze aanpak is dat er telkens eerst wordt gecontroleerd voor relevante *confounders* (verstorende variabelen) voordat het effect van andere predictoren, zoals kenmerken van het onderwijsleerproces, wordt onderzocht. De selectie van variabelen dient inhoudelijk onderbouwd te worden, en daarnaast moet er rekening gehouden worden met de bruikbaarheid (in statistische zin) van de variabelen die uit de Quickscan en interviews komen. Het streven bij de multilevelanalyses, die worden beschreven in paragraaf 8.4, was om tot een inhoudelijk aannemelijk en adequaat model te komen van hoe leerlingkenmerken en het onderwijsleerproces op de scholen de mondelinge taalvaardigheid van leerlingen beïnvloeden.

3 Quickscan

3.1 Inhoud en afname

De Quickscan werd per leerling ingevuld door de leerkracht en bestond uit twee delen. Het eerste deel bevat vragen over leerlingkenmerken, zoals informatie over de onderwijsbehoeften en het zorgprofiel van de leerling. Het tweede deel is een korte observatielijst voor Luisteren, Spreken en Gesprekken, die gebaseerd is op de doelen uit Passende Perspectieven en de Cito observatielijst voor vso dagbesteding mondelinge taal. De observatielijst is gericht op gedrag dat de leerkracht in dagelijkse situaties kan waarnemen en heeft een opbouw qua niveau zodat deze gemakkelijk te scoren was voor leerlingen met uiteenlopende uitstroomprofielen. Het invullen van de Quickscan nam per leerling ongeveer 10 tot 15 minuten in beslag. Sommige onderdelen zijn door vrij veel leerkrachten om privacyredenen niet ingevuld, en 7 van de 68 scholen hebben de Quickscan in het geheel niet ingevuld.

3.1.1 Leerlingkenmerken

De bevraagde leerlingkenmerken in de Quickscan zijn grotendeels gebaseerd op de onderwijs- en ondersteuningsbehoeften van leerlingen voor mondelinge taalvaardigheid die opgesomd zijn in het addendum Mondeling taalvaardigheid in s(b)o van SLO (2016, p. 24). In het addendum werden de volgende kenmerken genoemd:

- Profielschets of zorgprofiel;
- Uitstroomprofiel;
- Protectieve (of stimulerende) en belemmerende factoren;
- Pedagogische en didactische behoeften;
- Mate of niveau van gewenste ondersteuning;
- Taal-spraakontwikkeling;
- Opleidingsniveau ouders en thuis(taal)situatie;
- Hulpmiddelen.

De try-outscholen speelden een belangrijke rol in het bepalen welke leerlingkenmerken uiteindelijke zijn opgenomen in de Quickscan en de wijze waarop de kenmerken zijn bevraagd. Er is gestreefd om de belasting voor de scholen zoveel mogelijk beperkt te houden om uitval te voorkomen. Uitgangspunt was daarom dat de Quickscan gemakkelijk in te vullen moest zijn zonder het leerlingdossier te raadplegen. Zo is bijvoorbeeld niet gevraagd naar het opleidingsniveau van de ouders, omdat dit vaak onbekend is bij de leerkracht en de scholen aangaven dat dit erg veel tijd zou kosten om op te zoeken.

De Quickscan is een papieren vragenlijst met meerkeuzevragen. Aan de scholen is gevraagd om de Quickscan in te vullen voor alle leerlingen van wie de school verwachtte dat zij aan het einde van het schooljaar zouden uitstromen. Alleen de volgende leerlingen waren uitgezonderd:

- Leerlingen met een gediagnosticeerde Taal-ontwikkelingsstoornis (TOS)
- Leerlingen met een beperkt gehoor*
- Leerlingen met selectief mutisme
- Leerlingen met afasie
- Leerlingen die minder dan 2 jaar in Nederland verblijven

* Leerlingen die met behulp van een gehoorapparaat goed kunnen meekomen in de klas met mondelinge instructies konden wel meedoen aan het onderzoek.

Leerlinggegevens
Variabelen: Geboortedatum (**QS_V_GEB01**), thuis(taal)situatie (**QS_V_THU01 – QS_V_THUI04**), herkomst (**QS_V_HER01**)

In de Quickscan werd als eerste gevraagd naar het geslacht en de geboortedatum van de leerling. Hierna volgden vier vragen over de onderwijsloopbaan- en behoeften van de leerling. Daarna is gevraagd of de leerling in Nederland is geboren (1 = Ja, 2 = Nee) en wat de thuis(taal)situatie van de leerling is. De thuis(taal)situatie was een multiple respons-vraag waarbij aangekruist kon worden welke thuistaal of thuistalen de leerling heeft: Nederlands, Nederlands dialect, Fries, of Anders (0 = niet aangekruist, 1 = wel aangekruist).

Onderwijsloopbaan en -behoeften
Variabelen: Uitstroomprofiel (**QS_V_UIT01** – **QS_V_UIT18**) en Verblijfsduur sbo/so (**QS_V_VER01**)

Er is ook gevraagd naar welk onderwijstype de leerling naar verwachting zou uitstromen. Hierbij is een onderscheid gemaakt tussen schooltypes en tevens naar het onderscheid tussen regulier voortgezet onderwijs en voortgezet speciaal onderwijs. Leerkrachten konden de onderstaande uitstroomprofielen aankruisen (0 = niet aangekruist, 1 = wel aangekruist):

1. Regulier vwo
2. Regulier havo
3. Regulier vmbo-tl
4. Regulier vmbo-gt
5. Regulier vmbo-kb
6. Regulier vmbo-bb
7. Praktijkonderwijs
8. Vso arbeid
9. Vso dagbesteding
10. Geen uitstroom
11. Vso vwo
12. Vso havo
13. Vso vmbo-tl
14. Vso vmbo-gt
15. Vso vmbo-kb
16. Vso vmbo-bb
17. Anders,
18. … namelijk

Naast het uitstroomprofiel is ook gevraagd sinds welk schooljaar de leerling op het so/sbo zit (1 = 2009-2010, 2 = 2010-2011, 3 = 2011-2012, 4 = 2012-2013, 5 = 2013-2014, 6 = 2014-2015, 7 = 2015-2016, 8 = 2016-2017, 9 = 2017-2018).

Stimulerende en belemmerende factoren
Variabelen: **QS_V_STI01** – **QS_V_STI08**

Er is gevraagd naar acht factoren die zijn gebruikt in het NRO-onderzoek naar de keuze van een passende leerroute (Hollenberg, Keuning en Meijer, 2016). In dat onderzoek is samen met een groep scholen uit alle stimulerende en belemmerende factoren, die zij in hun ontwikkelingsperspectieven opnemen, een aantal factoren gekozen die volgens de scholen het meest relevant zijn. In de Quickscan is een korte definitie opgenomen van elk van de acht factoren:

1. Inhibitie
2. Emotieregulatie
3. Planning en organisatie
4. Zelfstandigheid
5. Motivatie
6. Concentratie en aandacht
7. Schoolwelbevinden
8. Thuissituatie

De leerkracht vulde per factor in of deze belemmerend (1), neutraal (2) of stimulerend (3) was.

Onderwijsbehoeften
Variabelen: **QS_V_OND01** – **QS_V_OND03**

Hierbij werd onderscheid gemaakt tussen drie behoeften zoals uitgewerkt door de CED-groep in het passend lesmodel. Deze driedeling is een vertaling naar de onderwijscontext van de drie basisbehoeften die van belang zijn voor intrinsieke motivatie (verbondenheid, competentie en autonomie; Deci & Ryan, 2000) en luidt als volgt:

1. Behoefte aan extra structuur
2. Behoefte aan stimulering zelfvertrouwen
3. Behoefte aan meer verantwoordelijkheid

Voor elk van de drie onderwijsbehoeften vulde de leerkracht in of de leerling er extra behoefte aan heeft ten opzichte van andere leerlingen op school (0 = niet aangekruist, 1 = wel aangekruist).

Zorgprofiel
Het zorgprofiel werd geïnventariseerd door te vragen naar de aanwezigheid van stoornissen, een IQ ver boven of onder het gemiddelde, en met behulp van drie schalen uit COOL Speciaal (Ledoux e.a., 2012): externaliserend probleemgedrag, internaliserend probleemgedrag en leerachterstand.

A Stoornissen (variabelen: **QS_V_STO01** – **QS_V_STO09**)
De school kon voor onderstaande stoornissen aangeven of deze waren gediagnosticeerd (1) of werden vermoed (2) bij de leerling:

1. ADHD/ADD
2. ODD/CD
3. Autisme en verwante stoornissen (zoals PDD-NOS)
4. Angst- en/of stemmingsstoornis
5. Hechtingsstoornis
6. OCD/dwanghandelingen
7. Gilles de la Tourette
8. Andere psychiatrische stoornis,
9. … namelijk

B IQ ver boven of onder het gemiddelde (variabelen: **QS_V_EIQ01** en **QS_V_EIQ02**)
De leerkracht kon aangeven of vastgesteld was dat de leerling een IQ had boven 130 of onder 70 (0 = niet vastgesteld, 1 = vastgesteld):

1. IQ lager dan 70
2. IQ hoger dan 130

C Probleemgedrag
Variabelen: Externaliserend probleemgedrag (**QS_V_EXT01** – **QS_V_EXT05)**, Internaliserend probleemgedrag (**QS_V_INT01** –**QS_V_INT07**) en leerachterstanden (**QS_V_LEE01** – **QS_V_LEE03**)

Voor 12 verschillende gedragingen die problematisch zijn, konden de leerkrachten aangeven of deze niet (1), in lichte mate (2), in behoorlijke mate (3) of in ernstige mate (4) aanwezig waren.
Externaliserend probleemgedrag was een schaal bestaande uit vijf gedragingen*:

1. Overactief, impulsief gedrag
2. Agressief gedrag tegenover klasgenoten
3. Gedrag dat niet past bij de leeftijd**
4. Sterk storend gedrag voor anderen
5. De leerling is zelf een pester.

* In de oorspronkelijke schaal waren er ook twee items die naar stoornissen (ADHD en ODD) vroegen.
** Het oorspronkelijke item was “Bizar gedrag, of gedrag dat niet past bij de leeftijd”.

Internaliserend probleemgedrag was een schaal bestaande uit de volgende kenmerken:

1. Faalangstig
2. Angstig, neerslachtig, depressief
3. Teruggetrokken, komt niet tot contact met anderen
4. Gebrek aan zelfvertrouwen, negatief beeld van zichzelf
5. Gebrek aan weerbaarheid
6. Gebrek aan sociale vaardigheden
7. Wordt afgewezen en/of gepest door anderen

Verder werd gevraagd of de leerling een achterstand had op drie gebieden:

1. Achterstand in taal/lezen
2. Achterstand in rekenen
3. Achterstand in informatieverwerking

Bij elk gebied kon de leerkracht aangeven of de leerling geen achterstand had (1), minder dan 1 jaar achterstand had (2), 1-2 jaar achterstand had (3) of meer dan 2 jaar achterstand had (4).

3.1.2 Observatiepunten Luisteren, Spreken en Gesprekken

Het tweede gedeelte van de Quickscan is een korte observatielijst op basis van doelen uit Passende Perspectieven en de Cito observatielijst voor vso dagbesteding mondelinge taal. Het doel was om te weten te komen welke leerdoelen van Luisteren, Spreken en Gesprekken de leerling volgens de leerkracht beheerst. Er is naar onderstaande doelen gevraagd. Ter illustratie is per doel ook de omschrijving van leerroute 1 gegeven.

Luisteren (variabelen: **QS_V_LUI01** – **QS_V_LUI03**)

1. De inhoud van een gesproken tekst in eigen woorden weergeven
   - Kan in eigen woorden de hoofdlijnen noemen van een gesproken tekst met een abstract onderwerp dat wat verder van de belevingswereld afligt.

2. De hoofdpunten uit gesproken (nieuws)berichten halen
    - Kan zonder hulp de hoofdpunten halen uit een nieuwsbericht over een concreet of abstract onderwerp.
3. Een oordeel over een tekst(deel) of tv- of radioprogramma verwoorden
    - Kan zonder hulp een oordeel verwoorden over een tekst of tv-programma met een abstracter verhaal dat wat verder van de belevingswereld afligt.

Spreken en gesprekken voeren (variabelen: **QS_V_SPR01** – **QS_V_SPR07**)

1. Verstaanbaarheid
    - Spreekt vrijwel altijd verstaanbaar voor zowel bekenden als onbekenden.
2. Een verhaal vertellen
    - Kan een duidelijk en samenhangend verhaal vertellen over een abstract, niet-alledaags onderwerp.
3. Bij het gespreksonderwerp blijven of afronden
    - Rondt vrijwel altijd eerst het oude gespreksonderwerp af voordat hij een nieuw onderwerp aansnijdt en reageert adequaat als hij wordt bijgestuurd.
4. Een gesprek beginnen
    - Schat in een vrije situatie goed in wat een gepast moment is om een gesprek te starten.
5. Een gesprek op gang houden
    - Kan gesprekken over abstracte, niet-alledaagse onderwerpen zelf op gang houden.
6. Gesprekspartner volgen
    - Reageert adequaat binnen een gesprek over een abstract, niet-alledaags onderwerp.
7. Opmerkingen

De vragen werden beantwoord met behulp van een vierpuntsschaal, waarbij 1 staat voor een lage vaardigheid en 4 voor een hoge vaardigheid. Bij een oordeel over een tekst(deel) of tv- of radioprogramma verwoorden kon de leerkracht bijvoorbeeld kiezen uit: (1) Kan geen oordeel verwoorden bij een concreet onderwerp; (2) Kan met hulp een oordeel verwoorden bij een concreet onderwerp; (3) Kan zonder hulp een oordeel verwoorden bij een abstract onderwerp dichtbij de belevingswereld; en (4) Kan zonder hulp een oordeel verwoorden bij een abstract onderwerp ver van de belevingswereld. Het eerste en tweede schaalpunt (1 en 2) passen qua inhoud bij leerroute 3, het derde schaalpunt (3) bij leerroute 2 en het vierde (4) bij leerroute 1.

De beschreven variabelen zijn niet gebruikt bij de inhoudelijke analyses (zie Hoofdstuk 8), maar kunnen eventueel als achtergrondinformatie gebruikt worden. Verder is een variabele aangemaakt die gebruikt kan worden om te onderzoeken in hoeverre de leerroute die leerkrachten impliciet hebben gekozen bij het invullen van de lijst overeenstemt met het referentieniveau dat we verwachten bij de leerlingen. Daarbij zijn Luisteren, en Spreken en Gesprekken samengenomen, omdat leerlingen in de praktijk ook slechts naar één niveau kunnen uitstromen. De leerroute is bepaald door uit te gaan van de score die het vaakste werd gekozen. Indien de meeste observatiepunten met een 4 (hoge vaardigheid) beoordeeld waren, dan was de verwachte leerroute dus 1 (*N* = 136; 10.1 procent). Als de meeste observatiepunten met een 3 beoordeeld waren, dan was de verwachte leerroute 2 (*N* = 555; 41.2 procent) en als de meeste met een 1 of 2 beoordeeld waren dan was de verwachte leerroute 3 (*N* = 480; 35.6 procent). Indien evenveel observatiepunten met een 4 of 3 beoordeeld waren, dan is de verwachte leerroute gecodeerd als 1,5 (tussen leerroute 1 en 2 in; N = 16; 1.2 procent) en als 1 of 2 en 3 even vaak voorkwamen, dan is de verwachte leerroute gecodeerd als 2,5 (tussen leerroute 2 en 3; *N* = 67; 5.0 procent). Deze indikking is in het datebestand terug te vinden onder **QS_I_LRO_A**.

3.2 Constructie en beschrijving indices

In de Quickscan zijn er drie soorten indices, namelijk eenvoudige indices, samengestelde indices en indices voor abstracte concepten.

- ***Eenvoudige indices*** komen overeen met één respons of transformatie van één respons. Dit is van toepassing bij geslacht, herkomst, verblijfsduur, de drie onderwijsbehoeften (behoefte aan extra structuur, behoefte aan stimulering zelfvertrouwen, behoefte aan meer verantwoordelijkheid), veelvoorkomende stoornissen en leerachterstand in taal/lezen, achterstand in rekenen en achterstand in informatieverwerking. Voor elke index is een nieuwe variabele aangemaakt.
- ***Samengesteld indices*** zijn geconstrueerd door middel van een transformatie van meerdere responsen op één of meerdere vragen. Leeftijd, thuistaal, uitstroomprofiel, onderwijsbehoefte, aantal belemmerende en stimulerende factoren en IQ zijn samengesteld indices. Weinig voorkomende stoornissen zijn ook samengenomen in één variabele.

- ***Indices voor abstracte concepten*,** zoals vaardigheden, worden geconstrueerd met behulp van schaalprocedures. Schaalprocedures zijn gebruikt voor stimulerende en belemmerende factoren en probleemgedrag. Voor de schalen zijn factoranalyses uitgevoerd met varimax rotatie en voor elke schaal is de gemiddelde score uitgerekend.

3.2.1 Leerlinggegevens

De volgende indices zijn aangemaakt:

- De eenvoudige index Geslacht (**QS_I_GES_A**) is een hercodering van de variabele geslacht met de waarden 1 = Jongen en 2 = Meisje.
- Leeftijd (**QS_I_LFT_A**) is een samengestelde index die berekend is als het aantal dagen tussen de afnamedatum en de geboortedatum gedeeld door 365. Leeftijden kleiner dan 5 jaar zijn beschouwd als een invulfout en omgecodeerd naar "niet valide" (waarde 888; 10 gevallen).
- De eenvoudige index Herkomst (**QS_I_HER_B**) is een hercodering van de variabele herkomst met de waarden 0 = Niet Nederland (was 2) en 1 = Nederland.
- De samengestelde index Thuistaal (**QS_I_THU_A**) is berekend met behulp van vier variabelen. De thuis(taal)situatie was een multiple responsvraag, waarbij elk van de vier keuzes als dichotome variabele is weergegeven met 0 = niet gekozen en 1 = is gekozen. Indien geen thuistaal is gekozen dan zijn de variabelen leeg. Er is een nieuwe variabele Thuistaal berekend. Indien is aangegeven dat de thuistaal van de leerling Nederlands, een Nederlands dialect of Fries is, is deze variabele waar (1) anders onwaar (0, tenzij de vraag niet beantwoord was).

Ruim twee derde van de leerlingen in de onderzoeksgroep (*N* = 901; 66.9 procent) waren jongens, en een derde meisjes (*N* = 446; 33.1 procent) en de leerlingen waren tussen de 10 en 14 jaar oud. Gemiddeld waren de leerlingen 12.41 jaar oud (*SD* = 0,52). Bijna 95 procent van de leerlingen is geboren in Nederland en ruim 90 procent van de leerlingen spreekt thuis Nederlands, Fries of een Nederlands dialect (zie Tabel 3.1). Er waren geen significante verschillen tussen sbo- en so-scholen qua thuistaal ($\chi^2(1) = 1.27$; $p = 0.260$) en herkomst ($\chi^2(1) = 0.280$; $p = 0.596$). Op de so-scholen zaten wel relatief meer jongens (88.8 procent, $\chi^2(1) = 69.53$; $p < 0.001$) en de so-leerlingen waren significant ouder dan de sbo-leerlingen (zie Tabel 3.2).

Tabel 3.1. *Geslacht, herkomst en thuistaal van de leerlingen uitgesplitst naar schooltype*

Groep	Index		Freq.	%		Freq.	%	Totaal *N*
Totaal	Geslacht	Jongen	901	66.9	Meisje	446	33.1	1347
	Herkomst	Niet Nederland	67	5.6	Nederland	1136	94.4	1203
	Thuistaal	Niet Nederlands, Fries of Nederlands dialect als thuistaal	119	9.8	Nederlands, Fries of Nederlands dialect als thuistaal	1098	90.2	1217
sbo	Geslacht	Jongen	671	61.7	Meisje	417	38.3	1088
	Herkomst	Niet Nederland	55	5.7	Nederland	902	94.3	957
	Thuistaal	Niet Nederlands, Fries of Nederlands dialect als thuistaal	99	10.3	Nederlands, Fries of Nederlands dialect als thuistaal	865	89.7	964
so	Geslacht	Jongen	230	88.8	Meisje	29	11.2	259
	Herkomst	Niet Nederland	12	4.9	Nederland	234	95.1	246
	Thuistaal	Niet Nederlands, Fries of Nederlands dialect als thuistaal	20	7.9	Nederlands, Fries of Nederlands dialect als thuistaal	233	92.1	253

Tabel 3.2. *Leeftijd van de leerlingen*

Leeftijd	Min	Max	*M*	*SD*	*N*	*F*	*P*	*t*	*df*	*p*
Totaal	10.64	14.84	12.41	0.52	1305					
Sbo	10.64	14.84	12.45	0.50	1060	8.540	0.004	-4.97	341.81	0.000
So	10.84	13.75	12.26	0.55	245					

3.2.2 Onderwijsloopbaan en -behoeften

De volgende indices zijn aangemaakt:

- Uitstroomprofiel (**QS_I_UIT_A**) is een samengestelde index. De vraag naar het verwachte type onderwijs waarnaar de leerling zou gaan uitstromen was een multiple responsvraag, waarbij de aangevinkte keuze als 1 is gecodeerd en alle niet gekozen opties met 0. Hiervan is één categorische variabele gemaakt met de volgende indeling in categorieën : 0 = Andere uitstroom, 1 = vso dagbesteding of vso arbeid, 2 = praktijkonderwijs, 3 = vmbo-bb, 4 = vmbo-kb, 6 = vmbo-gt, 6 = havo/vwo. Indien meerdere opties waren gekozen (*N* = 49; 3.6 procent) is het hoogste uitstroomprofiel gekozen.

- De index Verblijfsduur (**QS_I_VER_A**) is een hercodering van de vraag sinds welk schooljaar de leerling op het so/sbo zit, zodat de waarde het aantal jaren op so/sbo weergeeft: (9=0) (8=1) (7=2) (6=3) (5=4) (4=5) (3=6) (2=7) (1=8).

- De onderwijsbehoeften (**QS_I_OND_A**, **QS_I_OND_B**, **QS_I_OND_C**) zijn eenvoudige indices. Voor elk van de drie onderwijsbehoeften (behoefte aan extra structuur, stimulering zelfvertrouwen en aan meer verantwoordelijkheid) heeft de leerkracht ingevuld of de leerling daaraan extra behoefte had ten opzichte van andere leerlingen op school (0 = Niet aangekruist, 1 = wel aangekruist). Indien geen van de onderwijsbehoeften waren aangekruist en de voorgaande vragen (naar Stimulerende en belemmerende factoren) en de erop volgende vragen (Thuistaal en Geboorteland) ook leeg waren (*N* = 105; 7.8 procent), dan zijn de vragen over onderwijsbehoeften gecodeerd als niet-beantwoord (999).

- Op basis van de vraag welke factoren belemmerend, neutraal of stimulerend zijn voor de leerling, hebben we drie indices berekend. Allereerst is het aantal stimulerende factoren (**QS_I_STI_B**) berekend (aantal factoren met de waarde 3) en het aantal belemmerende factoren (**QS_I_STI_C**; het aantal factoren met de waarde 1). Voor de stimulerende en belemmerende factoren is daarna ook een factoranalyse met varimax rotatie uitgevoerd. Er waren twee factoren met een eigenwaarde groter dan 1, maar de screeplot liet een knik zien na de eerste factor, wat duidt op één factor. De betrouwbaarheid van de schaal was redelijk goed (Cronbach's alpha = 0.78). Wel is in Tabel 3.3 te zien dat de laatste twee factoren (schoolwelbevinden en thuissituatie) een lagere factorlading (FL) hebben. Schoolwelbevinden werd het meest positief beoordeeld van alle factoren (*M* = 2.37).

Tabel 3.3. *Kenmerken en factorladingen van de stimulerende en belemmerende factoren*

Stimulerende en belemmerende factoren	*M*	*SD*	*N*	r_{it}	Cronbach's alpha zonder item	FL
Inhibitie	1.69	0.72	1139	0.49	0.75	0.53
Emotieregulatie	1.56	0.67	1139	0.42	0.76	0.41
Planning en organisatie	1.83	0.72	1139	0.52	0.75	0.62
Zelfstandigheid	2.03	0.74	1139	0.53	0.75	0.62
Motivatie	2.12	0.76	1139	0.61	0.73	0.75
Concentratie en aandacht	1.86	0.76	1139	0.58	0.73	0.73
Schoolwelbevinden	2.37	0.65	1139	0.36	0.77	0.36
Thuissituatie	2.14	0.73	1139	0.32	0.78	0.32

Per leerling is de gemiddelde schaalscore berekend; de score is opgenomen onder variabele **QS_I_STI_A**. Indien geen van de items beantwoord zijn, zijn de drie indices beschouwd als niet-beantwoord (999). Als meer dan 6 procent (5 of meer) items ontbreken, zijn de indices niet berekend, maar in plaats daarvan als ontbrekend gecodeerd (999). Dit was slechts bij 1 leerling het geval.

Als we de genoemde indices analyseren, zien we dat de leerlingen gemiddeld 3.68 jaar (*SD* = 1.96) op het so/sbo zaten (zie Tabel 3.6). Verder is in Tabel 3.4 te zien dat de leerlingen naar verwachting vooral naar het praktijkonderwijs (36.5 procent) en het vmbo-bb (30.9 procent) zouden gaan uitstromen. Er zijn echter duidelijke uitstroomverschillen tussen sbo en so ($\chi^2(6) = 200.73$; $p < 0.001$). Driekwart van de leerlingen in het sbo zou naar verwachting uitstromen naar het praktijkonderwijs (43.2 procent) of vmbo-bb (32.0 procent). In het so waren de verwachtingen over de uitstroomniveaus hoger. Van slechts 9.9 procent van de so-leerlingen werd verwacht dat zij zouden uitstromen naar het praktijkonderwijs, ruim een kwart naar vmbo-bb (26.5 procent), ruim een kwart naar vmbo-kb (26.5 procent) en 21.7 procent naar vmbo-gt.

Tabel 3.4. *Verwachte uitstroom naar het voortgezet onderwijs*

	sbo		so		Totaal	
Uitstroom	Freq.	%	Freq.	%	Freq.	%
Anders	67	6.7%	15	5.9%	82	6.5%
VSO dagbesteding of VSO arbeid	17	1.7%	7	2.8%	24	1.9%
Praktijkonderwijs	435	43.2%	25	9.9%	460	36.5%
vmbo-bb	322	32.0%	67	26.5%	389	30.9%
vmbo-kb	105	10.4%	67	26.5%	172	13.7%
vmbo-gt	47	4.7%	55	21.7%	102	8.1%
havo/vwo	13	1.3%	17	6.7%	30	2.4%
Totaal	1006	100.0%	253	100.0%	1259	100.0%

Wat betreft de onderwijsbehoeften had 55.8 procent van de leerlingen behoefte aan extra structuur, 55.1 procent aan stimulering van het zelfvertrouwen, en 38.0 procent aan meer verantwoordelijkheid (zie Tabel 3.5). In het so gold voor significant meer leerlingen (63.1 procent) dat zij behoefte hadden aan extra structuur ($\chi^2(1) = ;\ p < 0{,}01$) dan in het sbo (53.9 procent). Er waren geen significante verschillen tussen het so en sbo wat betreft de behoefte van de leerlingen aan extra stimulering van het zelfvertrouwen ($\chi^2(1) = 0.12$; $p = 0.731$) of extra verantwoordelijkheid ($\chi^2(1) = 0.35$; $p = 0.554$).

Tabel 3.5. *Aantal leerlingen met extra onderwijsbehoeften ten opzichte van leerlingen op de eigen school*

Groep	Onderwijsbehoefte		Frequentie	%		Frequentie	%	Totaal
Totaal	structuur	Nee	549	44.2%	Extra	693	55.8%	1242
	zelfvertrouwen	Nee	558	44.9%	Extra	684	55.1%	1242
	verantwoordelijkheid	Nee	770	62.0%	Extra	472	38.0%	1242
So	structuur	Nee	94	36.9%	Extra	161	63.1%	255
	zelfvertrouwen	Nee	117	45.9%	Extra	138	54.1%	255
	verantwoordelijkheid	Nee	154	60.4%	Extra	101	39.6%	255
Sbo	structuur	Nee	455	46.1%	Extra	532	53.9%	987
	zelfvertrouwen	Nee	441	44.7%	Extra	546	55.3%	987
	verantwoordelijkheid	Nee	616	62.4%	Extra	371	37.6%	987

De mate waarin er sprake was van stimulerende factoren (zie Tabel 3.6) was gemiddeld bijna neutraal (*M* = 1.94; *SD* = 0.45) op de driepuntsschaal van belemmerend (1) tot stimulerend (3). Op het so was de mate waarin er sprake was van stimulerende factoren significant lager dan in het sbo ($t_{1192} = -3{,}50$; $p < 0.001$). Van de negen factoren die stimulerend of belemmerend kunnen zijn, werden gemiddeld 2.54 factoren (*SD* = 2.03) belemmerend genoemd en 2.09 factoren stimulerend (*SD* = 2.03). Er werden significant meer factoren belemmerend gevonden dan stimulerend ($t_{1193} = 4.37$; $p<0.01$). Bij leerlingen in het so werden significant meer factoren als belemmerend aangemerkt dan bij leerlingen in het sbo ($t_{1192} = 5.20$; $p < 0.001$), maar evenveel factoren als stimulerend ($t_{1192} = -0.90$; $p = 0.369$).

Tabel 3.6. *Analyse naar verblijfsduur op het so/sbo en stimulerende en belemmerende factoren*

Groep	Index	Min	Max	*M*	*SD*	*N*
Totaal	Jaren in het so/sbo	0	8	3.68	1.96	1216
	Mate stimulerende factoren	1.00	3.00	1.94	0.45	1194
	Aantal belemmerende factoren	0	8	2.54	2.03	1194
	Aantal stimulerende factoren	0	8	2.09	2.03	1194
so	Jaren in het so/sbo	0	8	3.60	2.27	252
	Mate stimulerende factoren	1.00	3.00	1.85	0.44	255
	Aantal belemmerende factoren	0	8	3.12	2.09	255
	Aantal stimulerende factoren	0	8	1.99	1.84	255
sbo	Jaren in het so/sbo	0	8	3.69	1.87	964
	Mate stimulerende factoren	1.00	3.00	1.97	0.45	939
	Aantal belemmerende factoren	0	8	2.38	1.99	939
	Aantal stimulerende factoren	0	8	2.12	2.08	939

3.2.3 Zorgprofiel

De volgende indices zijn aangemaakt:

- **Probleemgedrag.** Een factoranalyse met varimax rotatie met al het probleemgedrag samen liet zowel in het screeplot als qua eigenwaardes drie factoren zien. In Tabel 3.7 is te zien dat de eerste factor het externaliserend probleemgedrag vertegenwoordigt, exclusief het item ”Gedrag dat niet past bij de leeftijd”. De tweede factor verwijst naar negatieve gevoelens en gedachten over zichzelf en bestaat uit de items: Faalangstig, Gebrek aan zelfvertrouwen, en Angstig-Neerslachtig-Depressief. De derde factor, ten slotte, verwijst naar internaliserend sociaal probleemgedrag. De volgende items vallen eronder: Gebrek aan weerbaarheid, Teruggetrokken, Gebrek aan sociale vaardigheden, Wordt afgewezen en/of gepest door anderen, en Gedrag dat niet past bij de leeftijd.

  Tabel 3.7. *Resultaten exploratieve factoranalyse Probleemgedrag*

	M	*SD*	*N*	r_{it}	*Alpha*	*FL1*	*FL2*	*FL3*
Externaliserend probleemgedrag								
Overactief, impulsief gedrag	2.03	1.01	1153	0.710	0.79	**0.81**	0.00	0.10
Agressief tegenover klasgenoten	1.37	0.62	1153	0.620	0.82	**0.64**	0.07	0.10
Sterk storend voor anderen	1.77	0.91	1153	0.780	0.74	**0.87**	-0.02	0.20
Is zelf een pester	1.41	0.67	1153	0.650	0.81	**0.70**	0.00	0.03
Gedrag dat niet past bij de leeftijd	1.77	0.89	1133	0.470	0.71	0.25	0.08	**0.55**
Internaliserend probleemgedrag								
Teruggetrokken	1.46	0.75	1133	0.450	0.72	-0.31	0.29	**0.55**
Gebrek aan weerbaarheid	1.86	0.90	1133	0.580	0.67	-0.16	0.43	**0.61**
Gebrek aan sociale vaardigheden	1.96	0.88	1133	0.600	0.66	0.31	0.19	**0.67**
Wordt afgewezen of gepest door anderen	1.38	0.65	1133	0.470	0.72	0.25	0.12	**0.51**
Faalangstig	2.04	0.87	1150	0.700	0.73	0.03	**0.82**	0.12
Angstig, neerslachtig, depressief	1.54	0.77	1150	0.590	0.83	0.04	**0.59**	0.37
Gebrek aan zelfvertrouwen	1.99	0.84	1150	0.740	0.68	0.05	**0.84**	0.22

  In Tabel 3.7 is verder te zien dat de betrouwbaarheid van de drie subschalen redelijk goed tot goed is. De schaal Externaliserend probleemgedrag (Exclusief het item “Gedrag dat niet past bij de leeftijd”; **QS_I_EXT_A**) heeft een interne consistentie (Cronbach’s alpha) van 0.84. De tweede schaal, Internaliserend negatief probleemgedrag (**QS_I_INT_A**), heeft een interne consistentie van 0.82 en de derde schaal, Internaliserend sociaal probleemgedrag (**QS_I_INT_B**), heeft een interne consistentie van 0.75. Voor de drie schalen is de gemiddelde score berekend op een vierpuntsschaal (1 = ”niet aanwezig”, 2 = ”in lichte mate”, 3 = ”in behoorlijke mate”, 4 = ”in ernstige mate”). Als meer dan 6 procent (8 of meer) van de items ontbreken, zijn geen schaalscores berekend, maar zijn ze als missing gecodeerd (999). Dit was het geval bij 16 leerlingen.

- **Stoornissen.** De leerkracht kon voor zeven stoornissen en de categorie “Andere psychiatrische stoornis” aangeven of deze was gediagnosticeerd (1) of werd vermoed (2) bij de leerling. Voor de

stoornissen die minstens 25 keer of vaker *gediagnosticeerd* waren, zijn eenvoudige indices berekend. Dit was bij de volgende drie stoornissen het geval (0 = Niet gediagnosticeerd, 1 = Gediagnosticeerd): ADHD/ADD (**QS_I_STO_A**), Autisme en verwante stoornissen (**QS_I_STO_B**) en Hechtingsstoornis (**QS_I_STO_C**). Indien één van de andere vier stoornissen was gediagnosticeerd (ODD/CD, Angst- en/of stemmingsstoornis, OCD/dwanghandelingen, Gilles de la Tourette) is deze ondergebracht onder de variabele Andere stoornis (**QS_I_STO_D**): 0 = Niet gediagnosticeerd, 1 = Gediagnosticeerd. Missings zijn gecodeerd als 999.

- **Leerachterstand.** De leerachterstanden zijn eenvoudige indices die gelijk zijn aan de oorspronkelijke items. Er is gevraagd naar de achterstand in taal/lezen (**QS_I_LEE_A**), de achterstand in rekenen (**QS_I_LEE_B**) en de achterstand in informatieverwerking (**QS_I_LEE_C**). De leerkracht kon aangegeven of de leerling geen achterstand had (1), minder dan 1 jaar achterstand had (2), 1-2 jaar achterstand had (3) of meer dan 2 jaar achterstand had (4).

- **IQ ver boven of onder het gemiddelde.** De antwoorden op de twee vragen over IQ (vastgesteld dat de leerling een IQ heeft hoger dan 130 en vastgesteld dat de leerling een IQ heeft lager dan 70) zijn samengevoegd tot één index IQ (**QS_I_EIQ_A**), waarbij 1 = IQ lager dan 70, 2 = IQ niet ver boven of onder gemiddeld en 3 = IQ hoger dan 130. Als de voorgaande vragen over probleemgedrag en leerachterstand niet beantwoord waren, is de vraag over IQ als niet beantwoord beschouwd en is de index als missing (999) gecodeerd.

Als we de genoemde indices analyseren, zien we dat al het probleemgedrag in lichte mate (score 2) aanwezig was (zie Tabel 3.8). Internaliserend emotioneel probleemgedrag was het meest sterk aanwezig (M = 2.06; SD = 0.78) en externaliserend probleemgedrag het minst sterk (M = 1.89; SD = 0.78). T-toetsen lieten zien dat internaliserend negatief probleemgedrag (M = 1.86; SD = 0.71) in significant sterkere mate aanwezig was dan externaliserend probleemgedrag (M = 1,65; SD = 0.67; t_{1166} = 7.74; p < 0.001) en internaliserend sociaal probleemgedrag (M = 1,69; SD = 0.58; t_{1169} = 9.11, p < 0,001). Er was geen significant verschil in de mate van externaliserend probleemgedrag en internaliserend emotioneel probleemgedrag (t_{1169} = -1.61; p = 0.107). Alle drie vormen van probleemgedrag waren in significant sterkere mate aanwezig bij so-leerlingen dan bij sbo-leerlingen.

Tabel 3.8. *Analyse probleemgedrag en leerachterstand*

Groep	Categorie	*Min*	*Max*	*M*	*SD*	*N*	*t**	*df*	*p*
Totaal	Externaliserend probleemgedrag	1.00	4.00	1.65	0.67	1170			
	Internaliserend negatief probleemgedrag	1.00	4.00	1.86	0.71	1170			
	Internaliserend sociaal probleemgedrag	1.00	4.00	1.69	0.58	1173			
	Leerachterstand in taal/lezen	1.00	4.00	2.93	1.11	1143			
	Leerachterstand in rekenen	1.00	4.00	3.10	1.02	1149			
	Leerachterstand in informatieverwerking	1.00	4.00	2.96	1.03	1098			
So	Externaliserend probleemgedrag	1.00	3.75	1.89	0.71	239	5.93	344.88	0.00
	Internaliserend negatief probleemgedrag	1.00	4.00	2.06	0.78	239	4.55	336.69	0.00
	Internaliserend sociaal probleemgedrag	1.00	3.80	1.92	0.60	239	7.02	1171	0.00
	Leerachterstand in taal/lezen	1.00	4.00	2.21	1.10	233	-11.39	345.35	0.00
	Leerachterstand in rekenen	1.00	4.00	2.18	1.07	233	-15.08	313.61	0.00
	Leerachterstand in informatieverwerking	1.00	4.00	2.18	1.05	229	-12.95	327.55	0.00
Sbo	Externaliserend probleemgedrag	1.00	4.00	1.59	0.65	931			
	Internaliserend negatief probleemgedrag	1.00	4.00	1.81	0.68	931			
	Internaliserend sociaal probleemgedrag	1.00	4.00	1.63	0.56	934			
	Leerachterstand in taal/lezen	1.00	4.00	3.12	1.04	910			
	Leerachterstand in rekenen	1.00	4.00	3.33	0.87	916			
	Leerachterstand in informatieverwerking	1.00	4.00	3.16	0.92	869			

* Wanneer uit de F-toets bleek dat de varianties significant van elkaar verschillen, is de t-toets voor ongelijke varianties gebruikt.

Er is ook gevraagd naar de mate van leerachterstand in taal/lezen, rekenen en informatieverwerking. Zoals in Tabel 3.8 te zien is, hadden de leerlingen gemiddeld 1-2 jaar achterstand bij taal/lezen (M = 2.93; SD = 1.11), bij rekenen (M = 3.10; SD = 1.02) en bij informatieverwerking (M = 2.96; SD = 1.03). Leerlingen in het so hadden significant minder leerachterstand in taal/lezen, rekenen en informatieverwerking dan de leerlingen in het sbo. Wat betreft de stoornissen zien we in Tabel 3.9 dat ADHD/ADD bij ruim een kwart van de leerlingen was gediagnosticeerd (26.7 procent), autisme en verwante stoornissen bij 21.0 procent en

een hechtingsstoornis bij 3.2 procent. Bij 11.0 procent van de leerlingen is (ook) een andere stoornis gediagnosticeerd. Bij de leerlingen op het so zijn alle vier de stoornissen significant vaker gediagnosticeerd. Met name autisme was vaker gediagnosticeerd bij so-leerlingen: 30.8 procent vaker ($\chi^2(1) = 116.42$; $p < 0.001$). Verder wordt ADHD in het so bij 9.3 procent vaker gediagnosticeerd dan in het sbo ($\chi^2(1) = 8.89$; $p < 0.001$), een hechtingsstoornis 5.8 procent vaker ($\chi^2(1) = 22.11$; $p < 0.001$) en een andere stoornis 14.8 procent ($\chi^2(1) = 45.14$; $p < 0.001$) vaker.

Tabel 3.9. *Analyse gediagnosticeerde stoornissen*

Groep	Gediagnosticeerde stoornissen		Freq.	%		Freq.	%	Totaal
Totaal	ADHD/ADD	Niet	912	73.3%	Wel	333	26.7%	1245
	Autisme en verwante stoornissen	Niet	984	79.0%	Wel	261	21.0%	1245
	Hechtingsstoornis	Niet	1205	96.8%	Wel	40	3.2%	1245
	Andere stoornis	Niet	1108	89.0%	Wel	137	11.0%	1245
so	ADHD/ADD	Niet	168	65.9%	Wel	87	34.1%	255
	Autisme en verwante stoornissen	Niet	139	54.5%	Wel	116	45.5%	255
	Hechtingsstoornis	Niet	235	92.2%	Wel	20	7.8%	255
	Andere stoornis	Niet	197	77.3%	Wel	58	22.7%	255
sbo	ADHD/ADD	Niet	744	75.2%	Wel	246	24.8%	990
	Autisme en verwante stoornissen	Niet	845	85.4%	Wel	145	14.6%	990
	Hechtingsstoornis	Niet	970	98.0%	Wel	20	2.0%	990
	Andere stoornis	Niet	911	92.0%	Wel	79	8.0%	990

Ten slotte is bij 218 leerlingen (18.1 procent) een IQ lager dan 70 gediagnosticeerd en bij 6 leerlingen (0.5 procent) een IQ hoger dan 130 (zie Tabel 3.10). Bij sbo-leerlingen is significant vaker een IQ lager dan 70 vastgesteld ($\chi^2(2) = 35{,}69$; $p < 0{,}001$), namelijk bij 21.1 procent tegenover 6.3 procent bij de so-leerlingen.

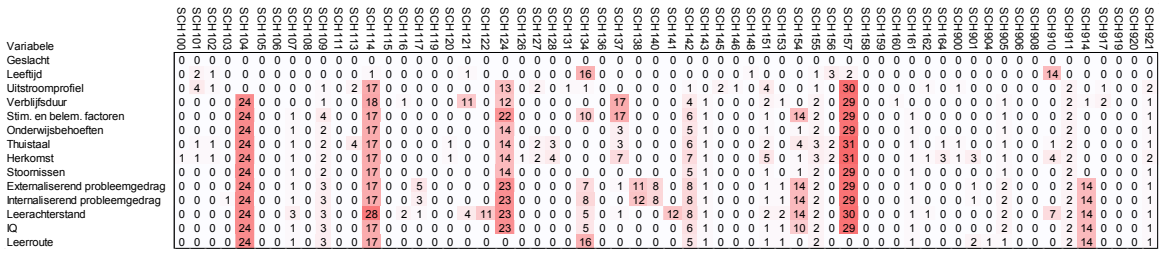
Tabel 3.10. *Analyse naar IQ*

	IQ vastgesteld lager dan 70		IQ niet vastgesteld lager dan 70 of hoger dan 130		IQ vastgesteld hoger dan 130		Totaal
Groep	Freq.	%	Freq.	%	Freq.	%	
Totaal	218	18.1%	980	81.4%	6	0.5%	1204
So	15	6.3%	221	92.1%	4	1.7%	240
Sbo	203	21.1%	759	78.7%	2	0.2%	964

3.3 Missende waarden

Omdat een aantal leerkrachten geen (al te) persoonlijke informatie over leerlingen wilde invullen, is het percentage ontbrekende waarnemingen bij de Quickscan vrij hoog (zie Tabel 3.10). Een plot van alle ontbrekende waarnemingen (zie Figuur 3.1) laat zien dat op sommige scholen systematisch geen informatie is ingevuld over de leerlingkenmerken. We beschouwen drie achtereenvolgende vragen met ontbrekende waarnemingen bij groepjes (van minstens 5) leerlingen op dezelfde school als indicatie dat de leerkracht weigerde deze informatie te verstrekken. Als de leerlingen waar dit voor geldt buiten beschouwingen worden gelaten, dan zien we dat het percentage van (incidentele) ontbrekende waarnemingen bij de Quickscan heel acceptabel is.

Figuur 3.1. *Ontbrekende waarden per onderdeel uitgesplitst naar school*



Tabel 3.10. Aantal leerlingen met ontbrekende waarnemingen per variabele

	Totaal		Exclusief leerlingen met systematisch ontbrekende waarnemingen	
Variabele	Aantal	%	Aantal	%
Geslacht	0	0.0%	0	0.0%
Leeftijd	42	3.1%	34	2.8%
Uitstroomprofiel	88	6.5%	27	2.2%
Verblijfsduur	131	9.7%	44	3.6%
Stimulerende en belemmerende factoren	153	11.4%	36	3.0%
Onderwijsbehoeften	105	7.8%	15	1.2%
Thuistaal	130	9.7%	37	3.0%
Herkomst	144	10.7%	54	4.4%
Stoornissen	102	7.6%	12	1.0%
Externaliserend probleemgedrag	177	13.1%	45	3.7%
Internaliserend probleemgedrag	177	13.1%	45	3.7%
Leerachterstand	204	15.1%	72	5.9%
IQ	143	10.6%	15	1.2%
Luisteren, Spreken en Gesprekken	93	6.9%	29	2.4%

3.4 Variabelen

Zoals uit de voorgaande paragrafen is af leiden, zijn op basis van de Quickscan de volgende 24 indices geconstrueerd en opgenomen in het databestand.

QS_I_GES_A Geslacht
QS_I_LFT_A Leeftijd
QS_I_THU_A Thuistaal
QS_I_HER_B Herkomst
QS_I_UIT_A Uitstroomprofiel
QS_I_VER_A Verblijfsduur in sbo/so
QS_I_STI_A Mate waarin er sprake is van stimulerende factoren
QS_I_STI_B Aantal stimulerende factoren
QS_I_STI_C Aantal belemmerende factoren
QS_I_OND_A Onderwijsbehoeften: behoefte aan extra structuur
QS_I_OND_B Onderwijsbehoeften: stimulering zelfvertrouwen
QS_I_OND_C Onderwijsbehoeften: meer verantwoordelijkheid
QS_I_EXT_A Externaliserend probleemgedrag
QS_I_INT_A Internaliserend negatief probleemgedrag
QS_I_INT_B Internaliserend sociaal probleemgedrag
QS_I_LEE_A Leerachterstand in taal/lezen
QS_I_LEE_B Leerachterstand in rekenen
QS_I_LEE_C Leerachterstand in informatieverwerking
QS_I_EIQ_A Laag of hoog IQ vastgesteld
QS_I_STO_A ADHD/ADD gediagnosticeerd
QS_I_STO_B Autisme of een verwante stoornis gediagnosticeerd
QS_I_STO_C Hechtingsstoornis gediagnosticeerd
QS_I_STO_D Andere stoornis gediagnosticeerd
QS_I_LRO_A Inschatting leerroute leerkracht

De oorspronkelijke variabelen zitten ook in het databestand (zie het overzicht hieronder). In paragraaf 3.1 is elke variabele uitgebreid omschreven.

QS_V_GEB01 Geboortedatum
QS_V_UIT01 Uitstroomprofiel: regulier vwo
QS_V_UIT02 Uitstroomprofiel: regulier havo
QS_V_UIT03 Uitstroomprofiel: regulier vmbo tl
QS_V_UIT04 Uitstroomprofiel: regulier vmbo gt
QS_V_UIT05 Uitstroomprofiel: regulier vmbo kb
QS_V_UIT06 Uitstroomprofiel: regulier vmbo bb
QS_V_UIT07 Uitstroomprofiel: praktijkonderwijs

QS_V_UIT08	Uitstroomprofiel: vso arbeid
QS_V_UIT09	Uitstroomprofiel: vso dagbesteding
QS_V_UIT10	Uitstroomprofiel: geen uitstroom
QS_V_UIT11	Uitstroomprofiel: vso vwo
QS_V_UIT12	Uitstroomprofiel: vso havo
QS_V_UIT13	Uitstroomprofiel: vso vmbo tl
QS_V_UIT14	Uitstroomprofiel: vso vmbo gt
QS_V_UIT15	Uitstroomprofiel: vso vmbo kb
QS_V_UIT16	Uitstroomprofiel: Vvso vmbo bb
QS_V_UIT17	Uitstroomprofiel: anders namelijk
QS_V_UIT18	Uitstroomprofiel: anders namelijk txt
QS_V_VER01	Sinds welk schooljaar zit de leerling op het so/sbo?
QS_V_STI01	Stimulerende en belemmerende factoren: Inhibitie
QS_V_STI02	Stimulerende en belemmerende factoren: Emotieregulatie
QS_V_STI03	Stimulerende en belemmerende factoren: Planning en organisatie
QS_V_STI04	Stimulerende en belemmerende factoren: Zelfstandigheid
QS_V_STI05	Stimulerende en belemmerende factoren: Motivatie
QS_V_STI06	Stimulerende en belemmerende factoren: Concentratie en aandacht
QS_V_STI07	Stimulerende en belemmerende factoren: Schoolwelbevinden
QS_V_STI08	Stimulerende en belemmerende factoren: Thuissituatie
QS_V_OND01	Onderwijsbehoeften: Structuur bieden
QS_V_OND02	Onderwijsbehoeften: Zelfvertrouwen stimuleren
QS_V_OND03	Onderwijsbehoeften: Verantwoordelijkheid geven
QS_V_THU01	Thuistaal: Nederlands
QS_V_THU02	Thuistaal: Nederlands dialect
QS_V_THU03	Thuistaal: Fries
QS_V_THU04	Thuistaal: anders
QS_V_HER01	Is de leerling in Nederland geboren?
QS_V_STO01	Stoornissen: ADHD/ADD
QS_V_STO02	Stoornissen: ODD/CD
QS_V_STO03	Stoornissen: Autisme en verwante stoornissen (zoals PDD-NOS)
QS_V_STO04	Stoornissen: Angst- en/of stemmingsstoornis
QS_V_STO05	Stoornissen: Hechtingsstoornis
QS_V_STO06	Stoornissen: OCD/dwanghandelingen
QS_V_STO07	Stoornissen: Gilles de la Tourette
QS_V_STO08	Stoornissen: Andere psychiatrische stoornis, namelijk
QS_V_STO09	Stoornissen: anders, namelijk txt
QS_V_EXT01	Externaliserend probleemgedrag: Overactief, impulsief gedrag
QS_V_EXT02	Externaliserend probleemgedrag: Agressief tegenover klasgenoten
QS_V_EXT03	Externaliserend probleemgedrag: Gedrag dat niet past bij de leeftijd
QS_V_EXT04	Externaliserend probleemgedrag: Sterk storend voor anderen
QS_V_EXT05	Externaliserend probleemgedrag: Is zelf een pester
QS_V_INT01	Internaliserend probleemgedrag: Faalangstig
QS_V_INT02	Internaliserend probleemgedrag: Angstig, neerslachtig, depressief
QS_V_INT03	Internaliserend probleemgedrag: Teruggetrokken
QS_V_INT04	Internaliserend probleemgedrag: Gebrek aan zelfvertrouwen, negatief zelfbeeld
QS_V_INT05	Internaliserend probleemgedrag: Gebrek aan weerbaarheid
QS_V_INT06	Internaliserend probleemgedrag: Gebrek aan sociale vaardigheden
QS_V_INT07	Internaliserend probleemgedrag: Wordt afgewezen en/of gepest door anderen
QS_V_LEE01	Leerachterstand: Achterstand in taal/lezen
QS_V_LEE02	Leerachterstand: Achterstand in rekenen
QS_V_LEE03	Leerachterstand: Achterstand in informatieverwerking
QS_V_EIQ01	IQ lager dan 70
QS_V_EIQ02	IQ hoger dan 130
QS_V_LUI01	De inhoud van een gesproken tekst in eigen woorden weergeven
QS_V_LUI02	de hoofdpunten uit gesproken (nieuws)berichten halen
QS_V_LUI03	Een oordeel over een tekst(deel) of tv- of radioprogramma verwoorden
QS_V_SPR01	Verstaanbaarheid
QS_V_SPR02	Een verhaal vertellen
QS_V_SPR03	Bij het gespreksonderwerp blijven of afronden
QS_V_SPR04	Een gesprek beginnen
QS_V_SPR05	Een gesprek op gang houden
QS_V_SPR06	Gesprekspartner volgen
QS_V_SPR07	Opmerkingen

4 Luisteren

4.1 Inhoud en afname

De luistervaardigheid van leerlingen is gemeten met behulp van items uit de Ankerset, PPON en het Cito-LVS. Om rekening te houden met niveauverschillen tussen leerlingen (en het so en sbo) zijn uit de itembank drie verschillende toetsversies samengesteld die gedeeltelijk overlappen. In tabel 4.1 staan de percentages (en aantallen) items per toetsversie voor elk van de genoemde bronnen. Er is bewust voor gekozen om de Ankerset het grootste gewicht te geven in elk van de toetsversies, omdat de prestatiestandaarden voor 1F en 2F dan beter over te brengen zijn van de Ankerset naar de toetsen in dit onderzoek; de prestatiestandaarden vormen een belangrijk onderdeel van de inhoudelijke analyses die in Hoofdstuk 8 gepresenteerd worden.

Tabel 4.1. *Herkomst items in percentages en (aantallen) uitgesplitst naar versie*

	Versie 1	Versie 2	Versie 3
Ankerset	57.1 (16)	41.7 (10)	45.0 (9)
Cito LVS/VSO DB	21.4 (6)	33.3 (8)	30.0 (6)
PPON	21.4 (6)	25.0 (6)	25.0 (5)
Totaal aantal items	28	24	20

Met het oog op het vergelijkbaar maken (of equivaleren) van de drie toetsversies is ervoor gezorgd dat verschillende groepen leerlingen gedeeltelijk dezelfde items maken. Figuur 4.1 laat zien hoe de overlap tussen de verschillende toetsversies precies gerealiseerd is. In Figuur 4.1 is bijvoorbeeld te zien dat het fragment “Melkmeisje” zowel in toetsversie 1 als in toetsversie 2 zat. Elke toets had een eigen boekje met items en een bijbehorende dvd met filmpjes.

Figuur 4.1. *Afnamedesign Luisteren*

			Toetsversie		
Fragment	**Herkomst**	***N* items**	1	2	3
Papier scheppen	Ankerset	4	■		
Ridders	Ankerset	5	■		
Dierentuin	Ankerset	3	■		
Pindakaas	Cito LVS	6	■		
Melkmeisje	Ankerset	4	■	■	
Bacterien	PPON	6	■	■	
Tekenares	Cito LVS	4		■	
Koningscobra	Ankerset	3		■	■
Zebra	Ankerset	3		■	■
Divers VSODB 1	VSODB	4		■	■
Kom in actie	Ankerset	3			■
Hoofdrekenen	PPON	5			■
Divers VSODB 2	VSODB	2			■

Om de toetsversies gericht aan leerlingen te kunnen toewijzen is aan leerkrachten gevraagd om de leerlingen te verdelen in drie (ongeveer even grote) groepen.

- De leerlingen met de hoogste (door de leerkracht ingeschatte) niveaus maakten toetsversie 1.
- De leerlingen met de laagste (door de leerkracht ingeschatte) niveaus maakten toetsversie 3.
- De middengroep maakte toetsversie 2.

Een uitzondering werd gevormd door scholen met (vrijwel) uitsluitend leerlingen die werken in leerroute 1. Deze leerlingen werden in twee groepen verdeeld, zodat de ene helft toetsversie 1 maakte en de andere helft versie 2. Bij scholen met (vrijwel) uitsluitend leerlingen in leerroute 3 werden de leerlingen eveneens in twee groepen verdeeld, zodat de ene helft toetsversie 2 maakte en de andere helft versie 3.

4.2 Missende waarden

Vanwege het gekozen afnamedesign hebben we bij het onderdeel Luisteren te maken met structurele missende waarden; niet aan alle leerlingen zijn alle items voorgelegd. Met behulp van item respons theorie (IRT) analyse kunnen de scores op de drie toetsversies op één vaardigheidsschaal worden afgebeeld, zodat ze onderling vergelijkbaar worden. Het is dan geen probleem meer dat elke leerling slechts een deelverzameling van de items heeft gemaakt. Ontbrekende antwoorden bij leerlingen binnen een toetsversie kunnen echter wel problematisch zijn. Gelukkig is in Tabel 4.2 te zien dat de meeste leerlingen alle items in hun toets hebben beantwoord. Slechts 28 van de 1414 leerlingen hebben twee of meer ontbrekende antwoorden. Een aantal van hen heeft maar een halve toets gemaakt, zoals in Tabel 4.2 dan te zien is aan een rij achtereenvolgende missende waarden. In de toets- en itemanalyses (TIA) en IRT-modellering zijn de leerlingen met meer dan 20 procent missende waarden niet meegenomen (21 leerlingen in totaal). Bij de overige leerlingen ging het om zo weinig ontbrekende antwoorden dat redelijkerwijs aangenomen kon worden dat zij het antwoord niet wisten. Daarom is in die gevallen het ontbrekende antwoord fout gerekend en dus een score van 0 geïmputeerd. Gegeven de kleine aantallen waarbij deze imputatie is toegepast (0.3 procent van de itemresponses), kan deze de uitkomsten nauwelijks beïnvloeden, zeker niet als het om uitkomsten op systeemniveau gaat zoals in dit onderzoek het geval is.

Tabel 4.2. *Overzicht missende waarden bij Luisteren, uitgesplitst naar toetsversie*

Aantal leerlingen	Versie 1	Versie 2	Versie 3	Totaal
Totaal	209	651	554	1414
Volledig ingevuld	196	600	516	1312
1 Ontbrekend antwoord	8	24	31	63
2 Ontbrekende antwoorden	1	7	3	11
3 tot 5 Ontbrekende antwoorden	3	3	1	7
6 tot 10 Ontbrekende antwoorden	0	6	0	6
Meer dan 10 ontbrekende antwoorden	1	11	3	15

4.3 Toets- en itemanalyse

Na het in kaart brengen van de missende waarden zijn de leerlingscores vergeleken met de scores die op basis van gokgedrag behaald hadden kunnen worden. Op vier leerlingen na presteerden alle leerlingen boven kansniveau. Deze vier leerlingen zijn buiten de analyses gelaten, omdat zij de toets mogelijk niet serieus gemaakt hebben. Op basis van de data van de overige leerlingen is een klassieke toets- en itemanalyse verricht voor elk van de toetsversies. De uitkomsten worden samengevat in Tabel 4.3. Vanwege het relatief grote aantal items wordt in Tabel 4.3 alleen samenvattende informatie gepresenteerd en niet de kenmerken van alle afzonderlijke items. In Tabel 4.3 is onder meer te zien dat de toewijzing van de toetsversies het beoogde effect heeft gehad, want de drie toetsversies waren ongeveer even moeilijk voor de leerlinggroepen die ze maakten: de gemiddelde proportionele score lag telkens dichtbij 0.680. Ook de laagste en hoogste *p*-waarden van individuele items liepen niet ver uiteen tussen de toetsversies. De moeilijkste items hadden een *p*-waarde van ongeveer 0.330 (de gokkans bij deze items was 0.25) en de makkelijkste items een *p*-waarde van 0.900 of hoger. Er waren geen items met een *p*-waarde onder gokkans, of items die door alle leerlingen correct werden beantwoord.

Tabel 4.3. *Samenvatting resultaten toets- en itemanalyse bij Luisteren*

Kenmerk	Versie 1	Versie 2	Versie 3
Aantal items	28	24	20
Gemiddelde prop. toetsscore	0.690	0.680	0.670
Laagste *p*-waarde items	0.360	0.310	0.360
Gemiddelde *p*-waarde items	0.690	0.690	0.670
Hoogste *p*-waarde items	0.910	0.995	0.990
Laagste r_{it}-waarde items	0.200	0.080	0.100
Gemiddelde r_{it}-waarde items	0.370	0.310	0.360
Hoogste r_{it}-waarde items	0.480	0.510	0.590
Cronbach's alpha	0.760	0.640	0.670

In Tabel 4.3 is verder te zien dat de item-totaalcorrelaties (r_{it}-waarde) in elk van de toetsversies gemiddeld rond de 0.350 lag. Er waren desalniettemin wel wat verschillen tussen de versies wat betreft de laagste en hoogste r_{it}-waarde. Enkele items uit het volgsysteem voor het VSO-Dagbesteding hadden een erg lage r_{it}-waarde (< 0.200). Deze items waren erg gemakkelijk en omdat zoveel leerlingen ze goed maakten, is het logisch dat de samenhang van de itemscore met de totaalscore bij deze items laag uitviel. In zo'n geval betekent de lage r_{it}-waarde niet dat er iets mis is met het item, maar alleen dat het weinig toevoegt aan de meetnauwkeurigheid bij de doelgroep. De interne consistentie, ten slotte, was het hoogste voor toetsversie 1, met Cronbach's alpha = 0.760, en het laagste voor toetsversie 2, met Cronbach's alpha = 0.640. De interne consistentie kan in dit geval echter beter niet worden geïnterpreteerd als een schatting van de toetsbetrouwbaarheid, omdat Cronbach's alpha bij niveau-gestratificeerde toetsen lager uit kan vallen naarmate de verschillende leerlinggroepen homogener zijn. In paragraaf 4.5 wordt de betrouwbaarheid onderzocht op een manier die wel rekening houdt met de toewijzing van de drie toetsversies.

4.4 Item respons theorie analyse

Om de toetsversies op één onderliggende meetschaal te plaatsen, zijn analyses verricht binnen het raamwerk van de item respons theorie. Item respons modellen verschillen in een aantal opzichten nogal sterk van de modellen uit de klassieke testtheorie (Verhelst, 1993). Binnen de item respons theorie staat de latente vaardigheid centraal in plaats van de ware score op een specifieke toets. Er wordt verondersteld dat de vaardigheid θ van een persoon niet begrensd is en dat de antwoorden op bepaalde items informatie geven over de vaardigheid van de persoon. De relatie tussen de latente vaardigheid en de itemantwoorden is gedefinieerd in een zogenaamde itemkarakteristieke functie. De itemkarakteristieke functie drukt de kans op het correct beantwoorden van item j uit als functie van θ. De kenmerken en het verloop van de itemkarakteristieke functie zijn gespecificeerd in een aantal randvoorwaarden: (a) de kans op een correct antwoord moet tussen 0 en 1 liggen, (b) de functie moet continu zijn, en (c) de functie moet strikt stijgend zijn binnen θ. Hoewel een groot aantal wiskundige functies uitgesloten wordt door deze drie vereisten, blijven er nog veel functies over die aan de gestelde eisen voldoen. Door één specifieke functie te kiezen, kan de theorie ingeperkt worden tot één speciaal 'geval', een zogenaamd item respons model.

De antwoorden van de leerlingen op de items bij het onderdeel Luisteren zijn geanalyseerd met het One-Parameter Logistic Model (OPLM) van Verhelst en Glas (1995). De item respons functie voor het OPLM wordt gegeven door de volgende vergelijking:

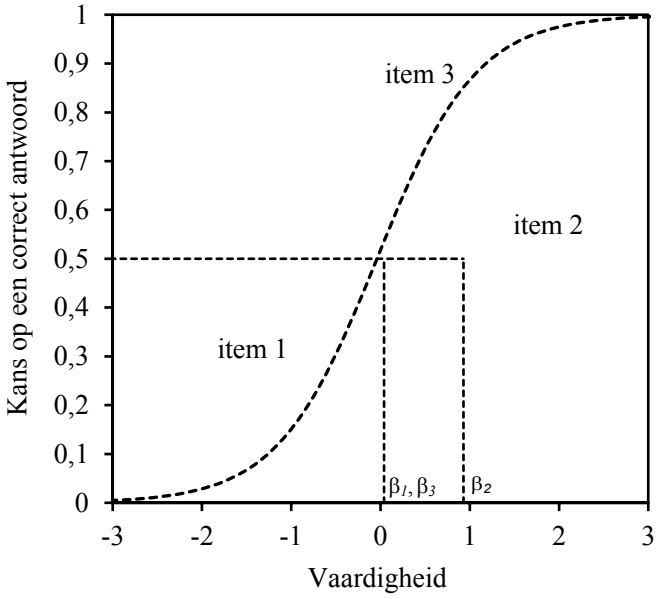
$$\mathrm{P}(X_j = 1 \mid \theta) = P_j(\theta) = \frac{\exp[a_j(\theta - \beta_j)]}{1 + \exp[a_j(\theta - \beta_j)]}, \qquad \text{voor } j = 1, \ldots, k$$

waarin $P_j(\theta)$ de kans is dat een random gekozen persoon met va~~ardigheid θ~~ item j correct beantwoordt, a_j en β_j getallen zijn die item j karakteriseren, k het aantal items in de toets is, en waar 'exp' een constante is met waarde 2.718. Er kan nagegaan worden dat functie $P_j(\theta)$ altijd tussen 0 en 1 ligt en dat $P_j(\theta)$ gelijk is aan 0.50 als de vaardigheid gelijk is aan β_j. De waarde voor β_j kan daarom ook geïnterpreteerd worden als de vaardigheid die vereist is om een kans van 50 procent te hebben op het correct maken van een item. Figuur 4.2 geeft drie verschillende item respons curven voor het OPLM. Zoals we kunnen zien, is voor item 2 meer vaardigheid nodig om die kans van 50 procent te bereiken dan voor item 1 en 3. In feite reflecteert parameter β_j dus de moeilijkheid voor item j of, meer formeel, de positie van de itemkarakteristieke curve in relatie tot de vaardigheidsschaal. Om deze reden wordt β_j in de literatuur vaak aangeduid als de moeilijkheids- of locatieparameter van item j. Figuur 4.2 laat verder zien de item respons curven voor item 1 en item 2 qua vorm hetzelfde zijn en dat de vorm van item 3 afwijkt. De afwijkende vorm van de curve voor item 3 wordt veroorzaakt door een andere waarde voor a_j. Er kan afgeleid worden dat item 3 in een bepaald vaardigheidsgebied beter discrimineert tussen personen, omdat de kans op het correct maken van het item snel toeneemt naarmate de vaardigheid hoger is. Bij item 1 en 2 neemt de kans op het correct maken van het item minder snel toe als de vaardigheid stijgt. Item 1 en 2 discrimineren in dat vaardigheidsgebied dus minder goed tussen personen.

Kenmerkend voor het OPLM is dat a_j niet uit de data geschat wordt, maar a priori als constante wordt ingevoerd. Door deze keuze kan de (gewogen) ruwe score direct uit de data berekend worden en behoort toepassing van de *conditional maximum likelihood* methode voor het schatten van de itemparameters tot de mogelijkheden. Een belangrijk voordeel van *conditional maximum likelihood* is dat er bij het schatten van de itemparameters geen veronderstelling hoeft te worden gedaan over de verdeling van de vaardigheid in de populatie. Bovendien doet het er niet toe hoe de steekproef getrokken is of hoe de modules in een incompleet design zijn samengesteld. Zelfs als een leerling meerdere keren meedoet aan het onderzoek of bepaalde modules systematisch bij bepaalde groepen leerlingen worden afgenomen, kunnen de

itemparameters adequaat geschat worden (Eggen, 2004). De schatting van de itemparameters vindt plaats met behulp van het computerprogramma OPLM (Verhelst, Glas & Verstralen, 1995). Dit programma voert tevens een aantal statistische toetsen uit op grond waarvan bepaald kan worden of het model een adequate beschrijving geeft van de data. Belangrijk zijn de zogenaamde itemgeoriënteerde *S*-toets en de overall *R1c*-toets. De *S*-toets is asymptotisch χ^2 verdeeld en is gebaseerd op de verschillen tussen de geobserveerde en verwachte proporties antwoorden in homogene scoregroepen. Een uniforme verdeling van *p*-waarden voor de *S*-toetsen in het interval [0,1] pleit voor passing van het model (zie Verhelst, Glas & Verstralen, 1995). De *R1c*-toets heeft dezelfde onderliggende rationale als de *S*-toets en soms wordt gesteld dat de modelpassing acceptabel is als de waarde niet groter is dan anderhalf keer het aantal vrijheidsgraden, maar het is belangrijk om zeer voorzichtig met deze vuistregel om te gaan.

Figuur 4.2. *Drie itemresponscurven voor het OPLM*



De item respons theorie analyse heeft in verschillende stappen plaatsgevonden. Voorafgaand aan de analyse is (ter controle) de dimensionaliteit van de drie luistertoetsen onderzocht aan de hand van een *Principal Axis* exploratieve factoranalyse op de matrix met tetrachorische correlaties tussen de itemscores binnen een toetsversie. De correlatiematrices lieten geen indicatie zien voor een tweede factor (of dimensie). Dit werd bevestigd door de factoranalyse. In de screeplots was telkens een duidelijke knik te zien na de eerste factor. Er is dan ook geconcludeerd dat de assumptie van unidimensionaliteit verdedigbaar is en het OPLM (in principe) gebruikt kan worden om de drie toetsversies op één meetschaal af te beelden. Als het OPLM een goede passing laat zien, wijst dat overigens ook op unidimensionaliteit.

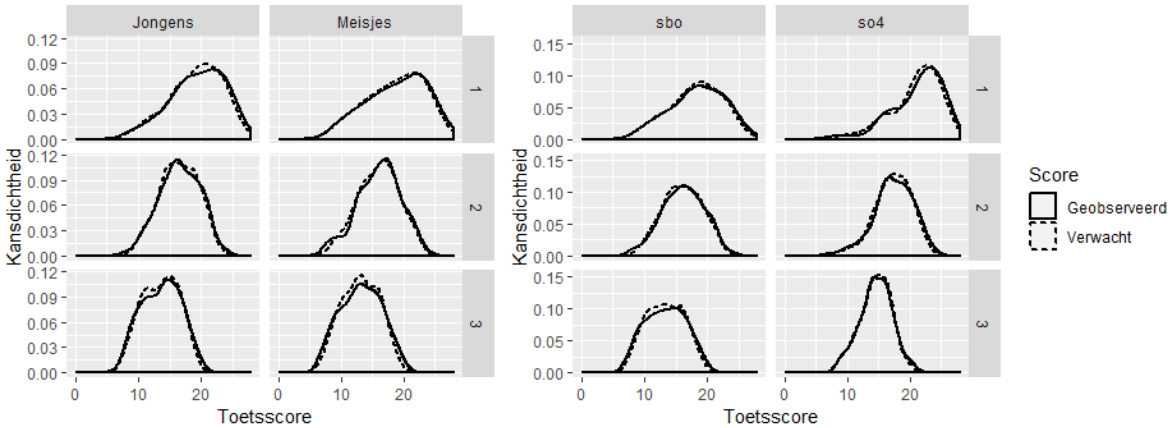
In een vervolgstap is nagegaan of het “eenvoudiger” Rasch model, dat wil zeggen een model waarin alle discriminatie-indices gelijk zijn, een goede beschrijving van de data kon geven, maar dat model paste niet goed, afgaande op onder andere de R1c-test voor globale modelpassing: $R1c = 531.8$ $df = 225$, $p < 0.01$. Voor een adequate beschrijving van de data moest het model dus rekening houden met verschillen in het discriminerend vermogen van de items. Er is nagegaan of een model waarbij de itemgewichten (oftewel de discriminatie-indices) een harmonisch gemiddelde van 2 hebben de data wel in voldoende mate zou beschrijven. Dit was het geval. Ten eerste bleek de verdeling van *p*-waarden (overschrijdingskansen) voor de *S*-toetsen redelijk uniform verdeeld te zijn. Ten tweede was de verhouding tussen de *R1c*-bijdrage en het aantal vrijheidsgraden acceptabel, $R1c = 290.1$, $df = 225$, $p < 0.01$. In dit licht is dan ook besloten om geen andere modellen te schatten, maar met dit model door te gaan en, in de derde en laatste stap, expliciet na te gaan of het model toereikend is voor verschillende groepen leerlingen.

Het computerprogramma OPLM biedt goede mogelijkheden om *Differential Item Functioning* (DIF) te onderzoeken. Er is gekeken of er sprake was van DIF tussen jongens en meisjes, tussen sbo-leerlingen en so-leerlingen, of tussen de niveaugroepen die de drie toetsversies maakten. Eén item leek op het oog iets moeilijker te zijn voor de leerlingen die toetsversie 2 maakten dan voor de leerlingen die toetsversie 1 maakten, maar de itemresponscurve voor de laatste groep was wel aanzienlijk grilliger omdat het een kleinere groep leerlingen betrof. Verder waren er twee items waarbij er sprake leek te zijn van DIF naar schooltype, maar de interpretatie van de statistische toetsen en de itemresponscurves werd ook hier bemoeilijkt door de ongelijke groepsgrootte; naar verhouding namen veel minder so-leerlingen deel aan het

onderzoek dan sbo-leerlingen. Dit gold ook voor het onderzoek naar DIF tussen jongens en meisjes, omdat de groep meisjes veel kleiner was. Daarop is besloten om de robuustheid van het OPLM in kaart te brengen.

De robuustheid van het OPLM is in kaart gebracht door te controleren hoe goed de geobserveerde scoreverdelingen van verschillende leerlinggroepen teruggeschat konden worden met behulp van het OPLM. We vergeleken dus de leerlingscores die we op grond van de modelparameters zouden verwachten met de geobserveerde waarden. Onder de verwachte leerlingscore verstaan we hier de ongewogen (en onafgeronde) IRT *true score* gebaseerd op de OPLM itemparameters (zie bijvoorbeeld Hambleton, Swaminathan & Rogers, 1991). In Figuur 4.3 zijn de geobserveerde en de verwachte (model-voorspelde) verdelingen van toetsscores geïllustreerd per toetsversie en leerlinggroep. Zoals we kunnen zien, leken de voorspelde scoreverdelingen sterk op de geobserveerde verdelingen. Al met al kan dan ook geconcludeerd worden dat als er al sprake was van DIF in één of enkele items, dit een verwaarloosbaar effect heeft, omdat het geschatte OPLM de gegevens van de verschillende leerlinggroepen voldoende adequaat (of voldoende nauwkeurig) beschrijft. Dit model is dan ook gebruikt als basis voor de uitkomstvariabelen (de prestatie op de luistertoets, en het behaalde referentieniveau) in alle (inhoudelijke) vervolganalyses.

Figuur 4.3. *Geobserveerde en verwachte scoreverdelingen per leerlinggroep en toetsversies*



4.5 Meetnauwkeurigheid

Binnen de item respons theorie wordt gebruikgemaakt van het statistische begrip informatie om de meetnauwkeurigheid van een toets lokaal uit te drukken (zie bijvoorbeeld Lindgren, 1976). De iteminformatiefunctie geeft aan welke bijdrage een item levert aan de meetnauwkeurigheid voor ieder punt op de vaardigheidsschaal waar het item betrekking op heeft. De iteminformatiefunctie van een dichotoom item bereikt zijn maximum wanneer $\theta = \beta_j$ en wordt kleiner naarmate het verschil tussen θ en β_j toeneemt. Met behulp van de informatiefuncties van de afzonderlijke items in een willekeurige toets t kan een zogenaamde toetsinformatiefunctie bepaald worden. Voor de toetsinformatiefunctie $I_t(\theta)$ geldt:

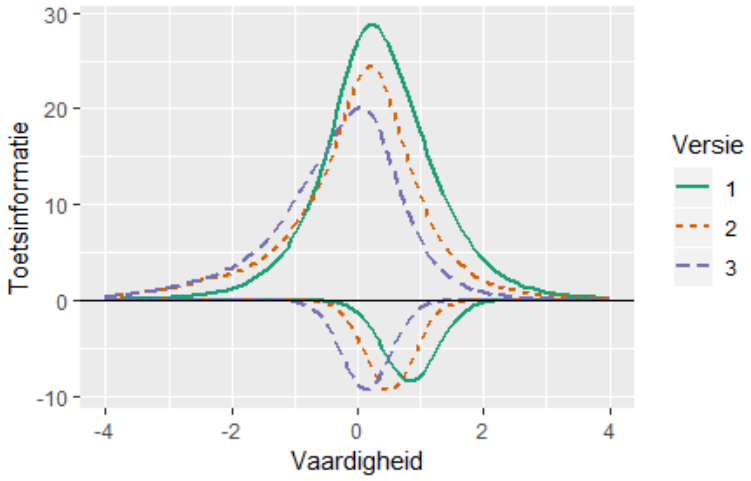
$$I_t(\theta) = \sum_{j=1}^{k_t} I_j(\theta)$$

De waarde van $I_t(\theta)$ hangt in de regel samen met het aantal items in de toets, de kwaliteit van de items, en de mate van overeenstemming tussen de moeilijkheidsgraad van de items en de vaardigheid van de leerling. Uit de toetsinformatiefunctie kan de standaardmeetfout van een individuele vaardigheidsschatting afgeleid worden:

$$SE_t(\hat{\theta}) = 1 \big/ \sqrt{I_t(\theta)}$$

De interpretatie van deze standaardmeetfout is hetzelfde als in de klassieke testtheorie. De toetsinformatiefunctie is per toetsversie weergegeven in Figuur 4.4. We zien dat de toetsversies in de beoogde richting verschilden in hoe nauwkeurig zij meten bij oplopende vaardigheidsniveaus. Onder de y-as is de geschatte populatieverdeling van de latente vaardigheid afgebeeld voor de leerlinggroepen die de drie toetsversies hebben gemaakt. De verschillende toetsversies sloten vrij goed aan bij de vaardigheid van de leerlinggroepen die ze gemaakt hebben, zoals we ook concludeerden op basis van de gemiddelde *p*-waarden van de toetsen. Psychometrisch gezien hadden de toetsversies iets moeilijker mogen zijn, maar

het werd in dit onderzoek belangrijk geacht om de leerlingen een gevoel van succes te geven; daarnaast maakt het voor onderzoek op systeemniveau (zoals dit onderzoek) ook niet veel uit als de vaardigheid van individuele leerlingen iets minder nauwkeurig gemeten wordt.



Figuur 4.4. *Toetsinformatiefunctie (en vaardigheidsverdeling onder de y-as) per toetsversie*

Binnen het raamwerk van de item respons theorie is het enigszins problematisch om de meetnauwkeurigheid te interpreteren, omdat er geen ondubbelzinnige criteria beschikbaar zijn om de toetsinformatiefunctie te beoordelen. De definitie uit de klassieke testtheorie van meetnauwkeurigheid is gemakkelijker te interpreteren, maar niet direct vergelijkbaar met de lokale maat voor meetnauwkeurigheid uit de item respons theorie. Wanneer we de variatie in standaardmeetfouten negeren en de variantie van de vaardigheidsverdeling vastzetten, wordt het echter mogelijk om de item respons theorie maat voor meetnauwkeurigheid te relateren aan de klassieke maat voor meetnauwkeurigheid. Verhelst, Glas, en Verstralen (1995) hebben in dit kader de *MAcc* coëfficiënt ontwikkeld. De *Macc* coëfficiënt kan opgevat worden als een generalisatie van coefficient alpha voor alle schattingen van θ.

Omdat er in dit onderzoek gebruikgemaakt is van het principe van niveau-gestratificeerd toetsen is de *MAcc* coëfficiënt eerst uitgerekend in de totale populatie voor alle items (dat wil zeggen, er is net gedaan alsof alle items in de itembank zijn afgenomen bij alle leerlingen). Vervolgens hebben we met behulp van de Spearman-Brown-formule voor testverlenging (of hier –verkorting) (Brown, 1910; Spearman, 1910) een schatting gemaakt van de betrouwbaarheid in het geval dat 28, 24, of 20 items zouden zijn afgenomen (de lengte van toetsversies 1, 2 en 3). De *MAcc* coëfficiënt in de totale populatie voor alle 52 geschaalde items bleek gelijk te zijn aan 0.860. De betrouwbaarheid van de drie toetsversies werd dan respectievelijk geschat op 0.770, 0.740 en 0.710. Volgens richtlijnen van Evers et al. (2010) is een betrouwbaarheid van 0.600 voldoende en één van 0.700 goed als het doel is om leerlingen te vergelijken op groepsniveau, zoals het geval is bij systeemevaluatie en terugrapportage op schoolniveau. We kunnen dus concluderen dat het gebruikte meetinstrument voor Luisteren een goede betrouwbaarheid heeft.

4.6 Variabelen

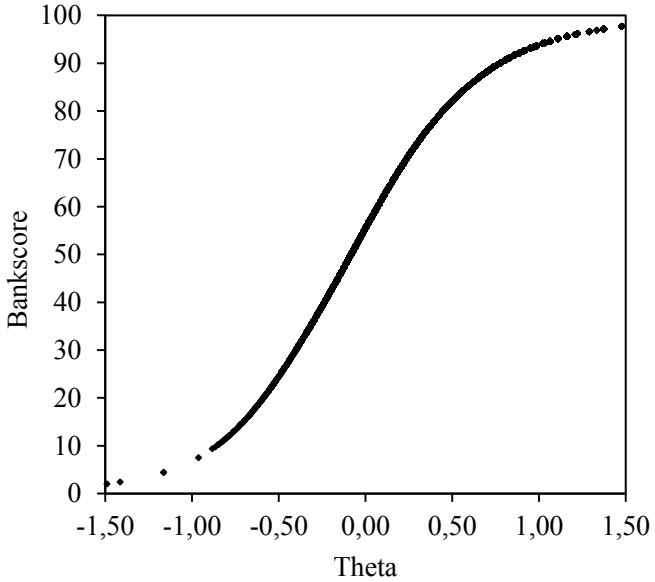
Als het OPLM geldt voor een itemverzameling kan de vaardigheid van leerlingen met elke willekeurige deelverzameling op dezelfde meetschaal geschat worden. De (gewogen) scores van de leerlingen op de drie toetsversies kunnen dus vertaald worden naar zogenaamde vaardigheidsscores die zonder problemen met elkaar vergeleken kunnen worden, ook als verschillende toetsen zijn afgenomen. Tevens is het mogelijk om op basis van de vaardigheidsscore van een leerling een zogenaamde bankscore te bepalen (zie bijvoorbeeld Hambleton, Swaminathan, & Rogers, 1991). Een belangrijk voordeel van bankscores is dat de scores geïnterpreteerd kunnen worden als een beheersingniveau. De bankscores liggen namelijk tussen 0 en 100 procent. Bij een bankscore van 70 kun je dus zeggen dat de leerling naar verwachting 70 procent van de items in de bank correct kan maken.

In het databestand staan ruwe scores, gewogen scores, vaardigheidsscores en bankscores (als percentage). De ruwe en gewogen scores zijn berekend op basis van de items in de toetsversie die de leerling gemaakt heeft. Dit betekent dat de maximale ruwe score voor versie 1 gelijk is aan 28, voor versie 2 gelijk aan 24 en voor versie 3 gelijk aan 20. De maximale gewogen scores zijn respectievelijk gelijk aan 58,

49 en 43. De bankscores zijn niet-lineaire transformaties van de vaardigheidsscores en kunnen worden uitgerekend via de itemparameters die verkregen zijn na schatting van het OPLM:

$$\pi(\theta) = \frac{1}{\sum_{j=1}^{k} \alpha_j} \sum_{j=1}^{k} \alpha_j p_j(\theta)$$

Figuur 4.5 geeft een visuele weergave van de relatie tussen de vaardigheidsscore θ en de bankscore π. We zien dat relatie tussen de vaardigheidsscores en de bankscores in het midden nagenoeg lineair is en aan de uiteinden van de meetschaal vlakker. Dit komt doordat leerlingen met een hoge vaardigheidsscore vanaf een bepaald vaardigheidsniveau (vrijwel) alle items correct zullen maken en leerlingen met een lage vaardigheidsscore (vrijwel) alle items fout. Bij het doen van analyses op systeemniveau heeft het de voorkeur om de bankscores uit het databestand te gebruiken, omdat het gebruik van vaardigheidsscores kan leiden tot een (sterke) overschatting van de populatievariantie. Bij het gebruik van bankscores speelt dit probleem niet.



Figuur 4.5. *Visuele weergave van de relatie tussen vaardigheidsscores en bankscores*

Samenvattend zijn in het databestand bij Luisteren de volgende variabelen opgenomen:

LUI_VERS	Toetsversie die de leerling heeft gemaakt
LUI_MISS	Percentage ontbrekende antwoorden
LUI_RSCO	Ruwe score op de gemaakte toetsversie
LUI_WSCO	Gewogen score op de gemaakte toetsversie
LUI_THET	Weighted Maximum Likelihood vaardigheidsscore
LUI_BANK	Verwachte bankscore
LUI_REFN	Referentieniveau (<1F/1F/2F)
LUI_WEIGHT	Gewicht om te corrigeren voor leerroute

Het behaalde referentieniveau is gebaseerd op een koppeling met de Ankersets. In Hoofdstuk 8 wordt deze koppeling uitgebreid beschreven en onderbouwd.

5 Spreken

5.1 Inhoud en afname

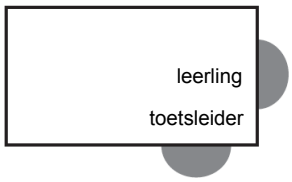
In aansluiting op het referentiekader Taal is een instrument ontwikkeld om de spreekvaardigheid van leerlingen te meten. Het instrument *Spreken* richtte zich op de volgende vijf kenmerken van de taakuitvoering.

Kenmerk	Uitwerking referentieniveau 1F
1. Samenhang	Maakt eigen gedachtegang voor de luisteraar begrijpelijk, hoewel de structuur van de tekst nog niet altijd klopt.
2. Afstemming op doel	Blijft trouw aan spreekdoel, soms met hulp van een ander.
3. Afstemming op publiek	Past het taalgebruik aan de luisteraar(s) aan. Kan gebruik maken van ondersteunende materialen om een voorbereide presentatie beter aan het publiek over te brengen.
4. Woordgebruik en woordenschat	Beschikt over voldoende woorden om te praten over vertrouwde situaties en onderwerpen, maar zoekt nog regelmatig naar woorden en varieert niet veel in woordgebruik.
5. Vloeiendheid, verstaanbaarheid en grammaticale beheersing	De uitspraak is duidelijk genoeg om de spreker te kunnen volgen, ondanks een eventueel accent, verkeerde intonatie, onduidelijke articulatie en/of haperingen. Redelijk accuraat gebruik van eenvoudige zinsconstructies. Houding, intonatie en mimiek ondersteunen het gesprokene. Pauzes, valse starts en herformuleringen komen af en toe voor.

Er zijn leerlingen die gemakkelijker spreken wanneer zij een feitelijk verhaal moeten (na)vertellen en leerlingen die beter tot hun recht komen wanneer zij meer fantasie moeten gebruiken. Om beide soorten leerlingen tegemoet te komen en tegelijkertijd bovengenoemde kenmerken te kunnen meten, is gekozen voor twee soorten taken (zie ook Hoofdstuk 2):

1. ***Navertellen van een nieuwsfragment uit het jeugdjournaal.*** De leerling ziet een kort fragment uit het jeugdjournaal en krijgt daarna de vraag om na te vertellen wat hij gezien en gehoord heeft. Deze opdracht is vergelijkbaar met bijvoorbeeld de spreektaken uit PPON, waarbij de leerling een verhaal moet lezen en navertellen. Het geschreven verhaal is in dit onderzoek echter vervangen door een filmfragment om zwakke lezers niet te benadelen. Om ook leerlingen met een zwakker werkgeheugen niet te benadelen, kregen alle leerlingen bij het navertellen een geheugensteuntje bestaande uit kernwoorden uit het filmfragment (zie Bijlage 7).
2. ***Vertellen van een verhaal aan de hand van plaatjes op dobbelstenen van Rory's Story Cubes***. De leerling krijgt negen dobbelstenen met daarop verschillende afbeeldingen. De leerling gooit de dobbelstenen en vertelt wat zij op de dobbelstenen ziet. Vervolgens kiest de leerling vijf dobbelstenen uit om een verhaal over te vertellen. Het verhaal moet beginnen met "Er was eens….". De leerling mocht het spel meerdere keren spelen. De eerste keer was om te oefenen en de tweede en derde keer werden gebruikt voor de beoordeling.

De afname van de twee spreektaken vond plaats in een aparte, zoveel mogelijk prikkelarme, ruimte. Elke afname werd begeleid door een getrainde toetsleider. De taken werden in de volgende opstelling afgenomen:



Op deze manier kon de toetsleider goed contact maken met de leerlingen. Om ervoor te zorgen dat de twee taken door álle leerlingen uitgevoerd konden worden, mocht de toetsleider, indien strikt noodzakelijk, kleine aanpassingen doorvoeren in de afname. In Bijlage 5 is te lezen welke aanpassingen toegestaan waren. De afname van de twee taken nam ongeveer 15 minuten per leerling in beslag.

Naast afnameleider vervulde de toetsleider tijdens de afname ook de rol van beoordelaar. Daarbij is uitgegaan van een model waarin de toetsleider voor vijftien verschillende aspecten, verdeeld over de verschillende kenmerken van de taakuitvoering, op een driepuntschaal (0 = onvoldoende, 1 = voldoende, 2 = goed) aangaf in hoeverre de leerling dit aspect liet zien. Concreet ging het om onderstaande vijftien aspecten. Ter illustratie is bij alle vijftien beoordelingsaspecten de omschrijving van de hoogste beoordelingscategorie weergegeven.

Uitvoering

1. Volledigheid
   - De leerling heeft (bijna) alle taalhandelingen van de opdracht uitgevoerd.
2. Inhoudelijke kwaliteit
   - De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten sterk.

Samenhang

3. Samenhang op macro- en mesoniveau: opbouw
   - De leerling laat een logische opbouw zien, bijvoorbeeld met zowel verbinding binnen als verbinding tussen inleiding, kern en afsluiting.
4. Samenhang op microniveau: signaalwoorden
   - De leerling gebruikt eenvoudige voeg- en verbindingswoorden correct.

Afstemming op doel

5. Spreekdoel
   - Filmfragment: De leerling kan de strekking van het filmpje duidelijk maken aan de gesprekspartner (die het filmpje niet gezien en gehoord heeft).
   - Dobbelspel: De leerling vertelt een verhaal waar hij alle plaatjes in verwerkt.

Afstemming op de gesprekspartner

6. Taalgebruik afstemmen
   - De leerling gebruikt consequent taal die bij de gesprekspartner past.
7. Gesprekspartner boeien
   - De leerling zet regelmatig aspecten als stiltes, wisselende intonatie/volume/spreeksnelheid en lichaamstaal bewust in om de gesprekspartner te boeien.
8. Non-verbale communicatie
   - Het non-verbale gedrag van de leerling zet zijn verbale boodschap kracht bij.

Woordgebruik en woordenschat

9. Woordgebruik
   - De uitwerking van de leerling is aantrekkelijk door voldoende variatie in woordgebruik.
10. Woordenschat
    - De getoonde woordenschat van de leerling is toereikend voor het uitvoeren van de opdracht.

Vloeiendheid, verstaanbaarheid en grammaticale beheersing

11. Vloeiendheid
    - De leerling spreekt redelijk vloeiend, al vertoont zijn uitwerking een enkele hapering en/of herformulering.
12. Uitspraak en intonatie
    - De leerling is goed verstaanbaar, ondanks een eventueel accent.
13. Grammatica: vervoeging
    - De leerling vervoegt alle werkwoorden over het algemeen correct.
14. Grammatica: verbuiging
    - De leerling verbuigt (combinaties met) naamwoorden meestal correct.
15. Grammatica: zinsbouw
    - De leerling maakt over het algemeen correcte enkelvoudige en samengestelde zinnen.

Voor valideringsdoeleinden werd er bovendien om een globaal oordeel gevraagd: *Hoe zou je de spreekvaardigheid van deze leerling globaal gezien typeren in termen van referentieniveau 1F? Matig-Voldoende-Goed*. Dit oordeel wordt niet meegenomen in de score van een individuele leerling. Het beoordelingsmodel is per taak ingevuld en alleen voor die leerlingen die de taak zelfstandig en volledig hebben uitgevoerd

De toetsleider vulde het beoordelingsmodel voor elke leerling zoveel mogelijk tijdens het spel in. Om ervoor te zorgen dat toetsleiders de beoordelingstaak betrouwbaar en hetzelfde konden uitvoeren zijn de volgende vier maatregelen genomen:

1. Elk aspect is uitgewerkt met concrete voorbeelden passend bij de spreektaken.
2. Alle toetsleiders hebben uitgebreid geoefend met het beoordelingsmodel tijdens een training op basis van filmfragmenten.
3. De afname werd opgenomen op video, zodat de toetsleider waar nodig achteraf de mogelijkheid had om de beoordeling te completeren of controleren.
4. Via een app-groep konden toetsleiders elkaar dilemma's rondom de beoordeling voorleggen.

In een flankerend onderzoek is empirisch nagegaan in hoeverre toetsleiders de beoordelingstaak adequaat hebben uitgevoerd. Hierover wordt in paragraaf 5.4 gerapporteerd.

5.2 Missende waarden

Bij het onderdeel *Spreken* is er sprake van twee taken die door alle leerlingen werden gemaakt. Dit betekent dat er geen sprake is van toetsversies die via item respons theorie modellering of een andere equivaleringstechniek vergelijkbaar gemaakt moeten worden. Wel kon er sprake zijn van missende waarden, bijvoorbeeld doordat de toetsleider:

- bij een taak één of meer beoordelingsaspecten "vergeten" was in te vullen;
- bij één of beide taken "niet te beoordelen" had ingevuld.

Bij de training van de beoordelaars is benadrukt dat alle taken volledig beoordeeld moesten worden en dat "niet te beoordelen" slechts in zeer uitzonderlijke gevallen mocht worden gekozen. Verkennende analyses lieten zien dat bij het jeugdjournaalfragment bij vier van de 1442 leerlingen "niet te beoordelen" is ingevuld en bij het dobbelspel bij elf van de 1442 leerlingen. Als we deze leerlingen buiten beschouwing laten, zien we dat het percentage missende waarde bij het jeugdjournaalfragment varieert van 0.28 tot 0.84 procent (*M* = 0.50 procent) en bij het dobbelspel van 0.91 tot 3.22 procent (*M* = 1.75 procent). Het is de vraag hoe het beste omgegaan kan worden met de missende waarden. Imputatie ligt voor de hand, maar deze oplossing is niet in alle gevallen goed te verdedigen.

- In het geval dat een toetsleider een beoordelingsaspect "vergeten" is in te vullen lijkt de aanname *missing at random* (zie Little & Rubin, 1987) verdedigbaar en is imputatie mogelijk, tenzij er zeer veel missende waarden zijn of in geval zij gerelateerd zijn aan specifieke toetsleiders. De data geven geen aanleiding om te veronderstellen dat er een relatie is tussen de toetsleider en het aantal missende waarden. Bij bijna alle 27 toetsleiders bedraagt het percentage missende waarden hooguit enkele procenten, zowel bij het jeugdjournaalfragment als bij het dobbelspel. Alleen bij toetsleider TL0017 is het percentage missende waarden met 11.11 procent bij het jeugdjournaalfragment en 12.35 procent bij het dobbelspel aan de hoge kant. Omdat het in absolute zin slechts om een beperkt aantal leerlingen gaat (27), is er niet direct reden om de toetsleider in de verdere analyses weg te laten. In plaats daarvan is ervoor gekozen om de "vergeten" beoordelingsaspecten te imputeren, zowel bij toetsleider TL0017 als bij de overige toetsleiders.
- Als een toetsleider "niet te beoordelen" heeft ingevuld is het lastiger om te bepalen wat een adequate aanpak is. *Missing at random* gaat dan niet op, waardoor imputatie lastig te verdedigen is. Omdat onduidelijk is waarom de leerling de taak volgens de toetsleider niet kon uitvoeren (dat werd niet genoteerd) zijn de betreffende leerlingen (14 in totaal) buiten alle analyses gelaten.

Tot nu toe is alleen op het niveau van de items (of beoordelingsaspecten) en de toetsleiders naar missende waarden gekeken. Het kan echter zo zijn dat er ook leerlingen zijn met extreem veel missende waarden. De verdeling ziet er als volgt uit wanneer beide taken worden samengenomen:

N aspecten beoordeeld	1	14	15	16	17	21	22	23	24	25	26	27	28	29	30
N leerlingen	1	1	3	1	1	2	1	1	2	6	8	11	47	133	1210

We zien dat verreweg de meeste leerlingen (1210) op alle aspecten beoordeeld zijn. Het is onduidelijk vanaf welk punt er aanleiding is om te twijfelen over de kwaliteit van een beoordeling. In dit onderzoek leggen we de afkapwaarde, net zoals eerder bij Luisteren, bij 80 procent. Dat betekent dat de leerlingen die op 23 of minder aspecten beoordeeld zijn (11 in totaal) niet worden meegenomen in de analyse. Lagere afkapwaarden waren vanzelfsprekend ook mogelijk geweest en soms is dat verdedigbaar (zie bijvoorbeeld Eekhout, 2018), maar omdat het in dit onderzoek ook bij een hoge afkapwaarde van 80 procent om slechts een zeer klein aantal leerlingen gaat, is daar niet voor gekozen. Bij de resterende leerlingen zijn de ontbrekende waarden geïmputeerd. Deze keuze wijkt af van de keuze die bij Luisteren is gemaakt (missing = 0), omdat er bij Spreken polytoom gescoord wordt en niet dichotoom zoals bij Luisteren het geval was. Bovengenoemde keuzes hebben concreet tot gevolg gehad dat van de oorspronkelijke groep met 1442 leerlingen er 1417 in de analyse worden meegenomen (1442 – 14 – 11) en dat in die dataset in totaal 334 waarden zijn geïmputeerd (0.786 procent).

In de literatuur zijn verschillende imputatietechnieken beschreven. Omdat het aantal missende waarden zeer klein is (0.786 procent), hebben we ervoor gekozen om gebruik te maken van een eenvoudige techniek die uitgaat van het idee dat de beste schatter voor een missende waarde in een reeks getallen het gemiddelde is van de wél geobserveerde getallen. We kiezen er dus voor in een lege cel een gemiddelde te imputeren. De vraag is alleen welk gemiddelde? Een lege cel in een datamatrix kan immers gezien worden als een ontbrekend getal in:

- de rij (dan zouden we het rijgemiddelde kunnen gebruiken),
- de kolom (dan zouden we het kolomgemiddelde kunnen gebruiken),
- de totale reeks cellen in de hele matrix (dan zouden we het totaalgemiddelde kunnen gebruiken).

Hier is ervoor gekozen om het gemiddelde van de geobserveerde waarden van de leerling, van het beoordelingsaspect en van de totale dataset te wegen naar de proportie van hun geobserveerde variantie in de dataset, en daarmee dus alle drie genoemde gemiddelden te betrekken in de schatting. Concreet zijn missende waarden Y dus geschat met:

Y = b1 x RijGem + b2 x KolGem + b3 x TotGem

waarin:

b1	=	proportioneel gewicht variantiecomponent leerlingen
b2	=	proportioneel gewicht variantiecomponent beoordelingsaspecten
b3	=	proportioneel gewicht variantiecomponent interactie/residu/error
RijGem	=	som van beschikbare waarnemingen in een rij gedeeld door hun aantal
KolGem	=	som van beschikbare waarnemingen in een kolom gedeeld door hun aantal
TotGem	=	som van alle beschikbare waarnemingen gedeeld door hun aantal

De imputatie is voor het jeugdjournaalfragment en het dobbelspel apart uitgevoerd. Andere imputatietechnieken hadden vanzelfsprekend ook gebruikt kunnen worden, maar vanwege het zeer kleine aantal missende waarden kan de techniek feitelijk geen invloed hebben op de uitkomsten. Geavanceerdere technieken bieden in dat geval geen voordelen boven de eenvoudigere techniek die in dit onderzoek gebruikt is.

5.3 Toets- en itemanalyse

Na het imputeren van de missende waarden is het instrument *Spreken* geanalyseerd binnen het raamwerk van de klassieke testtheorie. Eerst is het instrumentarium op itemniveau geanalyseerd. In Tabel 5.1 en 5.2 is voor respectievelijke het jeugdjournaalfragment en het dobbelspel per item (of beoordelingsaspect) de volgende informatie te vinden:

- N = Aantal leerlingen
- c0 = Percentage leerlingen met score 0
- c1 = Percentage leerlingen met score 2
- c2 = Percentage leerlingen met score 2
- p = Gemiddelde score op het item gedeeld door de maximaal haalbare score
- r_{it} = Correlatie tussen de scores op het item en de scores op de hele toets
- alpha = Cronbach's alpha wanneer het item wordt weggelaten uit de toets

We zien dat alle items zowel bij het jeugdjournaalfragment als het dobbelspel naar behoren functioneren. De *p*-waarden schommelen tussen 0.374 en 0.938 bij het jeugdjournaalfragment en tussen 0.469 en 0.963 bij het dobbelspel. Bij de r_{it} -waarden valt op dat deze ruimschoots boven de .30 liggen. Hoewel dit deels het gevolg is van de wijze waarop de items gescoord worden (i.e., driepuntsschalen in plaats van dichotome goed-foutscores), wijzen de waarden op zeer goed onderscheidende items (zie Evers et al., 2010, p. 40). Samenvattend laten de tabellen dan ook zien dat het beoordelingsmodel geschikt is om leerlingen met een zwakkere spreekvaardigheid te onderscheiden van leerlingen met een betere spreekvaardigheid. Dat kan zowel met behulp van het jeugdjournaalfragment als met behulp van het dobbelspel. Er is geen aanleiding om bepaalde items weg te laten of beoordelingcategorieën samen te voegen.

Tabel 5.1. *Itemindices instrument Spreken - jeugdjournaalfragment*

Item	N	c0	c1	c2	*p*	r_{it}	alpha
1	0.072	0.294	0.634	0.781	0.521	0.874	0.072
2	0.143	0.519	0.339	0.598	0.711	0.864	0.143
3	0.108	0.432	0.460	0.676	0.707	0.864	0.108
4	0.092	0.234	0.674	0.791	0.652	0.867	0.092
5	0.073	0.382	0.545	0.736	0.719	0.864	0.073
6	0.013	0.095	0.892	0.939	0.538	0.873	0.013
7	0.379	0.493	0.128	0.374	0.453	0.877	0.379
8	0.081	0.817	0.102	0.511	0.471	0.877	0.081
9	0.086	0.344	0.570	0.742	0.750	0.862	0.086
10	0.042	0.289	0.670	0.814	0.724	0.863	0.042
11	0.084	0.321	0.595	0.755	0.590	0.871	0.084
12	0.025	0.186	0.789	0.882	0.488	0.876	0.025
13	0.015	0.216	0.769	0.877	0.548	0.873	0.015
14	0.016	0.141	0.843	0.914	0.565	0.872	0.016
15	0.029	0.387	0.584	0.778	0.662	0.867	0.029

Tabel 5.2. *Itemindices instrument Spreken - dobbelspel*

Item	N	c0	c1	c2	*p*	r_{it}	alpha
1	1417	0.015	0.168	0.817	0.901	0.507	0.848
2	1417	0.094	0.526	0.380	0.643	0.690	0.836
3	1417	0.051	0.464	0.485	0.717	0.630	0.840
4	1417	0.080	0.252	0.668	0.794	0.591	0.842
5	1417	0.005	0.064	0.931	0.963	0.438	0.852
6	1417	0.010	0.128	0.862	0.926	0.510	0.847
7	1417	0.284	0.493	0.223	0.470	0.502	0.848
8	1417	0.063	0.781	0.156	0.547	0.481	0.849
9	1417	0.057	0.332	0.611	0.777	0.721	0.834
10	1417	0.029	0.252	0.719	0.845	0.678	0.837
11	1417	0.059	0.301	0.640	0.790	0.560	0.844
12	1417	0.036	0.190	0.774	0.869	0.503	0.848
13	1417	0.030	0.236	0.733	0.851	0.576	0.843
14	1417	0.017	0.210	0.773	0.878	0.517	0.847
15	1417	0.025	0.387	0.588	0.782	0.663	0.838

In de onderzoeksvragen is uitgegaan van Spreken als één vaardigheid en ook in het referentiekader is dat het geval. Impliciet wordt daarmee aangenomen dat leerlingen die goed zijn in het navertellen van feitelijke verhalen ook goed zijn in het vertellen van een fantasieverhaal en andersom. Als dat zo is, kan per leerling gemakkelijk één totaalscore voor Spreken uitgerekend worden door de score voor het jeugdjournaalfragment en de score voor het dobbelspel te sommeren. Er is nagegaan in hoeverre een dergelijke keuze tot een verlies aan informatie leidt. De correlatie tussen het jeugdjournaalfragment en het dobbelspel is gelijk aan 0.764. Als we deze correlatie corrigeren voor meetfout verkrijgen we een (latente) correlatie van 0.881. Deze waarde is te laag om met overtuiging over één onderliggende dimensie te spreken, maar ook weer niet zo laag dat grote groepen leerlingen op de ene taak extreem laag scoren en op de andere extreem hoog. Voor bijna driekwart van de leerlingen geldt dat de totaalscore op de ene taak ten hoogste drie punten hoger of lager ligt dan de totaalscore op de andere taak. Het lijkt dan ook verdedigbaar om voor alle leerlingen één score voor Spreken te berekenen.

Na de analyse op itemniveau is gekeken naar de betrouwbaarheid van het instrument. In de literatuur zijn verschillende methoden voorgesteld om de betrouwbaarheid te berekenen. Cronbach's alpha (Cronbach, 1951) wordt verreweg het meest gebruikt, maar er zijn methoden beschikbaar die een betere schatting geven van de werkelijke betrouwbaarheid. Een voorbeeld daarvan is Guttmans lambda2 (Guttman, 1945). Hoewel het verschil klein is, kan aangetoond worden dat de waarde van Guttmans lambda2 dichterbij de werkelijke betrouwbaarheid ligt dan Cronbach's alpha. Eventueel kan de betrouwbaarheid ook uitgedrukt worden in termen van de *greatest lower bound* (Ten Berge & Sočan, 2004). In principe geeft de *greatest lower bound* een nog betere (onder)schatting van de betrouwbaarheid, maar een nadeel is dat de schattingen te hoog kunnen uitvallen in steekproeven kleiner dan duizend leerlingen en bij toetsen van meer dan tien items (Sijtsma, 2009). Om deze reden is ervoor gekozen om Guttmans lambda2 als

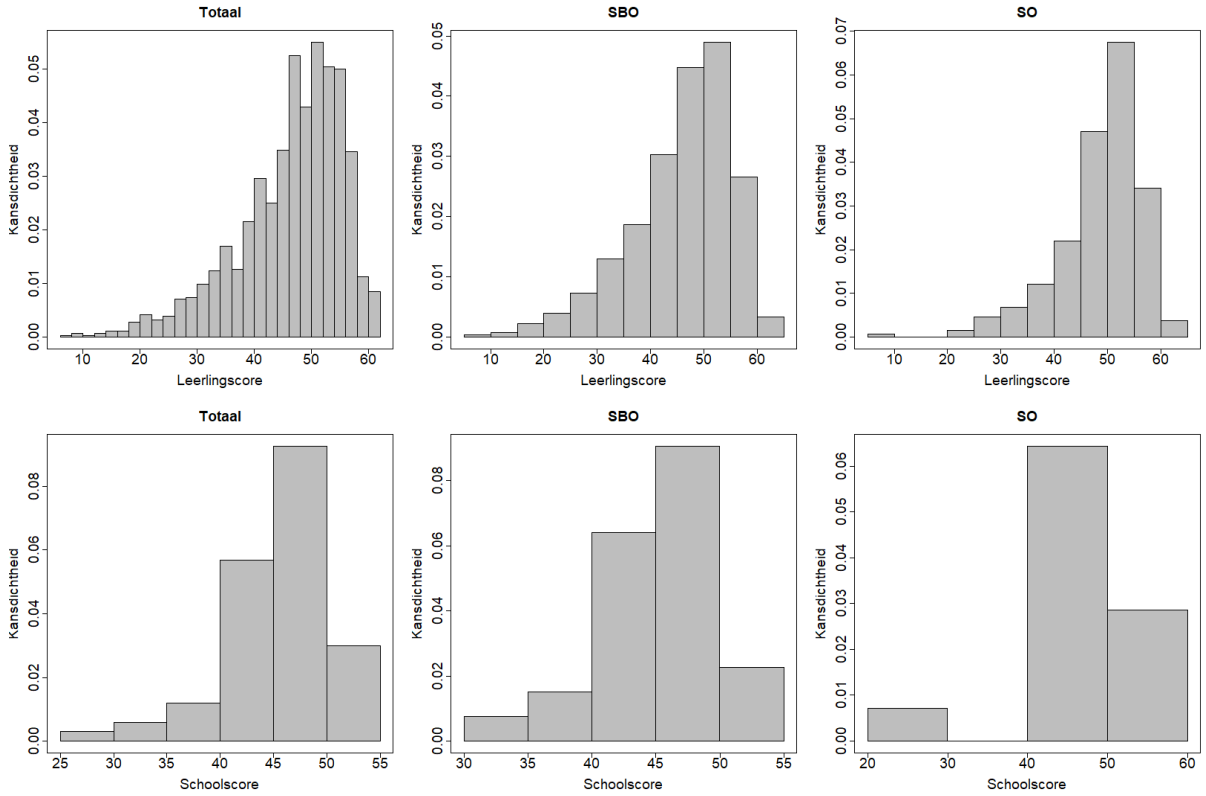
uitgangspunt te nemen en alleen voor de volledigheid ook Cronbach's alpha en de *greatest lower bound* te berekenen. De betrouwbaarheidscoëfficiënten zijn te vinden in Tabel 5.3. Hoewel besloten is om bij de beantwoording van de onderzoeksvragen in principe gebruik te maken van één totaalscore (d.w.z. de som over het jeugdjournaalfragment en het dobbelspel) is de betrouwbaarheid in Tabel 5.3 ook voor de twee taken apart berekend. We zien dat de betrouwbaarheid ver boven de .800 ligt, zowel voor de twee taken apart als voor het totaal. Dit betekent dat de betrouwbaarheid van het instrumentarium als 'goed' gekwalificeerd lijkt te kunnen worden (Evers et al., 2010), maar een belangrijke nuancering is wel dat de toetsleiders niet als foutenbron zijn meegenomen. Hier wordt apart aandacht aan besteed in paragraaf 5.4.

Tabel 5.3. *Toetsindices instrument Spreken*

Groepslid	min	max	N	M	SD	skew	kurt	alpha	lambda2	glb
Filmpje	0	30	1417	22.337	5.231	-1.031	3.907	0.877	0.880	0.931
Dobbelspel	0	30	1417	23.506	4.595	-1.022	3.999	0.852	0.854	0.916
Totaal	0	60	1417	45.843	9.231	-1.079	4.133	0.923	0.925	0.974

Naast betrouwbaarheidscoëfficiënten bevat Tabel 5.3 ook enkele maten die zicht geven op de verdeling van de (ruwe) scores. Het gaat om het gemiddelde, de standaarddeviatie, de *skewness* en de *kurtosis*. We zien onder meer dat leerlingen gemiddeld iets meer dan driekwart van het maximaal te behalen aantal punten wisten te behalen (45.843 ÷ 60), dat de scores iets linksscheef verdeeld zijn en dat de verdeling iets meer gepiekt is dan de normaalverdeling. Dat beeld wordt uiteraard bevestigd in de visuele weergave van de (ruwe) scoreverdelingen in Figuur 5.1. Wel zien we dat het beeld in het so iets extremer is dan in het sbo. Vermoedelijk wordt dit deels veroorzaakt door het grote verschil in leerlingaantal: 1153 sbo-leerlingen versus 264 so-leerlingen. Ook op schoolniveau lijken de verdelingen iets af te wijken van de normaalverdeling, maar vanwege het kleine aantal scholen is dat lastig met zekerheid te zeggen (53 sbo- en 14 so-scholen).

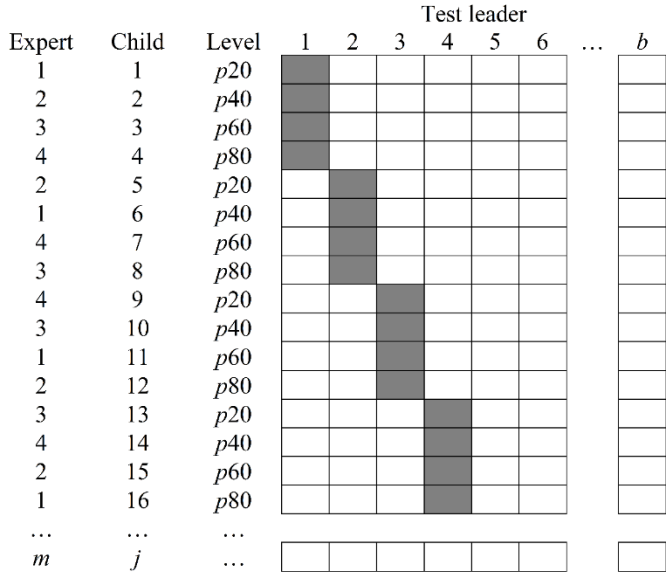
Figuur 5.1. *Scoreverdelingen op leerling- (boven) en schoolniveau (onder) uitgesplitst naar schooltype*



5.4 Beoordelaarsovereenstemming

In de toets- en itemanalyse in paragraaf 5.3 is een objectieve scoring verondersteld. Daar is in werkelijkheid geen sprake van. De betrouwbaarheid van de scores hangt ook af van de mate waarin de beoordelaars het beoordelingsmodel consistent toepassen. Hier is apart onderzoek naar gedaan door een deel van de leerlingen achteraf op basis van de beschikbare filmopnames door een tweede (expert)beoordelaar te laten beoordelen. In het design is ervan uitgegaan dat er b, $b = 1, \ldots, B$, toetsleiders waren en m, $m = 1, \ldots, M$, expertbeoordelaars. Van elke toetsleider b zijn J leerlingen (of sessies) gekozen en elke expertbeoordelaar m (her)beoordeelde $B \times J$ leerlingen. De leerlingen j zijn geselecteerd op basis van hun percentiele rang. We zijn uitgegaan van percentiel 20, 40, 60 en 80. Op deze manier was een goede spreiding geborgd en kon voorkomen worden dat we in analyses te maken hadden met nullen in de randtotalen. Figuur 5.2 geeft een schematische weergave van het design dat gehanteerd is in het onderzoek naar beoordelaarsovereenstemming.

Figuur 5.2. *Design voor het onderzoek naar beoordelaarsovereenstemming*



In totaal waren 27 toetsleiders betrokken bij dit onderzoek. Van die toetsleiders konden er 21 meegenomen worden in het onderzoek naar beoordelaarsovereenstemming. Van de overige 6 toetsleiders was geen of onvoldoende bruikbaar filmmateriaal beschikbaar voor een herbeoordeling. Na afloop van de herbeoordeling is op verschillende manieren gekeken naar het 'gebrek aan' overeenstemming met de norm (d.w.z., het expertoordeel). Het verschil tussen de expertbeoordelingen en de toetsleiderbeoordelingen is eerst gebruikt om de kwaliteit van de beoordelingen te evalueren. Concreet is het gebrek aan overeenstemming met de norm voor elk van de toetsleiders in kaart gebracht door de *Mean Absolute Error* te berekenen:

$$MAE_b = \frac{\sum_j \sum_i \left| s_{bji} - s_{mji} \right|}{N_j},$$

waarin s_{bji} en s_{mij} respectievelijk de beoordelingen van toetsleider *b* en expert *m* representeren voor leerling *j* op item *i*. De *Median Absolute Deviation* is vervolgens gebruikt als maat om "afwijkende" beoordelingen op te sporen. Om optimaal rekening te kunnen houden met een mogelijke asymmetrische verdeling van **MAE** is de *Median Absolute Deviation* van de mediaan gebaseerd op alle punten groter of gelijk aan de mediaan:

$$MAD = Mdn(\left| \mathbf{Y} - Mdn(MAE) \right|),$$

waarin

$$\mathbf{Y} = \left\{ MAE_b \in \mathbf{MAE} : MAE_b \geq Mdn(\mathbf{MAE}) \right\}.$$

Gegeven deze afstand is een toetsleider b als "afwijkend" beschouwd wanneer geldt dat:

$$\frac{MAE_b - Mdn(\mathbf{MAE})}{MAD} > 2.5$$

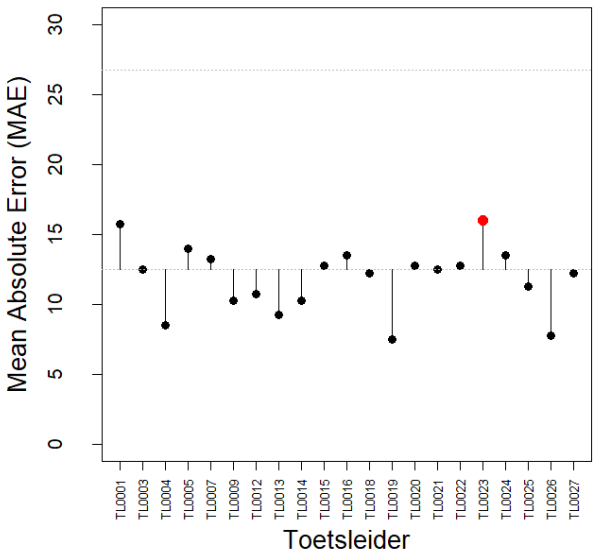
Een afkapwaarde van 2.5 is voorgesteld door Leys et al. (2013), maar andere afkapwaarden zijn vanzelfsprekend mogelijk. Na het vaststellen van het 'gebrek aan' overeenstemming met de norm is de overall kwaliteit van de beoordelingen geëvalueerd door Cohen's weighted kappa coefficient (κ) en Gower's

similarity coefficient (G_{xy}) te berekenen. Cohen's kappa is als volgt geïnterpreteerd: $\kappa < 0.200$ onvoldoende; $0.200 < \kappa < 0.400$ matig; $0.400 < \kappa < 0.600$ voldoende; $0.600 < \kappa < 0.800$ goed; $\kappa \geq 0.800$ uitstekend. Gower's similarity coefficient is als laag beschouwd wanneer $G_{xy} < 0.650$, als acceptabel wanneer $0.650 \leq G_{xy} < 0.800$, en als hoog wanneer $G_{xy} > 0.800$.

Figuur 5.3 laat per toetsleider de Mean Absolute Error (MAE) zien. De onderste grijze stippellijn representeert de mediaan van de MAE's van de toetsleiders. De bovenste grijze stippellijn is de mediaan van de MAE's voor een "oneindig" groot aantal ***random*** beoordelingen. Zoals we kunnen zien leken de beoordelingen van toetsleiders TL0004, TL0019 en TL0026 sterk op de beoordelingen van de experts. De beoordelingen van toetsleiders TL0001 en TL0023 weken daarentegen nogal af van de expertbeoordelingen. De MAE van toetsleider TL0023 is zelfs zo groot dat er gesproken kan worden van afwijkend gedrag ten opzichte van de andere twintig toetsleiders; de MAD voor deze toetsleider was groter dan 2.5. Cohen's weighted kappa coefficient (κ) and Gower's similarity coefficient (G_{xy}) wijzen op een matige overall kwaliteit: $\kappa = 0.345$ en $G_{xy} = 0.801$, waarbij opvalt dat de eerste maat een aanzienlijk positiever beeld geeft dan de tweede maat. Gemiddeld gezien zouden de expertbeoordelaars vermoedelijk iets lager en gevarieerder beoordeeld hebben ($M = 46.488$, $SD = 5.507$) dan de toetsleiders ($M = 43.667$, $SD = 9.494$); het verschil is statistisch significant, $t(166) = -2.356$, $p = 0.020$.

Het is de vraag of TL0023 wel of niet in analyses moet worden meegenomen. Wij hebben ervoor gekozen om de toetsleider wél mee te nemen, omdat het verschil in kwaliteit ten opzicht van bijvoorbeeld TL0001 zeer klein is. Bovendien is de vraag of een beoordeling "ter plekke" wel exact hetzelfde kan opleveren als een beoordeling achteraf op basis van filmopnamen. Het is goed denkbaar dat een beoordelingsaspect als "non-verbale communicatie" in een filmopname anders gepercipieerd wordt dan bij de feitelijke afname. Als dat het geval is, geven de eerder gepresenteerde resultaten mogelijk een te pessimistisch beeld van de werkelijkheid, en is er des te meer reden om voorzichtig te zijn met het weglaten van toetsleiders (of leerlingen).

Figuur 5.3. *Mean Absolute Error per toetsleider*

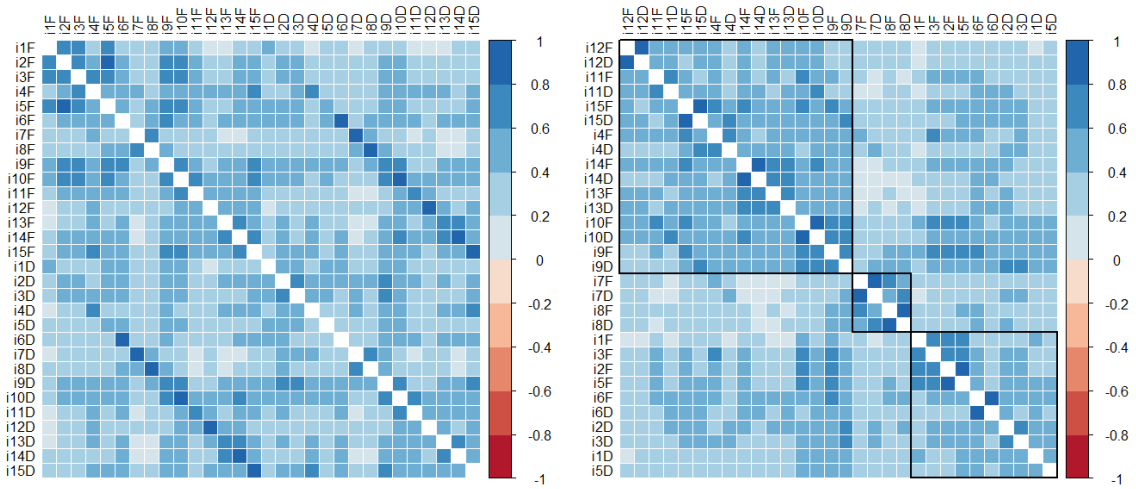


5.5 Dimensionaliteit

Het beoordelingsmodel omvat verschillende aspecten van spreekvaardigheid en hoewel er in paragraaf 5.3 vanuit is gegaan dat deze aspecten betekenisvol samen te voegen zijn tot één totaalscore, en dat ook het jeugdjournaalfragment en het dobbelspel samengenomen kunnen worden, is het belangrijk om inzicht te hebben in de dimensionele structuur van de vaardigheid. Als er bijvoorbeeld sprake is van een geringe samenhang tussen de eerder beschreven vijf kenmerken van de taakuitvoering kan er reden zijn om (ook) met deelscores te werken. De dimensionele structuur is eerst op itemniveau in kaart gebracht. Omdat het instrumentarium in een volledig afnamedesign is voorgelegd aan 1417 leerlingen zijn de polychorische correlaties tussen de items vrij gemakkelijk uit te rekenen. Figuur 5.4 geeft een visuele representatie van de matrix met polychorische inter-itemcorrelaties. In de Figuur 5.4 worden de correlaties weergegeven door

middel van een kleur; donkerder blauw betekent een hogere positieve correlatie en donkerder rood betekent een hogere negatieve correlatie.

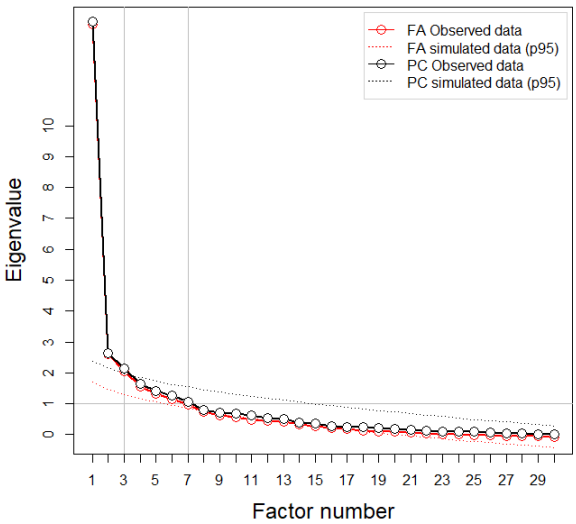
Figuur 5.4. *Visuele representatie polychorische inter-itemcorrelatiematrix Spreken*



In Figuur 5.4 is te zien dat de items allemaal positief met elkaar samenhangen (links) en dat een verkennende hiërarchische clusteranalyse (rechts) de correlaties tussen de itemparen ordent in drie verschillende clusters met items die (relatief) sterk met elkaar samenhangen. Het jeugdjournaalfragment (i1F … i15F) is in Figuur 5.4 op basis van de inter-itemcorrelaties niet te onderscheiden van het dobbelspel (i1D … i15D).

In aansluiting op de berekening van de inter-itemcorrelaties is op de polychorische correlatiematrix een exploratieve factoranalyse met varimax rotatie uitgevoerd. Om het aantal factoren enigszins beredeneerd te kiezen is gestart met een parallelanalyse zoals voorgesteld door Horn (1965). Het is een op simulatie gebaseerde methode waarin in essentie een willekeurige dataset wordt gegenereerd met hetzelfde aantal items en exact dezelfde scorerange. De eigenwaarden van de items in deze “junk” dataset worden vervolgens vergeleken met de eigenwaarden van de items in de werkelijke dataset. Alle factoren waarbij de eigenwaarde groter is dan de “toevallige” (gesimuleerde) eigenwaarden worden vervolgens behouden in de factoranalyse. Figuur 5.5. laat de uitkomsten van de parallelanalyse zien. We zien dat de parallelanalyse suggereert om drie componenten of zeven factoren te onderscheiden. Omdat de eerdere analyse (zie Figuur 5.4) geen enkele aanleiding geeft om zeven factoren te kiezen is in de factoranalyse voor drie gekozen.

Figuur 5.5. *Scree plot met de geobserveerde en gesimuleerde eigenwaarden zoals gevonden in (parallelle) factoranalyse en de principale componentenanalyse*



De uitkomsten van de factoranalyse zijn te vinden in Tabel 5.4. Voor de leesbaarheid zijn de ladingen kleiner dan 0.300 niet vermeld in Tabel 5.4. H2 en U2 representeren respectievelijk de communaliteit en

specifieke variantie van de gestandaardiseerde factorladingen zoals verkregen uit de correlatiematrix. De uitkomsten stemmen vrijwel volledig overeen met het beeld dat Figuur 5.4 gaf en zijn vrij gemakkelijk inhoudelijk te duiden naar Taalvorm, Taalinhoud en Taalgebruik (zie bijvoorbeeld Bloom & Lahey, 1978; Lahey, 1988):

- Onder **Taalvorm** wordt de grammatica van een taal verstaan, dat wil zeggen de wijze waarop klanken, woorden en zinnen worden gevormd. Het zijn de grammaticaregels met betrekking tot fonologie (klankleer), morfologie (woordvorming) en syntaxis (zinsleer).
- **Taalinhoud** heeft betrekking op het begrijpen en produceren van verhalen. Verhalen vertellen veronderstelt de vaardigheid om een eigen (non-) fictie verhaal over meerdere aaneensluitende uitingen vast te houden en daarbij de grote lijn duidelijk te maken.
- Bij **Taalgebruik** gaat het om de vraag hoe de leerling taal in zijn dagelijks leven inzet. Onder taalgebruik vallen onder meer het begrijpen en gebruiken van communicatieve functies en conversatieregels. Denk daarbij aan het leggen en onderhouden van sociale contacten, het vragen en geven van informatie, het maken van oogcontact of het initiëren en handhaven van een gespreksonderwerp

De uitkomsten geven ook enige aanleiding tot discussie. We zien onder meer dat sommige items bij het jeugdjournaalfragment net iets ander functioneerden dan bij het dobbelspel. Dat geldt bijvoorbeeld voor item 9. Op basis van het jeugdjournaalfragment zou het item onder Taalinhoud vallen en op basis van het dobbelspel onder Taalvorm. De factorladingen verschillen echter niet sterk, dus een keuze voor één van de twee lijkt verdedigbaar. Dat geldt ook voor de overige twijfelgevallen, dus los van de beperkte inhoudelijke discussie die gevoerd kan worden over de indeling van items wijst de analyse overtuigend op drie vrij sterk samenhangende factoren.

Tabel 5.4. *Resultaten exploratieve factoranalyse*

Item	Omschrijving	Factor 1 Taalvorm	Factor 2 Taalinhoud	Factor 3 Taalgebruik	H2	U2
i7F	De leerling zet regelmatig aspecten als stiltes, wisselende intonatie/volume/spreeksnelheid en lichaamstaal bewust in om de gesprekspartner te boeien.			0.770	0.640	0.360
i7D				0.870	0.780	0.220
i8F	Het non-verbale gedrag van de leerling zet zijn verbale boodschap kracht bij.			0.690	0.560	0.440
i8D				0.740	0.620	0.380
i1F	De leerling heeft (bijna) alle taalhandelingen van de opdracht uitgevoerd.		0.760		0.580	0.420
i1D			0.460		0.350	0.650
i2F	De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten sterk.		0.830		0.770	0.230
i2D		0.440	0.430	0.350	0.510	0.490
i3F	De leerling laat een logische opbouw zien, bijvoorbeeld met zowel verbinding binnen als verbinding tussen inleiding, kern en afsluiting.		0.770		0.690	0.310
i3D		0.400	0.340	0.400	0.440	0.560
i5F	De leerling kan de strekking van het filmpje duidelijk maken aan de gesprekspartner / De leerling vertelt een verhaal waar hij alle plaatjes in verwerkt.		0.830		0.790	0.210
i5D		0.350	0.360		0.310	0.690
i4F	De leerling gebruikt eenvoudige voeg- en verbindingswoorden correct.	0.550	0.350	0.320	0.530	0.470
i4D		0.580		0.360	0.480	0.520
i6F	De leerling gebruikt consequent taal die bij de gesprekspartner past.	0.510	0.380	0.340	0.520	0.480
i6D		0.490		0.310	0.380	0.620
i9F	De uitwerking van de leerling is aantrekkelijk door voldoende variatie in woordgebruik.	0.480	0.630		0.710	0.290
i9D		0.530	0.440	0.390	0.620	0.380
i10F	De getoonde woordenschat van de leerling is toereikend voor het uitvoeren van de opdracht.	0.530	0.630		0.690	0.310
i10D		0.640	0.480		0.670	0.330
i11F	De leerling spreekt redelijk vloeiend, al vertoont zijn uitwerking een enkele hapering en/of herformulering.	0.550	0.320		0.410	0.590
i11D		0.610			0.430	0.570
i12F	De leerling is goed verstaanbaar, ondanks een eventueel accent.	0.610			0.440	0.560
i12D		0.600			0.430	0.570
i13F	De leerling vervoegt alle werkwoorden over het algemeen correct.	0.760			0.610	0.390
i13D		0.750			0.590	0.410
i14F	De leerling verbuigt (combinaties met) naamwoorden meestal correct.	0.800			0.720	0.280
i14D		0.740			0.580	0.420
i15F	De leerling maakt over het algemeen correcte enkelvoudige en samengestelde	0.640	0.370		0.590	0.410
i15D		0.690			0.590	0.410

5.6 Variabelen

In het databestand zijn bij Spreken de volgende variabelen opgenomen:

SPR_TTLID Toetsleideridentificatie bij Spreken
SPR_MISS Percentage ontbrekende antwoorden
SPR_RSCO Toetsscore
SPR_REFN Referentieniveau (<1F/1F/2F)
SPR_WEIGHT Gewicht om te corrigeren voor leerroute

Bij de toetsscore-variabele is het de vraag hoe deze het beste tot stand kan komen. Het is het gemakkelijkst om de toetsscore voor een leerling gelijk te stellen aan de som van alle itemscores, maar in een recent peilingsonderzoek in groep 8 (zie Van Langen et al., 2017) zijn de meer technische beoordelingsaspecten (i9 tot en met i15) minder zwaar gewogen dan de meer inhoudelijke beoordelingsaspecten (i1 tot en met i8). Die keuze is in dit onderzoek overgenomen. Dit betekent concreet dat bij de berekening van de toetsscore aan items 1 tot en met 8 (zowel bij het filmfragment als het dobbelspel) gewicht 1.500 is toegekend ([30 scorepunten * 0.800] ÷ [8 items × 2 scorepunten]) en aan items 9 tot en met 15 gewicht 0.429 ([30 scorepunten * 0.200] ÷ [7 items × 2 scorepunten]). Het gevolg daarvan is dat “inhoud” 80 procent meetelt en “techniek” 20 procent. Alle toetsscores zijn afgerond naar gehele getallen. Over de wijze waarop het referentieniveau (**SPR_REFN**) van leerlingen is bepaald wordt in paragraaf 8.1.1 gerapporteerd. Het percentage ontbrekende antwoorden (**SPR_MISS**) is berekend door het aantal itemscores van een leerling te delen door het totale aantal items (2 × 15 = 30).

6 Gesprekken

6.1 Inhoud en afname

In aansluiting op het referentiekader Taal is een instrument ontwikkeld om de gespreksvaardigheid van leerlingen te meten. Het instrument Gesprekken richtte zich op de volgende vijf kenmerken van de taakuitvoering.

Kenmerk	Uitwerking referentieniveau 1F
1. Beurtnemen en bijdragen aan samenhang	Kan een kort gesprek beginnen, gaande houden en beëindigen
2. Afstemming op doel	Kan gesprekken voeren om informatie en meningen uit te wisselen, uitleg of instructie te geven en te volgen. Herkent gesprekssituaties en kan passende routines gebruiken.
3. Afstemming op gesprekspartner(s)	Kan de gesprekspartner(s) redelijk volgen tenzij ze voor onverwachte wendingen in het gesprek zorgen. Kan woorden ondersteunen met non-verbaal gedrag.
4. Woordgebruik en woordenschat	Beschikt over voldoende woorden om te praten over vertrouwde situaties en onderwerpen, maar zoekt nog regelmatig naar woorden en varieert niet veel in woordgebruik.
5. Vloeiendheid, verstaanbaarheid en grammaticale beheersing	De uitspraak is duidelijk genoeg om de spreker te kunnen volgen, ondanks een eventueel accent, verkeerde intonatie, onduidelijke articulatie en/of haperingen. Redelijk accuraat gebruik van eenvoudige zinsconstructies

Om bovengenoemde kenmerken te meten is gebruikgemaakt van een bestaand spel uit Duitsland, het zogeheten *Fischerspiel*. Leerlingen werken samen en overleggen om vissen te vangen en naar de haven te brengen voordat er storm komt. De instructies van het oorspronkelijke spel zijn wel iets aangepast. Zo is het aantal vissen dat leerlingen moeten vangen verminderd om het spel vlotter te laten verlopen. Ook zijn enkele regels weggelaten om ervoor te zorgen dat leerlingen van uiteenlopende niveaus het spel begrijpen en dus hun gespreksvaardigheid kunnen laten zien. In paragraaf 2.2.1 is een uitgebreide omschrijving van het *Fischerspiel* te vinden. De spelinstructie is te vinden in Bijlage 6.

De afname van het *Fischerspiel* vond plaats in een aparte, zoveel mogelijk prikkelarme, ruimte. Het spel werd in principe gespeeld in groepjes van vier leerlingen, maar indien gewenst kon het spel ook met twee of drie leerlingen gespeeld worden. De groepjes zijn vooraf *random* samengesteld, maar als sommige leerlingcombinaties volgens de leerkracht onhandig waren, mocht er (zeer beperkt) gewisseld worden. Ruzie tijdens de pauze kon bijvoorbeeld een reden zijn om een leerling in een ander groepje te plaatsen. Elke afname werd begeleid door een getrainde toetsleider. Tijdens het spel nam de toetsleider niet de rol in van deelnemer, maar die van coach: hij bewaakte het verloop van het spel. De toetsleider zorgde er bijvoorbeeld voor dat alle leerlingen ongeveer evenveel aan bod kwamen. Ook zorgde hij ervoor dat de leerlingen zich veilig en betrokken voelden. Om ervoor te zorgen dat de het spel door álle leerlingen gespeeld kon worden, mocht de toetsleider, indien strikt noodzakelijk, kleine aanpassingen doorvoeren in de afname. In Bijlage 5 is te lezen welke aanpassingen toegestaan waren. In totaal nam de afname van het *Fischerspiel* 30 tot 45 minuten in beslag, afhankelijk van het verloop van het spel.

Naast een coachende rol vervulde de toetsleider tijdens de afname ook de rol van beoordelaar. Daarbij is uitgegaan van een model waarin de toetsleider voor achttien verschillende aspecten, verdeeld over de verschillende kenmerken van de taakuitvoering, op een driepuntschaal (0 = onvoldoende, 1 = voldoende, 2 = goed) aangaf in hoeverre de leerling dit aspect liet zien. Concreet ging het om onderstaande achttien aspecten. Ter illustratie is bij alle achttien beoordelingsaspecten de omschrijving van de hoogste beoordelingscategorie weergegeven.

Uitvoering

1. Kwaliteit van de inhoud
   - De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten sterk.

Het woord nemen en bijdragen aan samenhang

2. Gesprek onderhouden
   - De leerling neemt regelmatig initiatief om het gesprek te beginnen, gaande te houden en/of af te sluiten (bijvoorbeeld door vragen te stellen of voorstellen te doen).
3. Passendheid van het woord nemen
   - De leerling neemt doorgaans op passende wijze het woord.
4. Inbreng gesprekspartners integreren
   - De leerling integreert bijdragen van de gesprekspartners in zijn eigen inbreng wanneer dat toegevoegde waarde heeft.

Afstemming op doel

5. Doelgerichtheid
   - De leerling neemt initiatief om een gezamenlijk gespreksdoel te behalen door de gesprekspartners te betrekken bij het gesprek, bijvoorbeeld met behulp van vragen.
6. Begrijpelijk maken
   - De leerling maakt de eigen gedachtegang begrijpelijk.

Afstemming op de gesprekspartners

7. Taalgebruik afstemmen
   - De leerling gebruikt consequent taal die bij de situatie past.
8. Non-verbale communicatie
   - Het non-verbale gedrag van de leerling zet zijn verbale boodschap kracht bij.
9. Actief luistergedrag
   - De leerling vertoont adequaat actief luistergedrag.
10. Reageren op gesprekspartner
    - De leerling geeft consequent gepaste reacties (verbaal/nonverbaal).

Woordgebruik en woordenschat

11. Woordgebruik
    - De inbreng van de leerling vertoont voldoende variatie in woordgebruik.
12. Woordenschat
    - De getoonde woordenschat van de leerling is toereikend om het gesprek te voeren.

Vloeiendheid, verstaanbaarheid en grammaticale beheersing

13. Vloeiendheid
    - De leerling spreekt redelijk vloeiend, al vertoont zijn uitwerking een enkele hapering en/of herformulering.
14. Uitspraak en intonatie
    - De leerling is goed verstaanbaar door zijn uitspraak, articulatie en intonatie, ondanks een eventueel accent.
15. Grammatica: vervoeging
    - De leerling vervoegt alle werkwoorden over het algemeen correct.
16. Grammatica: verbuiging
    - De leerling verbuigt (combinaties met) naamwoorden meestal correct.
17. Grammatica: zinsbouw
    - De leerling maakt over het algemeen correcte enkelvoudige en samengestelde zinnen.

Evaluatie

18. Evaluatie
    - De leerling geeft een helder oordeel over het spel.

Aspecten 1 en 18 maken geen onderdeel uit van de taakuitvoering, maar zijn desalniettemin wel toegevoegd aan het beoordelingsmodel om een vollediger en realistischer beeld van de gespreksvaardigheid te kunnen geven. Naast de genoemde aspecten was er bovendien sprake van een groepsbeoordeling en een globale evaluatie. Bij de groepsbeoordeling is gekeken naar de evenwichtigheid van het gesprek, de behoefte aan sturing en het gedrag van de toetsleider. Bij de globale evaluatie is de volgende vraag gesteld: *Hoe zou je de spreekvaardigheid van deze leerling globaal gezien typeren in termen van referentieniveau 1F? Matig-Voldoende-Goed*. Beoordelingsaspect 18, de groepsbeoordeling en de globale evaluatie worden niet meegenomen in de score van een individuele leerling. Ze waren voornamelijk bedoeld voor de controle van de data en de validering van uitkomsten.

De toetsleider vulde het beoordelingsmodel voor elke leerling zoveel mogelijk tijdens het spel in. Dat gebeurde na de eerste drie rondes van het spel, dus nadat alle leerlingen drie keer met de dobbelsteen hadden gegooid. Dit moment is gekozen, omdat uit de *try-outs* bleek dat leerlingen vanaf dat moment meer zelfstandig met het spel bezig zijn en de toetsleider daardoor meer in de gelegenheid is om de beoordeling

uit voeren. Drie spelrondes bleken bovendien ruim voldoende om een beeld te krijgen van de gespreksvaardigheid van de leerlingen. Om ervoor te zorgen dat toetsleiders de beoordelingstaak betrouwbaar en hetzelfde konden uitvoeren zijn de volgende vier maatregelen genomen:

1. Elk aspect is uitgewerkt met concrete voorbeelden passend bij het *Fischerspiel*.
2. Alle toetsleiders hebben uitgebreid geoefend met het beoordelingsmodel, tijdens een training, op basis van filmfragmenten.
3. Elk spel werd opgenomen op video, zodat de toetsleider waar nodig achteraf de mogelijkheid had om de beoordeling te completeren of controleren.
4. Via een app-groep konden toetsleiders elkaar dilemma's rondom de beoordeling voorleggen.

In een flankerend onderzoek is empirisch nagegaan in hoeverre toetsleiders de beoordelingstaak adequaat hebben uitgevoerd. Hierover wordt in paragraaf 6.4 gerapporteerd.

6.2 Missende waarden

Bij het onderdeel Gesprekken is er sprake van één taak (het spel) waar álle leerlingen in groepjes aan deelnemen. Dit betekent dat er geen sprake is van toetsversies die via item respons theorie modellering of een andere equivaleringstechniek vergelijkbaar gemaakt moeten worden. Wel kan er sprake zijn van missende waarden doordat een toetsleider:

- Bij een leerling één of meer beoordelingsaspecten "vergeten" is in te vullen;
- Bij één of meer beoordelingsaspecten "niet te beoordelen" heeft ingevuld.

Verkennende analyses lieten zien dat dit slechts in beperkt mate voorkwam. Het percentage missende waarden varieerde van 0.35 tot 2.11 procent. Gemiddeld was er sprake van 0.91 procent missende waarden. De beoordelingscategorie "niet te beoordelen" werd niet vaak gekozen. Gemiddeld werd de categorie in 2.58 procent van de gevallen gekozen, variërend van 1.20 procent bij beoordelingsaspect 14 tot maximaal 4.92 procent bij beoordelingsaspect 16. Het is de vraag hoe het beste omgegaan kan worden met de missende waarden. Imputatie ligt voor de hand, maar deze oplossing is niet in alle gevallen goed te verdedigen.

- In het geval dat een toetsleider een beoordelingaspect vergeten is in te vullen lijkt de aanname *missing at random* (zie Little & Rubin, 1987) verdedigbaar en is imputatie mogelijk, tenzij er zeer veel missende waarden zijn of gerelateerd zijn aan specifieke groepjes of toetsleiders. De data geven geen aanleiding om te veronderstellen dat er een relatie is tussen de toetsleider en het aantal missende waarden. Bij bijna alle 27 toetsleiders is het percentage missende waarden kleiner dan 1 procent en slechts bij 2 toetsleiders is het percentage missende waarden met 5.18 (TL0017) en 6.79 (TL0001) procent iets groter. Bij een aantal groepjes is het percentage missende waarden desalniettemin opvallend hoog. Van de in totaal 404 groepjes is er bij 4 sprake van meer dan 25 procent missende waarden (GRP10001143, GRP10001339, GRP10001368, GRP10001366). Deze groepjes (14 leerlingen) hebben we buiten de analyses gelaten, omdat de spelafname bij deze groepjes mogelijk niet optimaal is verlopen. De overige missende waarden zijn geïmputeerd.
- Als een toetsleider bij een beoordelingsaspect "niet te beoordelen" heeft ingevuld is het lastiger om te bepalen wat een adequate aanpak is. Toetsleiders hebben nadrukkelijk de instructie gekregen dat de categorie "niet te beoordelen" alleen in zeer uitzonderlijke gevallen gebruikt mocht worden. Als deze instructie gevolgd is, zal de beoordelingscategorie niet vaak gekozen zijn en is imputatie een praktische oplossing. In de data is te zien dat de beoordelingscategorie door geen enkele toetsleider consistent vaak is gekozen, maar bij een aantal specifieke groepjes wel (> 25 procent). Het gaat om groepjes GRP10001230, GRP10001301, GRP10001603 en GRP10001032. Deze groepjes (13 leerlingen) hebben we buiten de analyse gelaten. In alle resterende gevallen hebben we "niet te beoordelen" vervangen door een geïmputeerde score.

Tot nu toe zijn missende waarden en de beoordelingscategorie "niet te beoordelen" afzonderlijk beschouwd. Het kan zo zijn dat bij een leerling beide scenario's van toepassing waren en dat er daarom alsnog reden is om de leerling buiten de analyse te laten. De verdeling ziet er als volgt uit:

N aspecten beoordeeld	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
N leerlingen	0	0	1	3	3	2	2	1	6	1	2	6	11	36	33	78	1210

We zien dat verreweg de meeste leerlingen (1210) op alle aspecten beoordeeld zijn. Het is onduidelijk vanaf welk punt er aanleiding is om te twijfelen over de kwaliteit van een beoordeling. In dit onderzoek leggen we de afkapwaarde, net zoals eerder bij Spreken, bij 80 procent (zie paragraaf 5.2 voor een toelichting). Dat betekent dat de leerlingen die op 13 of minder aspecten beoordeeld zijn (38 in totaal) niet worden meegenomen in de analyse. Bij de overige leerlingen zijn de ontbrekende waarden geïmputeerd. Bovengenoemde keuzes hebben tot gevolg dat van de oorspronkelijke groep met 1422 leerlingen er 1357

in de analyse konden worden meegenomen (1422 – 14 – 13 – 38) en dat in die dataset 252 waarden zijn geïmputeerd (1.09 procent) op de manier zoals beschreven in paragraaf 5.2.

6.3 Toets- en itemanalyse

Na het imputeren van de missende waarden is het instrument Gesprekken, net zoals het instrument Spreken, geanalyseerd binnen het raamwerk van de klassieke testtheorie. Eerst is het instrumentarium op itemniveau geanalyseerd. In Tabel 6.1 zijn de resultaten te vinden. We zien dat alle items naar behoren (en min of meer vergelijkbaar) functioneerden. De *p*-waarden schommelen tussen 0.620 en 0.927, en bij de r_{it}-waarden valt op dat deze ruimschoots boven de .300 liggen. Samenvattend laat Tabel 6.1 zien dat het beoordelingsmodel geschikt is om leerlingen met een zwakkere gespreksvaardigheid te onderscheiden van leerlingen met een betere gespreksvaardigheid. Er is geen aanleiding om bepaalde items weg te laten of beoordelingcategorieën samen te voegen.

Tabel 6.1. *Itemindices instrument Gesprekken*

Item	n	c0	c1	c2	*p*	r_{it}	alpha
1	1357	0.133	0.492	0.374	0.620	0.671	0.876
2	1357	0.127	0.416	0.458	0.665	0.552	0.881
3	1357	0.077	0.268	0.655	0.789	0.521	0.882
4	1357	0.064	0.256	0.679	0.808	0.536	0.882
5	1357	0.143	0.444	0.413	0.635	0.641	0.877
6	1357	0.113	0.379	0.508	0.697	0.704	0.874
7	1357	0.033	0.252	0.715	0.841	0.592	0.879
8	1357	0.062	0.609	0.329	0.634	0.500	0.883
9	1357	0.088	0.413	0.499	0.705	0.648	0.877
10	1357	0.085	0.267	0.648	0.781	0.575	0.880
11	1357	0.040	0.205	0.755	0.858	0.705	0.874
12	1357	0.017	0.178	0.805	0.894	0.674	0.876
13	1357	0.029	0.217	0.753	0.862	0.594	0.879
14	1357	0.027	0.180	0.793	0.883	0.568	0.880
15	1357	0.007	0.149	0.844	0.918	0.512	0.882
16	1357	0.004	0.137	0.859	0.927	0.498	0.883
17	1357	0.010	0.267	0.723	0.856	0.609	0.879

Hoewel de resultaten in Tabel 6.1 er goed uitzien, verdienen de resultaten mogelijk om twee redenen nuancering. Ten eerste kan de dynamiek of het niveau van de groep invloed hebben gehad op de gespreksvaardigheid die individuele leerlingen hebben kunnen tonen. Ten tweede kan de volgorde waarin leerlingen tijdens het spel aan de beurt kwamen van invloed zijn geweest op de gespreksvaardigheid van individuele leerlingen. Bij de afname is zoveel mogelijk geprobeerd om voornoemde problemen te voorkomen:

a. Leerlingen zijn *random* toegewezen aan groepjes;
b. Er was sprake van één “opstartronde” en in totaal kwamen leerlingen gemiddeld acht keer aan de beurt. Op een gegeven moment weten leerlingen zeer waarschijnlijk niet meer wie precies is begonnen.
c. Het initiatief ligt steeds bij degene die de beurt heeft. De rest mag reageren; er is geen sprake van een vaste volgorde.
d. De verschillen tussen iets wel en niet kunnen zijn behoorlijk groot en bij elementen als grammaticabeheersing en zinsconstructie is het onwaarschijnlijk dat een leerling dat in het ene groepje wel kan en in een andere niet.

Om zicht te krijgen op de effectiviteit van voornoemde maatregelen is de analyse op itemniveau ook apart uitgevoerd voor het eerste, tweede, derde en vierde groepslid. Elk van de vier analyses resulteerde in vrijwel identieke uitkomsten. Deze conclusie werd bevestigd door een tweeniveau regressieanalyse met leerlingen genest in scholen, score als afhankelijke variabele en groepslidnummer als voorspeller. Groepsleden 2, 3 en 4 bleken niet significant hoger of lager te scoren dan groepslid 1: (groepslid 2-1) β = -0.657, z = -1.710; (groepslid 3-1) β = -0.638, z = -1.640; en (groepslid 4-1) β = -0.022, z = -0.050.

Een drieniveau regressieanalyse zonder voorspellers liet echter wel zien dat leerlingen binnen groepjes wat meer op elkaar leken dan leerlingen over groepjes heen. De proportie verklaarde variantie op groepjesniveau was gelijk aan 0.122 (zie voor een toelichting op de berekening Davis & Scott, 1995). Deze waarde is aan de hoge kant (als vuistregel wordt vaak aangehouden dat een multilevel analyse zinvol is als

de waarde 0.050 of hoger is, zie Snijders & Bosker, 2012) en kan een aanleiding vormen om in (sommige) analyses het groepsniveau mee te nemen. Echter, doordat het aantal groepjes per school in ons onderzoek zeer klein is (< 30; Maas & Hox, 2005), is dat in ons onderzoek niet zinvol te bewerkstelligen. Voor de beantwoording van onderzoeksvragen op systeemniveau lijkt het groepseffect geen groot probleem te zijn (uitkomsten worden geaggregeerd over groepen), maar in rapportages voor individuele leerlingen is het belangrijk om er bij de interpretatie op bedacht te zijn.

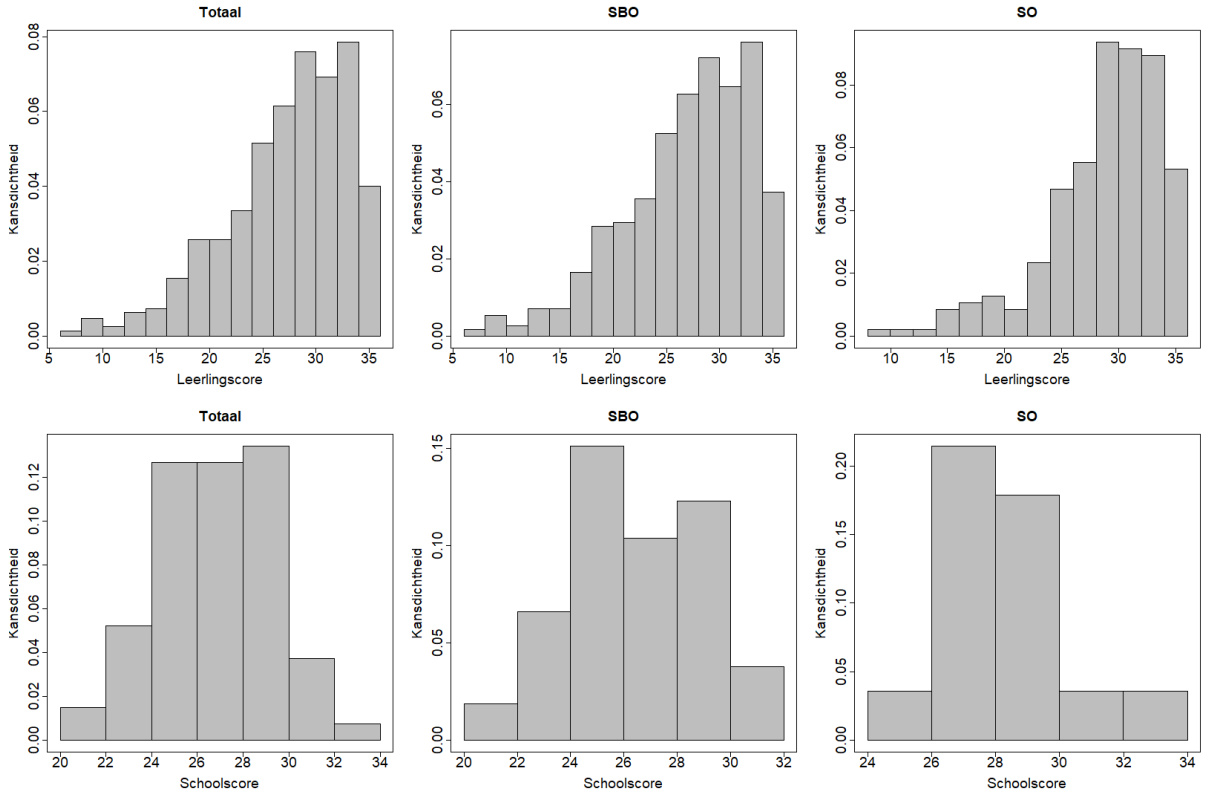
Na de analyse op itemniveau is gekeken naar de betrouwbaarheid van het instrument. Net als eerder bij Spreken is Guttmans lambda2 als uitgangspunt genomen en is daarnaast voor de volledigheid ook Cronbachs alpha en de *greatest lower bound* berekend. De betrouwbaarheidscoëfficiënten zijn te vinden in Tabel 6.2. We zien dat de betrouwbaarheid ver boven de .80 ligt. Dit betekent dat de betrouwbaarheid van het instrumentarium als 'goed' gekwalificeerd kan worden (Evers et al., 2010).

Tabel 6.2. *Toetsindices instrument Gesprekken*

Groepslid	min	max	N	M	SD	skew	kurt	alpha	lambda2	glb
1	0	34	374	27.064	5.895	-1.149	4.089	0.894	0.898	0.955
2	0	34	378	26.405	5.710	-0.851	3.484	0.879	0.883	0.951
3	0	34	362	26.436	5.674	-0.852	3.458	0.877	0.881	0.944
4	0	34	243	27.259	5.683	-0.887	3.070	0.893	0.896	0.954
totaal	0	34	1357	26.748	5.752	-0.938	3.563	0.885	0.889	0.942

Naast betrouwbaarheidscoëfficiënten bevat Tabel 6.2 ook enkele maten die zicht geven op de verdeling van de (ruwe) scores. Het gaat om het gemiddelde, de standaarddeviatie, de *skewness* en de *kurtosis*. We zien onder meer dat leerlingen gemiddeld ongeveer driekwart van het maximaal te behalen aantal punten weten te behalen (26.748 ÷ 34), dat de scores iets linksscheef verdeeld zijn en dat de verdeling iets meer gepiekt is dan de normaalverdeling. Dat beeld wordt uiteraard bevestigd in de visuele weergave van de (ruwe) scoreverdelingen in Figuur 6.1. Wel zien we dat het beeld in het so iets extremer is dan in het sbo. Vermoedelijk wordt dit deels veroorzaakt door het grote verschil in leerlingaantal: 1122 sbo-leerlingen versus 235 so-leerlingen. Op schoolniveau lijkt er sprake van normale verdelingen, maar het verloop is enigszins grillig vanwege het kleine aantal scholen (53 sbo- en 14 so-scholen).

Figuur 6.1. *Scoreverdelingen op leerling- (boven) en schoolniveau (onder) uitgesplitst naar schooltype*

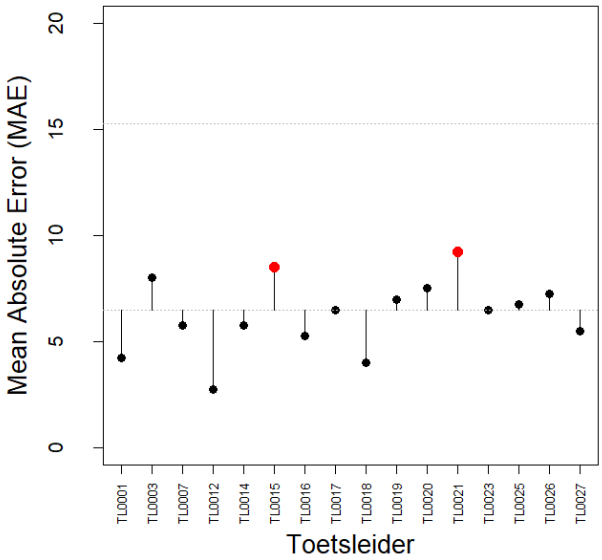


6.4 Beoordelaarsovereenstemming

In de toets- en itemanalyse in paragraaf 6.3 is een objectieve scoring verondersteld. In werkelijkheid hangt de betrouwbaarheid van de scores ook af van de mate waarin de beoordelaars het beoordelingsmodel consistent toepassen. Hier is apart onderzoek naar gedaan, op exact dezelfde manier als eerder bij Spreken (zie paragraaf 5.3 voor een uitgebreide toelichting). Van de 27 toetsleiders konden er 16 meegenomen worden in het onderzoek naar beoordelaarsovereenstemming. Van de overige 11 toetsleiders was geen of onvoldoende bruikbaar filmmateriaal beschikbaar voor een herbeoordeling. Eerst is voor elke toetsleider 'het gebrek aan' overeenstemming met de norm in kaart gebracht door de *Mean Absolute Error* te berekenen. De *Median Absolute Deviation* is vervolgens gebruikt als maat om "afwijkende" beoordelingen op te sporen. Na het vaststellen van het 'gebrek aan' overeenstemming met de norm is de overall kwaliteit van de beoordelingen geëvalueerd door Cohen's weighted kappa coefficient (κ) en Gower's similarity coefficient (G_{xy}) te berekenen.

Figuur 6.2 laat per toetsleider de Mean Absolute Error (MAE) zien. De onderste grijze stippellijn representeert, net zoals eerder bij Spreken, de mediaan van de MAE's van de toetsleiders. De bovenste grijze stippellijn is de mediaan van de MAE's voor een "oneindig" groot aantal ***random*** beoordelingen. Zoals we kunnen zien leken de beoordelingen van toetsleiders TL0001, TL0012 en TL0018 sterk op de beoordelingen van de experts. De beoordelingen van toetsleiders TL0003, TL0015 en TL0021weken daarentegen nogal af van de expertbeoordelingen. De MAE van toetsleiders TL0015 en TL0021 is zelfs zo groot dat er gesproken kan worden van afwijkend gedrag ten opzichte van de andere zestien toetsleiders; de MAD voor deze toetsleider was groter dan 2.5. Cohen's weighted kappa coefficient (κ) and Gower's similarity coefficient (G_{xy}) wijzen op een matige overall kwaliteit: $\kappa = 0.307$ and $G_{xy} = 0.815$, waarbij opvalt dat de eerst maat, net als eerder bij Spreken, een aanzienlijk positiever beeld geeft dan de tweede maat. Gemiddeld gezien maakte het niet uit of de expertbeoordelaars ($M = 26.328$, $SD = 5.252$) of de toetsleiders ($M = 27.078$, $SD = 4.487$) de beoordeling verrichten; $t(126) = -0.869$, $p = 0.387$.

Figuur 6.2. *Mean Absolute Error per toetsleider*

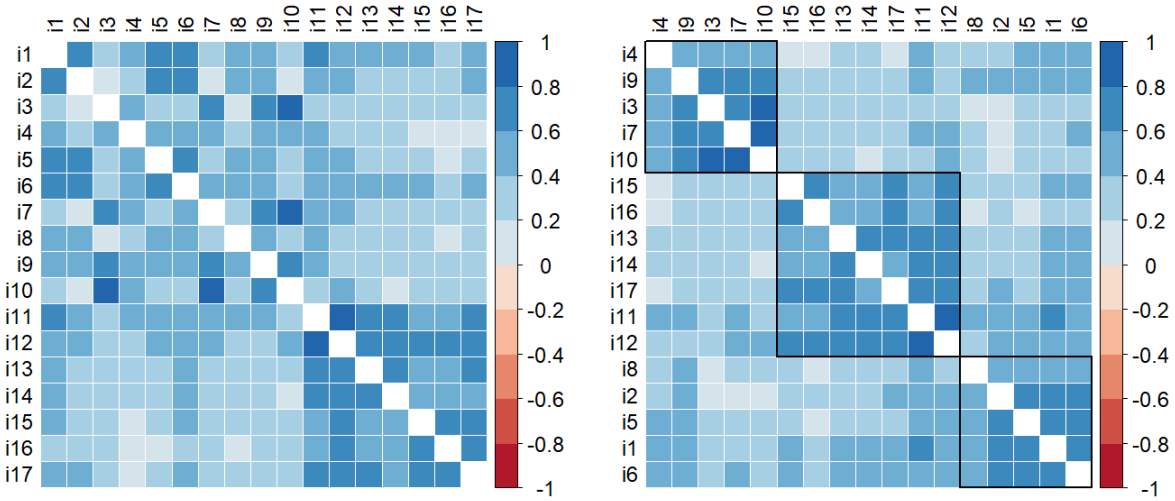


Het is de vraag of TL0015 (54 leerlingen) en TL0021 (24 leerlingen) wel of niet in analyses moet worden meegenomen. Wij hebben ervoor gekozen om de toetsleiders wél mee te nemen, omdat er geen duidelijke aanwijzingen zijn dat de toetsleiders anders beoordeeld hebben dan de experts zouden doen. Bovendien kunnen we net zoals bij Spreken de vraag stellen of een beoordeling "ter plekke" wel exact dezelfde informatie kan opleveren als een beoordeling achteraf op basis van filmopnamen. Het is goed denkbaar dat het spelelement (of de groepsdynamiek) niet altijd goed tot uitdrukking komt in de opnamen en dat kan de beoordeling achteraf bemoeilijken. Als dat het geval is, geven de eerder gepresenteerde resultaten mogelijk een te pessimistisch beeld van de werkelijkheid, en is er des te meer reden om huiverig te zijn voor het weglaten van toetsleiders (of leerlingen).

6.5 Dimensionaliteit

Het beoordelingsmodel omvat verschillende aspecten van gespreksvaardigheid, en hoewel er in paragraaf 6.3 vanuit is gegaan dat deze aspecten betekenisvol samen te voegen zijn tot één totaalscore, is het belangrijk om inzicht te hebben in de dimensionele structuur van de vaardigheid. De dimensionele structuur is eerst op itemniveau in kaart gebracht. Figuur 6.3 geeft een visuele representatie van de matrix met polychorische inter-itemcorrelaties zoals berekend op basis van de itemscores van 1357 leerlingen. In Figuur 6.3 is te zien dat de items allemaal positief met elkaar samenhangen (links) en dat een verkennende hiërarchische clusteranalyse (rechts) de correlaties tussen de itemparen ordent in drie verschillende clusters met items die (relatief) sterk met elkaar samenhangen.

Figuur 6.3. *Visuele representatie polychorische inter-itemcorrelatiematrix Gesprekken*



In aansluiting op de berekening van de inter-itemcorrelaties is op de polychorische correlatiematrix een exploratieve factoranalyse met varimax rotatie uitgevoerd. Een parallelanalyse suggereerde om drie factoren te kiezen en aangezien het correlogram ook in die richting wees, is die suggestie overgenomen. De uitkomsten van de factoranalyse zijn te vinden in Tabel 6.3.

Tabel 6.3. *Resultaten exploratieve factoranalyse*

		Factor 1 Taalvorm	Factor 2 Taalinhoud	Factor 3 Taalgebruik	H2	U2
i3	De leerling neemt doorgaans op passende wijze het woord.			0.870	0.790	0.210
i10	De leerling geeft consequent gepaste reacties (verbaal/nonverbaal).			0.910	0.880	0.120
i7	De leerling gebruikt consequent taal die bij de situatie past.			0.840	0.800	0.200
i9	De leerling vertoont adequaat actief luistergedrag.		0.470	0.630	0.660	0.340
i4	De leerling integreert bijdragen van de gesprekspartners in zijn eigen inbreng wanneer dat toegevoegde waarde heeft.		0.520	0.480	0.510	0.490
i8	Het non-verbale gedrag van de leerling zet zijn verbale boodschap kracht bij.		0.560		0.380	0.620
i6	De leerling maakt de eigen gedachtegang begrijpelijk.	0.400	0.680		0.690	0.310
i1	De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten sterk.	0.350	0.720		0.680	0.320
i2	De leerling neemt regelmatig initiatief om het gesprek te beginnen, gaande te houden en/of af te sluiten (bijvoorbeeld door vragen te stellen of voorstellen te doen).		0.830		0.740	0.260
i5	De leerling neemt initiatief om een gezamenlijk gespreksdoel te behalen door de gesprekspartners te betrekken bij het gesprek, bijvoorbeeld met behulp van vragen.		0.850		0.800	0.200
i11	De inbreng van de leerling vertoont voldoende variatie in woordgebruik.	0.680	0.480		0.740	0.260
i12	De getoonde woordenschat van de leerling is toereikend om het gesprek te voeren.	0.800	0.350		0.790	0.210
i13	De leerling spreekt redelijk vloeiend, al vertoont zijn uitwerking een enkele hapering en/of herformulering.	0.790			0.690	0.310
i14	De leerling is goed verstaanbaar door zijn uitspraak, articulatie en intonatie, ondanks een eventueel accent.	0.730			0.610	0.390
i15	De leerling vervoegt alle werkwoorden over het algemeen correct.	0.740			0.590	0.410
i16	De leerling verbuigt (combinaties met) naamwoorden meestal correct.	0.800			0.690	0.310
i17	De leerling maakt over het algemeen correcte enkelvoudige en samengestelde zinnen.	0.750			0.650	0.350

Voor de leesbaarheid zijn de ladingen kleiner dan 0.300 niet vermeld in Tabel 6.3. De uitkomsten stemmen volledig overeen met het beeld dat Figuur 6.3 gaf en zijn vrij gemakkelijk inhoudelijk te duiden. Net zoals bij Spreken lijkt er een driedeling te bestaan naar Taalvorm, Taalinhoud en Taalgebruik (zie voor een omschrijving paragraaf 5.5), al zijn er ook enkele items die minder gemakkelijk in te delen zijn. Dat is vooral het geval bij item 4 (De leerling integreert bijdragen van de gesprekspartners in zijn eigen inbreng wanneer dat toegevoegde waarde heeft) en item 9 (De leerling vertoont adequaat actief luistergedrag). Waar deze items theoretisch gezien vermoedelijk het sterkste een beroep doen op Taalgebruik, laat de factoranalyse duidelijk zien dat de items ook een beroep doen op Taalinhoud. Desalniettemin kan geconcludeerd worden dat het instrumentarium Gesprekken uiteen te leggen is in drie samenhangende factoren.

6.6 Extra onderdeel: Luisteren in een praktijksituatie

Het beoordelingsmodel Gesprekken bevatte ook een aantal aspecten die betrekking hadden op Luisteren. Het ging om onderstaande beoordelingsaspecten die op een driepuntsschaal werden gescoord (0 = lage vaardigheid, 2 = hoge vaardigheid). Ter illustratie is bij elk beoordelingsaspect de omschrijving van de hoogste beoordelingscategorie en een voorbeeldvraag weergegeven.

Begrijpen

1. Instructie toepassen
   - De leerling past de instructie van de leerkracht over het algemeen correct toe.
     Voorbeeldvraag: Leg je boot bij de haven met dezelfde kleur als je boot;
     Mag de boot hier varen?
2. Herhaling van instructie
   - De leerling begrijpt de instructie van de leerkracht in één keer.
     Voorbeeldvraag: Je hebt twee gegooid; Waar mag je dan allemaal heen varen?

Samenvatten

3. Mondeling samenvatten
   - De leerling kan een verhaal goed in grote lijnen navertellen.
     Voorbeeldvraag: Welke regels zijn er voor het brengen van vissen naar de haven?

Interpreteren

4. Meningen herkennen
   - De leerling benoemt de verschillende meningen over een onderwerp correct.
     Voorbeeldvraag: Na een spelronde - Welke meningen heb je gehoord over het weggeven of houden van de worp?

Deze beoordelingsaspecten zijn opgenomen, omdat in diverse onderzoeken die door Cito in het sbo en so zijn uitgevoerd door leerkrachten werd aangegeven dat zij het opvolgen van een instructie een belangrijk onderdeel vinden van luistervaardigheid. Ook gaven leerkrachten aan dat leerlingen niet altijd tot hun recht komen in een standaard toetssituatie met pen en papier. Door gericht vragen te stellen aan leerlingen tijdens het *Fischerspiel* kan worden nagegaan of leerlingen geluisterd hebben naar de instructie. Daarnaast kan door het stellen van vragen vastgesteld worden of een leerling de meningen van medespelers correct heeft begrepen. Aan alle vier leerlingen in het groepje zijn alle vier typen vragen gesteld. Het moment waarop dat gebeurde varieerde. De ene leerling kreeg een bepaalde vraag aan het begin van het spel, terwijl een andere leerling diezelfde vraag aan het einde van het spel kon krijgen.

Het extra onderdeel is als losse “toets” geanalyseerd binnen het raamwerk van de klassieke testtheorie. In Tabel 6.4 zijn de resultaten op itemniveau te vinden. We zien dat er in plaats van 1357 leerlingen (zie Tabel 6.1) er slechts 1346 konden worden meegenomen in de analyse, omdat bij 11 leerlingen helemaal niets was ingevuld. Bij de leerlingen die wel meegenomen zijn, was er af en toe sprake van een missende waarde. Deze waarden zijn geïmputeerd volgende de eerder beschreven werkwijze. De beoordelingsaspecten (of items) hebben verder goed gefunctioneerd. De *p*-waarden schommelen tussen 0.639 en 0.798 en de r_{it}-waarden zijn hoog te noemen. Hoewel het aantal items erg klein is lijkt de set als geheel wel een zinvolle indicatie te geven van de luistervaardigheid van leerlingen in praktijksituaties.

Tabel 6.4. *Itemindices onderdeel Luisteren bij het instrument Gesprekken*

Item	n	c0	c1	c2	*p*	r_{it}	alpha
1	1346	0.046	0.258	0.696	0.825	0.798	0.677
2	1346	0.053	0.425	0.522	0.735	0.793	0.682
3	1346	0.108	0.451	0.441	0.667	0.769	0.703
4	1346	0.110	0.503	0.387	0.639	0.697	0.760

Deze conclusie wordt bevestigd door de analyses die op toetsniveau hebben plaatsgevonden. In Tabel 6.5 is te zien dat de leerlingen behoorlijk hoge scores behaalden (ongeveer driekwart van het maximaal haalbare aantal punten) en dat de betrouwbaarheid als goed gekwalificeerd kan worden. Bovendien lijkt er geen relatie te bestaan tussen het groepslidnummer en de behaalde score. Dit neemt niet weg dat er een relatie kan bestaan tussen het moment van het stellen van de vraag en de moeilijkheid van de vraag. Naar dit aspect kan echter geen onderzoek gedaan worden, omdat toetsleiders tijdens het afnemen van het *Fischerspiel* niet systematisch hebben bijgehouden welke vraag wanneer gesteld is aan welke leerling. Dit was niet haalbaar en bovendien kan bij onderzoek op systeemniveau de vraag gesteld worden in hoeverre een eventuele relatie problematisch is. Zolang leerlingen niet systematisch bevoordeeld of benadeeld worden (wat in dit onderzoek ook niet het geval was) lijken geaggregeerde uitkomsten goed bruikbaar te zijn; het is met name in rapportages voor individuele leerlingen dat voorzichtigheid geboden is.

Tabel 6.5. *Toetsindices onderdeel Luisteren bij het instrument Gesprekken*

Groepslid	min	max	N	M	SD	skew	kurt	alpha	lambda2	glb
1	0	8	370	5.727	1.898	-0.786	3.319	0.767	0.745	0.792
2	0	8	374	5.580	1.976	-0.617	2.721	0.789	0.766	0.824
3	0	8	360	5.656	1.840	-0.535	2.648	0.737	0.720	0.806
4	0	8	242	6.079	1.756	-0.889	3.347	0.738	0.720	0.806
totaal	0	8	1346	5.730	1.886	-0.694	2.964	0.763	0.741	0.805

Hoewel onderzoek naar de luistervaardigheid van leerlingen in praktische situaties wetenschappelijk gezien zinvol is, is het de vraag hoe deze informatie het beste kan worden meegenomen bij het doen van uitspraken over het referentieniveau van leerlingen. Er kan immers maar één referentieniveau worden toegewezen aan een leerling en dit wordt (in de regel) gebaseerd op één score. Theoretisch gezien zijn er wel mogelijkheden om tot één score te komen die gebaseerd is op een (gecombineerde) pen-en-papier afname en een spelafname. De items zouden bijvoorbeeld meegenomen kunnen worden bij de IRT-modellering (zie paragraaf 4.4). In het design voegen we dan in feite vier regels (items) toe die van toepassing zijn op alle leerlingen ongeacht hun luistertoetsversie. Hoewel deze optie aantrekkelijk lijkt, omdat we na het schatten van het model direct per leerling één vaardigheidsscore kunnen bepalen waarin zowel het pen-en-papier als het spel-gedeelte zijn meegenomen, lijkt deze toch weinig kansrijk. We “vermengen” namelijk twee vaardigheden (terwijl we tegelijkertijd in een IRT-model expliciet unidimensionaliteit veronderstellen) en hebben naast dichotome scoring te maken met een (lastigere) polytome scoring. Ook het werken met een zogeheten *composite score* lijkt niet ideaal, omdat de beschikbare prestatiestandaarden voor 1F en 2F dan alleen gebruikt kunnen worden als allerlei extra aannames worden gedaan. Om deze redenen is ervoor gekozen om de praktijkscore voor luisteren in een aparte variabele op te nemen (**GSP_RSCO_LUI**) en die alleen te gebruiken bij het duiden en verklaren van leerlingprestaties (zie paragraaf 8.2). Bij het bepalen van het behaalde referentieniveau (zie paragraaf 8.1.1) is deze score niet gebruikt.

6.7 Variabelen

In het databestand zijn bij Gesprekken de volgende variabelen opgenomen:

GROEPSLID	Groepslidnummer bij Gesprekken
GRP_ID	Groepsidentificatie bij Gesprekken
GSP_TTLID	Toetsleideridentificatie bij Gesprekken
N_SPELERS	Aantal spelers bij Gesprekken
GSP_MISS	Percentage ontbrekende antwoorden
GSP_RSCO	Toetsscore
GSP_MISS_LUI	Percentage ontbrekende antwoorden bij het extra onderdeel Luisteren
GSP_RSCO_LUI	Toetsscore bij het extra onderdeel Luisteren
GSP_REFN	Referentieniveau (<1F/1F/2F)
GSP_WEIGHT	Gewicht om te corrigeren voor leerroute

Bij de toetsscore- variabele is het de vraag hoe deze het beste tot stand kan komen. Het is het gemakkelijkst om de toetsscore voor een leerling gelijk te stellen aan de som van alle itemscores, maar in een recent peilingsonderzoek in groep 8 (zie Van Langen et al., 2017) zijn de meer technische beoordelingsaspecten (i11 tot en met i17) minder zwaar gewogen dan de meer inhoudelijke beoordelingsaspecten (i1 tot en met i10). Die keuze is in dit onderzoek overgenomen. Dit betekent concreet dat bij de berekening van de toetsscore aan items 1 tot en met 10 gewicht 1.360 is toegekend ([34 scorepunten * 0.800] ÷ [10 items × 2 scorepunten]) en aan items 11 tot en met 17 gewicht 0.486 ([34 scorepunten * 0.200] ÷ [7 items × 2 scorepunten]). Het gevolg daarvan is dat “inhoud” 80 procent meetelt en “techniek” 20 procent. Alle toetsscores zijn afgerond naar gehele getallen. Over de wijze waarop het referentieniveau (**GSP_REFN**) van leerlingen is bepaald wordt in paragraaf 8.1.1 gerapporteerd. Het percentage ontbrekende antwoorden (**GSP_MISS** en **GSP_MISS_LUI**) is berekend door het aantal itemscores van een leerling te delen door het totale aantal items (17 of 4).

7 Interview

7.1 Inhoud en afname

Na afloop van de dataverzameling is per school, als afsluiting, een gesprek gevoerd met een vertegenwoordiging van de school. Het gesprek is met behulp van onderstaande lijst verwerkt tot variabelen. De variabelen zijn zowel voor luisteren als voor spreken en gesprekken aangemaakt (met uitzondering van het onderdeel Teamontwikkeling), omdat het onderwijs bij de ene vaardigheid mogelijk anders is ingericht dan bij de andere vaardigheid. Bij het aanmaken van het databestand viel op dat bij veel vragen de scores gelijk zijn voor de drie vaardigheden. In die gevallen zou slechts één variabele behouden kunnen worden, maar hier is niet voor gekozen, omdat het dan net lijkt alsof er geen rekening is gehouden met mogelijke verschillen tussen de drie vaardigheden. Bovendien zouden de variabelennamen binnen een onderwerp inconsistent worden. Om deze redenen zijn alle variabelen altijd per vaardigheid behouden, ook als het onderscheid tussen de drie vaardigheden redundant bleek te zijn.

Onderwijsaanbod en methodegebruik		
1	Wordt een methode gebruikt?	Nee-Ja
	Vrije invoer voor methode	---
2	Wordt de methode volledig gevolgd?	Nee-Ja
3	Zijn er afspraken over het gebruik van de methode?	Nee-Ja
4	Worden er aanvullende materialen gebruikt naast de (hoofd)methode?	Nee-Ja
	Vrije invoer voor aanvullende materialen	---
5	Wordt het onderwijs in X ingebed in de taalles? (bijvoorbeeld bij begrijpend lezen)	Nee-Ja
6	Wordt het onderwijs in X geïntegreerd in andere vakken? (bijvoorbeeld geschiedenis)	Nee-Ja
Doelgerichtheid en planmatigheid		
7	Heeft de school doelen opgesteld voor X?	Nee-Ja
8	Wordt gebruikgemaakt van een leerlijn?	Nee-Ja
	Vrije invoer voor leerlijnen	---
9	Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)	Nee-Ja
10	Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)	Nee-Ja
11	Wordt gebruikgemaakt van herhaald aanbod met feedback?	Nee-Ja
12	Wordt gebruikgemaakt van referentieniveaus?	Nee-Ja
13	Zijn er einddoelen voor schoolverlaters geformuleerd?	Nee-Ja
Onderwijstijd		
14	Wat is de hoeveelheid expliciete onderwijstijd op het rooster?	… Min
Monitoring en planning		
15	Worden leervorderingen in de lessen geregistreerd?	Nee-Ja
16	Worden leervorderingen in het schooljaar geregistreerd?	Nee-Ja
17	Wordt gebruikgemaakt van een leerlingadministratiepakket?	Nee-Ja
18	Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?	Nee-Ja
19	Wordt gebruikgemaakt van methodetoetsen?	Nee-Ja
20	Wordt gebruikgemaakt van zelfgemaakte instrumenten?	Nee-Ja
21	Wordt in de planningscyclus gebruikgemaakt van de leerresultaten van leerlingen?	Nee-Ja
22	Wordt gebruikgemaakt van groepsplannen?	Nee-Ja
Differentiatie		
23	Vindt er convergente differentiatie plaats?	Nee-Ja
24	Vindt er divergente differentiatie plaats?	Nee-Ja
25	Wordt er gewerkt met (twee of meer) niveaugroepen?	Nee-Ja
26	Is er differentiatie naar instructie?	Nee-Ja
27	Is er differentiatie naar leerstof ?	Nee-Ja
28	Is er differentiatie naar (streef-)doelen?	Nee-Ja
29	Zijn er leerlingen die extra ondersteuning krijgen door een specialist ***binnen*** de school?	Nee-Ja
	Vrije invoer specialist binnen de school	---

30	Zijn er leerlingen die extra ondersteuning krijgen door een specialist ***buiten*** de school?	Nee-Ja
	Vrije invoer specialist buiten de school	---
31	Zijn er leerlingen die extra ondersteuning krijgen op basis van een specifiek (interventie)programma?	Nee-Ja
	Vrije invoer over specifiek (interventie)programma	---

Teamontwikkeling

32	Is mondelinge taalvaardigheid benoemd in het school- of taalbeleidsplan?	Nee-Ja
33	Zijn er intervisiemomenten of wordt collegiale consultatie toegepast?	Nee-Ja
	Vrije invoer over intervisie en collegiale consultatie	---
34	Is er interne ondersteuning? (Bijvoorbeeld een taalcoördinator)	Nee-Ja
	Vrije invoer over interne ondersteuning	---
35	Is er externe ondersteuning? (Bijvoorbeeld een onderwijsadviseur)	Nee-Ja
	Vrije invoer over externe ondersteuning	---

Naast bovengenoemde vragen zijn ook enkele verdiepende vragen gesteld op basis van checklists. Ten eerste is bij scholen nagegaan welke activiteiten plaatsvinden op het gebied van luisteren, spreken en gesprekken. Voor Luisteren is de volgende checklist ingevuld:

- ☐ Kijken en luisteren naar tv-fragment (bijv. jeugdjournaal) en beantwoorden van vragen
- ☐ Luisterboek en beantwoorden van vragen
- ☐ Coöperatieve luisteropdrachten
- ☐ Voorlezen van boek en beantwoorden van vragen
- ☐ Naar gesproken tekst (tv/live) luisteren en aantekeningen maken
- ☐ Naar gesproken tekst (tv/live) luisteren en samenvatting maken
- ☐ Naar gesproken tekst (tv/live) luisteren en mening/oordeel/standpunt bepalen
- ☐ Instructies (praktisch) beluisteren en ze vervolgens uitvoeren
- ☐ Anders, namelijk _____

En voor Spreken en Gesprekken is gebruikgemaakt van onderstaande checklist:

- ☐ Spreekbeurten
- ☐ Kring/groepsgesprekken
- ☐ Verslag van persoonlijke belevenis
- ☐ Dialogen
- ☐ Samenwerkend leren
- ☐ Discussies
- ☐ (Zakelijke) boodschappen oefenen
- ☐ Verwoorden van meningen
- ☐ Verhaal vertellen volgens gespreksregels
- ☐ Deelname aan een weekafsluiting/toneelstuk/…
- ☐ Anders, namelijk _____

Ten tweede is bij Teamontwikkeling in kaart gebracht welke inhoudelijke onderwerpen zoal aan bod komen binnen het teamoverleg. Daarbij is de volgende checklist gebruikt:

- ☐ Gebruik van de methode voor mondelinge taalvaardigheid
- ☐ Onderwijs in mondelinge taalvaardigheid bij andere vakken dan taal
- ☐ Volgen van de ontwikkeling van leerlingen op het gebied van de mondelinge taalvaardigheden
- ☐ Uitkomsten van volgsysteem/registraties op het gebied van mondelinge taalvaardigheden
- ☐ Individuele leerlingen met spraak-taalstoornis en de ondersteuning daarvan door de leerkracht
- ☐ Individuele leerlingen met spraak-taalstoornis en de ondersteuning daarvan door de externe specialist
- ☐ Aanpak/begeleiding in de groep van leerlingen die blijkens het volgsysteem een achterstand hebben op het gebied van mondelinge taalvaardigheden
- ☐ De leerdoelen van de school op het gebied van mondelinge taalvaardigheden
- ☐ Vakdidactiek mondelinge taalvaardigheden
- ☐ Anders, namelijk _____

De gesprekken met de scholen namen in de regel een half uur tot een uur in beslag en zijn waar mogelijk opgenomen met een voicerecorder. In totaal zijn 62 scholen geïnterviewd. De geïnterviewden waren vooral leerkrachten of intern begeleiders. Bij zes scholen kon geen interview worden afgenomen, omdat zij daar niet toe bereid waren.

In paragrafen 7.2 tot en met 7.7 worden de variabelen uitgebreid beschreven. In de bijschrijving worden sbo en so samengenomen, omdat de focus hier ligt op het functioneren van het instrument; bij de inhoudelijke duiding van de uitkomsten in Hoofdstuk 8 worden het sbo en so wel los van elkaar beschouwd.

7.2 Onderwijsaanbod en methodegebruik

Bij dit onderdeel ging het om de vraag hoe gevarieerd het onderwijs in mondelinge taalvaardigheid op een school is en hoe het wordt ingebed in het onderwijsprogramma. Onderstaand overzicht laat zien welke onderwerpen tijdens het gesprek aan de orde zijn gesteld en onder welke variabelen de onderwerpen zijn terug te vinden in het databestand. Daarnaast is te zien op hoeveel scholen een onderwerp wel of niet van toepassing was. Van de 62 geïnterviewde scholen maken er 58 gebruik van een methode. Alle scholen gebruiken één en dezelfde methode voor alle drie vaardigheden. Uit het vrije invoerveld is af te leiden dat *Staal* en *Taalactief* (beiden op 16 scholen) het meest gebruikt worden, gevolgd door *Taal op maat* (11 scholen) en *Taalverhaal* (8 scholen). Scholen volgen de methode echter lang niet allemaal volledig (30 scholen doen dat niet) en op veel scholen zijn specifieke afspraken gemaakt over het gebruik van de methode (42 scholen in totaal). Bij een paar scholen ontbreekt informatie. In dat geval is code 999 toegekend.

Vraag	Variabele	Nee (0)	Ja (1)	Missend
Wordt een methode gebruikt?	INT_AANB_V1_gsp	4	58	0
	INT_AANB_V1_lui	4	58	0
	INT_AANB_V1_spr	4	58	0
Vrije invoer voor methode	*INT_AANB_V1_namelijk*			
Wordt de methode volledig gevolgd?	INT_AANB_V2_gsp	30	31	1
	INT_AANB_V2_lui	30	31	1
	INT_AANB_V2_spr	30	31	1
Zijn er afspraken over het gebruik van de methode?	INT_AANB_V3_gsp	17	42	3
	INT_AANB_V3_lui	17	42	3
	INT_AANB_V3_spr	17	42	3
Worden er aanvullende materialen gebruikt naast de hoofdmethode?	INT_AANB_V4_gsp	19	40	3
	INT_AANB_V4_lui	19	40	3
	INT_AANB_V4_spr	19	40	3
Vrije invoer voor aanvullende materialen	*INT_AANB_V4_namelijk*			
Wordt het onderwijs in X ingebed in de taalles?	INT_AANB_V5_gsp	27	32	3
	INT_AANB_V5_lui	26	33	3
	INT_AANB_V5_spr	26	33	3
Wordt het onderwijs in X geïntegreerd in andere vakken?	INT_AANB_V6_gsp	24	35	3
	INT_AANB_V6_lui	23	36	3
	INT_AANB_V6_spr	23	36	3

Een vervolgvraag over het onderwijsaanbod was of er wel (1) of geen (0) aanvullende materialen worden gebruikt naast de (hoofd)methode. Het kon bijvoorbeeld gaan om remediërend materiaal, activiteiten uit andere methoden, of (*online*) studiematerialen of *apps*. Op 40 van de 62 scholen wordt aanvullend materiaal gebruikt naast de methode, en op 19 scholen niet; voor 3 scholen ontbreekt het antwoord (code 999). Uit het vrije invoerveld is af te leiden welke aanvullende variabelen vaak gebruikt worden. Scholen gaven onder meer aan gebruik te maken van *Snappet*, *Nieuwsbegrip*, spelletjes, woordoefeningen en logopedisch materiaal. Luister-, spreek- en gesprekstaken worden dikwijls gekoppeld aan andere taken.

Een derde aspect van het onderwijsaanbod waar in de gesprekken naar gevraagd is, betreft de wijze waarop de vaardigheden luisteren, spreken en gesprekken ingebed worden in het onderwijs. De variabele **INT_AANB_V5** geeft aan of het onderwijs wordt ingebed in de taalles (bijvoorbeeld in combinatie met begrijpend lezen), en **INT_AANB_V6** of het wordt ingebed in andere vakken zoals geschiedenis. Bij elke vaardigheid geldt voor tenminste 16 scholen dat het onderwijs zowel in de taalles als in andere vakken wordt ingebed. Daarnaast is voor een aantal scholen (6 tot 8, afhankelijk van de vaardigheid) bij beide vragen *Nee* ingevuld.

Voorts is op basis van twee checklists in kaart gebracht welke luister- en spreekactiviteiten worden gebruikt in de lessen (zie ook paragraaf 7.1). Onderstaand overzicht laat zien welke activiteiten scholen gebruiken. Tevens is in het overzicht te zien onder welke variabele de activiteit is terug te vinden in het databestand. Als er activiteiten ter sprake kwamen die niet leken te vallen onder een van de omschrijvingen op de checklist, dan zijn deze apart genoteerd en ondergebracht in variabele **INT_AANB_ACT9_lui_namelijk** (in geval het om luisteren gaat) of **INT_AANB_ACT11_spr_namelijk** (in geval het om spreken/gesprekken

gaat). Voorbeelden van dergelijke activiteiten zijn boekbesprekingen en dramalessen. Van de 53 scholen waarvoor de checklist is ingevuld, gebruikt ongeveer de helft minimaal 5 van de 8 genoemde luisteractiviteiten en minimaal 7 van de 10 spreekactiviteiten. De meest gebruikte luisteractiviteit is het beantwoorden van vragen bij een tv-fragment en de meest gebruikte spreekactiviteit kring-groepsgesprekken. De minst gebruikte activiteiten van het lijstje bij luisteren zijn degenen waar de leerlingen schriftelijk werk moeten verrichten, zoals een samenvatting of aantekeningen maken. Bij spreken is het oefenen van zakelijke boodschappen of het vertellen van een verhaal volgens gespreksregels minder gebruikelijk dan de andere activiteiten.

Luisteractiviteit	**Variabele**	***Nee* (0)**	***Ja* (1)**	**Missend**
Kijken en luisteren naar tv-fragment (bijv. jeugdjournaal) en vragen beantwoorden	INT_AANB_ACT1_lui	5	48	9
Luisterboek en vragen beantwoorden	INT_AANB_ACT2_lui	27	26	9
Coöperatieve luisteropdrachten	INT_AANB_ACT3_lui	26	27	9
Voorlezen van boek en vragen beantwoorden	INT_AANB_ACT4_lui	8	45	9
Naar gesproken tekst (tv/live) luisteren en aantekeningen maken	INT_AANB_ACT5_lui	32	21	9
Naar gesproken tekst (tv/live) luisteren en samenvatting maken	INT_AANB_ACT6_lui	34	19	9
Naar gesproken tekst (tv/live) luisteren en mening/oordeel/standpunt bepalen	INT_AANB_ACT7_lui	18	35	9
Instructies (praktisch) beluisteren en ze vervolgens uitvoeren	INT_AANB_ACT8_lui	11	42	9
Vrije invoer voor andere activiteiten	*INT_AANB_ACT9_lui_namelijk*			

Spreekactiviteit	**Variabele**	**Nee (0)**	**Ja (1)**	**Missend**
Spreekbeurten	INT_AANB_ACT1_spr	8	45	9
Kring-/groepsgesprekken	INT_AANB_ACT2_spr	1	52	9
Verslag van persoonlijke belevenis	INT_AANB_ACT3_spr	10	43	9
Dialogen	INT_AANB_ACT4_spr	18	35	9
Samenwerkend leren	INT_AANB_ACT5_spr	10	43	9
Discussies	INT_AANB_ACT6_spr	15	38	9
(Zakelijke) boodschappen oefenen	INT_AANB_ACT7_spr	30	22	10
Verwoorden van meningen	INT_AANB_ACT8_spr	7	46	9
Verhaal vertellen volgens gespreksregels	INT_AANB_ACT9_spr	24	29	9
Deelname aan een weekafsluiting/toneelstuk/…	INT_AANB_ACT10_spr	22	31	9
Vrije invoer voor andere Spreekactiviteiten	INT_AANB_ACT11_spr_namelijk			

7.3 Doelgerichtheid en planmatigheid

Bij dit onderdeel ging het om de vraag hoe doelgericht en planmatig het onderwijs in mondelinge taalvaardigheid wordt ingevuld. Onderstaand overzicht laat zien welke onderwerpen tijdens het gesprek aan de orde zijn gesteld en onder welke variabelen de onderwerpen zijn terug te vinden in het databestand. Daarnaast is te zien op hoeveel scholen een onderwerp wel of niet van toepassing was. Er is onder meer gevraagd of de school expliciet doelen heeft opgesteld voor Luisteren, Spreken of Gesprekken. We zien dat er bij behoorlijk wat scholen (bij ongeveer een derde) geen expliciet beleid is waarbij zelf doelen zijn geformuleerd. Dat heeft niet automatisch tot gevolg dat ook in de dagelijkse lespraktijk niet aan leerdoelen wordt gewerkt. Op het overgrote deel van de scholen wordt namelijk wél gebruikgemaakt van een leerlijn (zie **INT_DOEL_V2**). Zoals blijkt uit het vrije invoerveld wordt de leerlijn van de lesmethode verreweg het vaakst gebruikt (> 30 keer) en soms worden ook de tussendoelen van SLO gebruikt (13 keer). De leerlijn van CED wordt nauwelijks gebruikt (6 keer) en er zijn ook bijna geen scholen die een eigen leerlijn hebben; slechts twee scholen hebben aangegeven dat te hebben.

Tijdens het gesprek is geprobeerd om ook verfijnder te weten te komen hoe scholen invulling geven aan mondelinge taalvaardigheid. Uit de derde vraag blijkt dat vrij veel scholen (43-44) doelgericht proberen te werken door aan elk doel concrete taalactiviteiten te koppelen (zie **INT_DOEL_V3**). Ook het expliciet

koppelen van taalactiviteiten aan doelen gebeurt vaak, namelijk op 39 tot 40 scholen (zie **INT_DOEL_V4**). Hieruit blijkt dat vrij veel scholen (29-31) zowel doelgericht als activiteitgericht werken. Scholen koppelen bijvoorbeeld eerst een doel aan activiteiten binnen een thema en voeren de activiteiten uit waarbij ze rekening houden met het doel. Herhaald aanbod met feedback is eveneens een methodiek die vaak (47-48) wordt toegepast (zie **NT_DOEL_V5**). Er wordt iets minder gebruikgemaakt van de referentieniveaus en ook zijn er relatief veel scholen die géén einddoelen voor schoolverlaters hebben geformuleerd.

Vraag	**Variabele**	***Nee* (0)**	***Ja* (1)**	**Missend**
Heeft de school (expliciet) doelen voor X?	INT_DOEL_V1_gsp	21	40	1
	INT_DOEL_V1_lui	20	41	1
	INT_DOEL_V1_spr	20	41	1
Wordt gebruikgemaakt van een leerlijn?	INT_DOEL_V2_gsp	6	52	4
	INT_DOEL_V2_lui	4	54	4
	INT_DOEL_V2_spr	5	53	4
Vrije invoer voor leerlijn	*INT_DOEL_V2_namelijk*			
Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)	INT_DOEL_V3_gsp	16	44	2
	INT_DOEL_V3_lui	17	43	2
	INT_DOEL_V3_spr	16	44	2
Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)	INT_DOEL_V4_gsp	19	40	3
	INT_DOEL_V4_lui	20	39	3
	INT_DOEL_V4_spr	19	40	3
Wordt gebruikgemaakt van herhaald aanbod met feedback?	INT_DOEL_V5_gsp	13	47	2
	INT_DOEL_V5_lui	13	47	2
	INT_DOEL_V5_spr	12	48	2
Wordt gebruikgemaakt van referentieniveaus?	INT_DOEL_V6_gsp	20	40	2
	INT_DOEL_V6_lui	20	40	2
	INT_DOEL_V6_spr	19	40	3
Zijn er einddoelen voor schoolverlaters geformuleerd?	INT_DOEL_V7_gsp	23	37	2

7.4 Onderwijstijd

Bij het interview is geprobeerd om zicht te krijgen op de hoeveel tijd die scholen besteden aan onderwijs in mondelinge taalvaardigheden. De vertegenwoordigers van de scholen gaven echter aan deze vraag niet goed te kunnen beantwoorden of alleen een hele grove inschatting te kunnen maken. Daardoor is voor slechts 4 van de 62 scholen een tijd per vaardigheid ingevuld, en bij 3 van deze scholen is exact hetzelfde getal ingevuld voor de drie vaardigheden. Bij 19 andere scholen is alleen de totale tijd ingevuld die wordt besteed aan mondelinge taalvaardigheid, uiteenlopend van 10 minuten tot 10 uur per week. Van de overige 39 scholen verkregen we geen enkele indicatie van de onderwijstijd. Scholen kunnen niet expliciet benoemen hoeveel tijd ze aan mondelinge taalvaardigheid besteden. Redenen die genoemd worden zijn dat het geïntegreerd wordt in het (taal)onderwijs, er zijn geen vast roostertijden, activiteiten rondom mondelinge taalvaardigheid worden bijvoorbeeld gekoppeld aan begrijpend lezen. Vanwege het grote aantal missende waarden en de onbetrouwbaarheid van de waarden die wel zijn genoemd, is er in het databestand geen variabele over onderwijstijd opgenomen.

7.5 Monitoring en planning

Bij dit onderdeel ging het om de vraag in hoeverre scholen leervorderingen in mondelinge taalvaardigheid volgen en aan de hand hiervan het onderwijs plannen. Onderstaand overzicht laat zien welke onderwerpen tijdens het gesprek aan de orde zijn gesteld en onder welke variabelen de onderwerpen zijn terug te vinden in het databestand. Daarnaast is te zien op hoeveel scholen een onderwerp wel of niet van toepassing was. De eerste twee vragen brengen algemeen in kaart of scholen de leervorderingen van hun leerlingen op het gebied van mondelinge taalvaardigheid registreren. Minder dan de helft (24-25) van de scholen houdt de leervorderingen in de lessen bij (**INT_VOLG_V1**), en hetzelfde geldt voor het systematisch bijhouden van de leervorderingen in de loop van het schooljaar (**INT_VOLG_V2**), wat op 28-30 scholen gebeurt. Als we kijken voor hoeveel scholen het antwoord op tenminste één van de twee vragen *Ja* is, dan blijkt dat bijna tweederde van de scholen aan monitoring doet. Een deel van deze scholen doet dit dus zowel in de lessen als systematisch op middellange termijn.

Vraag	Variabele	Nee (0)	Ja (1)	Missend
Worden leervorderingen in de lessen gemonitord en geregistreerd?	INT_VOLG_V1_gsp	38	24	0
	INT_VOLG_V1_lui	38	24	0
	INT_VOLG_V1_spr	37	25	0
Worden leervorderingen in het schooljaar gemonitord en geregistreerd?	INT_VOLG_V2_gsp	31	28	3
	INT_VOLG_V2_lui	31	28	3
	INT_VOLG_V2_spr	29	30	3
Wordt gebruikgemaakt van een leerling-administratiepakket?	INT_VOLG_V3_gsp	22	40	0
	INT_VOLG_V3_lui	22	40	0
	INT_VOLG_V3_spr	21	41	0
Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?	INT_VOLG_V4_gsp	31	21	10
	INT_VOLG_V4_lui	30	22	10
	INT_VOLG_V4_spr	31	21	10
Wordt gebruikgemaakt van methode toetsen?	INT_VOLG_V5_gsp	37	15	10
	INT_VOLG_V5_lui	37	15	10
	INT_VOLG_V5_spr	37	15	10
Wordt gebruikgemaakt van zelfgemaakte instrumenten?	INT_VOLG_V6_gsp	46	6	10
	INT_VOLG_V6_lui	46	6	10
	INT_VOLG_V6_spr	46	6	10
Wordt in de planning gebruikgemaakt van de leerresultaten van leerlingen?	INT_VOLG_V7_gsp	29	32	1
	INT_VOLG_V7_lui	29	32	1
	INT_VOLG_V7_spr	29	32	1
Wordt gebruikgemaakt van groepsplannen?	INT_VOLG_V8_gsp	48	13	1
	INT_VOLG_V8_lui	48	13	1
	INT_VOLG_V8_spr	48	13	1

Via een reeks andere vragen is geprobeerd om zicht te krijgen op de wijze waarop scholen leerresultaten bijhouden en registreren. In het overzicht is onder meer te zien dat tweederde (40-41) van de scholen een leerlingadministratiepakket gebruikt. Er wordt echter relatief weinig gebruikgemaakt van genormeerde toetsen of observatielijsten voor luisteren, spreken of gesprekken (slechts een derde van de scholen doet dat wel), en ook methodegebonden toetsen (15) of zelfgemaakte instrumenten (6) worden op weinig scholen gebruikt.

De leerresultaten van leerlingen worden in de planningscyclus gebruikt (zie **INT_VOLG_V7**). Het is daarbij op het eerste gezicht opvallend dat er scholen bij zitten die bij de algemenere vragen over monitoring en registratie (**INT_VOLG_V1** en **INT_VOLG_V2**) hebben aangegeven hier niet aan te doen. Dit lijkt inconsistent, maar de vraag kan breder worden opgevat, namelijk “in geval leerlingen moeite hebben met een activiteit of een toets, dan wordt dit niet genoteerd maar speelt de leerkracht er wel direct op in bij het bepalen van de volgende activiteit of les”. Onder die opvatting is het logisch dat hier *Ja* op geantwoord wordt zonder dat dit direct betekent dat er op de school ook systematisch wordt bijgehouden hoe de leerlingen presteren op mondelinge taalvaardigheid. Groepsplannen, ten slotte, worden nauwelijks gebruikt op de deelnemende scholen (zie **INT_VOLG_V8**).

7.6 Differentiatie

Bij dit onderdeel ging het om de vraag in hoeverre scholen differentiatie toepassen en leerlingen met leerproblemen ondersteunen. Tijdens de gesprekken is daarom nagegaan hoe de scholen in dit opzicht te werk gaan: wordt er bewust gekozen voor convergente of divergente differentiatie, wordt er gewerkt met niveaugroepen, en op wat voor manier wordt er differentiatie toegepast of extra ondersteuning geboden in de dagelijkse praktijk? Onderstaand overzicht laat zien welke onderwerpen tijdens het gesprek aan de orde zijn gesteld en onder welke variabelen de onderwerpen zijn terug te vinden in het databestand. Daarnaast is te zien op hoeveel scholen een onderwerp wel of niet van toepassing was.

Bij de vragen over convergente en divergente differentiatie zien we dat de eerste vorm (**INT_DIFF_V1**) wat vaker wordt toegepast dan de tweede vorm (**INT_DIFF_V2**): 35 tegenover 27 scholen. De vragen zijn onafhankelijk van elkaar gecodeerd, maar vertonen wel een sterke negatieve samenhang, $\chi^2(1)= 27.995$, $p < .001$, $\Phi_{Cramer} = 0.672$, omdat de meeste scholen (52) maar één van de twee vormen van differentiatie gebruiken. Er zijn vijf scholen die beide of geen van beide doen. Bij de overige vragen valt op dat minder dan de helft van de scholen (25) met niveaugroepen werkt (**INT_DIFF_V3**), dat er wel vaak (45-46) verlengde instructie of extra oefening wordt geboden aan leerlingen met een achterstand (**INT_DIFF_V4**), en dat differentiatie naar leerstof (**INT_DIFF_V5**) of doelen (**INT_DIFF_V6**) net zoals het werken met niveaugroepen nog minder gebruikelijk is (25). Ten slotte wordt op de meeste scholen extra ondersteuning

buiten de klas geboden, ofwel door een specialist in de school (39), ofwel door een specialist van buiten de school (35). Uit notities over deze extra ondersteuning blijkt dat het zeer vaak gaat om logopedisten, maar soms ook om ambulante begeleiding vanuit Kentalis of Auris. Veel scholen wisten niet of de extra ondersteuning geboden werd vanuit een specifiek (interventie)programma. Daardoor bevat variabele **INT_DIFF_V9** veel missende waarden (gecodeerd als 999).

Vraag	Variabele	Nee (0)	Ja (1)	Missend
Vindt er convergente differentiatie plaats?	INT_DIFF_V1_gsp	27	35	0
	INT_DIFF_V1_lui	27	35	0
	INT_DIFF_V1_spr	27	35	0
Vindt er divergente differentie plaats?	INT_DIFF_V2_gsp	35	27	0
	INT_DIFF_V2_lui	35	27	0
	INT_DIFF_V2_spr	35	27	0
Wordt er gewerkt met (twee of meer) niveaugroepen?	INT_DIFF_V3_gsp	37	25	0
	INT_DIFF_V3_lui	37	25	0
	INT_DIFF_V3_spr	37	25	0
Is er differentiatie naar instructie?	INT_DIFF_V4_gsp	17	45	0
	INT_DIFF_V4_lui	16	46	0
	INT_DIFF_V4_spr	17	45	0
Is er differentiatie naar leerstof?	INT_DIFF_V5_gsp	36	25	1
	INT_DIFF_V5_lui	36	25	1
	INT_DIFF_V5_spr	36	25	1
Is er differentiatie naar (streef)doelen?	INT_DIFF_V6_gsp	35	22	5
	INT_DIFF_V6_lui	35	22	5
	INT_DIFF_V6_spr	35	22	5
Zijn er leerlingen die extra ondersteuning krijgen door een specialist ***binnen*** de school?	INT_DIFF_V7_gsp	21	39	2
	INT_DIFF_V7_lui	21	39	2
	INT_DIFF_V7_spr	21	39	2
Vrije invoer over specialist binnen school	*INT_DIFF_V7_namelijk*			
Zijn er leerlingen die extra ondersteuning krijgen door een specialist ***buiten*** de school?	INT_DIFF_V8_gsp	25	35	2
	INT_DIFF_V8_lui	25	35	2
	INT_DIFF_V8_spr	25	35	2
Vrije invoer over specialist buiten school	*INT_DIFF_V8_namelijk*			
Zijn er leerlingen die extra ondersteuning krijgen op basis van een specifiek (interventie)programma?	INT_DIFF_V9_gsp	25	25	12
	INT_DIFF_V9_lui	25	24	13
	INT_DIFF_V9_spr	25	25	12
Vrije invoer over specifiek programma	*INT_DIFF_V9_namelijk*			

7.7 Teamontwikkeling

Bij dit onderdeel stond de teamontwikkeling op de school centraal. Tijdens de gesprekken is nagegaan of het team een gezamenlijk beleid heeft dat is vastgelegd (bijvoorbeeld in het schoolplan), of collega's elkaar raadplegen, of het team interne of externe ondersteuning krijgt, en hoeveel overleg er binnen het team plaatsvindt over het onderwijs in mondelinge taalvaardigheid. Onderstaand overzicht laat zien welke onderwerpen tijdens het interview aan de orde zijn gesteld en onder welke variabelen de onderwerpen zijn terug te vinden in het databestand.

Vraag	Variabele	Nee (0)	Ja (1)	Missend
Is mondelinge taalvaardigheid benoemd in het school- of taalbeleidsplan?	INT_TEAM_V1	26	31	5
Zijn er intervisiemomenten of wordt collegiale consultatie toegepast?	INT_TEAM_V2	27	34	1
Vrije invoer over intervisie en collegiale consultatie	*INT_TEAM_V2_namelijk*			
Is er interne ondersteuning? (Bijvoorbeeld een taalcoördinator)	INT_TEAM_V3	24	35	3
Vrije invoer over interne ondersteuning	*INT_TEAM_V3_namelijk*			
Is er externe ondersteuning? (Bijvoorbeeld een onderwijsadviseur)	INT_TEAM_V4	30	28	4
Vrije invoer over externe ondersteuning	*INT_TEAM_V4_namelijk*			

We zien dat ongeveer de helft van de scholen (31) een expliciet taalbeleid heeft dat is vastgelegd in een apart document of in het schoolplan (**INT_TEAM_V1**). Op iets meer dan de helft van de scholen (34) vindt intervisie of collegiale consultatie plaats op het gebied van het onderwijs in mondelinge taalvaardigheid (**INT_TEAM_V2**), en ook interne ondersteuning van het team (bijvoorbeeld door een taalcoördinator) is gebruikelijk op meer dan de helft van de scholen (**INT_TEAM_V3**). Het komt ook redelijk vaak voor dat gebruik wordt gemaakt van externe ondersteuning (**INT_TEAM_V4**). Het gaat dan onder meer om logopedisten, taalcoördinatoren, onderwijsadviseurs, en ambulante begeleiding (zie in het databestand variabelen **INT_TEAM_V3_namelijk** en **INT_TEAM_V4_namelijk**).

Aan de hand van een checklist is van negen mogelijke onderwerpen gevraagd (**INT_TEAM_OND1** tot en met **INT_TEAM_OND9**) of het team daarover overlegt. Onderstaand overzicht laat zien dat zorgleerlingen op veel scholen in het teamoverleg besproken worden: op 36 tot 40 scholen is dat het geval. Over vakdidactiek en leerlingvolgsysteemresultaten wordt op de meerderheid van de scholen niet overlegd: slechts 13 scholen overleggen wél over vakdidactiek en slechts 17 scholen doen datzelfde over leerlingvolgsysteemresultaten.

Bespreking in teamoverleg	Variabele	Nee (0)	Ja (1)	Missend
Gebruik van de methode voor MTV	INT_TEAM_OND1	23	32	7
Onderwijs in MTV bij andere vakken dan taal	INT_TEAM_OND2	22	33	7
Volgen van de ontwikkeling van leerlingen op MTV	INT_TEAM_OND3	33	22	7
Uitkomsten van volgsysteem/registraties op MTV	INT_TEAM_OND4	38	17	7
Individuele leerlingen met spraak-/taalstoornis en hun ondersteuning door de leerkracht	INT_TEAM_OND5	19	36	7
Individuele leerlingen met spraak-/taalstoornis en hun ondersteuning door de externe specialist	INT_TEAM_OND6	15	40	7
Aanpak/begeleiding van de groep leerlingen die blijkens het volgsysteem een achterstand hebben op MTV	INT_TEAM_OND7	33	22	7
De leerdoelen van de school op het gebied van MTV	INT_TEAM_OND8	31	24	7
Vakdidactiek MTV	INT_TEAM_OND9	42	13	7
Vrije invoer over teamoverleg/-ontwikkeling	INT_TEAM_OND_anders_namelijk			

7.8 Variabelen

Zoals uit de voorgaande paragrafen is af leiden, zijn voor het onderdeel Luisteren op basis van de interviews de volgende variabelen in het databestand opgenomen. De grijs gearceerde variabelen worden bij de inhoudelijke analyses alleen gebruikt voor inhoudelijke duiding; als voorspeller in een regressiemodel zijn de variabelen namelijk niet geschikt, ofwel omdat het om een vrije invoer variabele gaat, ofwel omdat er sprake is van een zeer klein aantal waarnemingen (< 10) in één van de twee (antwoord)categorieën.

INT_AANB_V1_lui	Wordt een methode gebruikt?
INT_AANB_V1_namelijk	Vrije invoer voor methode
INT_AANB_V2_lui	Wordt de methode volledig gevolgd?
INT_AANB_V3_lui	Zijn er afspraken over het gebruik van de methode?
INT_AANB_V4_lui	Worden er aanvullende materialen gebruikt naast de hoofdmethode?
INT_AANB_V4_namelijk	Vrije invoer voor aanvullende materialen
INT_AANB_V5_lui	Wordt het onderwijs in X ingebed in de taalles?
INT_AANB_V6_lui	Wordt het onderwijs in X geïntegreerd in andere vakken?
INT_DOEL_V1_lui	Heeft de school (expliciet) doelen voor X?
INT_DOEL_V2_lui	Wordt gebruikgemaakt van een leerlijn?
INT_DOEL_V2_namelijk	Vrije invoer voor leerlijn
INT_DOEL_V3_lui	Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)
INT_DOEL_V4_lui	Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)
INT_DOEL_V5_lui	Wordt gebruikgemaakt van herhaald aanbod met feedback?
INT_DOEL_V6_lui	Wordt gebruikgemaakt van referentieniveaus?

INT_DOEL_V7_lui	Zijn er einddoelen voor schoolverlaters geformuleerd?
INT_VOLG_V1_lui	Worden leervorderingen in de lessen gemonitord en geregistreerd?
INT_VOLG_V2_lui	Worden leervorderingen in het schooljaar gemonitord en geregistreerd?
INT_VOLG_V3_lui	Wordt gebruikgemaakt van een leerling- administratiepakket?
INT_VOLG_V4_lui	Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?
INT_VOLG_V5_lui	Wordt gebruikgemaakt van methode toetsen?
INT_VOLG_V6_lui	Wordt gebruikgemaakt van zelfgemaakte instrumenten?
INT_VOLG_V7_lui	Wordt in de planning gebruikgemaakt van leerresultaten van leerlingen?
INT_VOLG_V8_lui	Wordt gebruikgemaakt van groepsplannen?
INT_DIFF_V1_lui	Vindt er convergente differentiatie plaats?
INT_DIFF_V2_lui	Vindt er divergente differentie plaats?
INT_DIFF_V3_lui	Wordt er gewerkt met (twee of meer) niveaugroepen?
INT_DIFF_V4_lui	Is er differentiatie naar instructie?
INT_DIFF_V5_lui	Is er differentiatie naar leerstof?
INT_DIFF_V6_lui	Is er differentiatie naar (streef)doelen?
INT_DIFF_V7_lui	Extra ondersteuning krijgen door een specialist binnen de school?
INT_DIFF_V7_namelijk	Vrije invoer specialist binnen de school
INT_DIFF_V8_lui	Extra ondersteuning krijgen door een specialist buiten de school?
INT_DIFF_V8_namelijk	Vrije invoer specialist buiten de school
INT_DIFF_V9_lui	Extra ondersteuning krijgen op basis van een specifiek programma?
INT_DIFF_V9_namelijk	Vrije invoer over specifiek programma
INT_TEAM_V1	Is mondelinge taalvaardigheid benoemd in het school- of taalbeleidsplan?
INT_TEAM_V2	Zijn er intervisiemomenten of wordt collegiale consultatie toegepast?
INT_TEAM_V2_namelijk	Vrije invoer over intervisie en collegiale consultatie
INT_TEAM_V3	Is er interne ondersteuning? (Bijvoorbeeld een taalcoördinator)
INT_TEAM_V3_namelijk	Vrije invoer over interne ondersteuning
INT_TEAM_V4	Is er externe ondersteuning? (Bijvoorbeeld een onderwijsadviseur)
INT_TEAM_V4_namelijk	Vrije invoer over externe ondersteuning

Voor het onderdeel Spreken zijn de volgende variabelen beschikbaar.

INT_AANB_V1_spr	Wordt een methode gebruikt?
INT_AANB_V1_namelijk	Vrije invoer voor methode
INT_AANB_V2_spr	Wordt de methode volledig gevolgd?
INT_AANB_V3_spr	Zijn er afspraken over het gebruik van de methode?
INT_AANB_V4_spr	Worden er aanvullende materialen gebruikt naast de hoofdmethode?
INT_AANB_V4_namelijk	Vrije invoer voor aanvullende materialen
INT_AANB_V5_spr	Wordt het onderwijs in X ingebed in de taalles?
INT_AANB_V6_spr	Wordt het onderwijs in X geïntegreerd in andere vakken?
INT_DOEL_V1_spr	Heeft de school (expliciet) doelen voor X?
INT_DOEL_V2_spr	Wordt gebruikgemaakt van een leerlijn?
INT_DOEL_V2_namelijk	Vrije invoer voor leerlijn
INT_DOEL_V3_spr	Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)
INT_DOEL_V4_spr	Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)
INT_DOEL_V5_spr	Wordt gebruikgemaakt van herhaald aanbod met feedback?
INT_DOEL_V6_spr	Wordt gebruikgemaakt van referentieniveaus?
INT_DOEL_V7_spr	Zijn er einddoelen voor schoolverlaters geformuleerd?
INT_VOLG_V1_spr	Worden leervorderingen in de lessen gemonitord en geregistreerd?
INT_VOLG_V2_spr	Worden leervorderingen in het schooljaar gemonitord en geregistreerd?
INT_VOLG_V3_spr	Wordt gebruikgemaakt van een leerling- administratiepakket?
INT_VOLG_V4_spr	Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?
INT_VOLG_V5_spr	Wordt gebruikgemaakt van methode toetsen?
INT_VOLG_V6_spr	Wordt gebruikgemaakt van zelfgemaakte instrumenten?
INT_VOLG_V7_spr	Wordt in de planning gebruikgemaakt van leerresultaten van leerlingen?
INT_VOLG_V8_spr	Wordt gebruikgemaakt van groepsplannen?
INT_DIFF_V1_spr	Vindt er convergente differentiatie plaats?
INT_DIFF_V2_spr	Vindt er divergente differentie plaats?
INT_DIFF_V3_spr	Wordt er gewerkt met (twee of meer) niveaugroepen?
INT_DIFF_V4_spr	Is er differentiatie naar instructie?
INT_DIFF_V5_spr	Is er differentiatie naar leerstof?
INT_DIFF_V6_spr	Is er differentiatie naar (streef)doelen?
INT_DIFF_V7_spr	Extra ondersteuning krijgen door een specialist binnen de school?
INT_DIFF_V7_namelijk	Vrije invoer specialist binnen de school
INT_DIFF_V8_spr	Extra ondersteuning krijgen door een specialist buiten de school?

INT_DIFF_V8_namelijk	Vrije invoer specialist buiten de school
INT_DIFF_V9_spr	Extra ondersteuning krijgen op basis van een specifiek programma?
INT_DIFF_V9_namelijk	Vrije invoer over specifiek programma
INT_TEAM_V1	Is mondelinge taalvaardigheid benoemd in het school- of taalbeleidsplan?
INT_TEAM_V2	Zijn er intervisiemomenten of wordt collegiale consultatie toegepast?
INT_TEAM_V2_namelijk	Vrije invoer over intervisie en collegiale consultatie
INT_TEAM_V3	Is er interne ondersteuning? (Bijvoorbeeld een taalcoördinator)
INT_TEAM_V3_namelijk	Vrije invoer over interne ondersteuning
INT_TEAM_V4	Is er externe ondersteuning? (Bijvoorbeeld een onderwijsadviseur)
INT_TEAM_V4_namelijk	Vrije invoer over externe ondersteuning

Voor het onderdeel Gesprekken, ten slotte, zijn onderstaande variabelen beschikbaar.

INT_AANB_V1_gsp	Wordt een methode gebruikt?
INT_AANB_V1_namelijk	Vrije invoer voor methode
INT_AANB_V2_gsp	Wordt de methode volledig gevolgd?
INT_AANB_V3_gsp	Zijn er afspraken over het gebruik van de methode?
INT_AANB_V4_gsp	Worden er aanvullende materialen gebruikt naast de hoofdmethode?
INT_AANB_V4_namelijk	Vrije invoer voor aanvullende materialen
INT_AANB_V5_gsp	Wordt het onderwijs in X ingebed in de taalles?
INT_AANB_V6_gsp	Wordt het onderwijs in X geïntegreerd in andere vakken?
INT_DOEL_V1_gsp	Heeft de school (expliciet) doelen voor X?
INT_DOEL_V2_gsp	Wordt gebruikgemaakt van een leerlijn?
INT_DOEL_V2_namelijk	Vrije invoer voor leerlijn
INT_DOEL_V3_gsp	Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)
INT_DOEL_V4_gsp	Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)
INT_DOEL_V5_gsp	Wordt gebruikgemaakt van herhaald aanbod met feedback?
INT_DOEL_V6_gsp	Wordt gebruikgemaakt van referentieniveaus?
INT_DOEL_V7_gsp	Zijn er einddoelen voor schoolverlaters geformuleerd?
INT_VOLG_V1_gsp	Worden leervorderingen in de lessen gemonitord en geregistreerd?
INT_VOLG_V2_gsp	Worden leervorderingen in het schooljaar gemonitord en geregistreerd?
INT_VOLG_V3_gsp	Wordt gebruikgemaakt van een leerling- administratiepakket?
INT_VOLG_V4_gsp	Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?
INT_VOLG_V5_gsp	Wordt gebruikgemaakt van methode toetsen?
INT_VOLG_V6_gsp	Wordt gebruikgemaakt van zelfgemaakte instrumenten?
INT_VOLG_V7_gsp	Wordt in de planning gebruikgemaakt van leerresultaten van leerlingen?
INT_VOLG_V8_gsp	Wordt gebruikgemaakt van groepsplannen?
INT_DIFF_V1_gsp	Vindt er convergente differentiatie plaats?
INT_DIFF_V2_gsp	Vindt er divergente differentie plaats?
INT_DIFF_V3_gsp	Wordt er gewerkt met (twee of meer) niveaugroepen?
INT_DIFF_V4_gsp	Is er differentiatie naar instructie?
INT_DIFF_V5_gsp	Is er differentiatie naar leerstof?
INT_DIFF_V6_gsp	Is er differentiatie naar (streef)doelen?
INT_DIFF_V7_gsp	Extra ondersteuning krijgen door een specialist binnen de school?
INT_DIFF_V7_namelijk	Vrije invoer specialist binnen de school
INT_DIFF_V8_gsp	Extra ondersteuning krijgen door een specialist buiten de school?
INT_DIFF_V8_namelijk	Vrije invoer specialist buiten de school
INT_DIFF_V9_gsp	Extra ondersteuning krijgen op basis van een specifiek programma?
INT_DIFF_V9_namelijk	Vrije invoer over specifiek programma
INT_TEAM_V1	Is mondelinge taalvaardigheid benoemd in het school- of taalbeleidsplan?
INT_TEAM_V2	Zijn er intervisiemomenten of wordt collegiale consultatie toegepast?
INT_TEAM_V2_namelijk	Vrije invoer over intervisie en collegiale consultatie
INT_TEAM_V3	Is er interne ondersteuning? (Bijvoorbeeld een taalcoördinator)
INT_TEAM_V3_namelijk	Vrije invoer over interne ondersteuning
INT_TEAM_V4	Is er externe ondersteuning? (Bijvoorbeeld een onderwijsadviseur)
INT_TEAM_V4_namelijk	Vrije invoer over externe ondersteuning

Naast voornoemde variabelen staan ook alle variabelen die voortkomen uit de checklists in het databestand.

8 Resultaten

8.1 Beschrijving leerlingprestaties en onderwijsleerproces op scholen

Zoals in paragraaf 2.3 al is aangegeven, is één van de doelen van de peiling Mondelinge Taalvaardigheid om de prestaties van leerlingen te beschrijven in termen van de behaalde referentieniveaus. Om dat te kunnen doen, is voor de onderdelen Luisteren, Spreken en Gesprekken een prestatiestandaard nodig. In paragraaf 8.1.1 wordt beschreven hoe de prestatiestandaarden bepaald zijn. Vervolgens is in paragrafen 8.1.2 en 8.1.3 te lezen hoeveel leerlingen beschikken over referentieniveaus 1F en 2F. In paragraaf 8.1.4, ten slotte, karakteriseren we het onderwijsleerproces op sbo- en so-scholen.

8.1.1 Prestatiestandaarden voor Luisteren, Spreken en Gesprekken

Om de onderzoeksvragen te kunnen beantwoorden, zijn prestatiestandaarden voor referentieniveaus 1F en 2F nodig. De wijze waarop de prestatiestandaarden tot stand zijn gekomen, verschilt per onderdeel. Voor het onderdeel Luisteren is gebruikgemaakt van de landelijke Ankersets, terwijl CvTE voor de onderdelen Spreken en Gesprekken nieuwe prestatiestandaarden heeft ontwikkeld op basis van de materialen die in dit onderzoek gebruikt zijn. Hieronder wordt de totstandkoming van de prestatiestandaarden per onderdeel nader toegelicht.

Luisteren
Voor Luisteren is via de Ankersets een landelijk geldende prestatiestandaard beschikbaar. Daar is in dit onderzoek gebruik van gemaakt door 25 items uit de Ankerset (zie Hoofdstuk 4, Figuur 4.1) af te nemen bij de leerlingen in onze onderzoeksgroep. Doordat we deze items hebben afgenomen, wordt het mogelijk om de landelijk geldende prestatiestandaarden over te brengen naar de luistervaardigheidsschaal van dit onderzoek. Er zijn verschillende manieren om prestatiestandaarden over te brengen. De gemeenschappelijke items vormen in ons geval het uitgangspunt. Figuur 8.1 laat in een vereenvoudigde weergave zien welke items (of gegevens) beschikbaar zijn om de prestatiestandaarden over te brengen. Blokjes A en C zijn specifiek voor een bepaalde meetschaal (de Ankerset of onze luistervaardigheidsschaal) en blokje B geeft het overlappende deel weer. Omdat we in deze peiling IRT-modellering hebben gebruikt om de scores van de leerlingen op de drie toetsversies op één onderliggende meetschaal te plaatsen en zo onderling vergelijkbaar te maken, gebruiken we ook voor het overbrengen van de prestatiestandaarden IRT-modellering.

Figuur 8.1. *Gegevens voor overbrenging prestatiestandaarden Ankerset*

Meetschaal	A	B	C
1 Ankerset			
2 Luisteren peiling			

Net zoals onze luistertoetsen is ook de Ankerset met een IRT-model geanalyseerd. De grensscores (aantal items correct op een 1F- of 2F- itemset) voor de referentieniveaus zijn dan ook beschikbaar als waarden op één onderliggende latente vaardigheidsschaal. Concreet representeert bij de Ankerset een θ-waarde van 0.283 1F, en en een θ-waarde van 0.670 2F. Deze waarden zijn om te rekenen naar een zogeheten *True Score*, dat wil zeggen de onafgeronde verwachte ruwe score, gegeven θ, op een specifieke (deel)verzameling van items (zie ook paragraaf 4.6). Wij hebben de θ-waarden omgezet naar *True Scores* op blokje B. Vervolgens is nagegaan welke θ-waarden op onze luisterschaal corresponderen met deze *True Scores* op blokje B. Het zijn uiteindelijk deze θ-waarden die als prestatiestandaard zijn gebruikt. Door te werken met θ-waarden en *True Scores* met veel decimalen maken we de overbrenging zo nauwkeurig mogelijk. De bovenste helft van Tabel 8.1 geeft een samenvatting van het resultaat.

Een aanname bij de beschreven overbrenging is dat de gemeenschappelijke items (module B) vergelijkbaar functioneren bij de leerlingen in de peiling en de leerlingen die aan het Ankersetonderzoek hebben deelgenomen. Als er sprake is van DIF kan dit de overbrenging vertekenen of onnauwkeurig maken. Daarom hebben we een extra analyse uitgevoerd waarin we alle afnamedata van de peiling en van het ankeronderzoek samen in het computerprogramma OPLM analyseerden. Met behulp van de itemresponsecurves van de overlappende items konden we controleren of de items vergelijkbaar functioneerden in de verschillende leerlinggroepen. In drie van de 25 gevallen leek deze aanname

problematisch, omdat de items voor de leerlingen in deze peiling wat moeilijker of juist makkelijker leken te zijn dan voor de leerlingen uit het andere onderzoek. Om na te gaan in hoeverre het overbrengen van de prestatiestandaarden gevoelig is voor het functioneren van deze items, hebben we de overzetting opnieuw uitgevoerd zonder deze drie items tot de overlap te rekenen. Met andere woorden, module B in Figuur 8.1 omvatte nu 22 in plaats van 25 items. De onderste helft van Tabel 8.1 geeft een samenvatting van het resultaat.

Tabel 8.1. *Overbrenging prestatiestandaarden van de Ankersets naar de luistervaardigheidsschaal*

Niveau	Grens op de Ankerset	θ op schaal Ankerset	True score op module B	θ op schaal peiling	Bankscore Luisteren	% met *tenminste* dit niveau
25 Ankeritems als module B (overlap)						
1F	21/22 op 31 1F-items	0.283	16.152	0.452	65.393	45.94%
2F	24/25 op 30 2F-items	0.670	21.319	1.115	84.787	9.33%
22 Ankeritems als module B (overlap)						
1F	21/22 op 31 1F-items	0.283	14.457	0.447	65.177	45.94%
2F	24/25 op 30 2F-items	0.670	19.007	1.084	84.148	10.70%

In Tabel 8.1 is te zien dat het weinig uitmaakt of rekening gehouden wordt met de geconstateerde DIF. Door de drie items die DIF vertoonden uit de overlap te verwijderen, komen we uit op iets lagere (minder strenge) prestatiestandaarden voor 1F en 2F, maar het verschil is klein (het zit alleen in de decimalen van de bankscore bij de peiling). Als op basis van de twee overbrengingen nagaan wordt hoeveel leerlingen in de onderzoeksgroep 1F of 2F behalen, dan blijkt dat 98.64 procent van de leerlingen exact dezelfde classificatie krijgt (<1F, 1F, of 2F), en dat er evenveel leerlingen onder/boven de 1F-grens presteren. Het enige verschil is dat 1.36 procent van de leerlingen bij deze overbrenging 2F behaalt in plaats van 1F. Het maakt voor de resultaten dus weinig uit welke overbrenging gebruikt wordt. Omdat het verwijderen van "problematische" DIF-items op basis van de afnamedata gevoelig is voor toevalsresultaten en vanuit inhoudelijk oogpunt lastig te onderbouwen is (de drie items behoren ieder bij andere fragmenten en een verklaring voor de geobserveerde DIF is moeilijk te geven) kiezen we in dit onderzoek de meest transparante overbrenging van de prestatiestandaarden, dus degene waarbij gewoon alle 25 items worden meegenomen. Als we de prestatiestandaarden toepassen, vinden we in de onderzoeksgroep de volgende verdeling naar referentieniveaus: (<1F) 54.06 procent; (1F) 36.61 procent; (2F) 9.33 procent

Spreken

Voor het jeugdjournaalfragment en het dobbelspel heeft CvTE in opdracht van het Ministerie van OCW (nieuwe) prestatiestandaarden ontwikkeld. Er is niet één set prestatiestandaarden, maar er zijn er twee, namelijk één voor het jeugdjournaalfragment en één voor het dobbelspel (zie Hemker, Van Ingen & Linthorst, 2018). Dit bemoeilijkt het gebruik in dit onderzoek, omdat het jeugdjournaalfragment en het dobbelspel in dit onderzoek zijn samengevoegd tot één totaalscore. Er kan ook met twee scores worden gewerkt, maar dat is onpraktisch, omdat leerlingen in de praktijk wél of niet beschikken over een bepaald referentieniveau; inhoudelijke nuanceringen (bijvoorbeeld naar pen-en-papier en spelvorm) zijn niet mogelijk of gebruikelijk. Om deze reden heeft CvTE in addendum ook prestatiestandaarden beschikbaar gesteld die betrekking hebben op een gecombineerde afname van het jeugdjournaalfragment en het dobbelspel. Concreet komt het erop neer dat in dat geval de som van de afzonderlijke grensscores gebruikt kan worden: (1F) 20 + 21 = 41, (2F) 27 + 27 = 54. Als we de prestatiestandaarden toepassen, vinden we in de onderzoeksgroep de volgende verdeling naar referentieniveaus: (<1F) 29.71 procent; (1F) 55.82 procent; (2F) 14.47 procent

Gesprekken

Net zoals bij de taken voor Spreken heeft CvTE ook voor het *Fischerspiel* (nieuwe) prestatiestandaarden ontwikkeld. Prestatiestandaard 1F is geplaatst bij 28 van de 34 scorepunten. Prestatiestandaard 2F is geplaatst bij 31 van de 34 scorepunten. Als we de prestatiestandaarden toepassen, vinden we in de onderzoeksgroep de volgende verdeling naar referentieniveaus: (<1F) 57.18 procent; (1F) 17.32 procent; (2F) 25.50 procent

8.1.2 Prestatieniveau leerlingen

Na bepaling van de prestatiestandaarden is in kaart gebracht op welk referentieniveau schoolverlaters in het sbo en so functioneren bij de onderdelen Luisteren, Spreken en Gesprekken. Omdat in het so de verschillende leerroutes onvoldoende evenwichtig vertegenwoordigd waren in de onderzoeksgroep (zie paragraaf 2.1.3), worden de uitkomsten voor het so gewogen. De weging is gebaseerd op de variabelen **LUI_WEIGHT**, **SPR_WEIGHT** en **GSP_WEIGHT**. Tabel 8.2 bevat per vaardigheid enkele beschrijvende statistieken. Er is een uitsplitsing gemaakt naar onderwijssoort.

***Tabel 8.2**. Behaalde referentieniveaus bij Luisteren, Spreken en Gesprekken*

		sbo		so	
Onderdeel	Niveau	Frequentie	Percentage	Frequentie	Percentage
Luisteren	Lager dan 1F	672	58.9	71	33.6
	1F	387	33.9	97	46.0
	2F	82	7.2	43	20.4
	Totaal	1141	100	211	100
Spreken	Lager dan 1F	366	31.7	49	21.5
	1F	621	53.9	150	65.8
	2F	166	14.4	29	12.7
	Totaal	1153	100	228	100
Gesprekken	Lager dan 1F	657	58.6	101	51.3
	1F	189	16.8	39	19.8
	2F	276	24.6	57	28.9
	Totaal	1122	100	197	100

Zoals we in Tabel 8.2 kunnen zien, hebben de meeste uitstromers in het sbo bij Luisteren referentieniveau 1F nog niet behaald (58.9 procent). Van de leerlingen die al wel zo ver zijn, functioneert 33.9 procent op niveau 1F en 7.2 procent op niveau 2F. In het so ziet het beeld er behoorlijk anders uit. De meerderheid van de leerlingen presteert daar op niveau 1F (46.0 procent) of hoger (20.4 procent). Bij Spreken presteren de leerlingen op een hoger niveau dan bij Luisteren. De overgrote meerderheid behaalt referentieniveau 1F (sbo: 53.9 procent, so: 65.8 procent) of 2F (sbo: 14.4 procent, so: 12.7 procent). Bij Gesprekken zijn de prestaties in termen van behaalde referentieniveaus aanzienlijk lager dan bij Spreken. Zowel in het sbo als so functioneert de meerderheid van de leerlingen op een niveau lager dan 1F (sbo: 58.6 procent, so: 51.3 procent). Wel is het zo dat er ook wat meer leerlingen het 2F-niveau behalen bij Gesprekken dan bij Spreken of Luisteren. De prestaties van de leerlingen lopen bij Gesprekken dus meer uiteen.

8.1.3 Prestatieniveaus in relatie tot achtergrondkenmerken

Van de leerlingen zijn verschillende achtergrondkenmerken beschikbaar. In deze paragraaf gaan we na in hoeverre de leerlingprestaties gerelateerd lijken te zijn aan deze kenmerken. We kijken eerst naar de variabelen geslacht, herkomst en thuistaal. Vervolgens gaan we in op de kenmerken die betrekking hebben op de onderwijsloopbaan en het zorgprofiel van de leerling. Leerlingen met ontbrekende waarden op de achtergrondvariabelen worden in deze analyses buiten beschouwing gelaten, waardoor de exacte leerlinggroep, waarvoor we de percentages per behaald referentieniveau uitrekenen, ietwat kan verschillen per analyse. In de volgende analyses is dan ook **geen** weging naar leerroute toegepast, want de gewichten voor Luisteren, Spreken en Gesprekken zijn alleen correct als deze worden toegepast voor de gehele groep leerlingen bij het betreffende onderdeel.

Geslacht, herkomst en thuistaal

Tabellen 8.3 tot en met 8.5 laten voor Luisteren, Spreken en Gesprekken zien hoeveel leerlingen over een bepaald referentieniveau beschikken als we een uitsplitsing maken naar geslacht (jongen / meisje), herkomst (Nederland / Niet-Nederland) en thuistaal (Nederlands / Niet-Nederlands). In Tabel 8.3 kunnen we onder meer zien dat bij Luisteren relatief meer meisjes (60 procent) dan jongens (51.2 procent) op een niveau lager dan 1F functioneren. In diezelfde tabel is bovendien te zien dat leerlingen met een Nederlandse herkomst vaker het 1F-niveau behalen dan leerlingen met een niet-Nederlandse herkomst (37.4 procent versus 24.2 procent). Datzelfde geldt voor leerlingen met Nederlands als thuistaal (38.0 procent versus 26.1 procent). De overige waarden in Tabellen 8.3 tot en met 8.5 zijn op eenzelfde wijze te interpreteren. In algemene zin kan gesteld worden dat het beeld in de verschillende groepen (totaal, sbo en

so) vaak hetzelfde is, maar niet altijd; het verschil in de behaalde referentieniveaus Luisteren tussen jongens en meisjes lijkt bijvoorbeeld iets anders in het sbo als in het so.

Tabel 8.3. *Behaalde referentieniveaus bij Luisteren naar geslacht, herkomst en thuistaal*

			lager dan 1F		1F		2F	
		N	freq	%	freq	%	freq	%
totaal	Jongen	932	477	51.2	361	38.7	94	10.1
	Meisje	460	276	60.0	148	32.2	36	7.8
	Herkomst niet Nederland	66	45	68.2	16	24.2	5	7.6
	Herkomst Nederland	1036	549	53.0	387	37.4	100	9.7
	Niet Nederlands als thuistaal	115	79	68.7	30	26.1	6	5.2
	Nederlands als thuistaal	996	516	51.8	378	38.0	102	10.2
so	Jongen	220	75	34.1	109	49.5	36	17.2
	Meisje	31	6	19.4	13	41.9	12	36.0
	Herkomst niet Nederland	11	4	36.4	4	36.4	3	27.3
	Herkomst Nederland	207	69	33.3	100	48.3	38	18.4
	Niet Nederlands als thuistaal	17	7	41.2	8	47.1	2	11.8
	Nederlands als thuistaal	208	67	32.2	100	48.1	41	19.7
sbo	Jongen	712	402	56.5	252	35.4	58	8.1
	Meisje	429	270	62.9	135	31.5	21	5.6
	Herkomst niet Nederland	55	41	74.5	12	21.8	2	3.6
	Herkomst Nederland	829	480	57.9	287	34.6	62	7.5
	Niet Nederlands als thuistaal	98	72	73.5	22	22.4	4	4.1
	Nederlands als thuistaal	788	449	57.0	278	35.3	61	7.7

Tabel 8.4. *Behaalde referentieniveaus bij Spreken naar geslacht, herkomst en thuistaal*

			lager dan 1F		1F		2F	
		N	freq	%	freq	%	freq	%
totaal	Jongen	968	275	28.4	558	57.6	135	13.9
	Meisje	449	146	32.5	233	51.9	70	15.6
	Herkomst niet Nederland	60	23	38.3	30	50.0	7	11.7
	Herkomst Nederland	1064	309	29.0	611	57.4	144	13.5
	Niet Nederlands als thuistaal	109	47	43.1	51	46.8	11	10.1
	Nederlands als thuistaal	1028	289	28.1	600	58.4	139	13.5
so	Jongen	233	49	21.0	155	66.5	29	12.4
	Meisje	31	6	19.4	15	48.4	10	32.3
	Herkomst niet Nederland	12	3	25.0	9	75.0	0	
	Herkomst Nederland	224	46	20.5	148	66.1	30	13.4
	Niet Nederlands als thuistaal	19	5	26.3	11	57.9	3	15.8
	Nederlands als thuistaal	224	46	20.5	151	67.4	27	12.1
sbo	Jongen	735	226	30.7	403	54.8	106	14.4
	Meisje	418	140	33.5	218	52.2	60	14.4
	Herkomst niet Nederland	48	20	41.7	21	43.8	7	14.6
	Herkomst Nederland	840	263	31.3	463	55.1	114	13.6
	Niet Nederlands als thuistaal	90	42	46.7	40	44.4	8	8.9
	Nederlands als thuistaal	804	243	30.2	449	55.8	112	13.9

Tabel 8.5. *Behaalde referentieniveaus bij Gesprekken naar geslacht, herkomst en thuistaal*

			lager dan 1F		1F		2F	
		N	freq	%	freq	%	freq	%
totaal	Jongen	919	511	55.6	166	18.1	242	26.3
	Meisje	437	264	60.4	69	15.8	104	23.8
	Herkomst niet Nederland	59	35	59.3	9	15.3	15	25.4
	Herkomst Nederland	999	573	57.4	174	17.4	252	25.2
	Niet Nederlands als thuistaal	105	71	67.6	13	12.4	21	20.0
	Nederlands als thuistaal	967	543	56.2	172	17.8	252	26.1
so	Jongen	192	104	51.6	42	20.3	61	29.5
	Meisje	22	14	59.1	4	14.8	9	33.3
	Herkomst niet Nederland	8	4	50.0	3	37.5	1	12.5
	Herkomst Nederland	194	101	52.1	37	19.1	56	28.9
	Niet Nederlands als thuistaal	18	10	55.6	5	27.8	3	16.7
	Nederlands als thuistaal	191	100	52.4	35	18.3	56	29.3
sbo	Jongen	712	407	57.2	124	17.4	181	25.4
	Meisje	410	250	61.0	65	15.9	95	23.2
	Herkomst niet Nederland	51	31	60.8	6	11.8	14	27.5
	Herkomst Nederland	805	472	58.6	137	17.0	196	24.3
	Niet Nederlands als thuistaal	87	61	70.1	8	9.2	18	20.7
	Nederlands als thuistaal	776	443	57.1	137	17.7	196	25.3

Onderwijsloopbaan

Om een beeld te krijgen van de referentieniveaus van leerlingen met verschillende uitstroomprofielen is in Tabel 8.6 voor het onderdeel Luisteren per leerroute weergegeven hoeveel leerlingen de referentieniveaus 1F en 2F behalen. Onder leerroute 1 vallen de uitstroomprofielen vmbo-gl/tl, havo en vwo, onder leerroute 2 vallen vmbo-bb/kb en onder leerroute 3 vallen vso dagbesteding/arbeid en praktijkonderwijs. De aantallen en percentages worden gegeven voor de totale onderzoeksgroep en voor sbo en so afzonderlijk. Zoals we in Tabel 8.6 zien, behalen leerlingen met leerroute 1 in de regel niveau 1F (47.0 procent) of 2F (44.3 procent). Voorts zien we dat leerlingen in leerroute 2 relatief het vaakst op niveau 1F presteren (50.4 procent) of op een niveau lager dan 1F (40.0 procent); slechts 9.6 procent van de leerlingen met leerroute 2 behaalt het niveau 2F. Voor leerlingen met leerroute 3, ten slotte, geldt dat zij meestal onder het 1F-niveau presteren bij Luisteren (81.1 procent). Bovendien behaalt geen van deze leerlingen in de onderzoeksgroep het 2F-niveau. Deze patronen zijn vergelijkbaar tussen het so en sbo.

Tabel 8.6. *Behaalde referentieniveaus bij Luisteren naar leerroute*

			lager dan 1F		1F		2F	
	Leerroute	N	freq	%	freq	%	freq	%
totaal	1	115	10	8.7	54	47.0	51	44.3
	2	502	201	40.0	253	50.4	48	9.6
	3	455	369	81.1	86	18.9	0	0.0
so	1	67	6	9.0	32	47.8	29	43.3
	2	117	44	37.6	61	52.1	12	10.3
	3	27	19	70.4	8	29.6	0	0
sbo	1	48	4	8.3	22	45.8	22	45.3
	2	386	158	40.9	192	49.7	36	9.3
	3	428	350	81.8	78	18.2	0	0.0

In Tabel 8.7 staan dezelfde resultaten, maar nu voor het onderdeel Spreken. Hier valt op dat leerlingen met elk van de drie leerroutes relatief het vaakste presteren op niveau 1F. Bij leerroute 1 en 2 liggen de percentages leerlingen die lager presteren dan het 1F-niveau, of juist op het 2F-niveau, dicht bij elkaar. Bij leerroute 3 zien we dat relatief veel leerlingen het 1F-niveau niet behalen (41.5 procent), en het aantal

leerlingen dat niveau 2F behaalt, is erg klein (7.6 procent). Deze globale percentages zijn sterk vergelijkbaar als we apart naar de verdeling binnen het sbo en so kijken.

Tabel 8.7. *Behaalde referentieniveaus bij Spreken naar leerroute*

			lager dan 1F		1F		2F	
	Leerroute	N	freq	%	freq	%	freq	%
totaal	1	126	19	15.1	78	61.9	29	23.0
	2	535	111	20.7	331	61.9	93	17.4
	3	446	185	41.5	227	50.9	34	7.6
so	1	69	10	14.5	46	66.7	13	18.8
	2	128	25	19.5	88	68.8	15	11.7
	3	31	12	38.7	18	58.1	1	3.2
sbo	1	57	9	15.8	32	56.1	16	28.1
	2	307	86	28.0	143	46.6	78	25.4
	3	415	173	41.7	209	50.4	33	7.9

Tabel 8.8, ten slotte, toont de behaalde referentieniveaus per leerroute voor het onderdeel Gesprekken. Het percentage leerlingen dat presteert op het niveau 1F, is binnen de drie leerroutes vergelijkbaar: 16.4 procent in leerroute 3 en 18.8 procent in leerroutes 1 en 2. Voor de leerlingen met leerroute 1 geldt dat ongeveer evenveel van hen onder het 1F-niveau presteren of juist op het 2F-niveau (38.5 procent en 42.7 procent, respectievelijk). Bij de andere twee leerroutes zien we echter dat er relatief meer leerlingen onder het 1F-niveau presteren (respectievelijk 53.8 procent en 64.1 procent bij leerroute 2 en 3). Net als bij de onderdelen Luisteren en Spreken zien we grofweg dezelfde verhoudingen binnen het sbo als binnen het so, wanneer we de twee onderwijstypen apart analyseren.

Tabel 8.8. *Behaalde referentieniveaus bij Gesprekken naar leerroute*

			lager dan 1F		1F		2F	
	Leerroute	N	freq	%	freq	%	freq	%
totaal	1	117	45	38.5	22	18.8	50	42.7
	2	500	269	53.8	94	18.8	137	27.4
	3	422	271	64.2	69	16.4	82	19.4
so	1	64	23	35.9	15	23.4	26	40.6
	2	105	66	62.9	17	16.2	22	20.9
	3	28	15	53.6	6	21.4	7	25.0
sbo	1	53	22	41.5	7	13.2	24	45.3
	2	395	203	51.4	77	19.5	115	29.1
	3	394	256	65.0	63	16.0	75	19.0

Zorgprofiel

Om de behaalde referentieniveaus te relateren aan de verschillende zorgprofielen van de leerlingen kijken we naar het gemiddelde en de standaardafwijking op de variabelen over probleemgedrag en leerachterstanden. In Tabel 8.9 staan de resultaten van deze analyse voor het onderdeel Luisteren. Eerst is voor de drie schalen van externaliserend en internaliserend probleemgedrag de gemiddelde score berekend op de vierpuntsschaal (1 = 'niet aanwezig', 2 = 'in lichte mate', 3 = 'in behoorlijke mate', 4 = 'in ernstige mate'), apart voor de leerlinggroepen met een prestatieniveau onder 1F, op 1F, of op 2F. Vervolgens is voor de leerachterstanden hetzelfde gedaan (1 = 'geen achterstand', 2 = 'minder dan een jaar', 3 = '1-2 jaar achterstand', 4 = 'meer dan 2 jaar achterstand'). De gemiddelden voor externaliserend en internaliserend probleemgedrag zijn vergelijkbaar voor de totale groep leerlingen die verschillende referentieniveaus behalen. Er lijkt dus weinig verband te zijn tussen de vormen van probleemgedrag en de luistervaardigheid van de leerlingen. Leerlingen in het so scoren overigens iets hoger op probleemgedrag dan leerlingen uit het sbo. Zoals te verwachten valt, zien we verder dat het behalen van de referentieniveaus 1F en 2F negatief samenhangt met de geconstateerde leerachterstand van de leerling in taal/lezen. Hetzelfde patroon zien we ook voor leerachterstanden in rekenen of informatieverwerking. Over

het algemeen geldt dat de leerlingen met meer geconstateerde leerachterstanden een lagere luistervaardigheid aantonen. Bij de leerlingen in het sbo spelen leerachterstanden overigens een grotere rol dan bij de leerlingen in het so.

Tabel 8.9. *Gemiddelde zorgprofiel per referentieniveau Luisteren*

		lager dan 1F		1F		2F	
		M	*SD*	*M*	*SD*	*M*	*SD*
totaal	Externaliserend probleemgedrag	1.61	0.66	1.68	0.66	1.63	0.60
	Internaliserend negatief probleemgedrag	1.81	0.66	1.88	0.73	1.85	0.69
	Internaliserend sociaal probleemgedrag	1.69	0.56	1.63	0.59	1.73	0.56
	Leerachterstand in taal/lezen	3.25	0.98	2.77	1.09	1.95	1.06
	Leerachterstand in rekenen	3.45	0.80	2.86	1.04	2.07	1.07
	Leerachterstand in informatieverwerking	3.35	0.85	2.71	1.00	1.84	0.92
so	Externaliserend probleemgedrag	1.86	0.73	1.89	0.68	1.82	0.71
	Internaliserend negatief probleemgedrag	1.96	0.73	2.17	0.81	1.98	0.82
	Internaliserend sociaal probleemgedrag	1.94	0.57	1.89	0.64	1.95	0.54
	Leerachterstand in taal/lezen	2.60	1.11	2.21	1.09	1.49	0.75
	Leerachterstand in rekenen	2.71	1.08	2.09	1.02	2.07	1.07
	Leerachterstand in informatieverwerking	2.66	1.02	2.08	.099	1.41	0.67
sbo	Externaliserend probleemgedrag	1.58	0.65	1.77	0.68	1.50	0.47
	Internaliserend negatief probleemgedrag	1.61	0.65	1.77	0.68	1.76	0.61
	Internaliserend sociaal probleemgedrag	1.65	0.55	1.54	0.54	1.57	0.52
	Leerachterstand in taal/lezen	3.35	0.92	2.96	1.02	2.29	1.14
	Leerachterstand in rekenen	3.56	0.69	3.13	0.91	2.52	1.08
	Leerachterstand in informatieverwerking	3.45	0.77	2.93	0.91	2.14	0.96

Tabel 8.10. *Gemiddelde zorgprofiel per referentieniveau Spreken*

		lager dan 1F		1F		2F	
		M	*SD*	*M*	*SD*	*M*	*SD*
totaal	Externaliserend probleemgedrag	1.57	0.63	1.70	0.70	1.60	0.64
	Internaliserend negatief probleemgedrag	1.85	0.69	1.88	0.72	1.77	0.69
	Internaliserend sociaal probleemgedrag	1.74	0.60	1.67	0.57	1.57	0.54
	Leerachterstand in taal/lezen	3.16	1.09	2.84	1.10	2.71	1.14
	Leerachterstand in rekenen	3.34	0.95	2.97	1.02	2.93	1.13
	Leerachterstand in informatieverwerking	3.23	0.96	2.83	1.02	2.64	1.14
so	Externaliserend probleemgedrag	1.71	0.69	1.94	0.72	1.80	0.73
	Internaliserend negatief probleemgedrag	2.01	0.78	2.08	0.76	1.94	0.78
	Internaliserend sociaal probleemgedrag	1.99	0.64	1.88	0.57	1.87	0.59
	Leerachterstand in taal/lezen	2.35	1.17	2.21	1.10	2.04	0.10
	Leerachterstand in rekenen	2.45	1.16	2.12	1.03	2.00	1.18
	Leerachterstand in informatieverwerking	2.51	1.04	2.11	1.04	1.75	1.03
sbo	Externaliserend probleemgedrag	1.55	0.61	1.62	0.67	1.56	0.61
	Internaliserend negatief probleemgedrag	1.82	0.66	1.81	0.69	1.73	0.66
	Internaliserend sociaal probleemgedrag	1.69	0.58	1.60	0.55	1.51	0.51
	Leerachterstand in taal/lezen	3.31	1.01	3.04	1.02	2.86	1.12
	Leerachterstand in rekenen	3.50	0.82	3.25	0.86	3.14	1.01
	Leerachterstand in informatieverwerking	3.36	0.88	3.08	0.89	2.85	1.06

Tabel 8.10 toont de resultaten voor Spreken. Net als bij Luisteren zijn de gemiddelden voor externaliserend en internaliserend probleemgedrag sterk vergelijkbaar tussen de groepen leerlingen die verschillende referentieniveaus behalen, wat erop wijst dat probleemgedrag niet samen lijkt te hangen met Spreekvaardigheid. De verschillen in de gemiddelde leerachterstanden in taal/lezen, rekenen of informatieverwerking zijn allemaal in de richting die je zou verwachten, namelijk dat een hoger behaald

referentieniveau gepaard gaat met een lagere gemiddelde leerachterstand. De verschillen zijn echter kleiner dan we zagen bij Luisteren, dus spreekvaardigheid lijkt minder sterk gerelateerd te zijn aan de geconstateerde leerachterstanden op deze drie gebieden.

In Tabel 8.11, ten slotte, worden de resultaten voor het onderdeel Gesprekken gepresenteerd. De gemiddelden voor externaliserend en internaliserend probleemgedrag zijn net als bij de andere twee onderdelen weinig verschillend tussen de leerlinggroepen die onder 1F, op 1F of op 2F-niveau presteren. De gemiddelde leerachterstanden in taal/lezen, rekenen en informatieverwerking zijn lager naarmate leerlingen hogere prestatieniveaus bereiken, net als bij de andere onderdelen, maar de verschillen zijn bescheiden.

Tabel 8.11. *Gemiddelde zorgprofiel per referentieniveau Gesprekken*

		lager dan 1F		1F		2F	
		M	*SD*	*M*	*SD*	*M*	*SD*
totaal	Externaliserend probleemgedrag	1.67	0.68	1.60	0.60	1.61	0.66
	Internaliserend negatief probleemgedrag	1.88	0.72	1.84	0.75	1.81	0.67
	Internaliserend sociaal probleemgedrag	1.72	0.58	1.63	0.57	1.62	0.52
	Leerachterstand in taal/lezen	3.00	1.08	2.92	1.56	2.79	1.17
	Leerachterstand in rekenen	3.22	0.96	3.02	1.04	2.86	1.10
	Leerachterstand in informatieverwerking	3.07	0.98	2.84	1.06	2.69	1.11
so	Externaliserend probleemgedrag	1.89	0.69	1.78	0.62	1.82	0.77
	Internaliserend negatief probleemgedrag	2.08	0.79	1.96	0.74	2.01	0.82
	Internaliserend sociaal probleemgedrag	1.93	0.58	1.90	0.55	1.84	0.56
	Leerachterstand in taal/lezen	2.27	1.09	2.08	1.21	1.98	1.15
	Leerachterstand in rekenen	2.34	1.08	2.16	1.14	1.79	1.06
	Leerachterstand in informatieverwerking	2.25	1.02	2.19	1.19	1.75	1.04
sbo	Externaliserend probleemgedrag	1.63	0.67	1.56	0.59	1.55	0.61
	Internaliserend negatief probleemgedrag	1.84	0.70	1.81	0.75	1.55	0.61
	Internaliserend sociaal probleemgedrag	1.67	0.57	1.56	0.56	1.56	0.49
	Leerachterstand in taal/lezen	3.16	1.01	3.13	1.04	3.00	1.07
	Leerachterstand in rekenen	3.41	0.81	3.24	0.88	3.15	0.93
	Leerachterstand in informatieverwerking	3.25	0.87	3.01	0.96	2.95	0.98

8.1.4 Onderwijs op s(b)o scholen

In deze paragraaf karakteriseren we het onderwijsleerproces op de scholen voor elk van de drie mondelinge taalvaardigheden. Hierbij maken we onderscheid tussen de sbo- en so-scholen. We kijken naar de volgende aspecten: (a) Onderwijsaanbod & Methodegebruik, (b) Doelgerichtheid & Planmatigheid, (c) Monitoring & Planning, (d) Differentiatie, en (e) Teamontwikkeling.

Onderwijsaanbod en methodegebruik
Bij het schoolinterview is gevraagd naar het gebruik van lesmethoden voor de mondelinge taalvaardigheden. Zoals Tabel 8.12 laat zien, maken bijna alle scholen gebruik van een lesmethode om aan mondelinge taalvaardigheid te werken. Er is hierbij geen verschil tussen de vaardigheden Spreken, Luisteren en Gesprekken. Iets meer dan de helft van de so- en sbo-scholen volgt de methode volledig, en er zijn vaak afspraken gemaakt over het gebruik van de methode (so 66.7 procent; sbo 72.3 procent). In het so wordt de methode *Taalactief* het meest gebruikt en in het sbo de methode *Staal*. Ruim tweederde van de so- en sbo-scholen gebruikt bovendien aanvullende materialen naast de methode.

Tabel 8.12. *Methodegebruik*

		So		sbo	
Methodegebruik	Onderdeel	freq	%	freq	%
Er wordt een methode gebruikt	Luisteren	12	100.0	46	92.0
	Spreken	12	100.0	46	92.0
	Gesprekken	12	100.0	46	92.0
De methode wordt volledig gevolgd	Luisteren	7	58.3	24	49.0
	Spreken	7	58.3	24	49.0
	Gesprekken	7	58.3	24	49.0
Er zijn afspraken over het gebruik van methode	Luisteren	8	66.7	34	72.3
	Spreken	8	66.7	34	72.3
	Gesprekken	8	66.7	34	72.3
Er worden aanvullende materialen gebruikt naast de hoofdmethode	Luisteren	8	66.7	32	68.0
	Spreken	8	66.7	32	68.0
	Gesprekken	8	66.7	32	68.0

Als we kijken op welke momenten scholen aan mondelinge taalvaardigheid werken, zien we in het so (Tabel 8.13) dat er in de meeste gevallen zowel in de taalles als in andere vakken aandacht wordt besteed aan mondelinge taalvaardigheid (54.5 procent). In het sbo wordt het onderwijs in mondelinge taalvaardigheid daarentegen voornamelijk geïntegreerd in andere zaakvakken (35.4 tot 37.5 procent). Er zijn ook scholen die alleen in de taalles aan mondelinge taalvaardigheid werken (27.1 tot 29.2 procent). Een opvallend resultaat is dat er in het sbo 6 tot 8 scholen zijn (afhankelijk van de vaardigheid waar we naar kijken) die de mondelinge taalvaardigheid noch in de taalles, noch in andere zaakvakken integreren. Het lijkt erg onwaarschijnlijk dat er op deze scholen helemaal geen aandacht wordt besteed aan mondelinge taalvaardigheid, maar kan wel aanduiden dat dit niet structureel gebeurt of dat er geen bewust beleid is over het moment en de wijze waarop dit gebeurt

Tabel 8.13. *Integratie van onderwijs in mondelinge taalvaardigheid*

			Onderwijs MTV geïntegreerd in andere vakken							
			so				sbo			
			nee		ja		nee		ja	
			freq	%	freq	%	freq	%	freq	%
Onderwijs MTV ingebed in de taalles	Luisteren	nee	0	0.0	2	18.2	7	14.6	17	35.4
		ja	3	27.3	6	54.5	13	27.1	11	22.9
	Spreken	nee	0	0.0	2	18.2	6	12.5	18	37.5
		ja	3	27.3	6	54.5	14	29.2	10	20.8
	Gesprekken	nee	0	0.0	2	18.2	8	16.7	17	35.4
		ja	3	27.3	6	54.5	13	27.1	10	20.8

Tot slot is met behulp van een checklist in kaart gebracht welke luister- en spreekactiviteiten er in het onderwijs zoal worden gebruikt (Tabel 8.14). Bij de luisteractiviteiten zien we dat ‘kijken en luisteren naar een tv-fragment’ zowel in het so (88.9 procent) als het sbo (90.9 procent) het meeste wordt toegepast. Ook ‘voorlezen van een boek en vragen beantwoorden’ is een gangbare werkvorm (so 77.8 procent; sbo 86.4 procent), en hetzelfde geldt voor ‘beluisteren van instructies en ze uitvoeren’ (so 81.8 procent; sbo 66.7 procent). In het sbo lijkt de variatie in luisteractiviteiten groter dan in het so, maar hier moet wel bij aangetekend worden dat slechts een beperkt aantal so-scholen heeft deelgenomen aan dit onderzoek, zodat we voorzichtig moeten zijn met generaliseren naar de populatie van so-scholen. Bij de spreekactiviteiten valt op dat de activiteit ‘(zakelijke) boodschappen oefenen’ niet genoemd wordt in het so, maar wel in het sbo (50 procent). Verder zijn ‘kring-groepsgesprekken’, ‘spreekbeurten’ en ‘meningen verwoorden’ activiteiten die zowel in het so als het sbo veelvuldig worden toegepast.

Tabel 8.14. *Overzicht van gebruik van luister- en spreekactiviteiten per onderwijssoort*

	so		sbo	
	Freq	%	freq	%
Luisteractiviteit				
Kijken en luisteren naar tv-fragment en vragen beantwoorden	8	88.9	40	90.9
Luisterboek en vragen beantwoorden	2	22.2	24	54.4
Coöperatieve luisteropdrachten	2	22.2	25	56.8
Voorlezen van boek en vragen beantwoorden	7	77.8	38	86.4
Naar gesproken tekst luisteren en aantekeningen maken	0	0.0	21	47.7
Naar gesproken tekst luisteren en samenvatting maken	1	11.1	18	40.9
Naar gesproken tekst luisteren en mening/oordeel/standpunt bepalen	2	22.2	33	75.0
Instructies (praktisch) beluisteren en ze vervolgens uitvoeren	6	66.7	36	81.8
Spreekactviteit				
Spreekbeurten	7	77.8	38	86.4
Kring-/groepsgesprekken	9	100.0	43	97.7
Verslag van persoonlijke belevenis	6	66.7	37	84.1
Dialogen	5	55.6	30	68.2
Samenwerkend leren	6	66.7	37	84.1
Discussies	6	66.7	32	72.7
(Zakelijke) boodschappen oefenen	0	0.0	22	50.0
Verwoorden van meningen	7	77.8	39	88.6
Verhaal vertellen volgens gespreksregels	5	55.6	24	54.5
Deelname aan een weekafsluiting/toneelstuk/…	3	33.3	28	63.6

Doelgerichtheid en planmatigheid

Tabel 8.15 laat zien in hoeverre de so- en sbo-scholen hun onderwijs in mondelinge taalvaardigheid doelgericht en planmatig vormgeven. Ruim tweederde van de so- en sbo-scholen geeft aan expliciete doelen geformuleerd te hebben voor Luisteren, Spreken en Gesprekken. Verreweg de meeste scholen maken gebruik van een leerlijn (100 procent van de so-scholen, 91.5 procent van de sbo-scholen). Op ruim de helft van de so-scholen (58.3 procent) en driekwart van de sbo-scholen (75.0 tot 77.1 procent) wordt doelgericht gewerkt in die zin dat bij elk leerdoel taalactiviteiten worden gezocht. Op vergelijkbare percentages van de scholen wordt activiteitgericht gewerkt in die zin dat er leerdoelen worden gezocht en gekoppeld aan taalactiviteiten. Samengenomen impliceren deze aantallen dat er op veel scholen (24 in het sbo, 5 in het so) soms een doelgerichte en soms een activiteitgerichte aanpak gevolgd wordt, maar dus niet één van beiden exclusief. De doelen komen herhaaldelijk aan bod met tussentijdse feedback op het merendeel van de so-scholen (83.3 procent) en sbo-scholen (77.1 tot 79.2 procent). De helft van de so-scholen (50 procent), en iets meer dan de helft van de sbo-scholen (64.6 procent), blijkt einddoelen te hebben geformuleerd voor schoolverlaters.

Tabel 8.15. *Overzicht van doelgerichtheid naar onderwijssoort*

		so			sbo		
Doelgerichtheid	Onderdeel	N	freq	%	N	freq	%
De school heeft (expliciet) doelen	Luisteren	12	8	66.7	49	33	67.3
	Spreken	12	8	66.7	49	33	67.3
	Gesprekken	12	8	66.7	49	32	65.3
Er wordt gebruikgemaakt van een leerlijn	Luisteren	11	11	100.0	47	43	91.5
	Spreken	11	11	100.0	47	42	91.5
	Gesprekken	11	11	100.0	47	41	87.3
Aan elk doel zijn taalactiviteiten gekoppeld (doelgerichte aanpak)	Luisteren	12	7	58.3	48	36	75.0
	Spreken	12	7	58.3	48	37	77.1
	Gesprekken	12	7	58.3	48	37	77.1

Tabel 8.15. *Vervolg*

		so			sbo		
Doelgerichtheid	Onderdeel	N	freq	%	N	freq	%
Aan elke taalactiviteit zijn doelen gekoppeld (activiteitgerichte aanpak)	Luisteren	12	7	58.3	47	32	68.1
	Spreken	12	7	58.3	47	33	70.2
	Gesprekken	12	7	58.3	47	33	70.2
Er wordt gebruikgemaakt van herhaald aanbod met feedback	Luisteren	12	10	83.3	48	37	77.1
	Spreken	12	10	83.3	48	38	79.2
	Gesprekken	12	10	83.3	48	37	77.1
Er wordt gebruikgemaakt van referentieniveaus	Luisteren	12	7	58.3	48	33	68.7
	Spreken	12	7	58.3	47	33	70.2
	Gesprekken	12	7	58.3	48	33	77.1
Er zijn einddoelen voor schoolverlaters geformuleerd	Luisteren	12	6	50	48	31	64.6
	Spreken	12	6	50	48	31	64.6
	Gesprekken	12	6	50	48	31	64.6

Monitoring en planning

Tabel 8.16 laat zien in hoeverre scholen de leervorderingen in mondelinge taalvaardigheid volgen en aan de hand hiervan het onderwijs plannen. Leervorderingen worden relatief iets vaker gedurende het schooljaar (op de middellange termijn) gemonitord dan direct in de lessen: dat geldt zowel voor het so (41.7 procent versus 33.3 procent) als voor het sbo (48.9 – 53.2 procent versus 20.0 – 42.0 procent). Hiervoor gebruikt ongeveer een derde van de sbo-scholen een leerlingadministratiepakket. Het is opvallend dat ongeveer een kwart van de sbo-scholen wel gebruikmaakt van een leerlingadministratiepakket voor mondelinge taalvaardigheid, maar tegelijkertijd ook zegt de leervorderingen niet systematisch te volgen. In het so geven 4 van de 12 (33.3 procent) scholen aan de leervorderingen in het jaar te monitoren en te registreren en gebruik te maken van een leerlingadministratiepakket, maar er zijn ook scholen die de leervorderingen niet volgen en wél een leerlingadministratiepakket gebruiken (50.0 procent).

Tabel 8.16. *Overzicht monitoring en planning naar onderwijssoort*

		so			sbo		
Monitoring en planning	Onderdeel	N	freq	%	N	freq	%
Leervorderingen worden in de lessen gemonitord en geregistreerd	Luisteren	12	4	33.3	50	20	40.0
	Spreken	12	4	33.3	50	21	42.0
	Gesprekken	12	4	33.3	50	20	40.0
Leervorderingen worden in het schooljaar gemonitord en geregistreerd	Luisteren	11	5	41.7	47	23	48.9
	Spreken	11	5	41.7	47	25	53.2
	Gesprekken	11	5	41.7	47	23	48.9
Er wordt gebruikgemaakt van een leerlingadministratiepakket	Luisteren	12	10	83.3	50	30	60.0
	Spreken	12	10	83.3	50	31	62.0
	Gesprekken	12	10	83.3	50	30	60.0
Er wordt gebruikgemaakt van genormeerde toetsen of observatielijsten	Luisteren	11	3	27.3	41	19	46.3
	Spreken	11	3	27.3	41	18	43.9
	Gesprekken	11	3	27.3	41	18	43.9
Er wordt gebruikgemaakt van methodetoetsen	Luisteren	11	5	45.4	41	10	24.4
	Spreken	11	5	45.4	41	10	24.4
	Gesprekken	11	5	45.4	41	10	24.4

Tabel 8.16. *Vervolg*

		so			sbo		
Monitoring en planning	Onderdeel	N	freq	%	N	freq	%
Er wordt gebruikgemaakt van zelfgemaakte instrumenten	Luisteren	11	2	18.2	41	4	9.8
	Spreken	11	2	18.2	41	4	9.8
	Gesprekken	11	2	18.2	41	4	9.8
Er wordt in de planning gebruikgemaakt van de leerresultaten van leerlingen	Luisteren	12	5	41.7	49	27	55.1
	Spreken	12	5	41.7	49	27	55.1
	Gesprekken	12	5	41.7	49	27	55.1
Er wordt gebruikgemaakt van groepsplannen	Luisteren	12	1	8.3	49	12	24.5
	Spreken	12	1	8.3	49	12	24.5
	Gesprekken	12	1	8.3	49	12	24.5

In Tabel 8.16 zien we ook hoe vaak verschillende soorten toetsen worden gebruikt voor monitoring. Het gebruik van genormeerde toetsen of observatielijsten is zowel in het so (25 procent) als in het sbo gangbaar (46.3 en 43.6 procent). In vergelijking met het sbo wordt in het so relatief vaker gebruikgemaakt van methodetoetsen (45.5 procent). Op iets **minder** dan de helft van de so-scholen (41.7 procent) en iets **meer** dan de helft van de sbo-scholen (55.1 procent) worden de leerresultaten in de planning van het onderwijs gebruikt. Van groepsplannen, ten slotte, wordt weinig gebruikgemaakt door so-scholen (8.3 procent) en sbo-scholen (24.5 procent).

Differentiatie

In Tabel 8.17 staan de aantallen en percentages scholen die verschillende vormen van differentiatie toepassen of leerlingen met leerproblemen extra ondersteunen. In het so wordt vooral gewerkt met convergente differentiatie (83.3 procent) en minder vaak met divergente differentiatie (25.0 procent), maar soms wordt dan wel aangegeven dat er van niveaugroepen gebruik wordt gemaakt (50.0 procent in totaal). In het sbo wordt zowel convergente als divergente differentiatie vaak toegepast (50.0 procent versus 48.0 procent). In het so wordt vaker (50.0 procent) gewerkt met niveaugroepen dan in het sbo (38.0 procent). Van de 19 sbo-scholen die met niveaugroepen werken, passen 10 scholen convergente differentiatie toe en 8 scholen divergente differentiatie. In zowel het so als het sbo wordt vaak gedifferentieerd naar instructie (75.0 procent versus 72.0 tot 74.0 procent). In het sbo vindt relatief vaker (44.9 en 43.7 procent) dan in het so ook differentiatie naar leerstof en (streef)doelen plaats (25.0 en 11.1 procent).

Tabel 8.17. *Overzicht differentiatie naar onderwijssoort*

		so			sbo		
Monitoring en planning	Onderdeel	N	freq	%	N	freq	%
Er vindt convergente differentiatie plaats	Luisteren	12	10	83.3	50	25	50.0
	Spreken	12	10	83.3	50	25	50.0
	Gesprekken	12	10	83.3	50	25	50.0
Er vindt divergente differentiatie plaats	Luisteren	12	3	25.0	50	24	48.0
	Spreken	12	3	25.0	50	24	48.0
	Gesprekken	12	3	25.0	50	24	48.0
Er wordt gewerkt met (twee of meer) niveaugroepen	Luisteren	12	6	50.0	50	19	38.0
	Spreken	12	6	50.0	50	19	38.0
	Gesprekken	12	6	50.0	50	19	38.0
Er is differentiatie naar instructie	Luisteren	12	9	75.0	50	37	74.0
	Spreken	12	9	75.0	50	36	72.0
	Gesprekken	12	9	75.0	50	36	72.0

Tabel 8.17. *Vervolg*

		so			sbo		
Monitoring en planning	Onderdeel	N	freq	%	N	freq	%
Er is differentiatie naar leerstof	Luisteren	12	3	25.0	49	22	44.9
	Spreken	12	3	25.0	49	22	44.9
	Gesprekken	12	3	25.0	49	22	44.9
Er is differentiatie naar (streef)doelen	Luisteren	9	1	9.1	48	21	43.7
	Spreken	9	1	9.1	48	21	43.7
	Gesprekken	9	1	9.1	48	21	43.7
Er zijn leerlingen die extra ondersteuning krijgen binnen de school	Luisteren	11	7	63.6	49	32	65.3
	Spreken	11	7	63.6	49	32	65.3
	Gesprekken	11	7	63.6	49	32	65.3

Teamontwikkeling

In Tabel 8.18 staan de kenmerken van het aspect van teamontwikkeling; hierbij is geen onderscheid gemaakt tussen de drie mondelinge taalvaardigheden. Het valt op dat sbo-scholen vaker mondelinge taalvaardigheid benoemd hebben in hun school- of taalbeleidsplan dan so-scholen (58.7 procent versus 36.4 procent). De vier so-scholen waarvoor dit wél geldt, bleken ook expliciete doelen geformuleerd te hebben voor Luisteren, Spreken en Gesprekken. Van de 27 sbo-scholen die aangeven mondelinge taalvaardigheid in het school- of taalbeleidsplan te benoemen, hebben er 21 expliciet doelen voor Luister- en Spreekvaardigheid en 22 voor Gespreksvaardigheid. In het sbo wordt vaker (65.9 procent) dan in het so genoemd dat er interne ondersteuning plaatsvindt (65.9 procent), terwijl in het so in vergelijking met het sbo juist vaker sprake is van externe ondersteuning (66.6 procent).

Tabel 8.18. *Overzicht teamontwikkeling met betrekking tot beleid en ondersteuning*

	So			sbo		
Beleid en ondersteuning	N	freq	%	N	freq	%
MTV is benoemd in het school- of taalbeleidsplan	11	4	36.4	46	27	58.7
Intervisiemomenten en collegiale consultatie	12	5	42.7	49	29	59.1
Interne ondersteuning (bijv. taalcoördinator)	12	4	33.3	47	31	65.9
Externe ondersteuning (bijv. onderwijsadviseur)	12	8	66.6	46	20	43.5

Door middel van een checklist is bij het interview ook gevraagd welke onderwerpen besproken worden in het teamoverleg (aangenomen dat dat er regelmatig is); deze staan in Tabel 8.19. Het onderwerp dat het meeste genoemd is als bespreekpunt in het overleg is de ondersteuning van individuele leerlingen met een spraak- of taalstoornis door de leerkracht (circa tweederde van de scholen) of door een externe specialist (ruim tweederde van de scholen). De aanpak van leerlingen die een achterstand in mondelinge taalvaardigheid hebben blijkens het leerlingvolgsysteem is geen agendapunt op het teamoverleg in de so-scholen, maar wordt wel besproken op bijna de helft van de sbo-scholen (48.9 procent). De vakdidactiek komt op slechts één so-school aan bod (10 procent) en op 12 sbo-scholen (26.7 procent). Het gebruik van de methode voor mondelinge taalvaardigheid staat daarentegen wat vaker op de agenda voor het overleg in zowel het so (40.0 procent van de scholen) als het sbo (62.2 procent van de scholen).

Tabel 8.19. *Overzicht van bespreekpunten in teamoverleg*

	So			sbo		
Bespreking in teamoverleg	N	freq	%	N	freq	%
Gebruik van methode MTV	10	4	40.0	45	28	62.2
Onderwijs in MTV in andere vakken	10	4	40.0	45	29	64.4
Volgen van de ontwikkeling op MTV	10	2	20.0	45	20	44.4
Uitkomsten van volgsysteem/registratie op MTV	10	2	20.0	45	15	33.3

Tabel 8.19. *Vervolg*

	So			sbo		
Bespreking in teamoverleg	N	freq	%	N	freq	%
Individuele leerlingen met spraak/taalstoornis en hun ondersteuning door de leerkracht	10	6	60.0	45	30	66.7
Individuele leerlingen met spraak/taalstoornis en hun ondersteuning door de externe specialist	10	7	70.0	45	33	73.3
Aanpak/begeleiding van de groep leerlingen die blijkens het volgsysteem een achterstand hebben op MTV	10	0	0.0	45	22	48.9
De leerdoelen van de school op het gebied van MTV	10	3	30.0	45	21	46.7
Vakdidactiek MTV	10	1	10.0	45	12	26.7

8.2 Duiding en verklaring leerlingprestaties

Een tweede doel van dit onderzoek was om de prestaties van leerlingen te relateren aan kenmerken van de taakuitvoering (onderzoeksvraag 2, zie Hoofdstuk 1). Zoals al eerder is aangegeven, wordt daartoe op itemniveau in kaart gebracht wat leerlingen op de grenzen <1F/1F en 1F/2F wel en (nog) niet beheersen. Daarnaast wordt de samenhang tussen de drie vaardigheden (d.w.z., Luisteren, Spreken en Gesprekken) en tussen subonderdelen van die vaardigheden onderzocht.

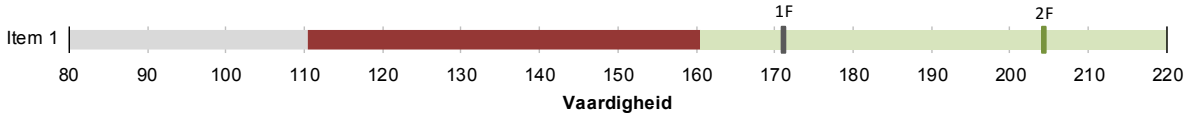
8.2.1 Beheersing op itemniveau

De luistertoets is geanalyseerd binnen het raamwerk van de item respons theorie. Met behulp van het item respons model kunnen we voorspellen wat de kans is op een juist antwoord gegeven de vaardigheid van een leerling. Andersom kan ook worden nagegaan welk niveau van vaardigheid benodigd is om een bepaalde kans te hebben op het correct maken van een item. Van deze tweede mogelijkheid is hier gebruikgemaakt. Beheersing is als volgt gedefinieerd:

Geen beheersing	kans op een correct antwoord kleiner dan 50 procent
Matige beheersing	kans op een correct antwoord ligt tussen de 50 en 80 procent
Goede beheersing	kans op een correct antwoord is groter dan 80 procent

Voor elk item is dus uitgerekend hoeveel vaardigheid nodig is om 50 of 80 procent kans te hebben op het correct maken van dat item en vervolgens is gekeken hoe dat zich verhoudt tot de hoeveelheid vaardigheid die nodig is om referentieniveau 1F of 2F aan te tonen. Figuur 8.2 laat grafisch zien hoe bij het onderdeel luisteren in kaart is gebracht welke items leerlingen op referentieniveau 1F en 2F in welke mate beheersen. Het groen gemarkeerde gedeelte in Figuur 8.2 reflecteert het vaardigheidsbereik waarin het item (goed) beheerst wordt, het rood gemarkeerde gedeelte het vaardigheidsbereik waarin het item matig beheerst wordt en het grijs gemarkeerde gedeelte, ten slotte, reflecteert het vaardigheidsbereik waarin het item nog helemaal niet beheerst wordt. Referentieniveaus 1F en 2F zijn ook in Figuur 8.2 afgebeeld. In dit voorbeeld kan geconcludeerd worden dat leerlingen die beschikken over referentieniveau 1F of hoger het (fictieve) item goed beheersen en toe zijn aan items van een hoger niveau.

Figuur 8.2. *Moeilijkheidsgraad van een (fictief) item in relatie tot referentieniveaus 1F en 2F op de (lineair getransformeerde) latente vaardigheidsschaal*



De onderdelen Spreken en Gesprekken zijn geanalyseerd binnen het raamwerk van de Klassieke Testtheorie. Binnen dat raamwerk is de beschreven werkwijze niet goed toe te passen, omdat een schaal waarop zowel de items als de vaardigheid worden afgebeeld, ontbreekt. Om toch op een vergelijkbare wijze in kaart te kunnen brengen welke items 1F en 2F leerlingen beheersen, is ook voor Spreken en Gesprekken een item respons theorie analyse verricht. Concreet is zowel voor Spreken (jeugdjournaalfragment en dobbelspel apart) als Gesprekken een *Partial Credit Model* geschat, waarbij wordt aangenomen dat (a) alle items gelijk discrimineren en (b) de items een beroep doen op één en dezelfde onderliggende dimensie. Hoewel een dergelijk (simplistisch) model in het licht van de gepresenteerde resultaten in Hoofdstukken 5 en 6 niet ideaal lijkt, volstaat het model om op globaal niveau

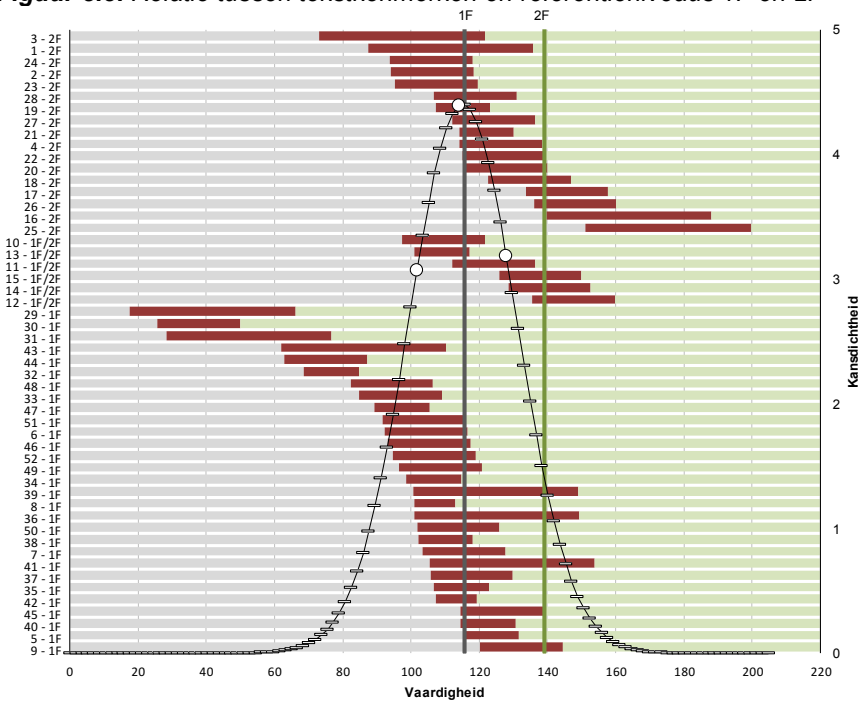
inzicht te krijgen in de relatie tussen de moeilijkheidsgraad van de items en de referentieniveaus. Vanwege de polytome scoring en het feit dat de parameters van een *partial credit* item niet op zichzelf te interpreteren zijn (zie Verhelst en Verstalen, 2008), is op basis van de parameters per item berekend welke vaardigheid nodig is om 50 of 80 procent van de maximale score te behalen, dus respectievelijk score 1 (50 procent) en score 1.6 (80 procent).

Om ook gevoel te krijgen bij de wijze van presteren van de leerlingen in het sbo en so, is de mate van beheersing niet alleen afgezet tegen de referentieniveaus, maar ook tegen de vaardigheidsverdeling van de totale onderzoeksgroep. Hoe dat is gedaan, is onder meer in Figuur 8.3 te zien. De vaardigheidsverdeling is weergegeven door een zwarte lijn; de ronde markeringspunten representeren respectievelijk (ongeveer) percentiel 20, percentiel 50 en percentiel 80. We zien in Figuur 8.3 dat de leerlingen die lager of gelijk aan percentiel 20 scoren veel items nog helemaal niet beheersen. Verder is uit het tweede markeringspunt (percentiel 50) af te leiden dat ongeveer de helft van de leerlingen onder 1F presteert en de andere helft boven 1F. Bij het beschrijven van de prestaties op itemniveau leggen we het accent op de referentieniveaus, maar indien gewenst kan op deze manier dus ook de relatie gelegd worden met de leerlingen. Daarbij is het wel belangrijk om de weergave te zien als impressie; de percentages die afgeleid kunnen worden uit de figuur kunnen iets afwijken van de percentages die te berekenen zijn op basis van het databestand.

Luisteren
Met het referentiekader in het achterhoofd is op verschillende manieren geprobeerd om de luistervaardigheid van 1F en 2F leerlingen inhoudelijk te duiden. Er is gekeken naar kenmerken van de tekst (1F of 2F), de taakuitvoering (begrijpen, interpreteren of samenvatten), de aard van de activiteit (luisteren of kijken en luisteren), en de gesprekvorm (monoloog of dialoog).

- ***Tekstkenmerken.*** De luisterfragmenten op 1F-niveau zijn niet lang en de luisterduur bedraagt 1 tot 6 minuten per tekst. Deze teksten hebben een eenvoudige opbouw en gaan over alledaagse, concrete onderwerpen en over onderwerpen die aansluiten bij de leefwereld van de leerling. De informatie is herkenbaar geordend met een duidelijk gebruik van verwijs-, verbindings- en signaalwoorden. De teksten hebben bovendien een lage informatiedichtheid: informatie wordt herhaald en er wordt niet te veel (nieuwe) informatie tegelijkertijd geïntroduceerd. De luisterfragmenten op 2F-niveau zijn daarentegen iets langer met een gemiddelde luisterduur van 8 minuten per tekst. Deze teksten hebben een heldere structuur en gaan over alledaagse onderwerpen en sluiten aan bij de leefwereld van de leerlingen of staan iets verder van de leerling af. In deze teksten worden verbanden en denkstappen duidelijk aangegeven en is sprake van een redelijke informatiedichtheid. De items bij de 1F- en 2F-luisterfragmenten zijn in principe op hetzelfde niveau geconstrueerd. Figuur 8.3 laat zien in hoeverre de leerlingen dat ook daadwerkelijk zo ervaren.

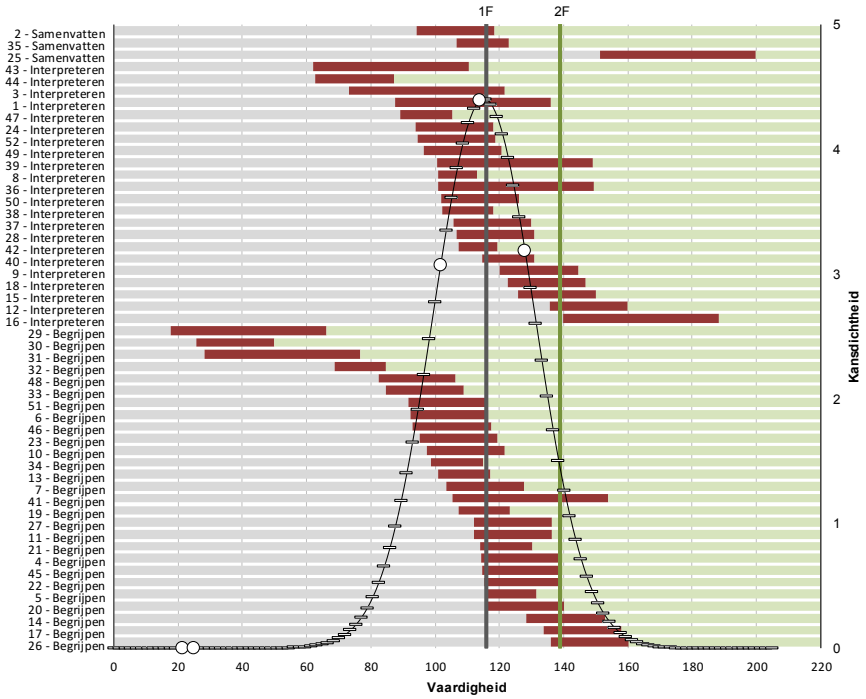
Figuur 8.3. *Relatie tussen tekstkenmerken en referentieniveaus 1F en 2F*



In Figuur 8.3 is te zien dat de 1F-items over het algemeen gemakkelijker zijn dan de 2F-items. De itemconstructie lijkt daarmee de bedoelde uitwerking te hebben gehad en bovendien kan voorzichtig geconcludeerd worden dat zaken als luisterduur en informatiedichtheid samenhangen met het referentieniveau waarover leerlingen beschikken. Leerlingen op niveau 2F kunnen langere teksten met een hogere informatiedichtheid aan dan leerlingen op niveau 1F.

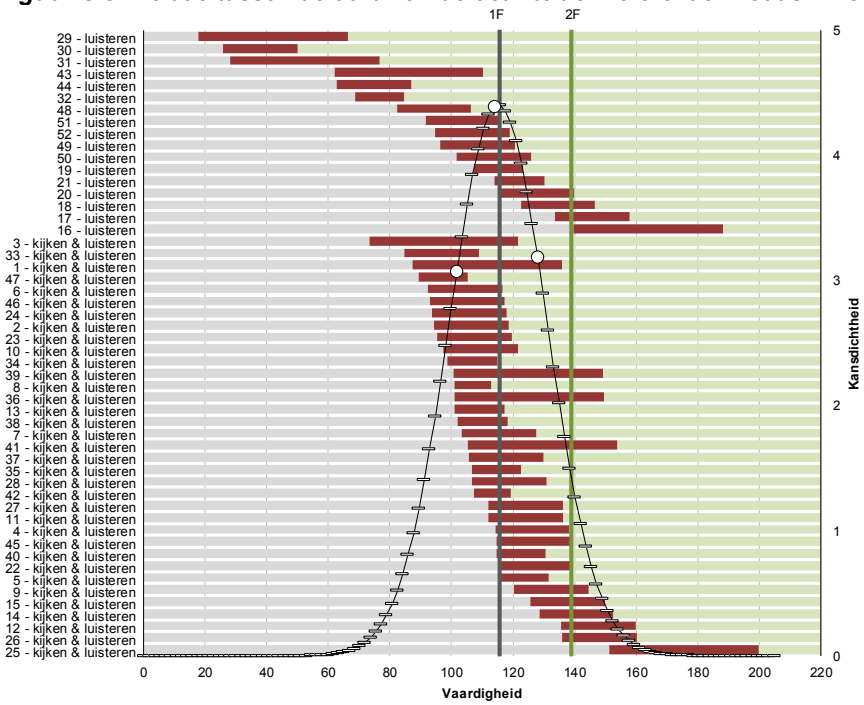
- ***Taakuitvoering.*** Omdat de leerlingen bij het subonderdeel Begrijpen vooral elementen reproduceren die letterlijk in de luisterfragmenten worden genoemd en ze bij de subonderdelen Interpreteren en Samenvatten een vertaalslag moeten maken van elementen uit de luisterfragmenten, lijkt het op het eerste gezicht aannemelijk dat Begrijpen eenvoudiger is dan Interpreteren en Samenvatten. Figuur 8.4 ondersteunt deze aanname echter niet. Op enkele extreme, incidentele uitschieters na is er weinig verschil in moeilijkheid tussen de items Begrijpen, Interpreteren en Samenvatten. Er zijn kennelijk zowel bij Begrijpen, Interpreteren en Samenvatten gemakkelijke, gemiddelde en moeilijke items te construeren op niveau 1F en 2F en dat ondersteunt het gegeven dat luisteren een concentrische vaardigheid is, waarbij leerlingen vanaf het prille begin álle te onderscheiden aspecten van het luisteren (of kenmerken van de taakuitvoering) gelijktijdig en geïntegreerd inzetten.

Figuur 8.4. *Relatie tussen kenmerken van de taakuitvoering en referentieniveaus 1F en 2F*



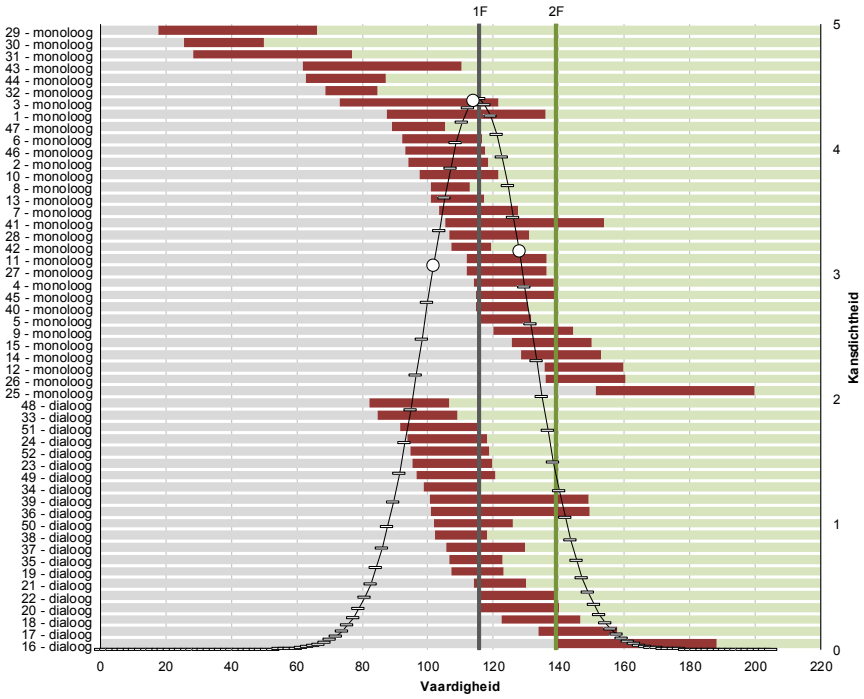
- ***Aard van de activiteit.*** Bij de meeste luisterfragmenten in deze peiling moesten de leerlingen kijken én luisteren naar teksten van televisie of internet. Dit betreft materiaal waarin beeld en geluid gecombineerd worden aangeboden. De leerlingen hebben echter ook fragmenten beluisterd waarbij beeld ontbrak. Dit betrof fragmenten van de radio en een zestal fragmenten met een luisterduur van een minuut of minder. Figuur 8.5 laat zien dat er meer variatie in moeilijkheid is in de luisterfragmenten waarnaar de leerlingen alleen luisteren dan in de luisterfragmenten waarnaar zij kijken én luisteren. In dat laatste geval liggen vrijwel alle items rond of tussen niveau 1F en 2F en zijn er geen extreem gemakkelijke of moeilijke items. De koppeling naar 1F en 2F is wel enigszins lastig te maken. Zowel voor luisteren als voor kijken en luisteren zijn vanwege de concentrische opbouw van de vaardigheid (zie ook taakuitvoering) gemakkelijke, gemiddelde en moeilijke items te construeren.

Figuur 8.5. *Relatie tussen de aard van de activiteit en referentieniveaus 1F en 2F*



- ***Gespreksvorm.*** De luisterfragmenten zijn te verdelen in monologen en dialogen. In Figuur 8.6 is te zien dat er bij de monologen (verhalen, berichten en documentaires) sprake is van een brede bandbreedte in de moeilijkheid van de items. Bij de dialogen (gesprekken in de vorm van interviews) is de bandbreedte veel smaller. In tegenstelling tot de dialogen horen bij de monologen relatief veel gemakkelijke items. Mogelijk zijn de leerlingen minder snel afgeleid bij het luisteren naar monologen, omdat ze gerichter luisteren, er is immers maar een persoon aan het woord. De luisterduur van de monologen is bovendien aanzienlijk korter dan die van de dialogen, zeker waar het gaat om de berichten en verhalen. Bij de berichten ligt dit voor de hand: in het dagelijks leven hebben berichten meestal ook een beperkte omvang. Bij verhalen ligt dit anders: de luisterduur daarvan kan in het dagelijkse en schoolse leven variëren van kort tot lang.

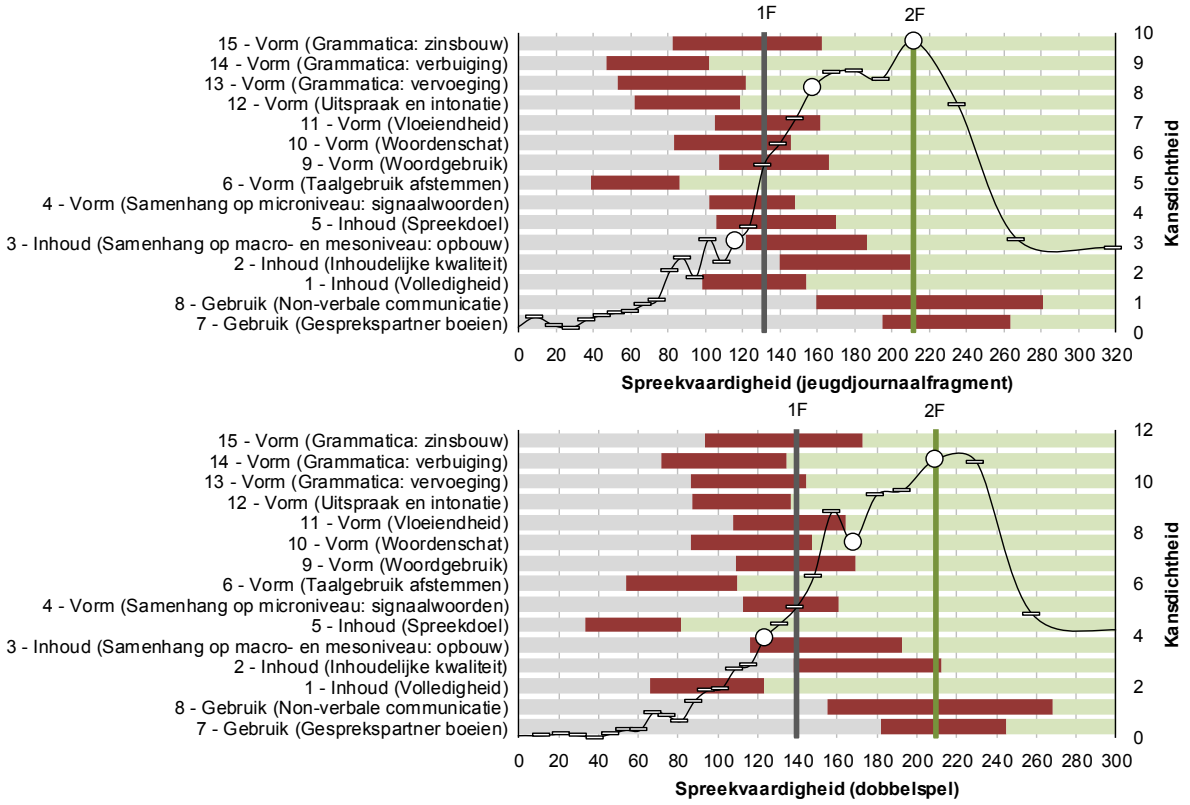
Figuur 8.6. *Relatie tussen de gespreksvorm en referentieniveaus 1F en 2F*



Spreken
Uit de data bij Spreken bleek dat de vijf kenmerken van de taakuitvoering zoals opgenomen in het referentiekader, te weten Samenhang, Afstemming op doel, Afstemming op publiek, Woordgebruik en woordenschat en Vloeiendheid, verstaanbaarheid en grammaticale beheersing niet empirisch te onderscheiden zijn. Wel liet factoranalyse zien dat er sprake is van drie factoren die we aangeduid hebben als: Taalvorm, Taalinhoud en Taalgebruik (zie paragraaf 5.5). We hebben geprobeerd om de spreekvaardigheid van de leerlingen die functioneren op niveau 1F of 2F inhoudelijk te duiden door de items te ordenen naar deze drie factoren en de vijf kenmerken van de taakvoering vervolgens mee te nemen bij de interpretatie. Figuur 8.7 laat zien welke items leerlingen op niveaus 1F en 2F al dan niet beheersen. De bovenste helft van Figuur 8.7 heeft betrekking op het jeugdjournaalfragment en de onderste helft op het dobbelspel.

In Figuur 8.7 (boven) is te zien dat leerlingen op niveau 1F nauwelijks tot geen moeite hebben met de Taalvorm als ze een monoloog houden op basis van het jeugdjournaalfragment. Bij het verbuigen van naamwoorden en het vervoegen van werkwoorden ondervinden 1F-leerlingen geen problemen. Ze zijn ook goed verstaanbaar en ze gebruiken consequent taal die bij de gesprekspartner past. Verder beschikken ze over voldoende woorden om over het filmpje te praten, maar het woordgebruik is nog niet erg gevarieerd. De taalinhoud is bij de 1F-leerlingen net wél of net niet voldoende op orde. De 1F-leerlingen lijken de strekking van het filmpje duidelijk te kunnen maken aan de gesprekspartner, maar de inhoudelijke uitwerking en opbouw is nog niet optimaal. Op niveau 1F lijkt het onderdeel Taalgebruik, waarbij het gaat om boeiend vertellen en de inzet van non-verbale communicatie, nog te hoog gegrepen. Dit kan te maken hebben met de gesprekvorm: bij het houden van een monoloog is het lastiger en minder natuurlijk om contact met de luisteraar te maken dan bij een dialoog. Het gaat echter om slechts twee items, waardoor het de vraag is of er sprake is van een representatief beeld. Bij leerlingen op niveau 2F vormt alleen het Taalgebruik nog een uitdaging. De aspecten die onder Taalvorm en Taalinhoud vallen, beheersen zij goed.

Figuur 8.7. *Relatie tussen de subonderdelen van Spreken en referentieniveaus 1F en 2F*



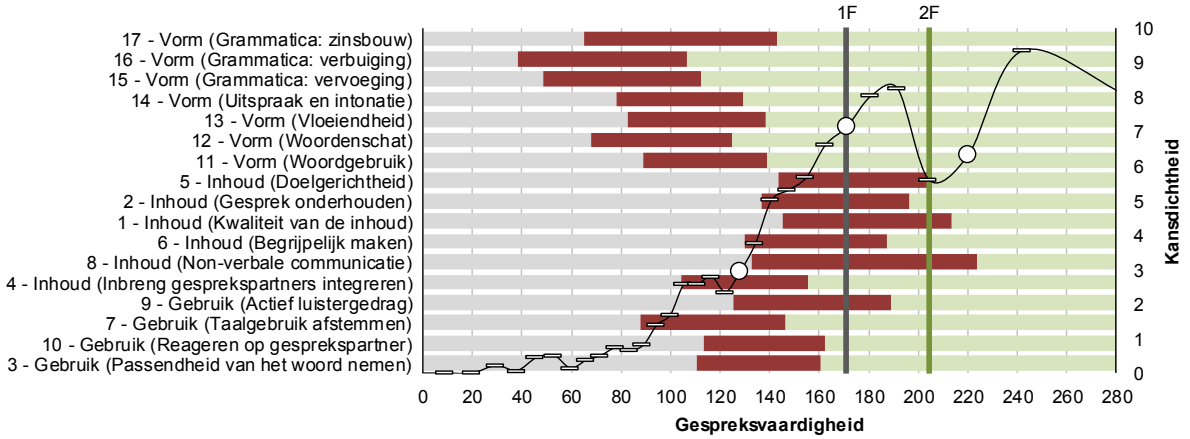
De onderste helft van Figuur 8.7 laat zien dat het de leerlingen op niveau 1F (en dus ook de leerlingen op niveau 2F) ook al aardig lukt om een monoloog in verhaalvorm (dobbelspel) te houden. Verder komen de elementen die wel en niet beheerst worden vrijwel één-op-één overeen met die bij het jeugdjournaalfragment. Dit betekent onder andere dat de leerlingen weinig moeite hebben met Taalvorm en dat het Taalgebruik aanzienlijk lastiger is voor de leerlingen. Bij Taalinhoud is het opmerkelijk dat item 1 en 5 bij het dobbelspel aanzienlijk beter beheerst worden dan bij het jeugdjournaalfragment. Concreet gaat het om de volgende punten: (Volledigheid – 1) De leerling heeft (bijna) alle taalhandelingen van de opdracht uitgevoerd, en (Spreekdoel – 5) De leerling kan de strekking van het filmpje duidelijk maken aan de

gesprekspartner (jeugdjournaalfragment) of de leerling vertelt een verhaal waar hij alle plaatjes in verwerkt (dobbelspel). Kennelijk slagen de leerlingen hier bij het creatief vertellen aan de hand van dobbelstenen aanzienlijk beter in dan bij het overbrengen van feiten aan de hand van een jeugdjournaalfragment.

Gesprekken
Ook uit de data bij Gesprekken bleek dat de vijf kenmerken van de taakuitvoering zoals opgenomen in het referentiekader, namelijk Beurten nemen en bijdragen aan samenhang, Afstemming op doel, Afstemming op de gesprekspartner(s), Woordgebruik en woordenschat en Vloeiendheid, verstaanbaarheid en grammaticale beheersing niet empirisch te onderscheiden zijn. Wel liet factoranalyse zien dat er sprake is van drie factoren die we aangeduid hebben als: Taalvorm, Taalinhoud en Taalgebruik (zie paragraaf 6.5). We hebben opnieuw geprobeerd om de gespreksvaardigheid van de leerlingen die functioneren op niveau 1F en 2F inhoudelijk te duiden door de items te ordenen naar deze drie factoren en bij de interpretatie vervolgens de vijf kenmerken van de taakvoering mee te nemen. Figuur 8.8 laat zien welke items leerlingen op niveaus 1F en 2F al dan niet beheersen.

Bij het *Fischerspiel* spelen de leerlingen een spel in groepjes waarbij er sprake is van een polyloog. De leerlingen moeten met hun medeleerlingen overleggen en informatie uitwisselen om tot oplossingen te komen. Daarbij gaat de Taalvorm de leerlingen op niveau 1F erg makkelijk af: de woordenschat van 1F-leerlingen is voldoende om het gesprek te kunnen voeren, ze spreken redelijk vloeiend en zijn goed verstaanbaar, ondanks een eventueel accent (zie Figuur 8.8). Ook de aspecten met betrekking tot grammatica (zinsbouw, verbuiging en vervoeging) zijn goed op orde. Taalinhoud blijkt voor 1F-leerlingen nog aanzienlijk lastiger te zijn. Het bijdragen aan samenhang door het beginnen, in stand houden en beëindigen van het gesprek is voldoende, maar kan beter. Verder is de gedachtegang van1F-leerlingen nog niet altijd begrijpelijk en is de inhoudelijke uitwerking niet per se sterk. 1F-leerlingen hebben hun Taalgebruik al wel redelijk onder de knie. Hun reacties zijn consequent gepast, ze nemen doorgaans op passende wijze het woord, en ze zijn in staat om de inbreng van de gesprekspartners te integreren in de eigen bijdrage. Het is opvallend dat de leerlingen de factor Taalgebruik in Gesprekken op niveau 1F (behoorlijk) goed beheersen en dat deze factor bij Spreken nog de nodige problemen oplevert. Wellicht heeft dit te maken met de gespreksvorm: bij het houden van een monoloog is het mogelijk lastiger en minder natuurlijk om contact met de luisteraar(s) te maken dan bij een polyloog. Waar Taalinhoud voor 1F-leerlingen nog enigszins problematisch is, is dat bij 2F-leerlingen niet meer het geval; zij beheersen alle beoordeelde aspecten in goede mate.

Figuur 8.8. *Relatie tussen de subonderdelen van Gesprekken en referentieniveaus 1F en 2F*



8.2.2 Samenhang tussen (sub)onderdelen

De samenhang tussen de drie vaardigheden, en alle subonderdelen van die vaardigheden, is in kaart gebracht met het computerprogramma MULTI (Kamphuis, 1993, 1998). In MULTI is het mogelijk om het meetmodel expliciet mee te nemen. In geval er sprake is van een item respons model wordt de correlatie direct berekend op basis van de (gewogen) somscores van de leerlingen en de itemparameters, zodat optimaal wordt gecorrigeerd voor onbetrouwbaarheid. Deze werkwijze is toegepast voor (alle subonderdelen van) luisteren. In geval het klassieke testmodel is gebruikt, kan de betrouwbaarheid (of de meetfoutvariantie) gespecificeerd worden. Dit is van toepassing bij (alle subonderdelen van) Spreken en Gesprekken. Omdat het nogal problematisch is om bij Spreken en Gesprekken de betrouwbaarheid realistisch te schatten (zie Hoofdstuk 5 en 6), is de betrouwbaarheid (bij gebrek aan beter) gelijkgesteld aan 0.80. Bij alle subonderdelen van Spreken en Gesprekken is de betrouwbaarheid vanwege het kleinere aantal items gelijkgesteld aan 0.70.

In het verlengde van paragraaf 8.2.1 is bij het in kaart brengen van de samenhang naar allerlei verschillende opdelingen gekeken. Eerst is gekeken naar de samenhang tussen de drie mondelinge taalvaardigheden. De (latente) correlatie tussen Luisteren en Spreken is gelijk aan 0.394, die tussen Luisteren en Gesprekken is 0.237 en die tussen Spreken en Gesprekken is 0.424. Deze correlaties zijn nogal laag. Dat betekent dat de drie mondelinge taalvaardigheden apart beschouwd moeten worden en dat vanuit de ene (mondelinge taal-) vaardigheid niet goed is af te leiden wat we op een andere (mondelinge taal-) vaardigheid van een leerling mogen verwachten. Dat de correlatie tussen Spreken en Gesprekken het hoogst is, ligt voor de hand omdat deze twee (deel)vaardigheid inhoudelijk dichter bij elkaar liggen dan Luisteren en Spreken of Luisteren en Gesprekken.

In een vervolgstap is nagegaan hoe de subonderdelen ***binnen*** de drie vaardigheden met elkaar samenhangen.

- ***Luisteren.*** Tabel 8.20 laat zien dat de correlaties tussen de subonderdelen van Luisteren zeer hoog zijn. De meeste correlaties zijn .850 of hoger en bij de correlaties die kleiner zijn, ligt de verklaring daarvoor vermoedelijk in het aantal items en/of in een *restriction of range*. Bij het subonderdeel Samenvatten horen bijvoorbeeld slechts drie items en dat aantal is dusdanig klein dat de correlatie niet goed uit te rekenen is. Het is onduidelijk waarom de correlatie tussen dialogen en monologen slechts .767 is en ook is onduidelijk waarom monologen slechts beperkt samenhangen (0.740) met items waarbij leerlingen hun luistervaardigheid moeten tonen zonder daarbij beeld te zien. Al met al ondersteunt Tabel 8.20 wel de eerdere keuze om luistervaardigheid te modelleren als een unidimensionaal construct (zie Hoofdstuk 4).

  Tabel 8.20. *Samenhang tussen subonderdelen van Luisteren.*

Nr	Schaal	1	2	3	4	5	6	7
1	Taakuitvoering - Begrijpen	1.000						
2	Taakuitvoering - Interpreteren	0.938	1.000					
3	Taakuitvoering - Samenvatten	0.768	0.788	1.000				
4	Activiteit - kijken & luisteren	0.929	0.991	0.848	1.000			
5	Activiteit - luisteren	0.972	0.853	0.707	0.830	1.000		
6	Gespreksvorm - dialoog	0.973	0.879	0.821	0.880	0.976	1.000	
7	Gespreksvorm - monoloog	0.863	0.976	0.741	0.972	0.740	0.767	1.000

- ***Spreken.*** De latente correlatie tussen feiten overbrengen (bij het jeugdjournaalfragment) en creatief vertellen (bij het dobbelspel) is gelijk aan 0.985. Deze waarde wijkt iets af van de waarde die in paragraaf 5.3 genoemd is, maar dat komt doordat hier is gerekend met een andere (lagere) betrouwbaarheid. Geconcludeerd kan worden dat navertellen (of feiten overbrengen) en creatief vertellen niet zinvol als aparte entiteiten onderscheiden kunnen worden. Tevens is gekeken naar de samenhang tussen de drie eerder benoemde factoren: Taalvorm, Taalinhoud en Taalgebruik. De eerste twee hangen behoorlijk sterk met elkaar samen (0.898), maar Taalgebruik lijkt een zelfstandig subonderdeel te vormen. De correlatie tussen Taalgebruik en Taalvorm is gelijk aan 0.562 en die tussen Taalgebruik en Taalinhoud is gelijk aan 0.589. Een dergelijk resultaat is niet vreemd als we bedenken dat Taalgebruik (in dit onderzoek) zeer specifiek gedefinieerd is in termen van non-verbaal gedrag en het bewust gebruik van stiltes, wisselende intonatie, volume en spreeksnelheid.

- ***Gesprekken.*** Ook bij Gesprekken is de samenhang tussen Taalvorm, Taalinhoud en Taalgebruik onderzocht. Het beeld wijkt nogal af van het beeld dat we vonden bij Spreken. De correlaties liggen tussen 0.516 (Taalgebruik versus Taalvorm) en 0.693 (Taalgebruik versus Taalinhoud). Voornoemde correlaties liggen weliswaar in lijn met die van Spreken, maar de correlatie tussen Taalinhoud en Taalvorm ligt ook in die range (0.688) en dat was bij Spreken niet zo. Vermoedelijk is het afwijkende beeld te verklaren vanuit verschillen in operationalisatie. Waar het non-verbale gedrag bij Gesprekken bijvoorbeeld vooral bij Taalinhoud hoorde en minder bij Taalgebruik (zie paragraaf 6.5) viel het non-verbale gedrag bij Spreken juist zeer duidelijk onder de factor Taalgebruik. Dergelijke nuances hebben hun weerslag op de correlatie, maar aan het gesignaleerde verschil in interne structuur bij Spreken en Gesprekken zijn daardoor niet direct allerlei (inhoudelijke) conclusies te verbinden.

Als derde is gekeken naar de samenhang tussen Luisteren naar instructie (in het samenwerkingsspel) en Luisteren naar fragmenten (bij de luistertoets) al dan niet ondersteunt met beeld. De correlaties zijn gelijk

aan 0.265 (met beeld) en 0.292 (zonder beeld) en daarmee als zeer laag aan te merken. Het is gissen naar een inhoudelijke verklaring. Mogelijk is de afnamesetting dusdanig verschillend (pen-en-papier toetsen versus een bordspel) dat de correlatie voornamelijk dat verschil reflecteert en minder een inhoudelijk verschil tussen subonderdelen van luisteren. In het licht van de correlaties die we eerder vonden binnen de luistertoets is het niet aannemelijk dat Luisteren naar **instructie** een geheel eigen entiteit vormt ten opzichte van Luisteren naar **radio, televisie of internet**.

Als vierde is gekeken naar de relatie tussen Taalvorm, Taalinhoud en Taalgebruik bij Spreken en Taalvorm, Taalinhoud en Taalgebruik bij Gesprekken. Eerder is al opgemerkt dat de correlationele structuur er bij Spreken anders uitziet dan bij Gesprekken, maar het is de vraag in hoeverre de drie factoren overkoepelend als afzonderlijke elementen te onderscheiden zijn. Zoals te zien is in Tabel 8.21 levert de correlatiematrix daarvoor geen bewijs. Taalvorm is nog het beste als afzonderlijk overkoepelend (sub)onderdeel te onderscheiden, maar verder vertonen de correlaties niet het patroon dat we zouden verwachten. De binnen-factorcorrelaties zijn bijvoorbeeld niet hoger dan de tussen-factorcorrelaties en dat zou je wel verwachten als er sprake is van drie te onderscheiden subonderdelen.

Tabel 8.21. *Samenhang tussen subonderdelen bij Gesprekken en Spreken*

		Spreken		
Vaardigheid	Schaal	Taalvorm	Taalinhoud	Taalgebruik
Gesprekken	Taalvorm	0.661	0.403	0.205
	Taalinhoud	0.434	0.390	0.430
	Taalgebruik	0.291	0.158	0.127

Ten slotte is een opdeling gemaakt naar (meer) theoretische/zakelijke taken enerzijds (de luistertoets en, in mindere mate, het navertellen van het filmfragment), en meer praktische/speelse taken anderzijds (het Fischerspiel en het dobbelspel). In deze vergelijking wordt ingestoken op de constatering van leerkrachten in het so en sbo dat leerlingen in praktische situaties ‘anders’ scoren dan wanneer zij een meer theoretische opdracht, bijvoorbeeld op papier, uitvoeren. In de correlatiematrix is inderdaad heel voorzichtig te zien dat **binnen**-taaktype-correlaties iets hoger zijn dan de **tussen**-taaktype-correlaties.
De correlatie tussen de theoretische/zakelijke taken is gelijk aan 0.415 en de correlatie tussen de praktische/speelse taken is gelijk aan 0.454. Als we de theoretische/zakelijke taken en de praktische/speelse taken met elkaar vergelijken vinden we correlaties van 0.384 (dobbelspel versus luistertoets), 0.237 (*Fischerspiel* versus luistertoets) en 0.431 (*Fischerspiel* versus navertellen van het jeugdjournaalfragment). Op één na zijn deze correlaties iets lager, maar de verschillen zijn niet erg groot. De vergelijking wordt bemoeilijkt doordat niet alleen het type taak, maar ook het type vaardigheid de correlatie beïnvloedt.

8.3 Trends in luisteren

Een derde doel bij deze peiling was om na te gaan hoe de prestaties van schoolverlaters in het sbo in 2018 zijn veranderd ten opzichte van eerdere peilingsjaren. Idealiter zouden we dit nagaan voor alle drie mondelinge taalvaardigheden, maar dat was niet mogelijk. Gespreksvaardigheid is in eerdere peilingsonderzoeken namelijk niet gemeten en Spreken is in deze peiling, om inhoudelijke redenen, dusdanig anders geoperationaliseerd dan in eerdere onderzoeken dat een vergelijking van scores niet zinvol is. Voor het onderdeel Luisteren kunnen we wel een vergelijking maken tussen de prestaties van de sbo-leerlingen in het huidige onderzoek en die in het peilingsonderzoek luistervaardigheid in 2007 (PPON2007; zie Krom, Van Berkel, Van der Schoot, Sijtstra, Hemker & Marsman, 2011). In paragraaf 8.3.1 lichten we toe hoe we met behulp van item respons theorie de leerlingscores uit de twee peilingen op één schaal hebben geplaatst om ze vergelijkbaar te maken. Vervolgens gaan we in paragraaf 8.3.2 in op de luistervaardigheid van de sbo-leerlingen in 2007 en in 2018.

8.3.1 Equivalering PPON2007 en Peil.Onderwijs

Net zoals in deze peiling is in PPON2007 een meetschaal voor luisteren gemaakt met behulp van het OPLM. Om een vergelijking mogelijk te maken, moet (a) een transformatie worden bepaald om scores van de ene schaal naar de andere te vertalen **óf** (b) een zogeheten *concurrent calibration* worden uitgevoerd, waarbij de datasets van de twee peilingen samen worden geanalyseerd en er één gezamenlijke (OPLM) meetschaal geconstrueerd wordt. Wij hebben voor de tweede aanpak gekozen, omdat de leerlingscores dan direct op dezelfde meetschaal staan en er mogelijkheden zijn om DIF tussen PPON2007 en deze

peiling in kaart te brengen. Dat laatste is onder meer belangrijk, omdat de validiteit van de equivalering afhangt van de mate waarin de (anker)items vergelijkbaar functioneren in de twee onderzoeken. Zowel verschillen in moeilijkheid (uniforme DIF) als verschillen in het onderscheidende vermogen (non-uniforme DIF) zijn relatief eenvoudig te beoordelen door in de OPLM-software de empirische itemresponscurves van verschillende groepen leerlingen afzonderlijk te plotten en deze met elkaar, en de model-voorspelde itemresponscurves, te vergelijken.

Voor de analyse is het databestand van deze peiling aangevuld met 7747 records uit de data van PPON 2007. Van de 7747 records waren er 1157 van sbo-leerlingen (aan het onderzoek namen destijds ook leerlingen uit leerjaar 5 tot en met 8 van het reguliere basisonderwijs deel). Omdat een aantal van de ankeritems in 2007 (ook) gemaakt werd door de leerlingen uit het reguliere basisonderwijs, namen we al deze leerlingen mee in de gezamenlijke kalibratie. Eerst hebben we de itemgewichten geschat, waarbij het harmonisch gemiddelde is ingesteld op 2. Enkele items uit de oude (PPON2007) dataset zijn vervolgens verwijderd in verband met een zeer zwak of zelfs negatief onderscheidend vermogen. Dit vormt voor onze analyse geen probleem, omdat het geen ankeritems betrof. Vervolgens hebben we de itemmoeilijkheden geschat met de CML-methode. De modelpassing van het OPLM was adequaat; de *R1c*-waarde van het OPLM was bijvoorbeeld kleiner dan anderhalf keer het aantal vrijheidsgraden ($R1c = 5691.6$, $df = 4505$, $p < .001$) en op itemniveau waren er geen ernstige modelschendingen te zien. Verder waren er geen aanwijzingen voor DIF bij de elf ankeritems; zowel qua moeilijkheid als onderscheidende vermogen functioneerden de ankeritems in 2007 en 2018 vergelijkbaar. Er is dan ook geconcludeerd dat het model goed bruikbaar is om de vaardigheidsverdelingen van de twee groepen sbo-leerlingen te vergelijken. Dat is gedaan door op basis van de itemparameters en de leerlingscores de latente vaardigheid van individuele leerlingen te bepalen met de WML-schatter. Met het oog op de interpretatie hebben we vervolgens de vaardigheidsscores omgezet naar de corresponderende *bankscores* op de 52 items van de peiling in 2018.

8.3.2 Prestatievergelijking 2007 en 2018

Bij het vergelijken van de prestaties van de sbo-leerlingen in 2007 en in 2018 is het van belang om rekening te houden met de samenstelling van de onderzoeksgroepen. In het peilingsonderzoek van 2007 ging de interesse uit naar 12- en 13-jarige sbo-leerlingen, terwijl in het onderzoek van 2018 schoolverlaters (en niet een bepaalde leeftijdsgroep) de doelgroep vormden. Zoals in Tabel 8.22 te zien is, namen aan PPON2007 relatief meer oudere sbo-leerlingen deel (13-jarigen) en aan Peil.Onderwijs 2018 juist meer jonge leerlingen (11-jarigen). Om een zinnige inschatting te maken van de trends, vergelijken we de leerlingprestaties daarom apart voor de leerlingen van 12 en 13 jaar. Dit zijn de twee leeftijdsgroepen waarvan we in beide datasets voldoende gegevens hebben.

Tabel 8.22. *Aantal sbo-leerlingen naar leeftijd in 2007 en 2018*

Leeftijd	*Aantal 2007*	*Aantal 2018*
10 jaar	2	4
11 jaar	22	188
12 jaar	661	649
13 jaar	461	131
14 jaar	11	3
Onbekend	1	166

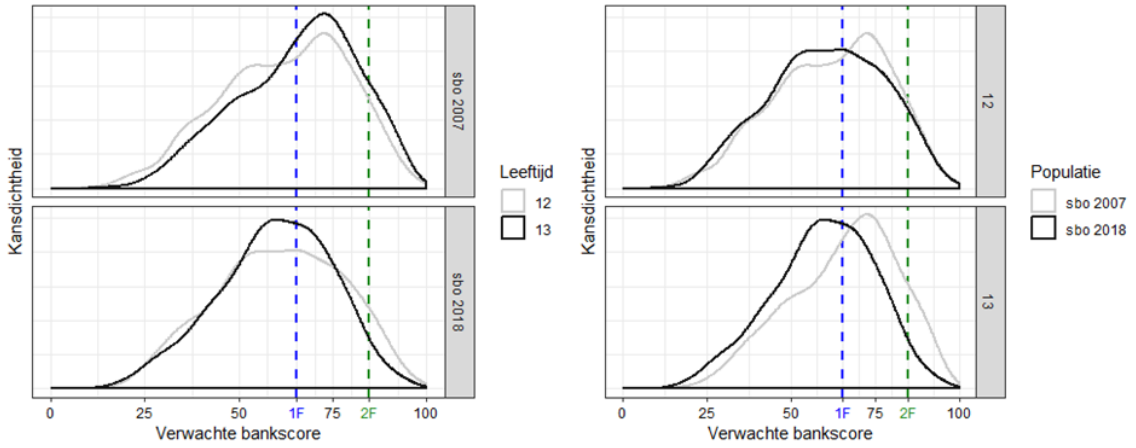
Tabel 8.23 geeft per leeftijdsgroep en peilingsjaar een aantal samenvattende statistieken en Figuur 8.9 toont de scoreverdelingen. De prestaties zijn weergegeven als (model-voorspelde) bankscores (percentage correct) op de 52 luisteritems van de peiling in 2018. Daarnaast is op basis van de bankscores nagegaan hoeveel procent van de leerlingen presteert op het 1F- of 2F-niveau. De bankscores die corresponderen met het 1F- en 2F-niveau zijn in Figuur 8.9 weergegeven met een blauwe en groene stippellijnen. De verdelingen van de leerlingprestaties worden in het linkerpaneel op zo'n manier naast elkaar gezet dat vooral goed te zien is hoe de twee leeftijdsgroepen binnen een peilingsjaar verschillen. In het rechterpaneel is daarentegen goed te zien hoe de prestaties tussen de peilingsjaren verschillen voor een leeftijdsgroep.

Tabel 8.23. *Vergelijking tussen leerlingprestaties in 2007 en 2018.*

	12 jarigen		13 jarigen	
Onderzoeksjaar	2007	2018	2007	2018
Aantal leerlingen	661	649	461	131
Gemiddelde	62.1	60.7	65.7	59.6
Standaardafwijking	16.9	16.8	16.0	15.0
% onder 1F	54.0	58.2	45.8	61.1
% 1F of 2F	46.0	41.8	54.2	38.9
% 2F	8.3	7.6	12.1	2.3
t-test van verschil tussen peilingsjaren	$t(1307.7) = 1.5$ $p = .12$		$t(220.4) = 4.1$ $p < .001$	

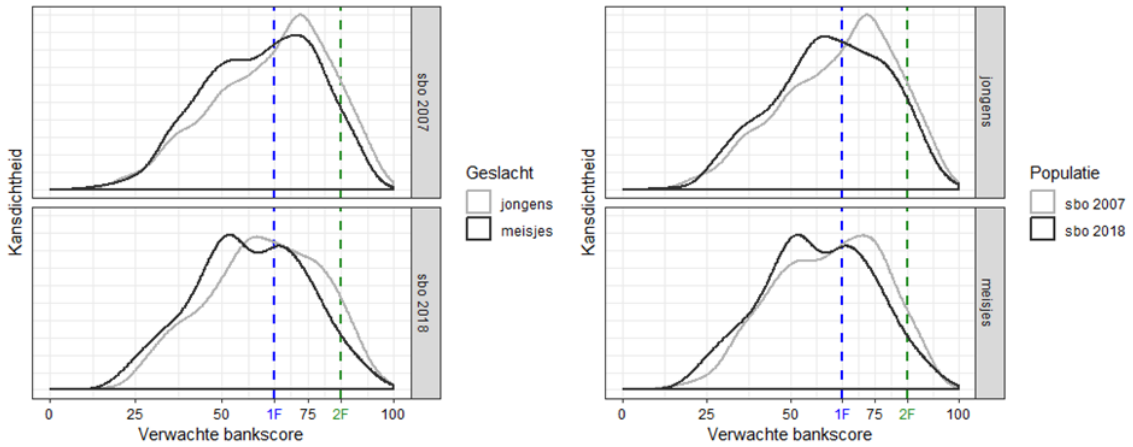
Zowel bij de 12- als 13-jarigen is de gemiddelde luistervaardigheid in 2018 lager dan in 2007. Voor de groep van 12-jarigen is dit verschil klein en statistisch niet significant; voor de 13-jarigen is het verschil beduidend groter (6 procent van de bankscore) en wel significant. In overeenstemming met deze trend in de gemiddelde vaardigheid van de populaties zien we dat in 2007 meer leerlingen voldeden aan de (hier gehanteerde) standaarden voor het 1F- en 2F-niveau.

Figuur 8.9. *Vaardigheidsverdeling van 11-, 12- en 13-jarigen in 2007 en 2018*



Naast leeftijd beschikken we ook over de variabelen geslacht, herkomst en thuistaal in de twee datasets. Vanwege definitieverschillen is een vergelijking naar herkomst en thuistaal niet betekenisvol, maar naar verschillen tussen jongens en meisjes kan wel gekeken worden. In Figuur 8.10 zijn de vaardigheidsverdelingen van de jongens en meisjes in de twee peilingen weergegeven. In de rapportage over PPON2007 werd vermeld dat meisjes in het sbo een iets lagere luistervaardigheid hadden dan jongens, alhoewel het verschil alleen statistisch significant was voor de subvaardigheid *Interpreteren* en niet voor *Begrijpen*. De algemene conclusie luidde dat het geslachtsverschil "… zo klein was dat er niet veel betekenis aan gehecht kan worden". In de onderzoeksgroep van 2018 zien we dat jongens significant beter presteren op Luistervaardigheid dan meisjes, $t(807.7) = 3.5$, $p < .001$. De gemiddelde bankscore van jongens was 62.1 en die van meisjes 58.3. Omdat de groep jongens aanzienlijk groter is dan de groep meisjes ($N = 593$ tegenover 375), gebruiken we voor de effectgrootte *Hedges's g*, een variant op *Cohen's d* die rekening houdt met ongelijke groepsgroottes. De effectgrootte van het verschil tussen jongens en meisjes is met $g = 0.23$ als klein te karakteriseren, en het resultaat lijkt daarmee redelijk in lijn te liggen met de bevindingen van PPON2007. Het linkerpaneel van Figuur 8.10 illustreert dat de verschillen tussen jongens en meisjes in de twee peilingsjaren behoorlijk vergelijkbaar waren. In het rechterpaneel is, andersom geredeneerd, te zien dat de trend vergelijkbaar is bij jongens en meisjes – namelijk, lagere prestaties in 2018.

Figuur 8.10. *Vaardigheidsverdeling van jongens en meisjes in 2007 en 2018*



8.4 Voorspellers op schoolniveau

Een vierde doel van dit onderzoek was het verkrijgen van inzicht in de samenhang tussen leerlingprestaties en kenmerken van het onderwijsleerproces. Zijn er, met andere woorden, verschillen in mondelinge taalvaardigheid op schoolniveau die verklaard worden door aspecten van het onderwijsleerproces? Om dit te onderzoeken hebben we voor elke uitkomstvariabele (Luisteren, Spreken en Gesprekken) multilevelmodellen geschat, waarin variabelen uit het schoolinterview en de Quickscan zijn gebruikt als predictoren. In paragraaf 8.4.1 beschrijven we eerst hoe we bij deze analyse zijn omgegaan met missende waarden. Vervolgens gaan we in paragraaf 8.4.2 in op de modelspecificaties en de procedure om tot een eindmodel te komen. In paragraaf 8.4.3, ten slotte, bespreken we de resultaten en inhoudelijke interpretatie van de modellen.

8.4.1 Imputatie van missende waarden

In het databestand is er sprake van missende waarden op leerling- en schoolniveau. Op leerlingniveau ontbreken er scores bij (a) leerlingen die niet aan alle onderdelen hebben meegedaan of (b) antwoorden op individuele vragen uit de Quickscan. Op schoolniveau was er soms geen interview, en soms konden niet alle variabelen gecodeerd worden op basis van het gesprek. In de hoofdstukken over de instrumenten (Hoofdstukken 3 tot en met 7) is aangegeven hoe vaak er sprake was van missende waarden op de onderdelen.

Voor het modelleren is het belangrijk hoe omgegaan wordt met de missende waarden. Als we niet imputeren, varieert het aantal datapunten tussen de modellen, omdat steeds alle leerlingen met ontbrekende waarden op predictoren worden verwijderd. Ook werken we dan onder de assumptie dat de ontbrekende waarden volledig toevallig ontbreken (*missing completely at random* oftewel MCAR; Schafer & Graham, 2002). Deze aanname is nogal streng, maar als deze onjuist is, worden de modelschattingen vertekend. Als we daarentegen de missende waarden in samenhang imputeren, dan zijn de schattingen correct onder de mildere assumptie dat het ontbreken *random* is ná conditioneren op de andere variabelen (*missing at random* of MAR; Schafer & Graham, 2002). Deze assumptie is een stuk aannemelijker, en als we *multipele imputatie* toepassen, houden we bovendien op een nette manier rekening met de onzekerheid in de modelschatting die ontstaat door het ontbreken van een deel van de datapunten.

Imputatie van deze dataset is complex omdat het hiërarchische data is, en omdat er waarden ontbreken op zowel leerling- als schoolniveau, en in continue en discrete variabelen. We hebben erg veel variabelen, maar ze zijn niet allemaal zomaar bruikbaar, door kleine categorieën of (bijna) perfecte onderlinge samenhang. Verder moeten we beoordelen welke waarden redelijkerwijs te imputeren zijn. Als een school bijvoorbeeld geen interview heeft gegeven, ontbreekt nagenoeg alle informatie op schoolniveau, waardoor het nauwelijks mogelijk is om met enige mate van zekerheid zinnige waarden te imputeren. Hetzelfde geldt wanneer de Quickscan voor een leerling niet (of nauwelijks) is ingevuld.

We hebben in dit onderzoek gebruikgemaakt van de methode van *joint modeling multipele imputatie*, die kan worden uitgevoerd met het R package *jomo (*Quartagno & Carpenter, 2019*).* In een studie van Audigier en collega's (2018) is deze methode vergeleken met een aantal andere methoden voor multipele imputatie bij hiërarchische data. Daarbij bespreken zij verschillende aspecten, waaronder:

- De (on)geschiktheid van de methoden voor discrete variabelen;
- De (on)mogelijkheid om datapunten op niveau 2 te imputeren;
- Het aantal clusters en de clustergrootte;
- De (on)mogelijkheid om random effecten over clusters toe te staan;
- De benodigde tijd om de imputatie uit te voeren.

De *joint modeling* methode lijkt op basis van deze overwegingen het meest geschikt voor onze data. Een bijkomend voordeel is dat er geen conditioneel model (per variabele) gespecificeerd hoeft te worden, maar alleen één lijst met alle relevante variabelen.

Bij de selectie van data voor het imputeren en het schatten van de modellen hebben we de volgende criteria gebruikt:

- Scholen zonder interview zijn verwijderd (6 complete scholen; 92 leerlingen);
- Leerlingen zonder Quickscan zijn verwijderd (6 complete scholen; 180 leerlingen);
- Leerlingen met zeer incomplete Quickscans zijn verwijderd (2 complete scholen; 132 leerlingen);

Er bleven 1131 leerlingen over (911 sbo; 220 so) en 54 scholen (43 sbo; 11 so). Voor iedere vaardigheid (dus 3x) hebben we 10 geïmputeerde datasets gemaakt, met een *burn in*-fase van 5000 en een interval van 1000 iteraties. Deze instellingen zijn gekozen na een *test-run* waarin de convergentie bekeken is. De imputatie is per vaardigheid apart gedaan, omdat het onderwijs kan verschillen tussen de vaardigheden (en wellicht is dit juist de reden dat het antwoord soms voor één vaardigheid ontbrak). De imputatie van schoolvariabelen baseren we daarom op *andere* variabelen *voor dezelfde* vaardigheid. Op leerlingniveau is wel altijd rekening gehouden met de samenhang tussen de vaardigheden (door ze alle drie mee te nemen). Tabel 8.24 geeft een overzicht van de variabelen die zijn geïmputeerd, of die als covariaten dienden bij het imputeren. Bij de interview-variabelen kan *lui*, *spr* of *gsp* gelezen worden waar *_ond* staat. De selectie van variabelen in Tabel 8.24 hebben we gebaseerd op inhoudelijke en statistische overwegingen. Deels betrof dit de keuze voor bepaalde schoolvariabelen als voorspellers in de multilevelmodellen, maar deels ook het meenemen van zoveel mogelijk informatie ten behoeve van het imputeren.

Tabel 8.24. *Overzicht van variabelen die zijn meegenomen bij het imputeren van missende waarden*

Variabele	Omschrijving	Type	Specificatie in jomo	% missing
SCH_ID	Schoolnummer		cluster-ID	---
LUI_BANK	Bankscore Luisteren	continu	imputeren, niv. 1	9.37
SPR_RSCO	Ruwe score Spreken	continu	imputeren, niv. 1	6.90
GSP_RSCO	Ruwe score Gesprekken	continu	imputeren, niv. 1	12.73
QS_I_UIT_A	Leerweg, als 1, 2, 3 of 4 =”anders”	polytoom	imputeren, niv. 1	2.03
QS_I_LRO_A	Leerkracht-inschatting uitstroomprofiel	polytoom	imputeren, niv. 1	2.39
QS_I_THU_A	Thuistaal	dichotoom	imputeren, niv. 1	2.92
QS_I_HER_B	Herkomst	dichotoom	imputeren, niv. 1	4.07
QS_I_LEE_A	Leerachterstand in taal/lezen	polytoom	imputeren, niv. 1	5.48
QS_I_LEE_C	Leerachterstand informatieverwerking	polytoom	imputeren, niv. 1	9.55
QS_I_STI_A	Mate waarin er sprake is van stimulerende factoren	continu	imputeren, niv. 1	3.09
QS_I_OND_A	Onderwijsbehoefte: extra structuur	dichotoom	imputeren, niv. 1	1.24
QS_I_OND_B	Onderwijsbehoefte: stimulering zelfvertrouwen	dichotoom	imputeren, niv. 1	1.24
QS_I_OND_C	Onderwijsbehoefte: meer verantwoordelijkheid	dichotoom	imputeren, niv. 1	1.24
QS_I_LIQ_A	Laag IQ (versus hoog/gemiddeld)	dichotoom	imputeren, niv. 1	1.15
QS_I_STO_A	ADHD/ADD gediagnosticeerd	dichotoom	imputeren, niv. 1	0.97
QS_I_STO_B	Autisme of een verwante stoornis gediagnosticeerd	dichotoom	imputeren, niv. 1	0.97
QS_I_STO_D	Andere stoornis gediagnosticeerd	dichotoom	imputeren, niv. 1	0.97
INT_AANB_V2_ond	Wordt de methode volledig gevolgd?	dichotoom	imputeren, niv. 2	2.12
INT_AANB_V5_ond	Wordt het onderwijs in X ingebed in de taalles?	dichotoom	imputeren, niv. 2	5.13
INT_AANB_V6_ond	Wordt het onderwijs in X geïntegreerd in andere vakken?	dichotoom	imputeren, niv. 2	5.13
INT_DOEL_V3_ond	Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)	dichotoom	imputeren, niv. 2	3.63

Tabel 8.24. *Vervolg*

Variabele	Omschrijving	Type	Specificatie in jomo	% missing
INT_DOEL_V4_ond	Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)	dichotoom	imputeren, niv. 2	4.86
INT_DOEL_V7_ond	Zijn er einddoelen voor schoolverlaters geformuleerd?	dichotoom	imputeren, niv. 2	3.01
INT_VOLG_V4_ond	Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?	dichotoom	imputeren, niv. 2	11.85
INT_VOLG_V9_ond	Worden leervorderingen in de lessen en/of in de loop van het schooljaar gemonitord en geregistreerd?	dichotoom	imputeren, niv. 2	0.71
INT_DIFF_V5_ond	Is er differentiatie naar leerstof?	dichotoom	imputeren, niv. 2	0.35
INT_DIFF_V6_ond	Is er differentiatie naar (streef)doelen?	dichotoom	imputeren, niv. 2	5.31
INT_TEAM_V1	Is mondelinge taalvaardigheid benoemd in het school- of taalbeleidsplan?	dichotoom	imputeren, niv. 2	10.43
INT_TEAM_V2	Zijn er intervisiemomenten of wordt collegiale consultatie toegepast?	dichotoom	imputeren, niv. 2	1.24
INT_TEAM_V3	Is er interne ondersteuning? (Bijvoorbeeld een taalcoördinator)	dichotoom	imputeren, niv. 2	3.98
INT_TEAM_V4	Is er externe ondersteuning? (Bijvoorbeeld een onderwijsadviseur)	dichotoom	imputeren, niv. 2	5.39
QS_I_GES_A	Geslacht	dichotoom	covariaat, niv. 1	---
SCH_TYPE	Schooltype (sbo of so-4)	dichotoom	covariaat, niv. 2	---
INT_DIFF_V1_ond	Vindt er convergente differentiatie plaats?	dichotoom	covariaat, niv. 2	---
INT_DIFF_V2_ond	Vindt er divergente differentie plaats?	dichotoom	covariaat, niv. 2	---
INT_DIFF_V3_ond	Wordt er gewerkt met niveaugroepen?	dichotoom	covariaat, niv. 2	---
INT_DIFF_V4_ond	Is er differentiatie naar instructie?	dichotoom	covariaat, niv. 2	---

8.4.2 Modelspecificatie en procedure

Het doel van de multilevelanalyses was onderzoeken hoe kenmerken van het onderwijsleerproces op de scholen zich verhouden tot de mondelinge taalvaardigheid van leerlingen. Dit kan waardevolle informatie opleveren over welke onderwijsvormen of beleidskeuzes een gunstige uitwerking lijken te hebben. Anders dan in de beschrijvende analyses in de voorgaande secties, gaat de interesse hier dus uit naar (mogelijke) causale effecten van onderwijskenmerken op de leerlingprestaties.

Met de schoolinterviews en de Quickscan hebben we een groot aantal variabelen op school- en leerlingniveau verkregen. In de multilevelanalyses is ons streven om plausibele causale verbanden te modelleren met een zorgvuldige selectie uit deze variabelen. Om misleidende conclusies over het onderwijsleerproces te voorkomen, moeten we in ieder geval *confounders* (verstorende variabelen) als predictoren in het model opnemen (zie bijvoorbeeld Elwert & Winship, 2014). Potentiële confounders zijn vroeg aanwezige leerlingkenmerken, zoals IQ of thuistaal, waarvan het aannemelijk is dat zij een causaal effect hebben gehad op de mondelinge taalvaardigheid. Hier moeten we voor controleren om de verschillen tussen scholen – en het effect van onderwijskenmerken – eerlijk te kunnen inschatten. In het INES-model dat bij PISA wordt gebruikt (*International Indicators of Education Systems*; zie OECD, 2003), worden dergelijke variabelen waarvoor men moet controleren *antecedenten* genoemd. Andere variabelen die bij het imputeren goed bruikbaar waren laten we juist achterwege bij het modelleren om vertekening en interpretatieproblemen te voorkomen. *Leerachterstand taal* of *uitstroomprofiel* zijn bijvoorbeeld geen aannemelijke oorzaken van mondelinge taalvaardigheid, maar andere uitkomstmaten die eraan verwant zijn of ten dele gevolgen ervan zijn. Controleren voor dergelijke *colliders* zou vertekening geven van de effecten waarin we geïnteresseerd zijn (*collider bias* of *overcontrol bias*; Elwert & Winship, 2014), en zou daarnaast afbreuk doen aan de interpreteerbaarheid van de modellen.[2]

[2] Een gebrekkige representativiteit van de steekproef naar een variabele (zoals in ons geval uitstroomprofiel) is voor een regressie-analyse in principe geen enkel probleem, tenzij er interactie-effecten zijn van die variabele. Interacties waar in het model geen rekening mee wordt gehouden vormen echter ook een probleem als een steekproef wel representatief is, en de enige oplossing is dan om de interacties te modelleren.

Bij de analyses hebben we voor iedere vaardigheid de volgende stappen doorlopen, waarbij telkens de variabelen uit de vorige stap(pen) werden meegenomen in het model:

1. fit een random-intercept model en bekijk de varianties op leerling- en schoolniveau;
2. controleer voor effecten van geslacht, herkomst en thuistaal;
3. schat de effecten van onderwijsbehoeften, stoornissen, stimulerende factoren, en laag IQ;
4. schat het model opnieuw na verwijderen van niet-significante leerlingkenmerken;
5. onderzoek (in verschillende modellen) het effect van aspecten van het onderwijsleerproces;
6. fit een eindmodel met de significante (en meest relevante) leerling- en schoolvariabelen.

Bij stap 5 hebben we meerdere modellen apart geschat, om verschillende aspecten van het onderwijsleerproces te kunnen onderzoeken. Het aantal van 54 scholen in de dataselectie vormt namelijk een reden dat we niet veel schoolvariabelen in één model kunnen opnemen. Er is voor gekozen om de volgende (setjes van) variabelen over het onderwijsleerproces mee te nemen:

- Aanbod: inbedding in de taalles (**INT_AANB_V5**) & inbedding in zaakvakken (**INT_AANB_V6**)
- Aantal luisteractiviteiten / spreekactiviteiten van de checklist dat wordt gebruikt
- Doelgerichtheid: einddoelen voor schoolverlaters (**INT_DOEL_V7**)
- Monitoring: gebruik genormeerde toetsen/observatielijsten (**INT_VOLG_V4**) & bijhouden leervorderingen (**INT_VOLG_V9**)
- Differentiatie: niveaugroepen (**INT_DIFF_V3**) & differentiatie naar streefdoelen (**INT_DIFF_V6**)
- Teamvorming: taalbeleid vastgelegd (**INT_TEAM_V1**) & interne ondersteuning (**INT_TEAM_V3**)
- Aantal onderwerpen van de checklist dat wordt besproken in het teamoverleg

De variabelen *Aantal luisteractiviteiten*, *Aantal spreekactiviteiten* en *Aantal onderwerpen teamoverleg* zijn afgeleid uit de ingevulde checklists, maar deze waren niet voor alle scholen beschikbaar en konden ook niet worden geïmputeerd, omdat het imputatie-algoritme dan te complex werd en slecht convergeerde. Deze variabelen zijn dus in aparte modellen alléén onderzocht voor die scholen, waarbij de betreffende checklist was ingevuld.

We hebben gebruikgemaakt van de *R* package merTools (Knowles & Frederick, 2018), die geschikt is voor het modelleren van meerdere datasets ontstaan door multipele imputatie en dan ook gemiddelde modelschattingen geeft. Als significantietoets van de modelparameters is gebruikgemaakt van de betrouwbaarheidsintervallen, die echter wel per dataset (dus 10 keer) moesten worden berekend. Predictoren op leerlingniveau zijn als significant beschouwd in een bepaalde dataset wanneer het 95%-betrouwbaarheidsinterval geheel boven of onder 0 lag. De conclusie over een bepaald effect kon dus uiteenlopen van “nooit significant” tot “in alle 10 datasets significant”. Voor predictoren op schoolniveau is het 90%-betrouwbaarheidsinterval gebruikt bij de significantietoetsing, om een wat hogere statistische *power* te hebben voor het detecteren van effecten op schoolniveau. Eenzelfde logica wordt in PISA gevolgd met betrekking tot effecten op landniveau.

8.4.3 Duiding uitkomsten multilevelanalyses

In Tabellen 8.25 tot en met 8.27 staan voor iedere vaardigheid de resultaten van de eerste vier stappen in de analyse, dus van de modellen met enkel een random intercept of predictoren op niveau 1. Van model 1 geven we de schatting (in kolom *Est*) van het gemiddelde intercept en de varianties op de twee niveaus. Bij de overige modellen staan naast de schattingen van de regressiecoëfficienten ook de bijbehorende standaardfouten (*SE*) en het aantal keer (in de 10 datasets) dat de coëfficiënten significant zijn bij tweezijdige toetsing – dat wil zeggen, het aantal keer dat het 95%-betrouwbaarheidsinterval geheel boven of onder 0 ligt. Met de varianties uit model 1 kunnen we de intraklasse-correlatie schatten. Deze is bij Luisteren en Gesprekken 0.15 en bij Spreken 0.20. Dit betekent dat zo’n 15 of 20 procent van de variantie op schoolniveau zit en 85 of 80 procent op leerlingniveau, respectievelijk. De gemiddelden verschillen tussen scholen lijken dus klein vergeleken bij de verschillen tussen individuele leerlingen.

In model 2 zijn eerst de antecedenten geslacht, herkomst en thuistaal onderzocht. Herkomst bleek bij alle vaardigheden niet significant. Vanaf model 3 is deze variabele weggelaten, omdat deze ook een erg kleine groep leerlingen van niet-Nederlandse herkomst betreft en sterk samenhangt met de variabele thuistaal. Thuistaal is bij Luisteren en Spreken wel een significante voorspeller; bij Gesprekken is de significantie twijfelachtig (verschillend tussen modellen en de 10 datasets). Geslacht, ten slotte, lijkt alleen voor het onderdeel Luisteren significant, waar meisjes wat lager scoren dan jongens (bankscore 2.3 lager). Van de andere leerlingkenmerken die we in model 3 toevoegen zijn er, zoals te zien is in Tabel 8.25, nog een aantal significant voorspellend bij Luisteren, namelijk de dichotome variabele *laag IQ (onder 70)*, de schaalscore van stimulerende/belemmerende factoren, en een stoornis van categorie B (autisme en aanverwante stoornissen). De indicatoren van onderwijsbehoeften en overige stoornissen waren niet significant. Model 4, wat we als benchmark gebruiken voor het onderzoeken van schoolpredictoren, bevat

daarom bij Luisteren geslacht, thuistaal, laag IQ, schaalscore stimulerende/belemmerende factoren, en autisme of een aanverwante stoornis. Het grootste effect is dat van een laag IQ: het model voorspelt dat leerlingen met een vastgesteld IQ van 70 of lager 15.4 punten lager scoren (in de bankscore). Het effect van stimulerende/belemmerende factoren is geschat op 5.2; merk op dat het voorspelde verschil tussen leerlingen kan oplopen tot zo'n 10.5 punten, omdat de scores op deze schaal ongeveer van -1 tot 1 lopen. Verder zien we dat leerlingen met autisme of een aanverwante stoornis iets beter presteren op het onderdeel Luisteren (4.3 punten). Gezamenlijk verklaren de antecedenten uit model 4 bij Luisteren 18.7 procent van de variantie op leerlingniveau en 36.7 procent van die op schoolniveau (zoals geschat in model 1). Blijkbaar zijn er flinke verschillen tussen de scholen in de groepssamenstelling wat betreft deze leerlingkenmerken.

Tabel 8.25. *Schattingen voor het intercept-only model en de modellen met leerlingkenmerken bij Luisteren*

	Model 1	Model 2			Model 3			Model 4		
Kenmerk	Est	Est	SE	sig	Est	SE	sig	Est	SE	sig
(Intercept)	62.8	56.1	2.7		60.7	2.1		61.8	1.9	
Geslacht (meisje)		-2.3	1.0	10	-2.0	1.0	9	-2.3	1.0	10
Herkomst (Nederland)		1.6	2.6	0						
Thuistaal (Nederlands)		6.5	2.0	10	3.5	1.6	9	3.8	1.6	10
Laag IQ (vastgesteld ≤ 70)					-15.3	1.2	10	-15.4	1.2	10
Behoefte extra structuur					0.8	1.0	0			
Behoefte stimulering zelfvertrouwen					0.1	0.9	0			
Behoefte meer verantwoordelijkheid					0.8	1.1	0			
Stimulerende factoren					5.9	1.2	10	5.2	1.1	10
Stoornis ADHD / ADD					1.3	1.0	0			
Stoornis autisme en aanverwant					4.4	1.2	10	4.3	1.2	10
Overige stoornissen					1.4	1.5	0			
Variantie niveau 1	231.5	227.4			188.3			188.2		
% verklaard op niveau 1 t.o.v. model 1		1.7%			18.6%			18.7%		
Variantie niveau 2	41.0	37.7			25.3			26.0		
% verklaard op niveau 2 t.o.v. model 1		8.1%			38.2%			36.7%		

Tabel 8.26. *Schattingen voor het intercept-only model en de modellen met leerlingkenmerken bij Spreken*

	Model 1	Model 2			Model 3			Model 4		
Kenmerk	Est	Est	SE	sig	Est	SE	sig	Est	SE	sig
(Intercept)	44.0	41.1	1.4		42.5	1.2		42.6	1.1	
Geslacht (meisje)		-0.4	0.5	0	-0.4	0.6	0			
Herkomst (Nederland)		-0.3	1.3	0						
Thuistaal (Nederlands)		3.5	1.1	10	2.3	1.0	10	2.5	0.9	10
Laag IQ (vastgesteld ≤ 70)					-5.5	0.7	10	-5.6	0.7	10
Behoefte extra structuur					0.5	0.6	0			
Behoefte stimulering zelfvertrouwen					-0.4	0.5	0			
Behoefte meer verantwoordelijkheid					0.2	0.6	0			
Stimulerende factoren					2.0	0.7	10	1.6	0.6	10
Stoornis ADHD / ADD					0.7	0.6	0			
Stoornis autisme en aanverwant					0.0	0.7	0			
Overige stoornissen					0.5	0.8	0			
Variantie niveau 1	67.3	66.4			61.9			61.8		
% verklaard op niveau 1 t.o.v. model 1		1.2%			8.0%			8.2%		
Variantie niveau 2	16.6	16.7			16.1			16.0		
% verklaard op niveau 2 t.o.v. model 1		-0.1%			3.5%			3.6%		

Tabel 8.27. *Schattingen voor het intercept-only model en de modellen met leerlingkenmerken bij Gesprekken*

	Model 1	Model 2			Model 3			Model 4		
Kenmerk	Est	Est	SE	sig	Est	SE	sig	Est	SE	sig
(Intercept)	25.4	24.0	1.0		25.1	1.0		24.7	0.8	
Geslacht (meisje)		0.3	0.4	0	-0.2	0.4	0			
Herkomst (Nederland)		0.2	1.0	0						
Thuistaal (Nederlands)		1.3	0.8	3	1.3	0.7	4	1.1	0.7	2
Laag IQ (vastgesteld ≤ 70)					-1.7	0.5	10	-1.6	0.5	10
Behoefte extra structuur					-0.4	0.5	1			
Behoefte stimulering zelfvertrouwen					0.0	0.4	0			
Behoefte meer verantwoordelijkheid					0.0	0.5	0			
Stimulerende factoren					1.4	0.9	10	1.6	0.7	10
Stoornis ADHD / ADD					-0.5	0.4	0			
Stoornis autisme en aanverwant					-0.7	0.5	1			
Overige stoornissen					0.2	0.6	0			
Variantie niveau 1	34.9	34.8			33.4			33.4		
% verklaard op niveau 1 t.o.v. model 1		0.3%			4.1%			4.3%		
Variantie niveau 2	6.2	6.3			7.7			8.0		
% verklaard op niveau 2 t.o.v. model 1		-1.8%			-24.6%			-29.2%		

Ook bij Spreken en Gesprekken blijken de indicatoren van onderwijsbehoeften en van stoornissen uit categorie A (ADHD of ADD) of D (overige) niet significant. Anders dan bij Luisteren is voor deze onderdelen ook een stoornis van categorie B (autisme en aanverwant) niet of nauwelijks voorspellend. Hierboven noemden we al dat geslacht bij deze vaardigheden nooit significant is, en ook het effect van thuistaal is kleiner en bij Gesprekken minder consistent significant over modellen en datasets heen. De uiteindelijke selectie van leerlingvariabelen in model 4 omvat daarom bij Spreken en Gesprekken iets minder kenmerken dan bij Luisteren, namelijk alleen thuistaal, laag IQ, en de schaalscore stimulerende/belemmerende factoren. Bij Spreken zien we een vergelijkbaar patroon als bij Luisteren wat betreft de omvang van de effecten van laag IQ en de schaalscore stimulerende/belemmerende factoren, namelijk dat een laag IQ meer “uitmaakt” dan de schaalscore. Bij het onderdeel Gesprekken zijn de twee effectschattingen daarentegen even groot (1.6), wat in combinatie met de scorerange (0 of 1 voor laag IQ, -1 tot 1 voor stimulerende/belemmerende factoren) impliceert dat een laag IQ minder bepalend is. Een kanttekening bij deze interpretatie is wel dat de laatste variabele alleen een dichotoom onderscheid maakt tussen een IQ onder óf boven 70.

Bij Spreken en Gesprekken verklaren de antecedenten veel minder variantie dan bij Luisteren. Bij Spreken verklaart model 4 namelijk 8.2 procent van de variantie op leerlingniveau en 3.6 procent van die op schoolniveau, en bij Gesprekken 4.3 procent op leerlingniveau en -29.2 procent op schoolniveau. De negatieve verklaarde variantie op schoolniveau bij Gesprekken wijst erop dat het intercept-only model de verschillen tussen scholen onderschat door geen rekening te houden met de leerlingkenmerken. Nadat we corrigeren voor thuistaal, laag IQ en de schaalscore stimulerende/belemmerende factoren zijn er blijkbaar *meer* onverklaarde gemiddeldenverschillen tussen scholen. Een (hypothetisch) voorbeeld wat kan helpen om dit te duiden, is dat twee scholen exact dezelfde gemiddelde score hebben maar in feite een hele andere samenstelling van de leerlinggroep qua IQ. Zodra we vervolgens in de overall analyse zien dat IQ invloed heeft op de leerlingprestaties en we hiervoor gaan corrigeren, wordt het juist opmerkelijk dat deze twee scholen gelijk presteren en wordt dit in het model dus een onverklaard verschil in intercept. De negatieve verklaarde variantie komt, met andere woorden, niet door een statistische anomalie maar laat zien dat het nuttig is om te corrigeren voor antecedenten. De variantieschatting uit model 4 dient als benchmark voor de modellen in stap 5, waar we deze variantie pogen te verklaren met kenmerken van het onderwijsleerproces.

De in stap 5 geschatte effecten van de verschillende onderwijskenmerken (niveau-2 predictoren) op de drie vaardigheden staan in Tabel 8.28. Om ruimte te besparen, vermelden we niet van ieder model alle parameters, maar alleen de parameter(s) die we met dat model onderzoeken. De resultaten bij de drie vaardigheden zijn naast elkaar in de tabel gezet om overeenkomsten en verschillen te benadrukken. De kolom *sig* geeft aan hoe vaak het 90%-betrouwbaarheidsinterval geheel boven of onder nul lag of, met

andere woorden, hoe vaak de niveau-2 predictor significant was bij tweezijdige toetsing. Merk op dat deze modellen altijd ook de leerlingkenmerken uit model 4 van de betreffende vaardigheid bevatten als niveau-1 predictoren. De parameterschattingen van die predictoren verschillen nauwelijks tussen de modellen.

Bij Luisteren zijn alleen het aantal activiteiten dat wordt gebruikt (van de checklist) en het bijhouden van leervorderingen significant, maar de geschatte effecten zijn beiden negatief – wat moeilijk te duiden is – en bescheiden in omvang. Het model met het aantal activiteiten als predictor is geschat op een kleinere groep scholen wegens ontbrekende gegevens; en als we op deze subset van de data een model schatten waarin we ook het monitoren van leervorderingen als predictor toevoegen, dan is die laatste weer niet significant. Om de conclusies op zoveel mogelijk data te baseren, ligt het meer voor de hand om als eindmodel te kiezen voor een aanpassing van model 5c (geschat op de hele dataset) waarbij we de niet-significante predictor (het gebruik van genormeerde toetsen/observatielijsten) verwijderen.

Bij het onderdeel Spreken zien we een significant positief effect van einddoelen voor schoolverlaters, en daarnaast een negatief effect van differentiatie naar streefdoelen waarvan de significantie twijfelachtig is (6 van de 10 keer). In het eindmodel voor deze vaardigheid gebruiken we deze beide predictoren op schoolniveau; de andere onderwijsaspecten zijn nooit significant. Bij Gesprekken, ten slotte, kiezen we voor model 5d als eindmodel omdat de twee predictoren uit dit model beiden significant zijn in tenminste 7 van de 10 datasets, en de andere onderwijskenmerken nooit significant zijn. Het werken met niveaugroepen lijkt gunstig te zijn, en differentiatie naar streefdoelen juist ongunstig; dit laatste zien we ook bij Spreken.

Tabel 8.28. *Schattingen van de niveau-2 predictoren uit de modellen in analysestap 5, per vaardigheid*

		Luisteren			Spreken			Gesprekken		
Model	Variabele	Est	SE	sig	Est	SE	sig	Est	SE	sig
5a	Inbedding in de taalles	0.7	2.1	0	0.0	1.3	0	-0.6	1.0	0
	Inbedding in zaakvakken	1.4	2.2	0	0.0	1.5	0	-0.1	1.0	0
5b	Aantal activiteiten (checklist)	-1.1	0.5	9	-0.3	0.3	0	-0.1	0.2	0
5c	Einddoelen voor schoolverlaters	0.1	2.0	0	2.4	1.3	10	0.5	1.0	0
5d	Genormeerde toetsen/observatielijsten	1.6	2.3	1	0.7	1.6	0	0.4	1.0	0
	Bijhouden leervorderingen	-3.4	2.2	8	0.3	1.4	0	-1.0	1.0	0
5e	Werken met niveaugroepen	-1.2	2.3	0	1.0	1.3	0	1.5	0.9	7
	Differentiatie naar streefdoelen	-1.8	2.7	1	-2.1	1.3	6	-2.0	0.9	10
5f	Taalbeleid vastgelegd	-0.3	2.4	0	0.6	1.3	0	-0.5	1.0	0
	Interne ondersteuning	2.1	2.5	1	1.2	1.3	0	0.6	1.0	0
5g	Aantal onderwerpen overleg (Checklist)	-0.3	0.4	0	0.1	0.3	0	0.1	0.2	0

In Tabel 8.29 geven we een overzicht van de geschatte parameters van het eindmodel per vaardigheid. Ook de variantie op leerling- en schoolniveau staat in de tabel. De variantie op leerlingniveau kan niet afnemen door het toevoegen van predictoren op schoolniveau, maar de variantie op schoolniveau kan worden vergeleken met die uit model 4 om na te gaan hoeveel variantie verklaard is door de onderwijskenmerken. Merk op dat de aantallen in de kolom *sig* in Tabel 8.29 voor de niveau-1 predictoren gebaseerd zijn het 95%-betrouwbaarheidsinterval en die voor de niveau-2 predictoren op het 90%-betrouwbaarheidsinterval.

Bij Luisteren zien we dat slechts 4.8 procent van de variantie op schoolniveau door het eindmodel wordt verklaard. Het model bevat dan ook slechts één schoolpredictor, waarvan het effect niet altijd significant is en qua omvang klein. De meest voor de hand liggende conclusie is daarom vooral dat we de verschillen op schoolniveau heel moeilijk kunnen verklaren vanuit de variabelen die we onderzocht hebben. De verschillen tussen leerlingen kunnen we iets beter verklaren (18.7 procent verklaarde variantie op niveau 1, zie Tabel 8.25) aan de hand van leerlingkenmerken, maar er blijft een grote onverklaarde variatie in de luistervaardigheid.

Bij Spreken en Gesprekken verklaren de eindmodellen meer variantie op schoolniveau, namelijk 10.0 procent en 8.7 procent, respectievelijk. Deze modellen bevatten dan ook ieder twee schoolpredictoren die overwegend of altijd significant zijn en grotere effecten hebben dan de schoolpredictor bij Luisteren (als we meewegen dat de scorerange bij Luisteren veel groter is dan bij Spreken en Gesprekken). Voor beide vaardigheden lijkt de conclusie dat differentiatie naar streefdoelen niet bevorderlijk is voor de gemiddelde mondelinge taalvaardigheid van de leerlingen. Voorzichtigheid is wel geboden bij de interpretatie, omdat we niet zeker kunnen weten of dit daadwerkelijk een causaal verband is. Er kunnen andere verklaringen voor zijn dat differentiatie naar streefdoelen negatief samenhangt met de gemiddelde leerlingprestaties. Deze kanttekening geldt ook voor het gevonden positieve effect van opgestelde einddoelen voor schoolverlaters op spreekvaardigheid, en het positieve effect van het werken met niveaugroepen op gespreksvaardigheid.

Tabel 8.29. *Schattingen van het eindmodel (analysestap 6) per vaardigheid*

	Luisteren			Spreken			Gesprekken		
Variabele	Est	SE	sig	Est	SE	sig	Est	SE	sig
(Intercept)	63.6	2.7		41.7	1.5		24.9	1.0	
Geslacht (meisje)	-2.3	1.0	10						
Thuistaal (Nederlands)	3.7	1.6	10	2.5	0.9	10	1.1	0.7	2
Laag IQ (vastgesteld ≤ 70)	-15.4	1.2	10	-5.6	0.7	10	-1.6	0.5	10
Stimulerende factoren	5.2	1.1	10	1.6	0.6	10	1.6	0.7	10
Stoornis autisme en aanverwant	4.4	1.2	10						
Einddoelen voor schoolverlaters				2.6	1.3	10			
Bijhouden leervorderingen	-2.8	2.2	7						
Werken met niveaugroepen							1.5	0.9	7
Differentiatie naar streefdoelen				-2.0	1.2	8	-2.0	0.9	10
Variantie niveau 1	188.2			61.8			33.4		
Variantie niveau 2	24.7			14.4			7.3		
% verklaarde variantie niveau 2 t.o.v. model 4	4,8%			10,0%			8.7%		

9 Samenvatting en conclusies

In het schooljaar 2017-2018 heeft een peilingsonderzoek naar de mondelinge taalvaardigheid van de schoolverlaters van het sbo en so (voormalig) cluster 4 plaatsgevonden. In opdracht van het Nationaal Regieorgaan Onderwijsonderzoek (NRO), en onder regie van de onderwijsinspectie, hebben Instituut Taalonderzoek en Taalonderwijs Amsterdam (ITTA) en Cito BV de peiling uitgevoerd. Bij de werving van scholen is samengewerkt met CED-groep. In het peilingsonderzoek zijn vier verschillende taken aan leerlingen voorgelegd, namelijk een luistertaak, een naverteltaak op basis van een jeugdjournaalfragment, een verteltaak op basis van het educatieve spel “Story Cubes”, en een gesprekstaak op basis van het bordspel “Fischerspiel”. Daarnaast was er een Quickscan om zicht te krijgen op de onderwijs- en ondersteuningsbehoeften van leerlingen en een interview om het onderwijsleerproces op sbo- en so-scholen te duiden. Aan het peilingsonderzoek hebben ruim 1500 leerlingen van 68 verschillende sbo- en so-scholen deelgenomen. De data zijn gebruikt om zeven verschillende onderzoeksvragen te beantwoorden. De onderzoeksvragen worden hieronder één voor één behandeld.

Op welk referentieniveau functioneren de schoolverlaters van het sbo en so?

Om de prestaties van leerlingen te kunnen beschrijven in termen van referentieniveaus 1F en 2F waren prestatiestandaarden nodig. De wijze waarop de prestatiestandaarden tot stand zijn gekomen, verschilde per onderdeel. Voor het onderdeel Luisteren is gebruikgemaakt van de landelijke Ankersets, terwijl CvTE voor de onderdelen Spreken en Gesprekken nieuwe prestatiestandaarden heeft ontwikkeld op basis van de materialen die in dit peilingsonderzoek gebruikt zijn. Bij **Luisteren** bleken de meeste leerlingen in het sbo referentieniveau 1F nog niet behaald te hebben (58.9 procent). Van de leerlingen die al wel zo ver waren, bleek 33.9 procent op 1F te functioneren en 7.2 procent op 2F. In het so zag het beeld er behoorlijk anders uit. De meerderheid van de leerlingen functioneerde daar op niveau 1F (46.0 procent) of hoger (20.4 procent). Bij **Spreken** functioneerden de leerlingen in termen van het referentiekader op een hoger niveau dan bij luisteren. De overgrote meerderheid functioneerde op referentieniveau 1F (sbo: 53.9 procent, so: 65.8 procent) of 2F (sbo: 14.4 procent, so: 12.7 procent). Bij **Gesprekken** was het beeld tegengesteld. Zowel in het sbo als so functioneerde de meerderheid van de leerlingen op een niveau lager dan 1F (sbo: 58.6 procent, so: 51.3 procent).

Wat beheersen leerlingen op referentieniveaus 1F en 2F al wel en wat nog niet?

Om de luister, spreek- en gespreksvaardigheid van leerlingen inhoudelijk te duiden, is gebruikgemaakt van de mogelijkheden die de item respons theorie ons daarvoor biedt. Per item (of beoordelingsaspect) is berekend welke vaardigheid nodig is om 50 of 80 procent van de maximale score te behalen. Beheersing is als volgt gedefinieerd: (**geen beheersing**) verwachte score kleiner dan 50 procent, (**matige beheersing**) verwachte score ligt tussen de 50 en 80 procent, (**goede beheersing**) verwachte score groter dan 80 procent.

- Bij **Luisteren** is nagegaan in hoeverre beheersing te relateren is aan kenmerken van de tekst (1F of 2F), de taakuitvoering (begrijpen, interpreteren of samenvatten), de aard van de activiteit (luisteren of kijken en luisteren) en de gesprekvorm (monoloog of dialoog). Leerlingen op niveau 2F bleken langere teksten met een hogere informatiedichtheid aan te kunnen dan leerlingen op niveau 1F. De taakuitvoering en de aard van de activiteit zeiden weinig over beheersing. Er waren items die leerlingen op niveau 1F en 2F zeer gemakkelijke vonden, maar ook items die zij zeer moeilijk vonden. Daaruit blijkt dat luisteren een concentrische vaardigheid is; vanaf het prille begin moeten leerlingen verschillende te onderscheiden aspecten van het luisteren gelijktijdig en geïntegreerd inzetten. Ten slotte bleken leerlingen monologen al eerder in de ontwikkeling aan te kunnen dan dialogen. Mogelijk komt dat doordat leerlingen bij monologen minder snel afgeleid zijn. De monologen waren in dit onderzoek bovendien aanzienlijk korter dan de dialogen, zeker waar het ging om berichten en verhalen.
- Bij **Spreken** is beheersing gerelateerd aan de in factoranalyses gevonden factoren Taalvorm, Taalinhoud en Taalgebruik. Leerling op niveau 1F bleken nauwelijks tot geen moeite te hebben met de Taalvorm als ze feiten moesten overbrengen aan de hand van een jeugdjournaalfragment. De taalinhoud was bij de 1F-leerlingen net wél of net niet voldoende op orde. Ze konden de strekking vaak al wel duidelijk maken aan de gesprekspartner, maar de inhoudelijke uitwerking en opbouw was nog niet optimaal. Op niveau 1F leek het onderdeel Taalgebruik, waarbij het ging om boeiend vertellen en de inzet van non-verbale communicatie, nog te hoog gegrepen. Ook voor leerlingen op niveau 2F vormde Taalgebruik nog een uitdaging. De aspecten die onder Taalvorm en Taalinhoud vielen, werden door deze leerlingen wel goed beheerst. Bij het creatief vertellen aan de hand van

het educatieve spel “Story Cubes” was het beeld identiek, maar bij Taalinhoud was het wel opvallend dat leerlingen op twee aspecten aanzienlijk beter scoorden. Bij het creatief vertellen waren leerlingen vollediger en bovendien was het voor de leerlingen geen enkel probleem om alle dobbelstenen in hun verhaal te verwerken. Bij het jeugdjournaalfragment ontbraken bepaalde elementen nogal eens in de samenvattingen van de leerlingen.

- Bij **Gesprekken** is beheersing op dezelfde wijze geduid als bij Spreken. De Taalvorm bleek de leerlingen op niveau 1F erg gemakkelijk af te gaan. De woordenschat van 1F-leerlingen was voldoende om het gesprek te kunnen voeren, ze spraken redelijk vloeiend en waren bovendien goed verstaanbaar, ondanks een eventueel accent. Ook de aspecten met betrekking tot grammatica (zinsbouw, verbuiging en vervoeging) waren goed op orde. Taalinhoud bleek voor 1F-leerlingen nog lastig te zijn. Zo was de gedachtegang niet altijd begrijpelijk een ook de inhoudelijke uitwerking kon beter. Het Taalgebruik, ten slotte, hadden 1F-leerlingen al redelijk onder de knie; reacties waren consequent gepast, ze namen op passende wijze het woord en waren in staat om de inbreng van de gesprekspartners te integreren in de eigen bijdrage. Het was opvallend dat de leerlingen de factor Taalgebruik in Gesprekken op niveau 1F (behoorlijk) goed beheersten en dat deze factor bij Spreken nog de nodige problemen opleverde. Wellicht heeft dit te maken met de gespreksvorm: bij het houden van een monoloog is het mogelijk lastiger en minder natuurlijk om contact met de luisteraar(s) te maken dan bij een polyloog. Waar Taalinhoud voor 1F-leerlingen nog enigszins problematisch bleek te zijn, was dat bij 2F-leerlingen niet meer het geval; zij beheersten alle beoordeelde aspecten in goede mate.

Hoe hangen de uitkomsten op Luisteren, Spreken en Gesprekken met elkaar samen?

De samenhang tussen de drie vaardigheden, en alle subonderdelen van die vaardigheden, is in kaart gebracht met het computerprogramma MULTI (Kamphuis, 1993, 1998). In MULTI is het mogelijk om het meetmodel expliciet mee te nemen en dat is bij het berekenen van de correlaties dan ook gedaan. De (latente) correlatie tussen Luisteren en Spreken was gelijk aan 0.394, die tussen Luisteren en Gesprekken was 0.237 en die tussen Spreken en Gesprekken was 0.424. Deze correlaties zijn nogal laag. Dat betekende dat de drie mondelinge taalvaardigheden apart beschouwd moesten worden en dat vanuit de ene (mondelinge taal-) vaardigheid niet goed was af te leiden wat we op een andere (mondelinge taal-) vaardigheid van een leerling mochten verwachten. De subonderdelen van Luisteren bleken zeer hoog te correleren. Deze waarneming ondersteunde de keuze om Luistervaardigheid te modelleren als een unidimensionaal construct. De correlatie tussen navertellen en creatief vertellen was eveneens zeer hoog; het jeugdjournaalfragment en het dobbelspel konden niet zinvol als aparte entiteiten onderscheiden worden. Bij Spreken leek Taalgebruik echter wel een zelfstandig onderdeel te vormen ten opzichte van Taalvorm en Taalinhoud. Bij Gesprekken was de correlatie tussen Taalvorm en Taalinhoud opvallend lager en dat resultaat wees erop dat bij Gesprekken ook deze twee subonderdelen zinvol van elkaar gescheiden kunnen worden. Een opdeling naar theoretische/zakelijke taken enerzijds en praktische/speelse taken anderzijds, ten slotte, leverde niet duidelijk nieuwe inzichten op. In de correlatiematrix was weliswaar heel voorzichtig te zien dat **binnen**-taaktype-correlaties iets hoger waren dan de **tussen**-taaktype-correlaties, maar de verschillen waren klein. Er waren, met andere woorden, geen duidelijke aanwijzingen dat leerlingen bij praktische opdrachten ‘anders’ scoorden dan bij theoretische opdrachten.

In hoeverre kunnen prestatieverschillen verklaard worden vanuit achtergrondkenmerken?

De relatie tussen achtergrondkenmerken van de leerlingen en hun prestaties op Luisteren, Spreken en Gesprekken is onderzocht door middel van beschrijvende analyses en multilevelanalyse. De Luistervaardigheid van leerlingen kan deels verklaard worden vanuit geslacht, thuistaal, een (zeer) laag IQ, een stoornis uit het autismespectrum of de mate waarin er stimulerende en belemmerende factoren van toepassing zijn (zoals de thuissituatie, het schoolwelbevinden, en motivatie of emotieregulatie). Meisjes presteren wat minder goed dan jongens op Luisteren. Bij Spreken en Gesprekken vonden we geen verschil tussen jongens en meisjes en lijken met name thuistaal, IQ en de mate van stimulerende/belemmerende factoren relevant om de prestatieverschillen te verklaren. Er was geen duidelijke indicatie dat mondelinge taalvaardigheid samenhangt met onderwijsbehoeften (zoals structuur) of probleemgedrag. Ze hangt wel samen met variabelen zoals uitstroomprofiel of leerachterstanden, maar dit zijn dan ook andere indicaties van leerprestaties, en niet zozeer achtergrondkenmerken van leerlingen die de prestaties kunnen verklaren. De achtergrondkenmerken verklaren bij Luisteren zo’n 19 procent van de scorevariantie op leerlingniveau, en bij Spreken en Gesprekken is de verklaarde variantie circa 8 procent of 4 procent, respectievelijk. We kunnen dus concluderen dat deze leerlingkenmerken samenhangen met de mondelinge taalvaardigheid, maar moeten ons tegelijk realiseren dat slechts een bescheiden deel van de prestatieverschillen hiermee is verklaard. Veel variatie in de leerlingprestaties op mondelinge taalvaardigheid blijft onverklaard.

Hoe kunnen scholen worden getypeerd als het gaat om kenmerken van het onderwijsleerproces en zijn welke verschillen zijn er tussen het sbo en het so?

Om zicht te krijgen op het onderwijsleerproces zijn interviews gehouden met betrokkenen op de scholen. Uit de interviews bleek dat vrijwel alle bevraagde scholen werken met een methode en dat zij daarnaast vaak aanvullende materialen gebruiken. Ongeveer de helft van de scholen, in zowel so als sbo, volgde de methode volledig. In het so was Taalactief de meest genoemde methode en in het sbo Staal. In het so werd mondelinge taalvaardigheid vaak geïntegreerd in de taalles aangeboden. In het sbo werd er daarentegen relatief vaak voor gekozen om mondelinge taalvaardigheid (ook) te integreren in andere vakken. Zowel de sbo- als de so-scholen gaven aan gebruik te maken van verschillende luisteractiviteiten. Er wordt vaak naar tv-fragmenten gekeken en geluisterd, er worden boeken voorgelezen en leerlingen krijgen mondelinge instructies zie zij moeten uitvoeren. Het was opvallend dat geen van de bevraagde so-scholen leerlingen aantekeningen laat maken bij het luisteren naar gesproken teksten. Het so en sbo verschilden nauwelijks in het type spreekactiviteiten dat wordt uitgevoerd. In het so wordt echter niet geoefend met (zakelijke) boodschappen, terwijl dat wel door de helft van de sbo-scholen genoemd werd. Het merendeel van de so- en sbo-scholen gaf aan doelgericht te werken door leerlijnen te gebruiken. Daarbij werd zowel een doelgerichte als activiteitgerichte aanpak toegepast. Om de leervorderingen te volgen werd in het so relatief vaker dan in het sbo gebruikgemaakt van methodetoetsen. Bij ongeveer de helft van de scholen werden de leerresultaten gebruikt bij de planning, maar dat resulteerde nauwelijks in groepsplannen. Slechts een kwart van de sbo-scholen gebruikte een groepsplan en in het so was er maar één school die dat deed. Voorts bleek in het so vooral convergent gedifferentieerd te worden, terwijl in het sbo het aantal scholen dat volgens convergente of divergente differentiatie werkte vrijwel gelijk was. In het so en het sbo werd vooral gedifferentieerd naar instructie. In het sbo werd daarnaast ook door iets minder dan de helft van de scholen gedifferentieerd naar leerstof en naar (streef)doelen. Dat werd minder vaak genoemd in het so.

In hoeverre hangen verschillen in het onderwijsleerproces samen met de prestaties van leerlingen?

In de multilevelanalyses zijn we van een aantal onderwijsaspecten nagegaan of deze samenhangen met de gemiddelde leerlingprestaties op de scholen. Bij Luisteren vonden we weinig overtuigende relaties tussen kenmerken van het onderwijsleerproces enerzijds en de leerlingprestaties anderzijds. Ook het “beste” model kon slechts zo’n 5 procent van de variantie op schoolniveau verklaren en was inhoudelijk lastig te duiden. Het bijhouden van leervorderingen lijkt negatief samen te hangen met de leerlingprestaties. Bij Spreken en Gesprekken konden we respectievelijk zo’n 10 procent en 9 procent van de variantie op schoolniveau verklaren. Bij Spreekvaardigheid blijken scholen die einddoelen hebben opgesteld voor schoolverlaters gemiddeld beter te presteren. Voor Gespreksvaardigheid lijkt het gunstig te zijn om te werken met niveaugroepen. Voor zowel Spreken als Gesprekken zien we verder dat differentiatie naar streefdoelen niet bevorderlijk lijkt te zijn voor de gemiddelde mondelinge taalvaardigheid. Voorzichtigheid is wel geboden bij de interpretatie van deze bevindingen, omdat we niet zomaar kunnen aannemen dat de onderwijskenmerken een causaal verband houden met de gemiddelde leerlingprestaties op de scholen. Er kunnen andere verklaringen zijn voor de gevonden relaties; het is bijvoorbeeld denkbaar dat een school begint met differentiëren naar streefdoelen, omdat er op de school een subgroep leerlingen is die een grote leerachterstand heeft vergeleken bij de rest. Ook zijn de genoemde percentages verklaarde variantie op schoolniveau bij alle onderdelen bescheiden. De verschillen in de mondelinge taalvaardigheid op de scholen zijn dus moeilijk te verklaren vanuit de onderzochte kenmerken van het onderwijsleerproces. Hierbij is het overigens wel goed om op te merken dat de gevonden variatie tussen scholen niet al te groot is (circa 15 à 20 procent van de totale variantie) vergeleken bij de diversiteit van individuele leerlingen (80 à 85 procent).

In hoeverre zijn de prestaties in het sbo in 2018 veranderd ten opzichte van 2007?

Voor het onderdeel Luisteren is een vergelijking gemaakt tussen dit peilingsonderzoek en PPON2007. Zowel bij de 12- als 13-jarigen was de gemiddelde luistervaardigheid in 2018 lager dan in 2007. Voor de 12-jarigen was dit verschil klein en statistisch niet significant, maar voor de 13-jarigen was het verschil beduidend groter en wel significant, $t(220.4) = 4.1$, $p < .001$. In overeenstemming met deze trend was te zien dat in 2007 meer leerlingen voldeden aan de standaarden voor het 1F- en 2F-niveau dan in 2018. In 2007 beschikte bijvoorbeeld 54.2 procent van de 13-jarigen over referentieniveau 1F of 2F, terwijl dat in 2018 maar 38.9 procent was. Er is gekeken of deze trend voor jongens en meisjes hetzelfde is. In de rapportage over PPON2007 werd vermeld dat meisjes in het sbo een iets lagere luistervaardigheid hadden dan jongens, alhoewel het verschil alleen statistisch significant was voor de subvaardigheid *Interpreteren* en niet voor *Begrijpen*. De algemene conclusie luidde dat het geslachtsverschil “… zo klein was dat er niet veel betekenis aan gehecht kan worden”. In 2018 presteerden jongens significant beter op luistervaardigheid dan meisjes, $t(807.7) = 3.5$, $p < .001$, maar het gestandaardiseerde verschil was opnieuw niet erg groot ($g = 0.23$). De teruggang in prestaties geldt dus zowel voor jongens als meisjes in ongeveer even grote mate.

10 Databestand

10.1 Algemene omschrijving

Het databestand bevat de gegevens van de 1535 leerlingen die aan ten minste één van de onderdelen van de gegevensverzameling hebben deelgenomen. In de voorgaande hoofdstukken is al een groot deel van de variabelen van het databestand geïntroduceerd. Enkele administratieve variabelen zijn nog niet besproken, maar de precieze betekenis van deze variabelen kan eenvoudig worden afgeleid uit de *variable labels* en *value labels* in het databestand. Onderstaande paragrafen geven een korte toelichting bij elke variabele in het databestand.

10.2 Administratieve variabelen

De volgende administratieve variabelen zijn beschikbaar:

LRL_ID	Leerlingidentificatie
SCH_ID	Schoolidentificatie
SCH_TYPE	SBO of SO
DUM_QSC	Is de Quickscan ingevuld voor de leerling?
DUM_INT	Is er een schoolinterview afgenomen?
DUM_LUI	Heeft de leerling deelgenomen aan het onderdeel luisteren?
DUM_SPR	Heeft de leerling deelgenomen aan het onderdeel spreken?
DUM_GSP	Heeft de leerling deelgenomen aan het onderdeel gesprekken voeren?
NTOETS	Op hoeveel onderdelen heeft de leerling gegevens (incl. de Quickscan)?
N_geinterviewden	Aantal vertegenwoordigers van de school aanwezig bij het interview
Rol_geinterviewde_1	Rol van de eerste geïnterviewde op de school
Rol_geinterviewde_2	Rol van de tweede geïnterviewde op de school
Rol_geinterviewde_3	Rol van de derde geïnterviewde op de school

10.3 Achtergrondvariabelen op schoolniveau

Op schoolniveau zijn de volgende achtergrondvariabelen (per vaardigheid) beschikbaar:

SCH_REG	In welke regio (Noord, Oost, Zuid, of West) staat de school?
SCH_STED	Mate van verstedelijking van de school
INT_AANB_V1	Wordt een methode gebruikt?
INT_AANB_V1_namelijk	Vrije invoer voor methode
INT_AANB_V2	Wordt de methode volledig gevolgd?
INT_AANB_V3	Zijn er afspraken over het gebruik van de methode?
INT_AANB_V4	Worden er aanvullende materialen gebruikt naast de hoofdmethode?
INT_AANB_V4_namelijk	Vrije invoer voor aanvullende materialen
INT_AANB_V5	Wordt het onderwijs ingebed in de taalles?
INT_AANB_V6	Wordt het onderwijs geïntegreerd in andere vakken?
INT_AANB_ACT1_lui	Luisteractiviteit: kijken+luisteren naar tv-fragment, vragen beantwoorden
INT_AANB_ACT2_lui	Luisteractiviteit: luisterboek en beantwoorden van vragen
INT_AANB_ACT3_lui	Luisteractiviteit: coöperatieve luisteropdrachten
INT_AANB_ACT4_lui	Luisteractiviteit: voorlezen van boek en beantwoorden van vragen
INT_AANB_ACT5_lui	Luisteractiviteit: aantekeningen maken bij gesproken tekst (tv/live)
INT_AANB_ACT6_lui	Luisteractiviteit: samenvatting maken van gesproken tekst (tv/live)
INT_AANB_ACT7_lui	Luisteractiviteit: mening/standpunt bepalen bij gesproken tekst (tv/live)
INT_AANB_ACT8_lui	Luisteractiviteit: instructies (praktisch) beluisteren en ze dan uitvoeren
INT_AANB_ACT9_lui_nl	Vrije invoer voor andere Luisteractiviteiten
INT_AANB_ACT1_spr	Spreekactiviteit: spreekbeurten
INT_AANB_ACT2_spr	Spreekactiviteit: kring-/groepsgesprekken
INT_AANB_ACT3_spr	Spreekactiviteit: verslag van persoonlijke belevenis

INT_AANB_ACT4_spr	Spreekactiviteit: dialogen
INT_AANB_ACT5_spr	Spreekactiviteit: samenwerkend leren
INT_AANB_ACT6_spr	Spreekactiviteit: discussies
INT_AANB_ACT7_spr	Spreekactiviteit: (zakelijke) boodschappen oefenen
INT_AANB_ACT8_spr	Spreekactiviteit: verwoorden van meningen
INT_AANB_ACT9_spr	Spreekactiviteit: verhaal vertellen volgens gespreksregels
INT_AANB_ACT10_spr	Spreekactiviteit: deelname aan een weekafsluiting/toneelstuk/…
INT_AANB_ACT11_spr_nl	Vrije invoer voor andere Spreekactiviteiten
INT_DOEL_V1	Heeft de school (expliciet) doelen?
INT_DOEL_V2	Wordt gebruikgemaakt van een leerlijn?
INT_DOEL_V2_namelijk	Vrije invoer voor leerlijn
INT_DOEL_V3	Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)
INT_DOEL_V4	Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)
INT_DOEL_V5	Wordt gebruikgemaakt van herhaald aanbod met feedback?
INT_DOEL_V6	Wordt gebruikgemaakt van referentieniveaus?
INT_DOEL_V7	Zijn er einddoelen voor schoolverlaters geformuleerd?
INT_VOLG_V1	Worden leervorderingen in de lessen gemonitord en geregistreerd?
INT_VOLG_V2	Worden leervorderingen in het schooljaar gemonitord en geregistreerd?
INT_VOLG_V3	Wordt gebruikgemaakt van een leerling- administratiepakket?
INT_VOLG_V4	Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?
INT_VOLG_V5	Wordt gebruikgemaakt van methodetoetsen?
INT_VOLG_V6	Wordt gebruikgemaakt van zelfgemaakte instrumenten?
INT_VOLG_V7	Wordt in de planning gebruikgemaakt van leerresultaten van leerlingen?
INT_VOLG_V8	Wordt gebruikgemaakt van groepsplannen?
INT_DIFF_V1	Vindt er convergente differentiatie plaats?
INT_DIFF_V2	Vindt er divergente differentie plaats?
INT_DIFF_V3	Wordt er gewerkt met (twee of meer) niveaugroepen?
INT_DIFF_V4	Is er differentiatie naar instructie?
INT_DIFF_V5	Is er differentiatie naar leerstof?
INT_DIFF_V6	Is er differentiatie naar (streef)doelen?
INT_DIFF_V7	Extra ondersteuning krijgen door een specialist binnen de school?
INT_DIFF_V7_namelijk	Vrije invoer specialist binnen de school
INT_DIFF_V8	Extra ondersteuning krijgen door een specialist buiten de school?
INT_DIFF_V8_namelijk	Vrije invoer specialist buiten de school
INT_DIFF_V9	Extra ondersteuning krijgen op basis van een specifiek programma?
INT_DIFF_V9_namelijk	Vrije invoer over specifiek programma
INT_TEAM_V1	Is mondelinge taalvaardigheid benoemd in het school- of taalbeleidsplan?
INT_TEAM_V2	Zijn er intervisiemomenten of wordt collegiale consultatie toegepast?
INT_TEAM_V2_namelijk	Vrije invoer over intervisie en collegiale consultatie
INT_TEAM_V3	Is er interne ondersteuning? (Bijvoorbeeld een taalcoördinator)
INT_TEAM_V3_namelijk	Vrije invoer over interne ondersteuning
INT_TEAM_V4	Is er externe ondersteuning? (Bijvoorbeeld een onderwijsadviseur)
INT_TEAM_V4_namelijk	Vrije invoer over externe ondersteuning
INT_TEAM_OND1	Teamoverleg: Gebruik van de methode voor mondelinge taalvaardigheid
INT_TEAM_OND2	Teamoverleg: Onderwijs in mtv bij andere vakken dan taal
INT_TEAM_OND3	Teamoverleg: Volgen van de ontwikkeling van leerlingen m.b.t. mtv
INT_TEAM_OND4	Teamoverleg: Uitkomsten van volgsysteem/registraties m.b.t. mtv
INT_TEAM_OND5	Teamoverleg: leerkracht-ondersteuning aan ll. met spraak-/taalstoornis
INT_TEAM_OND6	Teamoverleg: externe ondersteuning aan ll. met spraak-/taalstoornis
INT_TEAM_OND7	Teamoverleg: Aanpak ll.groep met achterstand o.b.v. leerlingvolgsysteem
INT_TEAM_OND8	Teamoverleg: De leerdoelen van de school op het gebied van mtv
INT_TEAM_OND9	Teamoverleg: Vakdidactiek mondelinge taalvaardigheden
INT_TEAM_OND_anders_nl	Vrije invoer voor opmerkingen over teamoverleg / teamontwikkeling

10.4 Achtergrondvariabelen op leerlingniveau

Op leerlingniveau zijn de volgende achtergrondvariabelen beschikbaar:

QS_I_GES_A	Geslacht
QS_I_LFT_A	Leeftijd
QS_I_THU_A	Thuistaal
QS_I_HER_B	Herkomst

QS_I_UIT_A	Uitstroomprofiel
QS_I_VER_A	Verblijfsduur in sbo/so
QS_I_STI_A	Schaalscore voor de mate van stimulerende vs. belemmerende factoren
QS_I_STI_B	Aantal stimulerende factoren
QS_I_STI_C	Aantal belemmerende factoren
QS_I_OND_A	Onderwijsbehoeften: behoefte aan extra structuur
QS_I_OND_B	Onderwijsbehoeften: stimulering zelfvertrouwen
QS_I_OND_C	Onderwijsbehoeften: meer verantwoordelijkheid
QS_I_EXT_A	Externaliserend probleemgedrag
QS_I_INT_A	Internaliserend negatief probleemgedrag
QS_I_INT_B	Internaliserend sociaal probleemgedrag
QS_I_LEE_A	Leerachterstand in taal/lezen
QS_I_LEE_B	Leerachterstand in rekenen
QS_I_LEE_C	Leerachterstand in informatieverwerking
QS_I_EIQ_A	Laag of hoog IQ vastgesteld
QS_I_STO_A	ADHD/ADD gediagnosticeerd
QS_I_STO_B	Autisme of een verwante stoornis gediagnosticeerd
QS_I_STO_C	Hechtingsstoornis gediagnosticeerd
QS_I_STO_D	Andere stoornis gediagnosticeerd
QS_I_LRO_A	Inschatting leerroute leerkracht
Geslacht	Geslacht (niet-numeriek)
QS_I_WEIG	Index Weigering om persoonlijke leerlinggegevens in te vullen bij de Quickscan
QS_V_GEB01	Geboortedatum
QS_V_UIT01	Uitstroomprofiel: regulier vwo
QS_V_UIT02	Uitstroomprofiel: regulier havo
QS_V_UIT03	Uitstroomprofiel: regulier vmbo tl
QS_V_UIT04	Uitstroomprofiel: regulier vmbo gt
QS_V_UIT05	Uitstroomprofiel: regulier vmbo kb
QS_V_UIT06	Uitstroomprofiel: regulier vmbo bb
QS_V_UIT07	Uitstroomprofiel: praktijkonderwijs
QS_V_UIT08	Uitstroomprofiel: vso arbeid
QS_V_UIT09	Uitstroomprofiel: vso dagbesteding
QS_V_UIT10	Uitstroomprofiel: geen uitstroom
QS_V_UIT11	Uitstroomprofiel: vso vwo
QS_V_UIT12	Uitstroomprofiel: vso havo
QS_V_UIT13	Uitstroomprofiel: vso vmbo tl
QS_V_UIT14	Uitstroomprofiel: vso vmbo gt
QS_V_UIT15	Uitstroomprofiel: vso vmbo kb
QS_V_UIT16	Uitstroomprofiel: Vvso vmbo bb
QS_V_UIT17	Uitstroomprofiel: anders namelijk
QS_V_UIT18	Uitstroomprofiel: anders namelijk txt
QS_V_VER01	Sinds welk schooljaar zit de leerling op het so/sbo?
QS_V_STI01	Stimulerende en belemmerende factoren: Inhibitie
QS_V_STI02	Stimulerende en belemmerende factoren: Emotieregulatie
QS_V_STI03	Stimulerende en belemmerende factoren: Planning en organisatie
QS_V_STI04	Stimulerende en belemmerende factoren: Zelfstandigheid
QS_V_STI05	Stimulerende en belemmerende factoren: Motivatie
QS_V_STI06	Stimulerende en belemmerende factoren: Concentratie en aandacht
QS_V_STI07	Stimulerende en belemmerende factoren: Schoolwelbevinden
QS_V_STI08	Stimulerende en belemmerende factoren: Thuissituatie
QS_V_OND01	Onderwijsbehoeften: Structuur bieden
QS_V_OND02	Onderwijsbehoeften: Zelfvertrouwen stimuleren
QS_V_OND03	Onderwijsbehoeften: Verantwoordelijkheid geven
QS_V_THU01	Thuistaal: Nederlands
QS_V_THU02	Thuistaal: Nederlands dialect
QS_V_THU03	Thuistaal: Fries
QS_V_THU04	Thuistaal: anders
QS_V_HER01	Is de leerling in Nederland geboren?
QS_V_STO01	Stoornissen: ADHD/ADD
QS_V_STO02	Stoornissen: ODD/CD
QS_V_STO03	Stoornissen: Autisme en verwante stoornissen (zoals PDD-NOS)
QS_V_STO04	Stoornissen: Angst- en/of stemmingsstoornis
QS_V_STO05	Stoornissen: Hechtingsstoornis
QS_V_STO06	Stoornissen: OCD/dwanghandelingen

QS_V_STO07	Stoornissen: Gilles de la Tourette
QS_V_STO08	Stoornissen: Andere psychiatrische stoornis, namelijk
QS_V_STO09	Stoornissen: anders, namelijk txt
QS_V_EXT01	Externaliserend probleemgedrag: Overactief, impulsief gedrag
QS_V_EXT02	Externaliserend probleemgedrag: Agressief tegenover klasgenoten
QS_V_EXT03	Externaliserend probleemgedrag: Gedrag dat niet past bij de leeftijd
QS_V_EXT04	Externaliserend probleemgedrag: Sterk storend voor anderen
QS_V_EXT05	Externaliserend probleemgedrag: Is zelf een pester
QS_V_INT01	Internaliserend probleemgedrag: Faalangstig
QS_V_INT02	Internaliserend probleemgedrag: Angstig, neerslachtig, depressief
QS_V_INT03	Internaliserend probleemgedrag: Teruggetrokken
QS_V_INT04	Internaliserend probleemgedrag: Gebrek aan zelfvertrouwen, negatief zelfbeeld
QS_V_INT05	Internaliserend probleemgedrag: Gebrek aan weerbaarheid
QS_V_INT06	Internaliserend probleemgedrag: Gebrek aan sociale vaardigheden
QS_V_INT07	Internaliserend probleemgedrag: Wordt afgewezen en/of gepest door anderen
QS_V_LEE01	Leerachterstand: Achterstand in taal/lezen
QS_V_LEE02	Leerachterstand: Achterstand in rekenen
QS_V_LEE03	Leerachterstand: Achterstand in informatieverwerking
QS_V_EIQ01	IQ lager dan 70
QS_V_EIQ02	IQ hoger dan 130
QS_V_LUI01	De inhoud van een gesproken tekst in eigen woorden weergeven
QS_V_LUI02	de hoofdpunten uit gesproken (nieuws)berichten halen
QS_V_LUI03	Een oordeel over een tekst(deel) of tv- of radioprogramma verwoorden
QS_V_SPR01	Verstaanbaarheid
QS_V_SPR02	Een verhaal vertellen
QS_V_SPR03	Bij het gespreksonderwerp blijven of afronden
QS_V_SPR04	Een gesprek beginnen
QS_V_SPR05	Een gesprek op gang houden
QS_V_SPR06	Gesprekspartner volgen
QS_V_SPR07	Opmerkingen

10.5 Variabelen voor de toetsondersdelen

Voor de verschillende toetsonderdelen zijn de volgende variabelen beschikbaar:

LUI_WEIGHT	Sampling weight so-leerlingen bij Luisteren (weging naar leerroutes 1/2/3)
LUI_VERS	Toetsversie die de leerling heeft gemaakt bij Luisteren
LUI_MISS	Percentage ontbrekende antwoorden bij Luisteren
LUI_RSCO	Ruwe score op de gemaakte toetsversie Luisteren
LUI_WSCO	Gewogen score op de gemaakte toetsversie Luisteren
LUI_THET	Weighted Maximum Likelihood vaardigheidsscore Luisteren
LUI_BANK	Verwachte bankscore bij Luisteren
LUI_REFN	Behaald referentieniveau (<1F/1F/2F) Luistervaardigheid
SPR_WEIGHT	Sampling weight so-leerlingen bij Spreken (weging naar leerroutes 1/2/3)
SPR_MISS	Percentage ontbrekende antwoorden bij Spreken
SPR_RSCO	Toetsscore bij Spreken
SPR_REFN	Behaald referentieniveau (<1F/1F/2F) Spreekvaardigheid
SPR_TTLID	Toetsleideridentificatie bij Spreken
GROEPSLID	Groepslidnummer bij Gesprekken
GRP_ID	Groepsidentificatie bij Gesprekken
GSP_TTLID	Toetsleideridentificatie bij Gesprekken
N_SPELERS	Aantal spelers bij Gesprekken
GSP_WEIGHT	Sampling weight so-leerlingen bij Gesprekken (weging naar leerroutes 1/2/3)
GSP_MISS	Percentage ontbrekende antwoorden bij Gesprekken
GSP_RSCO	Toetsscore bij Gesprekken
GSP_MISS_LUI	Percentage ontbrekende antwoorden bij het extra onderdeel Luisteren
GSP_RSCO_LUI	Toetsscore bij het extra onderdeel Luisteren
GSP_REFN	Behaald referentieniveau (<1F/1F/2F) Gespreksvaardigheid

11 Referenties

Alberts, N., de Haan, S. & Kuhlemeijer, H. (in voorbereiding). *Wetenschappelijke verantwoording toetsen VSO Dagbesteding en VSO Arbeid-PrO*. Arnhem: Cito

Audigier, V., White, I. R., Jolani, S., Debray, T. P. A., Quartagno, M., Carpenter, J., van Buuren, S., & Resche-Rigon, M. (2018). Multiple Imputation for Multilevel Data with Continuous and Binary Variables. *Statistical Science, Vol. 33-2*, pp. 160-183.

Bloom, L. & Lahey, M. (1978). *Language development and language disorders*. New York: Wiley & Sons.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296-322.

Cito (2010). *Checklist toetsconstructie Speciaal (basis)onderwijs*. Arnhem: Cito.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale: Erlbaum.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Davis, P. & Scott, A. (1995). The effect of interviewer variance on domain comparisons. *Survey Methodology, 21*, 99-106.

Deci, E.L. & Ryan, R.M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 319-338.

Eggen, T.J.H.M. (2004). *Contributions to the theory and practice of computerized adaptive testing*. Arnhem: Cito.

Eekhout, I. (2014). *Don't Miss Out! Incomplete data can contain valuable information*. PhD Dissertation.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology, Vol. 40*, pp. 31-53.

Evers, A., Lucassen, W., Meijer, R. & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam, NIP/COTAN.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*, 255–282

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hemker, B., Van Ingen, K. & Linthorst, R. (2018). *Rapportage Standaardbepalingen Peilingsonderzoek mondelinge taalvaardigheid so/sbo 2018*. Arnhem: Cito.

Hollenberg, J., Keuning, J. & Meijer, J. (2016). *Technisch rapport. Gebruik van toetsen bij het plannen van leerroutes*. Amsterdam: Kohnstamm Instituut.

Kamphuis, F. (1998). *Multi*. Arnhem: Cito

Kamphuis, F. (1993). Estimation and prediction of individual ability in longitudinal studies. In J.H.L. Oud & R.A.W. van den Blokland-Vogelesang (Eds.), *Advances in long and multivariate analysis in the behavioral sciences*. Nijmegen: ITS.

Keuning, J., & Hemker, B. (2014). Estimation of Error Components in Cohort Studies: A Cross-Cohort Analysis of Dutch Mathematics Achievement. *Educational Research and Evaluation, 20*, 199-209.

Knowles, J. E., & Frederick, C. (2018). merTools: Tools for Analyzing Mixed Effect Regression Models. R package version 0.4.1. https://CRAN.R-project.org/package=merTools.

Krom, R., Berkel, S. van, Schoot, F. van der, Sijtstra, J., Hemker, B., & Marsman, M. (2011). Balans van het luisteronderwijs in het basis- en speciaal basisonderwijs 4. Uitkomsten van de vierde peiling in 2007. PPON-reeks nr. 46. Arnhem: Cito.

Lahey, M. (1988). *Language disorders and language development*. New York: MacMillan.

Langberg, M., Leenders, E., & Koopmans, A. (2012). *Passende perspectieven taal. Overzichten van leerroutes*. Enschede: SLO.

Langberg, M, Jansma, S., Fortgens, C., & Van der Schie, V. (2012). *Passende perspectieven taal. Profielschetsen*. Enschede: SLO.

Ledoux, G., Roeleveld, J., Langen, A. van, Smeets, E. (2012). *Cool Speciaal. Inhoudelijk rapport* (Rapport 884, projectnummer 20476). Amsterdam: Kohnstamm Instituut.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*, 764-766.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison–Wesley.

Lindgren, B.W. (1976) *Statistical theory* (3rd ed.). New York: Macmillan.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86-92.

Meijerink (2009). *Referentiekader taal en rekenen*. Enschede: SLO.

OECD (2003). *Framework for the development of OECD Education Indicators*.
Quartagno, M., & Carpenter, J. (2019). jomo: A package for Multilevel Joint Modelling Multiple Imputation. https://CRAN.R-project.org/package=jomo.
Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.
SLO (2016). *Mondelinge taalvaardigheid in het basisonderwijs. Een domeinbeschrijving ten behoeve van peilingsonderzoek*. Enschede: SLO.
Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107–120.
Snijders, T. A. B., & Bosker, R. J. (2012). *Multi-level analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: SAGE Publications.
Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271-295.
Ten Berge, J.M.F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika, 69*, 613–625.
Van Berkel, S., Hilte, M., Van der Zanden, M. en Kamphuis, F. (2016). *Wetenschappelijke verantwoording Begrijpend luisteren groep 8*. Arnhem: Cito
Van Boxtel, H., Hollenberg, J., Horsels, L. & Kleintjes, F. (2012). *Wetenschappelijke verantwoording van de toetsen Taal voor zml-leerlingen*. Arnhem: Cito.
Van Boxtel, H., & Hemker, B.T. (2009). *Wetenschappelijke verantwoording van de Intelligentietest Eindtoets Basisonderwijs*. Arnhem: Cito.
Van Langen, A., Van Druten-Frietman, L., Wolbers, M., Teunissen, C., Strating, H., Dood, C., Geelen, A. & Binsbergen, M. (2017). *Peilingsonderzoek Mondelinge Taalvaardigheid in het basisonderwijs*. Nijmegen: KBA.
Verhelst, N.D. & Verstralen, H.H.F.M. (2008): Some Considerations on the Partial Credit Model. *Psicologica, 29*, 229-254.
Verhelst, N.D., & Glas, C.A.W. (1995). The one-parameter logistic model. In G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models. Foundations, recent developments, and applications*. New York: Springer-Verlag.
Verhelst N.D., Glas C.A.W., & Verstralen H.H.F.M. (1995). *OPLM: One-Parameter Logistic Model. Computer program and manual*. Arnhem: Cito.

Bijlage 1 Quickscan

Quickscan Peiling mondelinge taalvaardigheid sbo en so

17930

Schoolnummer

Naam Invuller
(in blokletters invullen)

Leerlinggegevens

Voornaam
(in blokletters invullen)

Achternaam
(in blokletters invullen)

Geslacht
Jongen ○ Meisje ○

Geboortedatum
__ __ - __ __ - __ __ __ __

Invuldatum
__ __ - __ __ - 2 0 1 8

quickscan_1709 Pag 1

Leerlingkenmerken peiling Mondelinge Taalvaardigheid sbo en so

17930

Voor het peilingsonderzoek hebben wij enkele gegevens van uw leerlingen nodig. Wilt u deze lijst invullen voor alle leerlingen van wie u er nu vanuit gaat dat zij eind van het schooljaar uitstromen?
Uitgezonderd zijn:

- leerlingen met een gediagnosticeerde Taal-ontwikkelingsstoornis (TOS)[1]
- leerlingen met een beperkt gehoor[2]
- leerlingen met selectief mutisme
- leerlingen met afasie
- leerlingen die minder dan 2 jaar in Nederland verblijven

Voor deze leerlingen die uitgezonderd zijn van deelname willen we vragen wel het voorblad en onderstaande vraag over het uitstroomprofiel in te vullen.

Uitstroomprofiel
Waar stroomt deze leerling naar verwachting na groep 8 naar uit?

- ☐ regulier vwo
- ☐ regulier havo
- ☐ regulier vmbo tl
- ☐ regulier vmbo gt
- ☐ regulier vmbo kb
- ☐ regulier vmbo bb
- ☐ Praktijkonderwijs
- ☐ vso Arbeid
- ☐ vso dagbesteding
- ☐ geen uitstroom
- ☐ anders, namelijk

- ☐ vso vwo
- ☐ vso havo
- ☐ vso vmbo tl
- ☐ vso vmbo gt
- ☐ vso vmbo kb
- ☐ vso vmbo bb

[1] TOS staat voor taalontwikkelingsstoornis. Dit is een stoornis in het leren van de moedertaal. De problemen die kinderen met TOS ervaren verschillen per kind. Ze kunnen bijvoorbeeld moeite hebben met:
- het leren van nieuwe woorden
- het begrijpen van mondelinge instructies
- het maken van grammaticaal goede zinnen

[2] De testonderdelen van de peiling zijn niet geschikt voor leerlingen met een beperkt gehoor. We weten dan namelijk niet of eventuele lagere prestaties toe te schrijven zijn aan een beperkte mondelinge taalvaardigheid, of dat een leerling de opgaven niet goed gehoord heeft. Heeft een leerling een gehoorapparaat, waardoor hij in de klas goed kan meekomen met mondelinge instructies? Dan kan hij uiteraard meedoen.

quickscan_1709 Pag 2

Verblijfsduur so/sbo
Sinds welk schooljaar zit de leerling op het so/sbo?

2009-2010	☐
2010-2011	☐
2011-2012	☐
2012-2013	☐
2013-2014	☐
2014-2015	☐
2015-2016	☐
2016-2017	☐
2017-2018	☐

17930

Stimulerende en belemmerende factoren
Geef bij elk van onderstaande factoren aan of deze belemmerend, neutraal of stimulerend voor de leerling is.

	toelichting	belemmerend	neutraal	stimulerend
Inhibitie	Vermogen impulsen te beheersen.	☐	☐	☐
Emotieregulatie	Emoties onderkennen, regelen en in goede banen leiden.	☐	☐	☐
Planning en organisatie	Bij een taak bekijken welke acties en middelen nodig zijn en bepalen in welke volgorde de acties uitgevoerd moeten worden.	☐	☐	☐
Zelfstandigheid	In staat zijn om onafhankelijk van een ander praktische taken uit te voeren en keuzes te maken.	☐	☐	☐
Motivatie	In hoeverre een leerling zonder aansporing en beloning aan het werk kan gaan en blijven.	☐	☐	☐
Concentratie en aandacht	Hoe lang en intensief de leerling de aandacht ergens op kan vestigen.	☐	☐	☐
Schoolwelbevinden	In hoeverre de leerling zich fijn op school voelt, een goede relatie met de leerkracht heeft en zich sociaal aanvaard voelt.	☐	☐	☐
Thuissituatie	In hoeverre de leerling vanuit huis voor school wordt gestimuleerd (zoals veel voorlezen, interesse voor school, culturele uitstapjes). Daarnaast in hoeverre ouders realistische verwachtingen hebben en toezicht houden op het gedrag van hun kind.	☐	☐	☐

Onderwijsbehoeften
Welke van onderstaande onderwijsbehoeften heeft de leerling extra ten opzichte van andere leerlingen op uw school?
Meerdere antwoorden mogelijk.

	Extra ten opzichte van leerlingen op eigen school
Structuur bieden	☐
Zelfvertrouwen stimuleren	☐
Verantwoordelijkheid geven	☐

17930

Thuis(taal)situatie

Wat is de thuistaal van de leerling?
- ☐ Nederlands
- ☐ Nederlands dialect
- ☐ Fries
- ☐ Anders

Is de leerling in Nederland geboren?
- ☐ ja
- ☐ nee

Stoornissen, beperkingen en kenmerken
Is er sprake van één of meerdere van onderstaande stoornissen, beperkingen en kenmerken?

17930

Stoornissen	Gediagnosticeerd	Vermoeden van
ADHD/ADD	☐	☐
ODD/CD	☐	☐
Autisme en verwante stoornissen (zoals PDD-NOS)	☐	☐
Angst- en/of stemmingsstoornis	☐	☐
Hechtingsstoornis	☐	☐
OCD/dwanghandelingen	☐	☐
Gilles de la Tourette	☐	☐
Andere psychiatrische stoornis, namelijk	☐	☐

Externaliserend probleemgedrag	niet aanwezig	in lichte mate	in behoorlijke mate	in ernstige mate
Overactief, impulsief gedrag	☐	☐	☐	☐
Agressief tegenover klasgenoten	☐	☐	☐	☐
Gedrag dat niet past bij de leeftijd	☐	☐	☐	☐
Sterk storend voor anderen	☐	☐	☐	☐
Is zelf een pester	☐	☐	☐	☐
Internaliserend probleemgedrag	**niet aanwezig**	**in lichte mate**	**in behoorlijke mate**	**in ernstige mate**
Faalangstig	☐	☐	☐	☐
Angstig, neerslachtig, depressief	☐	☐	☐	☐
Teruggetrokken, komt niet tot contact met anderen	☐	☐	☐	☐
Gebrek aan zelfvertrouwen, negatief beeld van zichzelf	☐	☐	☐	☐
Gebrek aan weerbaarheid	☐	☐	☐	☐
Gebrek aan sociale vaardigheden	☐	☐	☐	☐
Wordt afgewezen en/of gepest door anderen	☐	☐	☐	☐
Leerachterstand	**geen achterstand**	**achterstand minder dan 1 jaar**	**achterstand 1-2 jaar**	**achterstand meer dan 2 jaar**
Achterstand in taal/lezen	☐	☐	☐	☐
Achterstand in rekenen	☐	☐	☐	☐
Achterstand in informatieverwerking	☐	☐	☐	☐

IQ (ver) boven of beneden gemiddeld	Vastgesteld
IQ lager dan 70	☐
IQ hoger dan 130	☐

QuickScan Peiling Mondelinge Taalvaardigheid sbo en so

17930

Werkwijze
U vult de lijst in met de leerling in gedachten. Per onderdeel kruist u aan welke beschrijving het best bij de leerling past. Twijfelt u over een onderdeel? Observeer deze dan op een later moment in de klas.
Belangrijk: Kruis één vakje per regel aan.
Eventuele opmerkingen over het invullen van de quickscan kunt u kwijt in het opmerkingenveld (onderaan de laatste bladzijde van dit formulier).

Luisteren

	Onderdeel	→ Moeilijkheid →			
1	**De inhoud van een gesproken tekst in eigen woorden weergeven**	☐ Kan (nog) niet in eigen woorden de hoofdlijnen noemen van een gesproken tekst met een concreet onderwerp.	☐ Kan in eigen woorden de hoofdlijnen noemen van een gesproken tekst met een concreet onderwerp.	☐ Kan in eigen woorden de hoofdlijnen noemen van een gesproken tekst met een wat abstracter onderwerp dat wel dicht bij de leefwereld ligt.	☐ Kan in eigen woorden de hoofdlijnen noemen van een gesproken tekst met een abstract onderwerp dat wat verder van de leefwereld ligt.
2	**de hoofdpunten uit gesproken (nieuws)berichten halen**	☐ Kan (nog) niet de hoofdpunten halen uit een kort nieuwsbericht over een concreet onderwerp, ook niet met hulp.	☐ Kan met hulp de hoofdpunten halen uit een kort nieuwsbericht over een concreet onderwerp.	☐ Kan zonder hulp de hoofdpunten halen uit een nieuwsbericht over een concreet onderwerp en met hulp uit een nieuwsbericht met een abstracter onderwerp.	☐ Kan zonder hulp de hoofdpunten halen uit een nieuwsbericht over een concreet of abstract onderwerp.
3	**Een oordeel over een tekst(deel) of tv- of radioprogramma verwoorden**	☐ Kan (nog) geen oordeel verwoorden over een gesproken tekst of tv-programma met een concreet onderwerp, ook niet met hulp.	☐ Kan met hulp een oordeel verwoorden over een gesproken tekst of tv-programma met een concreet onderwerp.	☐ Kan zonder hulp een oordeel verwoorden over een gesproken tekst of tv-programma met een abstracter verhaal dat dicht bij de belevingswereld ligt.	☐ Kan zonder hulp een oordeel verwoorden over een tekst of tv-programma met een abstracter verhaal dat wat verder van de belevingswereld ligt.

Spreken en Gesprekken voeren

	Onderdeel	→ Moeilijkheid →			
4	**Verstaanbaarheid**	☐ Spreekt niet of nauwelijks; daardoor niet goed te beoordelen.	☐ Spreekt moeilijk verstaanbaar voor zowel bekenden als onbekenden.	☐ Spreekt vrijwel altijd verstaanbaar voor bekenden, maar niet voor onbekenden.	☐ Spreekt vrijwel altijd verstaanbaar voor zowel bekenden als onbekenden.
5	**Een verhaal vertellen**	☐ Kan (nog) niet een duidelijk en samenhangend verhaal vertellen over een concreet en vertrouwd onderwerp.	☐ Kan met hulp een duidelijk en samenhangend verhaal over een concreet en vertrouwd onderwerp.	☐ Kan een duidelijk en samenhangend verhaal vertellen over een concreet en vertrouwd onderwerp.	☐ Kan een duidelijk en samenhangend verhaal vertellen over een abstract, niet alledaags onderwerp.
6	**Bij het gespreks-onderwerp blijven of afronden**	☐ Rondt zelden het oude gespreksonderwerp af voordat hij een nieuw onderwerp aansnijdt, reageert zelden adequaat als hij wordt bijgestuurd.	☐ Rondt zelden het oude gespreksonderwerp af voordat hij een nieuw onderwerp aansnijdt, reageert meestal adequaat als hij wordt bijgestuurd.	☐ Rondt meestal eerst het oude gespreksonderwerp af voordat hij een nieuw onderwerp aansnijdt, en reageert adequaat als hij wordt bijgestuurd.	☐ Rondt vrijwel altijd eerst het oude gespreksonderwerp af voordat hij een nieuw onderwerp aansnijdt en reageert adequaat als hij wordt bijgestuurd.
7	**Een gesprek beginnen**	☐ Neemt zelden het initiatief in de communicatie of kan ondanks sturing niet op een gepast moment een gesprek starten.	☐ Vindt het in alle situaties lastig in te schatten wat een gepast moment is om een gesprek te starten; hij heeft hierbij sturing nodig.	☐ Schat in een bekende, gestandaardiseerde situatie in wat een gepast moment is om een gesprek te starten. Toelichting: Bijvoorbeeld tijdens een kringgesprek of een instructie.	☐ Schat in een vrije situatie in wat een gepast moment is om een gesprek te starten. Toelichting: Bijvoorbeeld tijdens de pauze.
8	**Een gesprek op gang houden**	☐ Kan een gesprek niet op gang houden, ook niet met hulp van de gesprekspartner.	☐ Kan een gesprek zelf meestal niet op gang houden, maar heeft hierbij hulp nodig van de gesprekspartner.	☐ Kan een gesprek over een concreet en vertrouwd onderwerp zelf op gang houden. Toelichting: Bijvoorbeeld door te knikken of een wedervraag te stellen.	☐ Kan gesprekken over abstracte, niet alledaagse onderwerpen zelf op gang houden. Toelichting: Bijvoorbeeld door te knikken of een wedervraag te stellen.

17930

9	**Gesprekspartner volgen**	☐ Reageert meestal niet adequaat binnen een gesprek over een concreet, vertrouwd onderwerp.	☐ Reageert meestal adequaat binnen een gesprek over een concreet, vertrouwd onderwerp. Bijvoorbeeld door passende antwoorden op vragen te geven of passende vragen te stellen. Soms heeft hij ((on)gevraagd)) herhaling nodig.	☐ Reageert adequaat binnen een gesprek over een concreet, vertrouwd onderwerp. Bijvoorbeeld door passende antwoorden op vragen te geven of passende vragen te stellen.	☐ Reageert adequaat binnen een gesprek over een abstract, niet alledaags onderwerp. Bijvoorbeeld door passende antwoorden op vragen te geven of passende vragen te stellen.

Opmerkingen

17930

Bijlage 2 Beoordelingsmodel Spreken

Spreken
Beoordelingsmodel
Peiling Mondelinge Taalvaardigheid in het SBO en SO

Schoolnummer

Naam Toetsleider
(in blokletters invullen)

Leerlinggegevens

Voornaam
(in blokletters invullen)

Achternaam
(in blokletters invullen)

Geslacht
Jongen Meisje

Afnamedatum
- - 2018

Draft

Preconditie		
0. Inhoudelijke adequaatheid		
	De leerling heeft de beoogde taak niet (zelfstandig/volledig) uitgevoerd. *Indien deze score wordt toegekend, wordt de leerling niet verder beoordeeld op dat onderdeel.*	De leerling heeft de beoogde taak uitgevoerd.
Filmpje	☐ stop	☐ Naar 1
Dobbelspel	☐ stop	☐ Naar 1

Uitvoering			
1. Volledigheid			
	De leerling heeft lang niet alle taalhandelingen van de opdracht uitgevoerd.	De leerling heeft de meerderheid van de taalhandelingen van de opdracht uitgevoerd.	De leerling heeft (bijna) alle taalhandelingen van de opdracht uitgevoerd.
Filmpje[1]	☐	☐	☐
Dobbelspel[2]	☐	☐	☐
2. Inhoudelijke kwaliteit[3]			
	De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten zwak.	De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten redelijk.	De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten sterk.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐

Samenhang			
3. Samenhang op macro- en mesoniveau: opbouw			
	De leerling laat geen herkenbare opbouw zien.	De leerling laat deels een opbouw zien, hoewel deze niet altijd logisch en/of correct is, bijvoorbeeld met inleiding en afsluiting, maar zonder verbinding tussen de onderdelen.	De leerling laat een logische opbouw zien, bijvoorbeeld met zowel verbinding binnen als verbinding tussen inleiding, kern en afsluiting.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐
4. Samenhang op microniveau: signaalwoorden			
	De leerling gebruikt geen voeg- en verbindingswoorden.	De leerling gebruikt eenvoudige voeg- en verbindingswoorden, maar niet altijd correct.	De leerling gebruikt eenvoudige voeg- en verbindingswoorden correct.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐

[1] De leerling vertelt wat hij gezien én gehoord heeft in het filmpje.

[2] De leerling begint een verhaal met 'Er was eens' en verwerkt de plaatjes van de dobbelstenen in zijn verhaal.

[3] Voorbeelden van inhoudsaspecten zijn: de complexiteit van de inhoud, de helderheid van chronologie en de mate van originaliteit.

Draft

Afstemming op doel			
5. Spreekdoel			
Filmpje	De leerling kan de strekking van het filmpje niet duidelijk maken aan de gesprekspartner (die het filmpje niet gezien en gehoord heeft).	De leerling kan de strekking van het filmpje deels duidelijk maken aan de gesprekspartner (die het filmpje niet gezien en gehoord heeft).	De leerling kan de strekking van het filmpje duidelijk maken aan de gesprekspartner (die het filmpje niet gezien en gehoord heeft).
	☐	☐	☐
Dobbelspel	De leerling vertelt een verhaal waar hij geen, of maar één of twee plaatjes in verwerkt.	De leerling vertelt een verhaal waar hij drie of vier plaatjes in verwerkt.	De leerling vertelt een verhaal waar hij alle plaatjes in verwerkt.
	☐	☐	☐

Afstemming op de gesprekspartner			
6. Taalgebruik afstemmen			
	De leerling gebruikt taal die niet bij de gesprekspartner past.	De leerling gebruikt af en toe taal die bij de gesprekspartner past, maar doet dat niet consequent.	De leerling gebruikt consequent taal die bij de gesprekspartner past.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐
7. Gesprekspartner boeien			
	De leerling zet geen aspecten als stiltes, wisselende intonatie/volume/spreeksnelheid en lichaamstaal in om de gesprekspartner te boeien.	De leerling zet in beperkte mate aspecten als stiltes, wisselende intonatie/volume/spreeksnelheid en lichaamstaal bewust in om de gesprekspartner te boeien.	De leerling zet regelmatig aspecten als stiltes, wisselende intonatie/volume/ spreeksnelheid en lichaamstaal bewust in om de gesprekspartner te boeien.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐
8. Non-verbale communicatie			
	Het non-verbale gedrag van de leerling past niet altijd bij zijn verbale boodschap.	Het non-verbale gedrag van de leerling past bij zijn verbale boodschap.	Het non-verbale gedrag van de leerling zet zijn verbale boodschap kracht bij.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐

Draft

Woordgebruik en woordenschat			
9. Woordgebruik			
	De uitwerking van de leerling is weinig aantrekkelijk door het eenzijdig en/of herhaald woordgebruik.	De uitwerking van de leerling is enigszins aantrekkelijk door een beperkte variatie in woordgebruik.	De uitwerking van de leerling is aantrekkelijk door voldoende variatie in woordgebruik.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐
10. Woordenschat			
	De getoonde woordenschat van de leerling is niet toereikend voor het uitvoeren van de opdracht.	De leerling zoekt regelmatig naar woorden, maar de getoonde woordenschat is toereikend voor het uitvoeren van de opdracht.	De getoonde woordenschat van de leerling is toereikend voor het uitvoeren van de opdracht.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐

Vloeiendheid, verstaanbaarheid en grammaticale beheersing			
11. Vloeiendheid			
	De leerling spreekt niet vloeiend, bijvoorbeeld met een groot aantal haperingen en/of herformuleringen.	De leerling spreekt redelijk vloeiend, al vertoont zijn uitwerking meerdere haperingen en/of herformuleringen.	De leerling spreekt redelijk vloeiend, al vertoont zijn uitwerking een enkele hapering en/of herformulering.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐
12. Uitspraak en intonatie[4]			
	De leerling is niet goed verstaanbaar vanwege zijn uitspraak, articulatie en/of intonatie.	De leerling is voldoende verstaanbaar, ondanks een eventueel accent.	De leerling is goed verstaanbaar, ondanks een eventueel accent.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐
13. Grammatica: vervoeging			
	De leerling vervoegt veelvoorkomende werkwoorden meestal niet correct.	De leerling vervoegt veelvoorkomende werkwoorden over het algemeen correct.	De leerling vervoegt alle werkwoorden over het algemeen correct.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐

[4] Bij voldoende luidheid

Draft

14. Grammatica: verbuiging			
	De leerling verbuigt (combinaties met) naamwoorden meestal niet correct.	De leerling verbuigt (combinaties met) naamwoorden soms correct.	De leerling verbuigt (combinaties met) naamwoorden meestal correct.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐

15. Grammatica: zinsbouw			
	De leerling maakt geen correcte enkelvoudige zinnen.	De leerling maakt over het algemeen correcte enkelvoudige zinnen, maar hij maakt geen samengestelde zinnen of doet dat niet correct.	De leerling maakt over het algemeen correcte enkelvoudige en samengestelde zinnen.
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐

Globaal oordeel			
16. Hoe zou je de spreekvaardigheid van deze leerling globaal gezien typeren in termen van referentieniveau 1F?			
	matig	voldoende	goed
Filmpje	☐	☐	☐
Dobbelspel	☐	☐	☐

Draft

Bijlage 3 Beoordelingsmodel Gesprekken

Samenwerkingsspel – Gesprekken

Peiling Mondelinge Taalvaardigheid in het SBO en SO

Beoordelingsmodel

Schoolnummer

Naam Toetsleider (in blokletters invullen)

Leerlinggegevens

Achternaam leerling 1 (lln 1) (in blokletters invullen)

Leerlingnummer lln 1

Achternaam leerling 2 (lln 2) (in blokletters invullen)

Leerlingnummer lln 2

Achternaam leerling 3 (lln 3 (in blokletters invullen)

Leerlingnummer lln 3

Achternaam leerling 4 (lln 4) (in blokletters invullen)

Leerlingnummer lln 4

Afnamedatum

- - 2018

17910

Pagina 1

lln 1	
lln 2	
lln 3	
lln 4	

Het gesprek als geheel (per groep beoordelen)				
1. Evenwichtigheid				
	De bijdragen aan het gesprek zijn niet evenwichtig verdeeld over de leerlingen.	De bijdragen aan het gesprek zijn niet helemaal evenwichtig verdeeld over de leerlingen.	De bijdragen aan het gesprek zijn evenwichtig verdeeld over de leerlingen.	Niet te beoordelen.
groepje	☐	☐	☐	☐
2. Behoefte aan sturing				
	In het gesprek is vaak sturing van de toetsleider nodig.	In het gesprek is af en toe sturing van de toetsleider nodig.	In het gesprek is geen sturing van de toetsleider nodig.	Niet te beoordelen.
groepje	☐	☐	☐	☐
3. Gedrag van de toetsleider				
	Het gedrag van de toetsleider heeft mogelijk een negatief effect op het verloop van het spel[1].	Het gedrag van de toetsleider heeft geen negatief effect op het verloop van het spel.		Niet te beoordelen.
groepje	☐	☐		☐
4. Bijzonderheden				
Zijn er bijzonderheden[1] tijdens de afname?				
Niet meegenomen in score individuele leerling.				

[1] Voorbeelden van mogelijke negatieve invloeden zijn: inhoudelijk feedback geven; onnodig ingrijpen in het gesprek; evaluatievragen stellen; non-verbale communicatie.

17910

lln 1	
lln 2	
lln 3	
lln 4	

Uitvoering				
5. Kwaliteit van de inhoud[2]				
	De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten zwak.	De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten redelijk.	De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten sterk.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

Het woord nemen[3] en bijdragen aan samenhang				
6. Gesprek onderhouden				
	De leerling neemt geen initiatief om het gesprek te beginnen, gaande te houden en/of af te sluiten.	De leerling neemt af en toe initiatief om het gesprek te beginnen, gaande te houden en/of af te sluiten.	De leerling neemt regelmatig initiatief om het gesprek te beginnen, gaande te houden en/of af te sluiten (bijvoorbeeld door vragen te stellen of voorstellen te doen).	
lln 1	☐	☐	☐	
lln 2	☐	☐	☐	
lln 3	☐	☐	☐	
lln 4	☐	☐	☐	
7. Passendheid van het woord nemen[4]				
	De leerling neemt vaak op niet-passende wijze de het woord.	De leerling neemt het woord, maar niet altijd op passende wijze.	De leerling neemt doorgaans op passende wijze het woord.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

[2] Voorbeelden van inhoudsaspecten zijn: de aanwezigheid en kwaliteit van de argumentatie; de helderheid van de voorstellen; de complexiteit van de inhoud.

[3] Met 'het woord nemen' wordt bedoeld: binnen het gesprek het woord nemen. Het gaat hier dus niet om 'het woord nemen' wanneer de leerling in het spel aan de beurt is.

[4] Voorbeelden van het woord nemen op passende wijze zijn:
- een zin beginnen wanneer de gesprekspartner een zin beëindigd heeft,
- interrumperen op juist moment (wanneer de bijdrage van de ander goed beoordeeld is),
- frases gebruiken om het woord te krijgen.

17910

lln 1	
lln 2	
lln 3	
lln 4	

8. Inbreng gesprekspartners integreren				
	De leerling houdt geen rekening met de bijdragen van anderen.	De leerling erkent de bijdragen van de gesprekspartners, maar gaat vervolgens met zijn eigen inhoud verder.	De leerling integreert bijdragen van de gesprekspartners in zijn eigen inbreng wanneer dat toegevoegde waarde heeft.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

Afstemming op doel				
9. Doelgerichtheid				
	De leerling levert geen actieve bijdrage aan het behalen van gezamenlijke gespreksdoelen[5].	De leerling levert een actieve bijdrage aan het behalen van gezamenlijke gespreksdoelen.	De leerling neemt initiatief om een gezamenlijk gespreksdoel te behalen door de gesprekspartners te betrekken bij het gesprek, bijvoorbeeld met behulp van vragen.	
lln 1	☐	☐	☐	
lln 2	☐	☐	☐	
lln 3	☐	☐	☐	
lln 4	☐	☐	☐	
10. Begrijpelijk maken				
	De leerling maakt de eigen gedachtegang niet begrijpelijk.	De leerling maakt de eigen gedachtegang redelijk begrijpelijk.	De leerling maakt de eigen gedachtegang begrijpelijk.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

[5] Voorbeelden van gezamenlijke gespreksdoelen zijn:
- besluiten of een speler een worp beter zelf kan houden of door kan geven
- besluiten welke vis er gevangen gaat worden en door wie
- besluiten wie de vis naar de haven gaat brengen

lln 1	
lln 2	
lln 3	
lln 4	

Afstemming op de gesprekspartners				
11. Taalgebruik afstemmen				
	De leerling gebruikt geen taal die bij de situatie past.	De leerling gebruik af en toe taal die bij de situatie past, maar doet dat niet consequent.	De leerling gebruikt consequent taal die bij de situatie past.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐
12. Non-verbale communicatie				
	Het non-verbale gedrag van de leerling past niet altijd bij zijn verbale boodschap.	Het non-verbale gedrag van de leerling past bij zijn verbale boodschap.	Het non-verbale gedrag van de leerling zet zijn verbale boodschap kracht bij.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐
13. Actief luistergedrag[6]				
	De leerling vertoont geen actief luistergedrag.	De leerling vertoont enigszins actief luistergedrag.	De leerling vertoont adequaat actief luistergedrag.	
lln 1	☐	☐	☐	
lln 2	☐	☐	☐	
lln 3	☐	☐	☐	
lln 4	☐	☐	☐	
14. Reageren op gesprekspartner				
	De leerling geeft vaak ongepaste reacties (verbaal/non-verbaal).	De leerling geeft af en toe gepaste reacties (verbaal/non-verbaal), maar reageert soms ook ongepast.	De leerling geeft consequent gepaste reacties (verbaal/non-verbaal).	
lln 1	☐	☐	☐	
lln 2	☐	☐	☐	
lln 3	☐	☐	☐	
lln 4	☐	☐	☐	

[6] Voorbeelden van actief luistergedrag zijn:
- een actieve zithouding,
- oogcontact,
- hoofdbewegingen (knikken) en/of klanken ter bevestiging ('hummen').

17910

In 1	
Iln 2	
Iln 3	
Iln 4	

Woordgebruik en woordenschat				
15. Woordgebruik				
	De inbreng van de leerling vertoont eenzijdig en/of herhaald woordgebruik.	De inbreng van de leerling vertoont een beperkte variatie in woordgebruik.	De inbreng van de leerling vertoont voldoende variatie in woordgebruik.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
Iln 1	☐	☐	☐	☐
Iln 2	☐	☐	☐	☐
Iln 3	☐	☐	☐	☐
Iln 4	☐	☐	☐	☐
16. Woordenschat				
	De getoonde woordenschat van de leerling is niet toereikend om het gesprek te voeren.	De leerling zoekt regelmatig naar woorden, maar de getoonde woordenschat is toereikend om het gesprek te voeren.	De getoonde woordenschat van de leerling is toereikend om het gesprek te voeren.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
Iln 1	☐	☐	☐	☐
Iln 2	☐	☐	☐	☐
Iln 3	☐	☐	☐	☐
Iln 4	☐	☐	☐	☐

Vloeiendheid, verstaanbaarheid en grammaticale beheersing				
17. Vloeiendheid				
	De leerling spreekt niet vloeiend, bijvoorbeeld met een groot aantal haperingen en/of herformuleringen.	De leerling spreekt redelijk vloeiend, al vertoont zijn uitwerking meerdere haperingen en/of herformuleringen.	De leerling spreekt redelijk vloeiend, al vertoont zijn uitwerking een enkele hapering en/of herformulering.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
Iln 1	☐	☐	☐	☐
Iln 2	☐	☐	☐	☐
Iln 3	☐	☐	☐	☐
Iln 4	☐	☐	☐	☐
18. Uitspraak en intonatie				
	De leerling is niet goed verstaanbaar vanwege zijn uitspraak, articulatie en/of intonatie.	De leerling is verstaanbaar genoeg door zijn uitspraak, articulatie en intonatie, ondanks een eventueel accent.	De leerling is goed verstaanbaar door zijn uitspraak, articulatie en intonatie, ondanks een eventueel accent.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
Iln 1	☐	☐	☐	☐
Iln 2	☐	☐	☐	☐
Iln 3	☐	☐	☐	☐
Iln 4	☐	☐	☐	☐

lln 1	
lln 2	
lln 3	
lln 4	

19. Grammatica: vervoeging				
	De leerling vervoegt veelvoorkomende werkwoorden meestal niet correct.	De leerling vervoegt veelvoorkomende werkwoorden over het algemeen correct.	De leerling vervoegt alle werkwoorden over het algemeen correct.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

20. Grammatica: verbuiging				
	De leerling verbuigt (combinaties met) naamwoorden meestal niet correct.	De leerling verbuigt (combinaties met) naamwoorden soms correct.	De leerling verbuigt (combinaties met) naamwoorden meestal correct.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

21. Grammatica: zinsbouw				
	De leerling maakt geen correcte enkelvoudige zinnen.	De leerling maakt over het algemeen correcte enkelvoudige zinnen, maar hij maakt geen samengestelde zinnen of doet dat niet correct.	De leerling maakt over het algemeen correcte enkelvoudige en samengestelde zinnen.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

lln 1	
lln 2	
lln 3	
lln 4	

Evaluatie				
22. Evaluatie				
	De leerling geeft geen helder oordeel over het spel.	De leerling geeft een enigszins helder oordeel over het spel.	De leerling geeft een helder oordeel over het spel.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

Globaal oordeel				
23. Hoe zou je de gespreksvaardigheid van deze leerling globaal gezien typeren in termen van het referentieniveau 1F?				
	matig	voldoende	goed	
lln 1	☐	☐	☐	
lln 2	☐	☐	☐	
lln 3	☐	☐	☐	
lln 4	☐	☐	☐	

17910

Samenwerkingsspel – Luisteren

Beoordelingsmodel

lln 1	
lln 2	
lln 3	
lln 4	

Begrijpen				
24. Instructie toepassen		**Bijvoorbeeld af te leiden uit:** *Leg je boot bij de haven met dezelfde kleur als je boot;* *Mag de boot hier varen?* *Je hebt 2 gegooid. Waar mag je dan allemaal heen varen?* *Mag je hier de vis vangen?* *Mogen de boten zo deze vis vangen?*		
	De leerling past de instructie van de leerkracht meestal niet correct toe.	De leerling past de instructie van de leerkracht soms correct toe.	De leerling past de instructie van de leerkracht over het algemeen correct toe.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

25. Herhaling van instructie		**Bijvoorbeeld af te leiden uit:** *Leg je boot bij de haven met dezelfde kleur als je boot;* *Mag de boot hier varen?* *Je hebt 2 gegooid. Waar mag je dan allemaal heen varen?* *Mag je hier de vis vangen?* *Mogen de boten zo deze vis vangen?*		
	De leerling heeft altijd een herhaling van de instructie nodig.	De leerling heeft soms herhaling van de instructie nodig.	De leerling begrijpt de instructie van de leerkracht in één keer.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

Samenvatten				
26. Mondeling samenvatten		*Opdracht Mondeling samenvatten*		
	De leerling kan een verhaal niet in grote lijnen navertellen.	De leerling kan een verhaal redelijk in grote lijnen navertellen.	De leerling kan een verhaal goed in grote lijnen navertellen.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

17910

lln 1	
lln 2	
lln 3	
lln 4	

Interpreteren				
27. Meningen herkennen		*Opdracht Meningen herkennen*		
	De leerling benoemt de verschillende meningen over een onderwerp niet correct.	De leerling benoemt de verschillende meningen over een onderwerp redelijk correct.	De leerling benoemt de verschillende meningen over een onderwerp correct.	*De inbreng van de leerling is onvoldoende om dit te kunnen beoordelen.*
lln 1	☐	☐	☐	☐
lln 2	☐	☐	☐	☐
lln 3	☐	☐	☐	☐
lln 4	☐	☐	☐	☐

Globaal oordeel				
28. Hoe zou je de luistervaardigheid van deze leerling globaal gezien typeren in termen van het referentieniveau 1F?				
	matig	voldoende	goed	
lln 1	☐	☐	☐	
lln 2	☐	☐	☐	
lln 3	☐	☐	☐	
lln 4	☐	☐	☐	

17910

Bijlage 4 Interviewleidraad

Interviewleidraad Peilingsonderzoek Mondelinge taalvaardigheid sbo en so

Voorbereiding

Afspraak plannen met gesprekspartners
Het interview wordt gehouden met een vertegenwoordiging van de school die een beeld kunnen schetsen van het onderwijs in mondelinge taalvaardigheid. Deze vertegenwoordiging bestaat bij voorkeur uit 3 personen die zicht hebben op wat er schoolbreed aan mondelinge taalvaardigheid wordt gedaan en wat er concreet in de klassen gebeurt. In elk geval moet ook de leerkracht van de eindgroep (waarin peiling heeft plaatsgevonden) deelnemen. Overleg van tevoren met de contactpersoon wie naast deze leerkracht ook aan het gesprek zal deelnemen.

Het interview houden we met een vertegenwoordiging van die school die een goed beeld kan schetsen van de manier waarop het onderwijs in mondelinge taalvaardigheid op uw school is ingericht. Onze vragen gaan onder andere over leerdoelen, type taken en activiteiten, leertijd, evaluatie en aansluiting bij onderwijsbehoeften.

We vragen u van tevoren al wat informatie te verzamelen.

Documenten opvragen
Voorafgaand aan het interview worden de schoolgids en het schoolplan van de school via de website opgezocht zodat de interviewer al een globaal beeld kan vormen van het onderwijs in mondelinge taalvaardigheid op de school. Dit biedt al aanknopingspunten voor het interview. Het gaat om het snel scannen van de informatie: wordt er informatie gegeven over het onderwijs in mondelinge taalvaardigheid.
Bij het inplannen van het interview wordt gevraagd het taalbeleidsplan, indien aanwezig, van tevoren toe te sturen.

Voorbereiding interview
Het interview heeft het karakter van een open gesprek dat wordt gevoerd aan de hand van een checklist. Het is van belang de instructies bij (door)vragen en de scoringsinstructie van te voren goed door te nemen.

Opname gesprek
Het gesprek wordt met een voicerecorder opgenomen. Vermeld dit aan de start van het interview.

Algemene richtlijnen voor het interview
- De checklist moet volledig worden ingevuld.
- Stel open vragen. Daarbij is het van belang dat er niet aangestuurd wordt op sociaal-wenselijke antwoorden.
- Parafraseer en vat samen om te controleren of je het antwoord juist begrepen hebt.
- Laat de gesprekspartners zelf vertellen en vraag om verduidelijking door vragen te stellen als:
    - Waar blijkt dat uit?
    - Kunt u daar voorbeelden van noemen?
- De gesprekspartners kunnen een verschillend beeld schetsen. Vraag door om in kaart te brengen welk antwoord het best de situatie voor de school als geheel schetst door te parafraseren en samen te vatten. Noteer op het interviewformulier als er op onderdelen geen overeenstemming wordt bereikt.

Inhoudelijke aandachtspunten voor het interview
- Maak duidelijk onderscheid tussen schoolbrede aanpak en de aanpak van de specifieke leerkracht
-

Richtlijnen voor het interview

INTRO: Met dit gesprek willen we een goede beschrijving van het onderwijsleerproces op het gebied van mondelinge taalvaardigheid bij u op school geven. Het gaat om een beschrijving van het onderwijsleerproces en niet om een beoordeling van de kwaliteit. Wat doet u aan MTV en hoe?
Het is van belang dat we allereerst een goed beeld krijgen wat er schoolbreed gebeurt, dus niet zozeer in de specifieke groep van schoolverlaters tenzij het expliciet gevraagd wordt.
Verschillende aspecten van het onderwijsleerproces zullen we bespreken, maar om te beginnen wil ik u een algemene vraag stellen:

Wilt u schetsen hoe bij u op school – schoolbreed - aan de ontwikkeling van mondelinge taalvaardigheden bij leerlingen wordt gewerkt?

Met de openingsvraag krijgt de interviewer een globaal beeld van wat er op school gebeurt. Dit geeft aanknopingspunten om de checklist deels al in te vullen. Maak aantekeningen.

Vervolgens vraag je door om de checklist volledig in te vullen.

Het is belangrijk dat de leerkracht uit groep 8 weet dat het niet om zijn/haar persoonlijk handelen in de klas gaat. Het gaat er om een totaalbeeld te schetsen.

DOELGERICHTHEID

PARAFRASEREN: Ik hoor u zeggen.../Vat ik het goed samen als... > brug slaan naar volgende onderdeel.

INTRO: Leerlingen zijn eigenlijk de hele dag bezig met mondelinge taal. Ze luisteren naar uitleg en instructie, ze beantwoorden vragen, overleggen met elkaar en er zijn kringgesprekken, enzovoorts. Taal is de hele dag een middel om dingen te leren. Dat is belangrijk en nuttig, want door taal te gebruiken verbeteren leerlingen hun mondelinge taalvaardigheid. Ik zou nu echter graag met u dieper ingaan op de doelen die u met de leerlingen op het gebied van luisteren, spreken en gesprekken wilt bereiken en de wijze waarop.

Wilt u schetsen in welke mate *doelgericht* aan mondelinge taalvaardigheden wordt gewerkt?
Werken leerkrachten aan doelen mondelinge taalvaardigheid en zo ja, welke?
Doorvragen (voorbeeldvragen):

- U zegt dat de leerlijn uit de methode wordt gevolgd. Geldt dat voor alle domeinen (luisteren, spreken en gesprekken)?
- Wordt er met leerlijnen/doelen gewerkt? Zo ja, welke of waar zijn ze op gebaseerd?
- Geldt dit voor de subdomeinen van mondelinge taalvaardigheid (luisteren, spreken en gesprekken)?
- Op welke wijze worden leerlingen gestimuleerd deze doelen ook daadwerkelijk te behalen (is er sprake van herhaald aanbod met feedback?)
- Heeft de school doelen geformuleerd voor de mondelinge taalvaardigheden die leerlingen aan het einde van de schoolperiode moeten beheersen?
- Zo ja, is het team goed bekend met die doelen en in hoeverre vormen de doelen de basis voor het onderwijs in mondelinge taalvaardigheden in de groepen? Kunt u voorbeelden noemen?
- In hoeverre zijn de activiteiten aan talige doelen gekoppeld. Kunt u hier concrete voorbeelden van noemen?

Laat de gesprekspartners concrete voorbeelden noemen of vastgelegde doelen zien. Vraag om een uitdraai van de doelen als deze zijn vastgelegd.

Het is steeds van belang een goed onderscheid te maken tussen het werken met *taal als middel* (luisteren naar instructies, vragen beantwoorden, samenwerking, kringgesprek) of als *doel* (talig doel dat nagestreefd wordt).

- Benadruk tijdens het volledige gesprek als interviewer dat het gaat om talige doelen voor de domeinen van mondelinge taalvaardigheid. Vraag naar concrete voorbeelden om er zeker van te zijn dat dit ook zo begrepen wordt.

Zowel invullen voor luisteren als voor spreken/gesprekken

		Toelichting voor de interviewer
De school/het team heeft doelen mondelinge taalvaardigheden	☐ ja	Het gaat om doelen op schoolniveau dus niet de doelen die binnen de les (bijv. aan hand van methode) worden gesteld
	☐ nee	
Welke leerlijnen/doelen	☐ geen leerlijn	Er is geen leerlijn vastgelegd
	☐ leerlijn methode	De methode wordt genoemd: vraag door naar de wijze waarop
	☐ leerlijn CED	De leerlijn CED wordt genoemd: vraag door naar wijze waarop
	☐ eigen leerlijn	Vraag naar deze leerlijn: laat zien en vraag om een kopie
	☐ tussendoelen SLO	Deze tussendoelen kunnen op zich staan of op basis hiervan is een eigen leerlijn binnen school opgesteld
	☐ anders,...	Noteer wat er genoemd wordt
Aan elk doel voor mondelinge taalvaardigheid zijn taalactiviteiten gekoppeld (doelgericht)	☐ ja	Er zijn bij de doelen uit de gebruikte leerlijnen activiteiten gekoppeld; geredeneerd vanuit de doelen.
	☐ enigszins/deels	Dit geldt voor een aantal taaldoelen, maar niet voor alle; leerkrachten missen een aantal taaldoelen (mogelijkheden om eraan te werken, etc.)
	☐ nee	
Aan elke taalactiviteit zijn doelen voor mondelinge taalvaardigheid gekoppeld (activiteitgericht)	☐ ja	Er zijn activiteiten geselecteerd waaraan doelen gekoppeld zijn; geredeneerd vanuit de activiteiten i.t.t. doelen.
	☐ enigszins/deels	Dit geldt voor een aantal taalactiviteiten, maar niet voor alle.
	☐ nee	
Herhaald aanbod met feedback	☐ ja	De leerling oefent herhaald met doelgerichte activiteiten en krijgt feedback om ontwikkeling te stimuleren
	☐ deels	Er is herhaling van een aantal doelgerichte activiteiten maar sommige activiteiten komen maar incidenteel of eens per jaar aan bod (bijv. spreekbeurt, etc).
	☐ nee	
Gebruik referentieniveaus	☐ ja	Er wordt gewerkt met de referentieniveaus: er wordt aantoonbaar gewerkt aan het bereiken van de referentieniveaus mondelinge taalvaardigheden
	☐ redelijk/deels	Er wordt deels gewerkt met de referentieniveaus
	☐ nee	Referentieniveaus op het gebied van mondelinge taalvaardigheden worden niet gehanteerd.
Doorgaande lijn mondelinge taalvaardigheid	☐ ja	Er wordt aantoonbaar gewerkt aan mondelinge taalvaardigheden in een doorgaande lijn: doelen voor elk jaar gespecificeerd, leerkrachten zijn op de hoogte van de doelen per leerjaar of van doelen voor/na groep waarin hij/zij lesgeeft

	☐ redelijk/deels	Er zijn globaal doelen voor onderdelen vastgelegd; de leerkrachten zijn globaal op de hoogte van de doelen uit voorgaande en volgende jaren; er is duidelijk wat leerlingen uiteindelijk moeten beheersen
	☐ nee	Er is geen sprake van een doorgaande leerlijn in mondelinge taalvaardigheden
Einddoelen voor schoolverlaters	☐ ja	Leerkrachten weten wat leerlingen aan het einde van de basisschooltijd in s(b)o moeten kunnen en kennen op het gebied van mondelinge taalvaardigheden
	☐ deels	Er is een globale beschrijving maar weinig specifiek/concreet
	☐ nee	

Onderwijsaanbod: materiaal, activiteiten en onderwijstijd

PARAFRASEREN: Ik hoor u zeggen.../Vat ik het goed samen als... > brug slaan naar volgende onderdeel.

INTRO: Zojuist ging het over de doelen voor de subdomeinen van mondelinge taalvaardigheid. Ik zou nu graag dieper met u willen ingaan op het onderwijsaanbod waarmee u deze doelen wilt bereiken.

Hoe ziet het lesaanbod voor de leerdoelen op het gebied van luisteren, spreken en gesprekken er op jullie school uit?

Doorvragen (voorbeeldvragen):
- Op welke manier werkt u met de methode voor mondelinge taalvaardigheid, dat wil zeggen: voor luisteren, spreken en gesprekken?
- U zegt dat u keuzes uit de methode maakt. Geldt dat voor luisteren, spreken en gesprekken?
- Is deze werkwijze schoolbreed ingevoerd? Dus: werken alle groepen op deze wijze/met deze methode?
- Komen doelgerichte activiteiten herhaald terug?

Het is steeds van belang een goed onderscheid te maken tussen het werken met *taal als middel* (luisteren naar instructies, vragen beantwoorden, samenwerking, kringgesprek) of als *doel* (talig doel dat nagestreefd wordt).

- Benadruk tijdens het volledige gesprek als interviewer dat het gaat om talige doelen voor de domeinen van mondelinge taalvaardigheid. Vraag naar concrete voorbeelden om er zeker van te zijn dat dit ook zo begrepen wordt.

		Toelichting voor de interviewer
Methodegebruik	☐ geen methode	Noteer bijzonderheden/opmerkingen.
	☐ taalactief	
	☐ taal op maat	
	☐ staal	
	☐ taaltrapeze	
	☐ taalverhaal	
	☐ anders,...	Noteer welke methode genoemd wordt.
Taken en activiteiten		Loop met de gesprekspartners de lijst taak en type activiteiten door. Geef ruimte om ook andere activiteiten te noemen.
Wijze van gebruik methode mondelinge taalvaardigheid (luisteren, spreken en gesprekken)	☐ methode wordt gevolgd	De methode is leidend en wordt in zijn geheel gevolgd en uitgemaakt.
	☐ methode wordt deels gevolgd	De methode wordt grotendeels/in principe gevolgd, maar er

		worden keuzes gemaakt: niet alles wordt gebruikt.
	☐ enkele onderdelen gekozen	Er worden soms onderdelen/activiteiten uit de methode gebruikt.
Afspraken over gebruik methode mondelinge taalvaardigheid (luisteren, spreken en gesprekken)	☐ ja	Er zijn schoolbreed afspraken over het aanbod (inhoud en aantal lessen)
	☐ deels	De methode is in principe leidend, maar leerkrachten zijn vrij om zelf te bepalen welke lesstof gegeven wordt
	☐ nee	Leerkrachten zijn vrij om zelf het onderwijsaanbod te bepalen
Aanvullende materialen	☐ geen aanvullend materiaal	
	☐ remediërend materiaal, nl.:	
	☐ activiteiten uit andere methode dan de hoofdmethode	
	☐ anders, nl.: ...	
Inbedding van aanbod (concrete voorbeelden kunnen genoemd worden)	☐ in de taalmethode/tijdens de taallessen	Er wordt doelgericht aan mondelinge taalvaardigheden gewerkt tijdens de taallessen op het rooster
	☐ geïntegreerd in andere (zaak)vakken	Er wordt doelgericht aan mondelinge taalvaardigheden gewerkt bij alle (zaak)vakken; er worden activiteiten uitgevoerd . Vraag naar concrete voorbeelden.
	☐ niet	Er wordt de hele dag aan mondelinge taalvaardigheid gewerkt maar niet doelgericht (geen talige doelen bij activiteiten, maar taal is onderdeel van activiteiten als middel – naar elkaar luisteren bij activiteiten en instructie, praten tijdens kringgesprek etc.)
Onderwijstijd mondelinge taalvaardigheden	☐ talige doelen centraal:...	Noteer op basis van wat er in het rooster staat en in overleg hoeveel **minuten per week** er doelgericht aan luisteren, spreken en gesprekken wordt besteed. Het gaat om de tijd die expliciet aan mondelinge taalvaardigheden wordt besteed. Het gaat niet om de tijd waarin taal als middel wordt gebruikt.
Onderwijstijd mondelinge taalvaardigheden schoolverlaters	☐ talige doelen centraal:...	

Beoordelen en volgen

PARAFRASEREN: Ik hoor u zeggen.../Vat ik het goed samen als... > brug slaan naar volgende onderdeel.

INTRO: Zojuist hebben we het gehad over het onderwijsaanbod. Nu zou ik graag met u ingaan op de wijze waarop u vaststelt of de doelen ook behaald zijn en hoe de ontwikkeling verloopt.

Wordt de ontwikkeling in mondelinge taalvaardigheden gevolgd, in kaart gebracht, geregistreerd en gevolgd?
Doorvragen (voorbeeldvragen):
- Van welke instrumenten wordt gebruikgemaakt?
- Hoe wordt gecontroleerd of de doelen voor mondelinge taalvaardigheden behaald zijn?
- Is er vastgelegd op welke manier en op welke momenten de ontwikkeling in luisteren, spreken en gesprekken in kaart wordt gebracht?
- U zegt, dat u de ontwikkeling in kaart brengt. Heeft u daarmee een goed beeld van de mondelingen vaardigheden van de leerlingen, en van de verschillen tussen leerlingen, van de vaardigheden die ze nog niet of nog niet goed genoeg beheersen?

		Toelichting voor de interviewer
Leervorderingen mondelinge taalvaardigheden worden *in de lessen* geregistreerd/bijgehouden voor alle leerlingen	☐ altijd	Er wordt na een les/activiteit altijd beoordeeld in hoeverre de leerling het doel heeft behaald
	☐ deels/soms (enkele activiteiten)	Er wordt na een les/activiteit af en toe beoordeeld in hoeverre de leerling het doel heeft behaald
	☐ nee	
Leervorderingen mondelinge taalvaardigheden worden in het schooljaar (middellange termijn) systematisch geregistreerd/bijgehouden voor alle leerlingen	☐ ja structureel	Er wordt op vaste, afgesproken momenten in het schooljaar getoetst
	☐ ja incidenteel	Er wordt incidenteel, terloops (niet op vaste momenten) getoetst.
	☐ nee	Geen toets
Leervorderingen mondelinge taalvaardigheden worden voor zorgleerlingen in kaart gebracht (evt. door zorgspecialist/logopedist, etc.)	☐ ja	Het gaat hier om toetsen anders dan wat er schoolbreed/in de klas gebeurt. Aspecten van MTV worden alleen getoetst bij leerlingen die achterstand op mondelinge taalvaardigheid laten zien.
	☐ nee	
Hoe/waarmee de vorderingen worden bijgehouden	☐ genormeerde toetsen	
	☐ volginstrument(en) methode	De toetsen of observatiepunten die in de methode worden aangereikt om vordering te bepalen.
	☐ zelfgemaakte instrumenten	Vraag naar concrete voorbeelden

	☐ observatielijsten/punten	Overige observatielijsten/punten die niet uit de methode komen (bijvoorbeeld combilist, boek 'Opbrengstgericht werken mondelinge taalvaardigheid', etc.)
Gebruik van groepsplannen voor mondelinge taalvaardigheden voor alle leerlingen	☐ ja	Vraag om een voorbeeld; vraag om een kopie
	☐ nee	
Systematisch gebruik van resultaten voor het afstemmen van het aanbod	☐ ja	De resultaten worden gebruikt om het onderwijsaanbod/groepsplan aan te passen
	☐deels	De resultaten worden voor sommige leerlingen gebruikt om onderwijsaanbod/ondersteuning aan te passen
	☐ nee	
Gebruik van leerlingadministratiepakket	☐ ja	Noteer welk systeem/pakket wordt gebruikt
	☐ deels	Voor onderdelen binnen de vaardigheid
	☐ nee	
Zicht op vorderingen en verschillen tussen leerlingen bij gebruik van toetsen	☐ ja	Laat concrete voorbeelden noemen (hoe, waar blijkt dat uit)
	☐ deels	Niet voor alle domeinen
	☐ nee	

Differentiatie

PARAFRASEREN: Ik hoor u zeggen.../Vat ik het goed samen als... > brug slaan naar volgende onderdeel.

INTRO: Nu wil ik graag met u ingaan op de manier waarop er in de klassen gedifferentieerd wordt bij mondelinge taalvaardigheid. Het gaat er om een goed beeld te schetsen van de manier waarop er rekening gehouden wordt met verschillen tussen leerlingen. Het gaat niet zozeer om differentiatie op basis van algemene onderwijsbehoeften, maar om het afstemmen op het gebied van mondelinge taalvaardigheden om de leerdoelen binnen dit domein te bereiken.

Wilt u schetsen hoe er rekening wordt gehouden met verschillen tussen leerlingen in leerlingbehoeften op het gebied van mondelinge taalvaardigheid?

Doorvragen:
- Werken leerkrachten met niveaugroepen?
- Werken leerkrachten schoolbreed met groepsplannen voor luisteren, spreken/gesprekken ten behoeve van differentiatie?
- Is de instructie en leerstof afgestemd op de verschillende niveaus?
- Zijn er aangepaste (streef)doelen voor specifieke leerlinggroepen?

<u>Aandachtspunt voor de interviewer</u>
Vraag goed door om na te gaan dat er daadwerkelijk gedifferentieerd wordt op basis van verschillen in mondelinge taalvaardigheid om de ontwikkeling te stimuleren. Het gaat dus niet om algemene aanpassingen als 'eenvoudiger spreken', 'de instructie nog eens in andere woorden herhalen', ' bieden van visuele ondersteuning bij het geven van instructie'.

		Toelichting voor de interviewer
Niveaugroepen bij mondelinge taal	☐ nee (1 groep)	Er zijn geen verschillende niveaugroepen
	☐ ja: 2 of 3 groepen	Er zijn in het algemeen per klas 2 of 3 niveaugroepen
	☐ ja: 4 of meer groepen	Er zijn 4 of meer groepen
Differentiatie naar instructie in mondelinge taalvaardigheid; *verlengde instructie & extra oefening op onderdelen aan leerlingen met achterstand*	☐ ja	De instructie wordt aangepast op het niveau van de leerlingen; er is o.a. onderscheid tussen basisinstructie en verlengde instructie: leerlingen met zwakke mondelinge taalvaardigheden krijgen verlengde instructie/oefening
	☐ nee	De instructie is voor alle leerlingen hetzelfde
Differentiatie naar leerstof in doelen mondelinge	☐ ja	Leerlingen werken op niveau waarbij er keuzes in de leerstof

taalvaardigheid		en oefeningen worden gemaakt
	☐ nee	De leerstof en oefeningen zijn voor alle leerlingen hetzelfde
Differentiatie naar (streef)doel mondelinge taalvaardigheid	☐ ja	De (streef)doelen verschillen voor (groepen) leerlingen en dit is (beredeneerd/onderbouw) vastgelegd: school moet inzichtelijke maken hoe/op basis van welke keuzes de streefdoelen voor mondelinge taalvaardigheid voor (specifieke) groepen leerlingen zijn aangepast.
	☐ nee	De (streef)doelen zijn voor alle leerlingen gelijk
Convergente/divergente differentiatie	☐ convergente differentiatie	Er wordt uitgegaan van aanbod voor de hele klas met verlengde instructie
	☐ divergente differentiatie	Leerlingen werken op hun eigen niveau, in homogene groepen
	☐ geen differentiatie	
Extra ondersteuning buiten de klas	☐ ja, nl....	Noteer welke extra ondersteuning er geboden wordt op het gebied van mondelinge taalvaardigheden
	☐ nee, geen extra ondersteuning	
Indien extra ondersteuning: door wie	☐ specialist in school	Specialist binnen school die ondersteuning biedt/interventie uitvoert op gebied van mondelinge taalvaardigheden
	☐ specialist van buiten school	Specialist buiten school die ondersteuning biedt/interventie uitvoert op gebied van mondelinge taalvaardigheden
Indien extra ondersteuning: specifiek programma	☐ ja, nl.:...	Noteer met welke programma('s) wordt gewerkt.
	☐ nee	
Indien extra ondersteuning plaatsvindt: aantal schoolverlaters met extra ondersteuning		Noteer het aantal leerlingen in groep 8 (schoolverlaters) dat extra ondersteuning heeft ontvangen op het gebied van mondelinge taalvaardigheden

Teamontwikkeling
PARAFRASEREN: Ik hoor u zeggen.../Vat ik het goed samen als... > brug slaan naar volgende onderdeel. INTRO: Graag zou ik ook willen weten in welke mate en de manier waarop er aandacht wordt besteed binnen het team aan mondelinge taalvaardigheid. **Kunt u schetsen op of en hoe het onderwijs in mondelinge taalvaardigheden op de agenda van het team staat? Wat komt er zoal aan de orde? Wat bespreken jullie?**

Doorvragen (voorbeeldvragen):
- Staan er afspraken/uitgangspunten voor mondelinge taalvaardigheden beschreven in het schoolplan/taalbeleidsplan?
- Is er ruimte om bij elkaar te gaan kijken?
- Is er extra ondersteuning voor leerkrachten, intern dan wel extern?
- Neem de teamoverleggen van het afgelopen schooljaar, van september tot nu, in gedachten: is mondelinge taalvaardigheid besproken? Wat is er inhoudelijk aan de orde geweest?

		Toelichting voor de interviewer
Schoolplan/taalbeleidsplan	☐ beschreven in schoolplan	Er wordt aantoonbaar aandacht geschonken aan het onderwijs in mondelinge taalvaardigheden in het schoolplan
	☐ taalbeleidsplan	Er is een taalbeleidsplan waarin aantoonbaar aandacht wordt geschonken aan het onderwijs in mondelinge taalvaardigheden in het schoolplan
	☐ nee	
Intervisie en/of collegiale consultatie op het gebied van onderwijs in mondelinge taalvaardigheid en/of leerlingen met problemen met mondelinge taalvaardigheden	☐ ja, structureel	Het vindt structureel, gepland, volgens afspraak plaats
	☐ ja, incidenteel	Het vindt plaats op verzoek, niet structureel
	☐ nee	
	☐ anders, nl.:	Er wordt een andere vorm van coaching genoemd
Ondersteuning intern (bijv. taalcoördinator) op het gebied van onderwijs in mondelinge taalvaardigheid en/of leerlingen met problemen met mondelinge taalvaardigheden	☐ ja	Er is iemand intern aangewezen om leerkrachten te ondersteunen bij het vormgeven van onderwijs in mondelinge taalvaardigheden.
	☐ nee	
	☐ anders, nl.:	Er is niet een persoon, maar er is bijvoorbeeld een taalwerkgroep van leerkrachten, etc. Noteer.
Ondersteuning extern (bijv. onderwijsadviseur) op het gebied van onderwijs in mondelinge taalvaardigheid en/of leerlingen met problemen met mondelinge taalvaardigheden	☐ ja	Er is iemand extern aangetrokken/ingehuurd om leerkrachten te ondersteunen bij het vormgeven van onderwijs in mondelinge taalvaardigheden.
	☐ nee	
	☐ anders, nl.:	
Inhoudelijke onderwerpen die aan bod zijn gekomen binnen het teamoverleg	☐	Loop met de gesprekspartners de lijst inhoudelijke onderwerpen (teamoverleg) door. Geef ruimte om ook andere onderwerpen te noemen.

Behoefte aan (extra) verantwoordelijkheid en keuzes

De aanpassingen zoals beschreven bij passende perspectieven vormen de basis voor onderstaande schema's. Uit de daar genoemde adviezen zijn adviezen geselecteerd die van toepassing zijn op de taken uit de peiling. Ook zijn enkele punten, specifiek gericht op de peiling toegevoegd. (√ = standaard gedaan; + = kan de toetsleider/spelleiderextra doen)

	Samenwerkingsspel		Spreektaken		Luistertaken	
Instructie	√	Instructie afgewisseld met doe-opdrachtjes	√	Heldere korte instructies	√	Heldere korte instructies
	√	Leerling kan keuzes maken zoals kleur van de boot,				
	+	Laat de leerling meehelpen met het klaarleggen van het spel, bijvoorbeeld door de winden allemaal naast het spel neer te leggen.				
Vorm en tijd	√	Voldoende afwisseling in opdrachten	√	Voldoende afwisseling in opdrachten	√	Voldoende afwisseling in opdrachten
			√	Leerling kan in eigen tempo werken		
	+	Zorg indien nodig voor een extra pauze of doe wat kleine beweegoefeningen tussen de opdrachten door.	+	Zorg indien nodig voor een extra pauze of doe wat kleine beweegoefeningen tussen de opdrachten door.	+	Zorg indien nodig voor een extra pauze of doe wat kleine beweegoefeningen tussen de opdrachten door.
	+	Geef de leerling tijdens het spel een extra taak, zoals dat hij bijhoudt hoeveel winden er al zijn omgedraaid of zorgt dat de vissen netjes op het eiland liggen.				
Inhoud	√	Leerling mag elke keer als hij aan de beurt is een moment om een keuze te maken (wat moet met de worp gebeuren)			√	Deel van de fragmenten bevat een filmpje
	+	Vraag de leerling na afloop hoe hij het spel vond				

Behoefte aan (extra) zelfvertrouwen

De aanpassingen zoals beschreven bij passende perspectieven vormen de basis voor onderstaande schema's. Uit de daar genoemde adviezen zijn adviezen geselecteerd die van toepassing zijn op de taken uit de peiling. Ook zijn enkele punten, specifiek gericht op de peiling toegevoegd. (√ = standaard gedaan; + = kan de toetsleider/spelleiderextra doen)

	Samenwerkingsspel		Spreektaken		Luistertaken	
Instructie	√	Geen toetssetting, dus geen toetsdruk	√	Niet spreken over een 'toets'	√	Niet spreken over een 'toets'
	√	Voldoende tijd om de regels te oefenen en vertrouwd te raken met het spel				
	+	Geef complimenten over hoe de leerling bezig is met het spel, zoals dat wat hij met de worp heeft verzonnen een goed idee is	+	Geef complimenten bij de instructie, zoals "dat kan je inderdaad zeggen, goed gedaan"	+	Geef complimenten bij de instructie, zoals "zo moet je inderdaad je antwoord geven, dat doe je goed"
			+	Vraag voor dat de echte taken beginnen of alles duidelijk is	+	Vraag voor dat de echte opgaven beginnen of alles duidelijk is
Vorm en tijd	√	Korte, overzichtelijke taken	√	Korte, overzichtelijke taken	√	Korte, overzichtelijke taken
					√	Bij het stellen van de vragen luistertekst opgedeeld in delen.
	+	Geef de leerling regelmatig een compliment over hoe hij bezig is	+	Geef na elke taak een compliment over hoe de leerling bezig is.	+	Geef na elke luistertekst een compliment over hoe de leerling aan het werk is.
Inhoud	+	Spelleider herhaalt goede argumenten van de leerling als bevestiging				
Omgeving	√	Vertrouwd persoon die het spel begeleidt	√	Vertrouwd persoon die de taken afneemt	√	Vertrouwd persoon die de taken afneemt
	√	Spelleider zoekt een rustige omgeving	√	Toetsleider zoekt een rustige omgeving	√	Toetsleider zoekt een rustige omgeving
	+	Zorg (extra) voor een groepje waarin alle leerlingen zich prettig en vertrouwd voelen				
Hulp-middelen						

Dyslexie

De aanpassingen zoals beschreven bij passende perspectieven vormen de basis voor onderstaande schema's. Uit de daar genoemde adviezen zijn adviezen geselecteerd die van toepassing zijn op de taken uit de peiling. Ook zijn enkele punten, specifiek gericht op de peiling toegevoegd. (√ = standaard gedaan; + = kan de toetsleider/spelleiderextra doen)

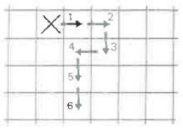
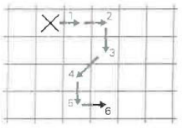
	Samenwerkingsspel		Spreektaken		Luistertaken	
Instructie	√	Alle instructies in gesproken tekst.	√	Alle instructies in gesproken tekst.	√	Alle instructies in gesproken tekst.
Vorm en tijd					+	Zorg indien nodig voor een extra pauze
					+	Lees de antwoordalternatieven voor.
Hulp-middelen	√	Visuele ondersteuning bij stappenplan				

Bijlage 6 Instructie Gesprekstaak

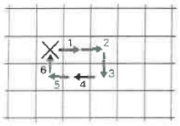
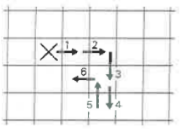
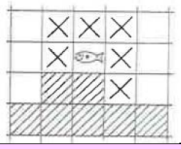
Spelinstructie

INSTRUCTIE OM VOOR TE LEZEN	INSTRUCTIE VOOR DE SPELLEIDER
Intro	
We gaan zo een spel spelen. Kijk, dit is het spel [laat de doos zien]. Bij dit spel hoort een uitleg. Die duurt best even, maar het is wel heel belangrijk. Alleen dan kun je het spel goed spelen. Af en toe stel ik je een vraag over de uitleg, dus let goed op. Ná de uitleg mag je vragen stellen als er iets niet duidelijk is. [Pak de doos uit en leg het bord op tafel met de gekleurde havens aan jouw kant]. Op dit bord staat een eiland en een zee. Jullie zijn nu allemaal vissers en gaan samen tegen het spel spelen. Dus jullie winnen allemaal, of verliezen allemaal. Samen gaan jullie zoveel mogelijk vissen vangen [houd het zakje met de visjes omhoog] en naar de haven brengen. Je krijgt zo een eigen boot [houd het zakje met de boten omhoog] en een eigen haven. Maar pas op: er komt storm! Jullie hebben het spel samen gewonnen, wanneer alle boten en vissen veilig in de havens gebracht zijn vóórdat de storm komt. Elkaar helpen en samenwerken is dus heel belangrijk bij dit spel.	
	Opdracht mondeling samenvatten (leerling 1) Vraag aan één leerling ('leerling 1'): → Wat gaat we dus doen in het spel?
Voorbereiding	
→ We gaan spelen met de blauwe, rode, gele en groene boot. Kijk welke haven voor jou het dichtst bij is. Pak het bootje in die kleur. [Houd voor jezelf een grijs of zwart bootje bij de hand om een bepaalde handeling voor te doen]	
→ Leg je boot bij de haven met dezelfde kleur als je boot.	
Wie als eerste jarig is mag beginnen. → Overleg met elkaar wie als eerste jarig is.	N.B. Bij deze opdracht gaat het niet om de juistheid van de uitkomst, maar om het voeren van het gesprek hierover.
Daarna mogen jullie om de beurt dobbelen, met de klok mee.	Wijs de richting van de klok aan, beginnend met de leerling die mag beginnen.
De speciale dobbelsteen	
Met de dobbelsteen kun je 1, 2, 3, 4, ster en wind gooien	en laten zien
1-4: je mag met je boot 1-4 vakjes verder varen (precies zoveel vakjes als op de dobbelsteen	

24

staat) Dus als je 3 gooit, mag je 3 vakjes verder varen.	
ster: je mag zelf kiezen of je 1-6 vakjes verder vaart	Laat de ster nog een keer zien
wind: je draait een windplaatje om aan de rand van het bord	Laat de wind nog een keer zien. Laat ook zien waar de windplaatjes liggen en hoe je die omdraait.
→ We gaan dit even uitproberen. Gooi nu om beurten met de dobbelsteen. Vertel wat je hebt gegooid en wat je mag doen als je dat gooit.	Laat ieder ten minste één keer gooien en vertellen. Stop er daarna mee als duidelijk is dat ze het begrepen hebben.
Varen	
De boten kunnen op alle blauwe vakjes, dus donkerblauwe én lichtblauwe vakjes, varen, maar niet op het eiland.	
Op 1 vakje mag niet meer dan 1 boot tegelijk zijn; een boot en een vis mogen wel samen op 1 vakje zijn.	
Boten mogen niet over elkaar heen, maar moeten om elkaar heen varen.	
→ We gaan weer even wat uitproberen. Ik laat wat dingen zien. En stel jullie dan om beurten een vraag. Alleen degene aan wie ik de vraag stel geeft antwoord. De anderen zijn dan stil.	1. Laat zien: 2 boten op één vakje. Vraag: Mag dit? 2. Laat zien: 1 boot en 1 vis op één vakje. Vraag: Mag dit? 3. Laat zien: 1 boot mag 3 vakjes varen en vaart over een ander boot heen. Vraag: Mag dit? 4. Laat zien: 1 boot mag 4 vakjes varen en vaart over een hoekje van het eiland. Vraag: Mag dit?
Je mag recht vooruit, recht achteruit en recht opzij varen. Je mag niet schuin varen. Kijk, zo mag je varen. En zo mag je varen. En zo mag je NIET varen. Je mag dus WEL recht vooruit, recht achteruit en recht opzij varen.	Laat zien hoe je wel en niet mag varen. Wel recht vooruit. Wel recht opzij. Niet schuin (diagonaal). Mogelijk:  Niet mogelijk, omdat 4 diagonaal is: 
Je mag niet in één beurt heen en weer varen. [Doe dit even voor].	Je mag niet in één beurt twee keer dezelfde route varen, ook niet door vooruit en achteruit varen (zie voorbeeld).

25

	Mogelijk:  Niet mogelijk, omdat 4 en 5 dezelfde route is. 
Een vis vangen	
Je mag een vis vangen als je boot *naast* een vakje met een vis is. Dus niet *op* het vakje met de vis. Het mag ook schuin naast het vakje van de vis zijn.	Voorbeeld: de vakjes met een X zijn naast het vakje met de vis; dit zijn de toegestane opties. 
→ Mag je hier de vis vangen?	Leg een boot in de buurt van een vis met één stip. Vraag: mag je hier de vis vangen? Herhaal dit voor elke speler één keer. Stel toegestane en niet-toegestane opties voor.
Als je boot naast het vakje met een vis is, mag je de vis omdraaien.	Laat zien hoe je een vis (met één stip) omdraait. En laat zien hoe je die in je boot laadt.
Als het een kleine vis is (met één stip), dan mag je de vis meteen in je boot laden.	
Om een grote vis te vangen zijn twee boten nodig. Daarvoor moet je dus samenwerken.	Laat de vissen zien en vertel per vis hoeveel boten er nodig zijn om ze te vangen. - vis met 1 stip 1 boot - vis met 2 stippen 2 boten
	Opdracht mondeling samenvatten (leerling 2) Vraag aan leerling 2: → Hoe mag je dus de vissen vangen?
Als je een vis niet alleen kunt vangen, kun je twee dingen doen: 1. Naast de vis wachten tot de boot van een andere speler ook naast de vis ligt. Spreek met elkaar af wie helpt om welke vis te vangen. Als er genoeg boten zijn om de vis te vangen, spreek je samen af in welke boot de vis gaat. 2. Als je niet wilt wachten kun je in de volgende beurt weer verder varen. Je moet de vis dan weer omdraaien en terug leggen waar hij lag.	

26

→ Mogen de boten zo deze vis vangen?	Leg twee boten in de buurt van een vis met twee stippen. Vraag: Mogen de boten zo deze vis vangen? Herhaal dit voor elke speler één keer, met verschillende posities. Stel toegestane en niet-toegestane opties voor, bijvoorbeeld: • Eén boot staat op een hokje links naast een vis; de tweede boot staat niet op een hokje direct rechts naast de vis, maar er nog één hokje van af (niet toegestaan). • Eén boot staat op een hokje schuin boven de vis; de tweede boot staat op een hokje direct onder de vis (wel toegestaan)
Elke boot mag 3 vissen vervoeren. Het maakt niet uit hoe groot de vissen zijn.	
Vissen aan land brengen	
Een boot kan de vissen die hij vervoert alleen in zijn eigen haven aan land brengen. Om in de haven te komen moet je precies het juist aantal gooien. Je mag niet minder vakjes varen dan je hebt gegooid met de dobbelsteen. Je mag wel eerst nog bochten varen.	
Als alle vissen aan land gebracht zijn en alle boten in de havens zijn, hebben jullie samen gewonnen. Als alle 10 windplaatjes zijn omgedraaid, stormt het zo hard dat alle boten zinken. Als er dan nog een boot op zee is of er zijn nog vissen in het water, hebben jullie samen het spel verloren.	
	Opdracht mondeling samenvatten (leerling 3) Vraag aan leerling 3: → Welke regels zijn er voor het brengen van vissen naar de haven?
Als je gegooid hebt	
Nu vertel ik hoe je dit spel echt samen speelt. *[Noem de naam van degene die mag beginnen]* Gooi maar. Als je hebt gegooid met de dobbelsteen, mag je kiezen of je zelf gaat varen of dat je de worp aan iemand anders geeft. Dan mag die ander varen met zijn eigen boot.	Als een leerling de worp heeft weggegeven, gaat het spel daarna in de oude volgorde verder.
Vertel als je gegooid hebt of je van plan bent de worp zelf te houden of weg te geven.	

27

Bijlage 7 Geheugensteuntje spreektaak (jeugdjournaal)

Babyotter

Alaska

Strand

Niet gegeten

Koud

Vacht

Dierenopvang

Terug naar het wild