Technisch rapport Peil.Mondelinge taalvaardigheid 2023-2024

Expertisecentrum Nederlands
KBA Nijmegen
Stichting Cito, divisie CitoLab

Technisch rapport Peil.Mondelinge taalvaardigheid 2023-2024

Martine Gijsel, Nicole Swart, Emmi Stuij, Marian Bruggink (Expertisecentrum Nederlands)
Tessa Jenniskens, Annemarie van Langen (KBA Nijmegen)
Friso Koman, Emmelien van der Scheer (Stichting Cito, divisie CitoLab)

Met medewerking van: Cito BV, Victor Schrm en Ake Eimers (KBA Data) en Femke Scheltinga (iTTA).

Februari 2025

Inhoud

1 Inleiding 5
1.1 Onderzoeksvragen 5
1.2 Leeswijzer 7

2 Instrumentarium 8
2.1 Theoretische uitgangspunten 8
2.2 Luisteren 8
2.3 Spreken 10
2.4 Gesprekken 12
2.5 Overige instrumenten 14
2.6 Het afnamedesign 17

3 Steekproef, werving, respons, dataverzameling 21
3.1 Inleiding 21
3.2 Werving 22
3.3 Respons 23

4 Schaling luisteren, spreken en gesprekken voeren 30
4.1 Aantallen en ontbrekende waarden 30
4.1.1 Luistervaardigheid 30
4.1.2 Spreekvaardigheid 33
4.1.3 Gespreksvaardigheid 35
4.2 Klassieke analyses 37
4.2.1 Luistervaardigheid 37
4.2.2 Spreekvaardigheid 39
4.2.3 Gespreksvaardigheid 41
4.3 Beoordelaarsonderzoek 43
4.3.1 Spreekvaardigheid 45
4.3.2 Gespreksvaardigheid 47
4.4 Samenhang subdomeinen 49
4.4.1 Luistervaardigheid 50
4.4.2 Spreekvaardigheid 51
4.4.3 Gespreksvaardigheid 52
4.4.4 Samenhang hoofddomeinen 53
4.5 Schaling en DIF 53
4.5.1 Luistervaardigheid 53
4.5.2 Spreekvaardigheid 60
4.5.3 Gespreksvaardigheid 66

5 Schaling vragenlijsten 72
5.1 Inleiding 72
5.2 Leerlingvragenlijst 73
5.3 Quickscan 74
5.4 Leerkrachtvragenlijst 76
5.5 Schoolleidersvragenlijst 80

6 Resultaten luisteren, spreken en gesprekken voeren 82
6.1 Prestatiestandaarden 82
6.1.1 Luistervaardigheid 82
6.1.2 Spreekvaardigheid 85
6.1.3 Gespreksvaardigheid 86
6.2 Trendanalyses 88
6.2.1 Luistervaardigheid 88
6.2.2 Spreekvaardigheid 92
6.2.3 Gespreksvaardigheid 96
6.3 Beheersing vaardigheid op itemniveau 99
6.3.1 Luistervaardigheid 99
6.3.2 Spreekvaardigheid 104
6.3.3 Gespreksvaardigheid 106

7 Resultaten vragenlijsten 109
7.1 Inleiding 109
7.2 Leerlingvragenlijst 109
7.3 Quickscan 118
7.4 Leerkrachtvragenlijst 122
7.5 Schoolleidersvragenlijst 136

8 Woordenschattoets 146
8.1 Inhoud en afname 146
8.2 Resultaten 147

9 Samenhang prestaties en kenmerken van leerlingen, leerkrachten en onderwijsleerproces 150
9.1 Inleiding 150
9.2 Geselecteerde variabelen 151
9.3 Imputatie 152
9.4 Opzet multilevel analyses 153
9.5 Resultaten multilevel analyses 155
9.5.1 Regulier basisonderwijs (bao) 155
9.5.2 Speciaal basisonderwijs (sbo) 164
9.5.3 Speciaal onderwijs (so) 171

10 Samenvatting 172

10.1 Leerlingprestaties 172
10.2 Onderwijsleerproces op scholen 175
10.3 Samenhang kenmerken en vaardigheid 177
10.3.1 Basisonderwijs 177
10.3.2 Speciaal basisonderwijs 179

11 Literatuur 180

12 Bijlagen 182
12.1 Bijlage 1. Aantallen en p-waarden per item 182
12.2 Bijlage 2: Resultaten multilevel modellen speciaal onderwijs 184

1 Inleiding

Dit is het technisch rapport bij het peilingsonderzoek Mondelinge taalvaardigheid. Dit peilingsonderzoek heeft in het schooljaar 2023-2024 plaatsgevonden bij leerlingen van groep 8 in het basisonderwijs (bao) en schoolverlaters in het speciaal basisonderwijs (sbo) en speciaal onderwijs (so). Het peilingsonderzoek is uitgevoerd in opdracht van het Nationaal Regieorgaan Onderwijsonderzoek en de Inspectie van het Onderwijs (verder: inspectie) door een consortium bestaande uit Expertisecentrum Nederlands, KBA Nijmegen en CitoLab.

Bij het peilingsonderzoek horen twee technische rapportages. Ten eerste de onderhavige rapportage, geschreven door het uitvoerend consortium. Dit betreft een uitgebreide beschrijving en verantwoording van het instrumentarium, de wijze van dataverzameling en de eerste gegevensverwerking en -analyse ter beantwoording van de onderzoeksvragen. Ten tweede het technisch rapport dat geschreven is door onderzoekers van de inspectie. Dit rapport dient ter verantwoording van de gegevensverwerking die is uitgevoerd ten behoeve van het publieksrapport.

Het doel van het peilingsonderzoek Mondelinge taalvaardigheid is om vast te stellen wat de gespreks-, spreek- en luistervaardigheid van leerlingen aan het eind van het (speciaal) basisonderwijs en speciaal onderwijs is. Daarnaast is vastgesteld hoe het onderwijsleerproces voor mondelinge taalvaardigheid er op de deelnemende scholen uitziet en geeft het een beeld van de kenmerken van leerlingen, leerkrachten en scholen die gerelateerd zijn aan de mondelinge taalvaardigheid van leerlingen. De gespreks-, spreek- en luistervaardigheid van leerlingen is gemeten met verschillende taken. Daarnaast zijn vragenlijsten afgenomen bij de leerlingen, hun leerkrachten en schoolleiders om informatie te verzamelen over achtergrondkenmerken en het onderwijsleerproces. Met de resultaten kunnen op stelselniveau uitspraken worden gedaan over de mondelinge taalvaardigheid van leerlingen. Daarnaast bieden ze aanknopingspunten voor een dialoog over de inhoud en kwaliteit van het mondelinge taalonderwijs in de verschillende typen onderwijs.

1.1 Onderzoeksvragen

De onderzoeksvragen die in deze technische rapportage centraal staan zijn:
Leerlingprestaties
1. Hoeveel procent van de leerlingen einde basisonderwijs en van de schoolverlaters speciaal basisonderwijs en speciaal onderwijs beheerst de referentieniveaus (1F en 2F) voor gesprekken voeren, spreken en luisteren en wat is de vaardigheidsverdeling op de onderliggende schaal?

2. Hoe kan de prestatie van leerlingen op de referentieniveaus voor luisteren, spreken en gesprekken voeren worden getypeerd naar kenmerken van de taakuitvoering?

3. Wat is de onderlinge samenhang in prestaties op de drie subdomeinen?

Beschrijving van het onderwijsleerproces

4. Hoe kan het onderwijs in mondelinge taalvaardigheid op de deelnemende scholen worden getypeerd (in elk geval het luister- en spreekklimaat in de klas, de aanwezigheid van geconcretiseerde leerdoelen en expliciete instructie voor gesprekken voeren, luisteren en spreken, het volgen van de vaardigheid van de leerlingen op deze subdomeinen, de aan deze subdomeinen intentioneel bestede onderwijstijd en de integratie van het onderwijs in deze subdomeinen met zowel andere taaldomeinen als andere leergebieden)?

Samenhang prestaties met leerling-, leerkracht- en schoolkenmerken (incl. onderwijsleerproces)

5. Hoeveel variabiliteit in de leerlingprestaties is toe te schrijven aan scholen (klassen) en hoeveel aan leerlingen?

6. Welke verschillen in leerlingprestaties zijn er naar algemene leerlingkenmerken (in elk geval geslacht, leeftijd, opleidingsniveau ouders, advies vervolgonderwijs, thuistaal, vaardigheid begrijpend lezen en woordenschat)?

7. Welke verschillen in leerlingprestaties zijn er naar domeinspecifieke leerlingkenmerken (in elk geval taalrijkheid van de thuisomgeving, type taalactiviteiten dat de leerling thuis onderneemt en ervaren spreekvrijheid)?

8. Welke verschillen in leerlingprestaties zijn er naar algemene leerkrachtkenmerken (in elk geval geslacht, leeftijd, opleidingsniveau, omvang van de aanstelling en leservaring)?

9. Welke verschillen in leerlingprestaties zijn er naar domeinspecifieke leerkrachtkenmerken (in elk geval vakinhoudelijke en vakdidactische kennis, interactievaardigheden en voorbeeldgedrag op het gebied van mondelinge taalvaardigheid)?

10. Welke verschillen in leerlingprestaties zijn er naar algemene schoolkenmerken (in elk geval schoolweging (alleen in bao), taalachtergrond van de leerlingpopulatie, regio, urbanisatiegraad en schoolgrootte)?

11. Welke verschillen in leerlingprestaties zijn er naar kenmerken van het onderwijs in mondelinge taalvaardigheid (in elk geval het luister- en spreekklimaat in de klas, de aanwezigheid van geconcretiseerde leerdoelen en expliciete instructie voor gesprekken voeren, luisteren en spreken, het volgen van de vaardigheid van de leerlingen, de intentioneel bestede onderwijstijd en de integratie van het onderwijs met zowel andere taaldomeinen als andere leergebieden)?

12. Welke kenmerken van leerlingen, leerkrachten, scholen en van het onderwijs in mondelinge taalvaardigheid zijn in onderlinge samenhang mogelijke verklarende variabelen voor verschillen in leerlingprestaties?

Trends over tijd in leerlingprestaties

13. Welke verschillen zijn er tussen de leerlingprestaties einde bao, sbo en so anno 2024 en de leerlingprestaties in Peil.M ondelinge taalvaardigheid einde bao 2016-2017 en Peil.Mondelinge taalvaardigheid einde s(b)o 2017-2018, ook gecorrigeerd voor eventuele verschillen in achtergrondkenmerken van de leerlingpopulaties?

1.2 Leeswijzer

In de volgende hoofdstukken beschrijven we eerst het instrumentarium inclusief het afnamedesign (hoofdstuk 2) en de werkwijze die is gehanteerd (hoofdstuk 3). We geven de kenmerken van de steekproef, beschrijven hoe de werving heeft plaatsgevonden, de respons en de wijze van dataverzameling. In de hoofdstukken 4 en 5 worden respectievelijk de schaling van de spreek-, luister- en gesprekstaken en van de vragenlijsten beschreven. In hoofdstuk 6 worden de resultaten op de verschillende taken toegelicht. Hierbij wordt antwoord gegeven op onderzoeksvragen 1 t/m 3 en vraag 13 (trends over tijd). In hoofdstuk 7 worden de resultaten van de vragenlijsten beschreven. Hiermee wordt antwoord gegeven op onderzoeksvraag 4. Hoofdstuk 8 laat vervolgens de resultaten op de woordenschattoets zien. In hoofdstuk 9 wordt de samenhang tussen de leerlingprestaties en de algemene en domeinspecifieke leerling-, leerkracht- en schoolkenmerken en de kenmerken van het onderwijsleerproces beschreven. Hiermee wordt antwoord gegeven op onderzoeksvragen 5 t/m 12. Hoofdstuk 10 vat ten slotte de resultaten van het peilingsonderzoek samen.

2 Instrumentarium

Ten behoeve van deze peiling zijn verschillende spreek-, luister- en gesprekstaken gebruikt en vragenlijsten ontworpen. Enkele taken zijn afkomstig uit de vorige peiling Mondelinge taalvaardigheid einde speciaal (basis)onderwijs 2017-2018 (Keuning et al., 2019). Andere instrumenten zijn speciaal voor dit peilingsonderzoek ontworpen. Dit hoofdstuk beschrijft hoe de verschillende taken eruitzien, voor respectievelijk luisteren, spreken en gesprekken voeren. Ook beschrijven we kort de inhoud van de vragenlijsten. Tot slot beschrijven we het afnamedesign.

2.1 Theoretische uitgangspunten

Bij de instrumentontwikkeling vormde het Referentiekader Taal het uitganspunt. Met de nieuw ontwikkelde taken moeten leerlingen op de verschillende kenmerken van de taakuitvoering, zoals deze in het referentiekader voor spreken, gesprekken voeren en luisteren beschreven staan, te beoordelen zijn. Dat betekent dat leerlingen bijvoorbeeld bij een gesprekstaak de kans moeten krijgen of uitgelokt moeten worden om een gesprek te beginnen, gaande te houden en te beëindigen. Ook zijn de taken zo ontwikkeld dat leerlingen met een taalniveau onder 1F, op 1F en op 2F kunnen deelnemen en op die niveaus hun prestaties kunnen laten zien.

Bij het ontwikkelen van de spreektaak (betoog) is daarnaast als uitgangspunt genomen dat het doel en het publiek van de spreektaak anders zijn dan het doel en publiek van de naverteltaak. Het doel van de naverteltaak is informeren en het publiek is de toetsleider. Bij de nieuwe spreektaak is daarom gekozen voor ‘overtuigen’ als doel, met een fictieve leerkracht als publiek.

In de nieuw ontworpen gesprekstaak (raadsel) staat de gezamenlijke constructie van ideeën centraal en kunnen leerlingen uitgebreid laten zien hoe zij kunnen ingaan op en voortbouwen op elkaars ideeën, hoe ze kunnen redeneren, argumenten kunnen geven, kritisch naar redenaties en argumenten van anderen kunnen luisteren en daarop kunnen reageren. De gesprekstaak heeft daarmee kenmerken van een dialogisch gesprek (Van der Veen et al., 2015).

2.2 Luisteren

De luistervaardigheid van leerlingen is onderzocht met behulp van een luistertoets. De luistertoets is hoofdzakelijk samengesteld met items uit de vorige peiling in het s(b)o (Keuning et al., 2019) en het basisonderwijs (Van Langen et al., 2017). Deze items zijn aangevuld met items uit PPON 2007 en het Cito LVS 2.0. De luistertoets bestond in totaal uit 43 unieke items, verdeeld over 12 fragmenten. De items kunnen worden onderverdeeld over de in het referentiekader Taal en Rekenen (2019) genoemde kenmerken van de taakuitvoering: begrijpen, interpreteren en samenvatten. In Tabel 2.2.1 is weergegeven

hoe de items verdeeld zijn over deze kenmerken van de taakuitvoering. Ook is de verdeling weergegeven wat betreft de aard van de luisteractiviteit, de gespreksvorm en het ingeschatte referentieniveau.

Tabel 2.2.1: Overzicht over de verdeling van de luistertoets items, de aard van de luisteractiviteit, de gespreksvorm en de inschatting van het niveau.

Type item	Kenmerk	Aantal items
Kenmerken van de taakuitvoering	Begrijpen	19
	Interpreteren	23
	Samenvatten	1
Aard van de luisteractiviteit	Luisteren	9
	Kritisch luisteren	3
	Kijken en luisteren	29
	Kritisch kijken en luisteren	2
Gespreksvorm	Dialoog	15
	Monoloog	28
Niveau inschatting	1F 1F/2F 2F	18 5 20

De fragmenten werden vervolgens verdeeld over vier versies, waarbij fragmenten in meerdere versies voorkwamen. Bij de toewijzing van items aan versies is rekening gehouden met de moeilijkheidsgraad en de totale lengte van de versie. Op deze manier kon worden aangesloten bij de uiteenlopende doelgroepen van de peiling en bleef er voldoende tijd over voor de uitvoering van de andere taken. In paragraaf 4.1 wordt de samenstelling van de luistertoets en de verdeling van de items over de versies in detail beschreven. In Tabel 2.2.2 is de samenstelling samenvattend weergegeven.

Tabel 2.2.2: Het aantal items en de afnameduur per versie van de luistertoets.

Versie	N items	Afnameduur
Versie 1	20	40-45 minuten
Versie 2	20	35-40 minuten
Versie 3	15	35-40 minuten
Versie 4	11	30-35 minuten

De afnameprocedure was voor alle versies hetzelfde. De toetsleider leidde de afname klassikaal volgens van een protocol, waarbij leerlingen na het luisteren of kijken van een fragment 30 seconden de tijd kregen om antwoord te geven op de vraag. Leerlingen gaven hun antwoord in een digitale omgeving. Vervolgens liet de toetsleider het volgende fragment horen.

2.3 Spreken

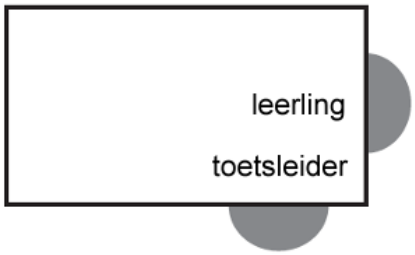
De spreekvaardigheid van leerlingen is onderzocht met behulp van twee spreektaken: naverteltaak en betoog. De spreektaken zijn gericht op de volgende vijf kenmerken van de taakuitvoering uit het Referentiekader Taal en Rekenen (2019):

Kenmerk	Uitwerking referentieniveau 1F
Samenhang	Maakt eigen gedachtegang voor de luisteraar begrijpelijk, hoewel de structuur van de tekst nog niet altijd klopt.
Afstemming op doel	Blijft trouw aan spreekdoel, soms met hulp van een ander.
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.
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.
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.

De naverteltaak is afkomstig uit de vorige peiling Mondelinge taalvaardigheid einde speciaal (basis)onderwijs 2017-2018 (Keuning et al., 2019). Deze taak diende in de huidige peiling als ankertaak. De taak betoog is in het kader van deze peiling ontwikkeld. Hierna worden de taken kort beschreven.

- Naverteltaak (ankertaak). De leerling bekeek met een koptelefoon op een fragment van 20 seconden uit het jeugdjournaal over een baby-otter. Na afloop vertelde de leerling aan de toetsleider wat hij heeft gezien en gehoord. Alle leerlingen kregen bij het navertellen een geheugensteuntje met kernwoorden uit het fragment, zodat leerlingen met een zwakker werkgeheugen niet benadeeld werden. De afname van deze taak nam ongeveer 10 minuten in beslag.
- Betoog. De leerling werd aan de hand van vragen van een fictieve leerkracht voorbereid op het houden van een betoog over hoe het jaarlijkse schoolreisje eruit zou moeten zien. Het doel was dat het gekozen schoolreisje leuk is voor iedereen en goed voor de natuur is. Tijdens het houden van hun betoog kregen leerlingen de antwoorden op de vragen uit de voorbereiding te zien. Leerlingen hielden het betoog voor de fictieve leerkracht op het scherm. De afname van deze taak nam ongeveer 15 minuten in beslag.

De spreektaken zijn individueel afgenomen door een getrainde toetsleider. De afname van de spreektaken vond plaats in een aparte ruimte met zo weinig mogelijk ruis en prikkels. De naverteltaak werd afgenomen in de volgende opstelling:



In deze opstelling kon de toetsleider goed contact maken met de leerlingen tijdens zijn rol als toehoorder. Tijdens de afname van het betoog volgde de leerling zelfstandig de instructies van een fictieve leerkracht op een laptop. Ook het betoog zelf hield de leerling tegen die fictieve leerkracht. De toetsleider hield zich op de achtergrond, maar hielp de leerling bij vragen of problemen.

De spreektaken werden ter plekke beoordeeld door de toetsleider met behulp van een beoordelingsformulier. Hiervoor is gebruikgemaakt van het beoordelingsformulier behorende bij de ankertaak (naverteltaak), die met kleine aanpassingen geschikt is gemaakt voor de beoordeling van de nieuw ontwikkelde taak (betoog). Voor de beoordeling is uitgegaan van een model waarin de toetsleider voor verschillende aspecten, verdeeld over de verschillende kenmerken van de taakuitvoering, op een driepuntsschaal (0 = onvoldoende; 1 = voldoende; 2 = goed) aangaf in hoeverre dit aspect werd geobserveerd in de spreekvaardigheid van de leerling. Concreet ging het om de onderstaande aspecten. Ter illustratie is bij de beoordelingsaspecten de hoogste beoordelingscategorie weergegeven.

Uitvoering

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

Samenhang

- 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.
- Samenhang op microniveau: signaalwoorden
  De leerling gebruikt eenvoudige voeg- en verbindingswoorden correct.

Afstemming op doel

- Spreekdoel
  - Naverteltaak: De leerling kan de strekking van het filmpje duidelijk maken aan de gesprekspartner (die het filmpje niet heeft gezien en gehoord).
  - Betoog: De leerling kan het voorstel overtuigend overbrengen aan de gesprekspartner.
- *Alleen bij betoog:* Argumenten
  De leerling geeft passende, zelfbedachte argumenten.

Afstemming op de gesprekspartner

- Taalgebruik afstemmen
  De leerling gebruikt consequent taal die bij de gesprekspartner past.
- Gesprekspartner boeien
  De leerling zet regelmatig stiltes, wisselende intonatie/volume/spreeksnelheid en lichaamstaal bewust in om de gesprekspartner te boeien.

• Non-verbale communicatie Het non-verbale gedrag van de leerling zet zijn verbale boodschap kracht bij.
Woordgebruik en woordenschat
• Woordgebruik De uitwerking van de leerling is aantrekkelijk door voldoende variatie in woordgebruik. • Woordenschat De getoonde woordenschat van de leerling is toereikend voor het uitvoeren van de opdracht.
Vloeiendheid, verstaanbaarheid en grammaticale beheersing
• Vloeiendheid De leerling spreekt redelijk vloeiend, al vertoont zijn uitwerking een enkele hapering en/of herformulering. • Uitspraak en intonatie De leerling is goed verstaanbaar door zijn uitspraak, articulatie en intonatie, ondanks een eventueel accent. • Grammatica: vervoeging De leerling vervoegt alle werkwoorden over het algemeen correct. Grammatica: verbuiging De leerling verbuigt (combinaties met) naamwoorden meestal correct. • Grammatica: zinsbouw De leerling maakt over het algemeen correcte enkelvoudige en samengestelde zinnen.

Het beoordelingsformulier is alleen ingevuld voor leerlingen die de beoogde taak zelfstandig en volledig hebben uitgevoerd.

2.4 Gesprekken

De gespreksvaardigheid van leerlingen is onderzocht met behulp van twee gesprekstaken: samenwerkingsspel en raadsel. De spreektaken zijn gericht op de volgende vijf kenmerken van de taakuitvoering uit het Referentiekader Taal en Rekenen (2019):

Kenmerk	Uitwerking referentieniveau 1F
1. Beurt nemen 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 de 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.

Samenwerkingsspel is afkomstig uit de vorige peiling Mondelinge taalvaardigheid einde speciaal (basis)onderwijs 2017-2018 (Keuning et al., 2019). Deze taak diende in deze peiling als ankertaak. Raadsel is in het kader van deze peiling ontwikkeld. Hierna worden de taken kort beschreven.

- **Samenwerkingsspel (ankertaak).** Voor deze taak is gebruikgemaakt van het *Fischerspiel*, een bestaand bordspel uit Duitsland. Bij dit spel moeten leerlingen samenwerken en overleggen om vissen te vangen en naar de haven te brengen voordat er storm komt. Deze gesprekstaak werd afgenomen in groepjes van drie of vier leerlingen. De afname nam ongeveer 45 minuten in beslag.
- **Raadsel.** Bij deze taak moeten leerlingen samen tot een oplossing voor een raadsel komen. Het raadsel ging over een jongen, Maik, die in een wat vreemde situatie is beland en verschillende voorwerpen bij zich had. Elke leerling kreeg één van die voorwerpen met daarbij een aanwijzing op een briefje. De leerlingen zochten, in overleg met elkaar en door daarbij gebruik te maken van de aanwijzingen, naar een mogelijke verklaring voor het raadsel. Deze gesprekstaak werd afgenomen in groepjes van drie leerlingen. De afname nam ongeveer 30 minuten in beslag.

De gesprekstaken zijn afgenomen door een getrainde toetsleider in een aparte ruimte met zo weinig mogelijk ruis en prikkels. De toetsleider had een passieve rol en liet de leerlingen zoveel mogelijk op een natuurlijke manier met elkaar het gesprek voeren. In enkele gevallen kon de toetsleider sturing geven aan het gesprek of ingrijpen, bijvoorbeeld als leerlingen de opdracht niet snapten, met elkaar in conflict raakten of als een leerling na een tijdje nog helemaal niet had deelgenomen aan het gesprek.

De taken werden ter plekke beoordeeld door de toetsleider met behulp van een beoordelingsformulier. Hiervoor is gebruik gemaakt van het beoordelingsformulier behorende bij de ankertaak (samenwerkingsspel), die met kleine aanpassingen geschikt is gemaakt voor de beoordeling van de nieuw ontwikkelde taak (raadsel). Voor de beoordeling is uitgegaan van een model waarin de toetsleider voor verschillende aspecten, verdeeld over de verschillende kenmerken van de taakuitvoering, op een driepuntsschaal (0 = onvoldoende; 1 = voldoende; 2 = goed) aangaf in hoeverre dit aspect terug te zien was in de gespreksvaardigheid van de leerling. Er was ook de mogelijkheid om 'Niet te beoordelen' in te vullen, indien een leerling niet genoeg bijdrage had in het gesprek om een aspect te kunnen beoordelen. Concreet ging het om de onderstaande aspecten. Ter illustratie is bij de beoordelingsaspecten de hoogste beoordelingscategorie weergegeven.

Uitvoering
• Kwaliteit van de inhoud De inhoudelijke uitwerking van de leerling is op de meeste inhoudsaspecten sterk.
Het woord nemen en bijdragen aan samenhang
• 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). • Passendheid van het woord nemen De leerling neemt doorgaans op passende wijze het woord. • Inbreng gesprekspartners integreren De leerling integreert bijdragen van gesprekspartners in zijn eigen inbreng wanneer dat toegevoegde waarde heeft.
Afstemming op doel
• Doelgerichtheid De leerling neemt initiatief om een gezamenlijk gespreksdoel te behalen door de gesprekspartners te betrekken bij het gesprek, bijvoorbeeld met behulp van vragen. • Begrijpelijk maken De leerling maakt de eigen gedachtegang begrijpelijk.
Afstemming op de gesprekspartners

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

Woordgebruik en woordenschat

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

Vloeiendheid, verstaanbaarheid en grammaticale beheersing

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

Naast bovenstaande aspecten is ook beoordeeld hoe leerlingen het spel evalueren (*De leerling geeft een helder oordeel over het spel*). Ook bevatte het beoordelingsformulier een groepsevaluatie. Toetsleiders konden bij dat onderdeel aangeven hoe evenwichtig de bijdragen aan het gesprek verdeeld waren over de leerlingen en of er sturing van de toetsleider nodig was in het gesprek. De individuele evaluatie en de groepsevaluatie zijn niet meegenomen in de score van een individuele leerling.

2.5 Overige instrumenten

De peilingsonderzoeken hebben als doel een betrouwbaar en valide beeld te geven van de vaardigheden in een bepaald domein of vakgebied van leerlingen in een bepaalde onderwijsfase, zoals het einde van het po of het tweede leerjaar van het vo. Daarnaast dienen zij inzicht te geven in de kenmerken die mogelijk samenhangen met het vaardigheidsniveau van deze leerlingen. Het gaat dan om algemene en domeinspecifieke kenmerken van leerlingen, leerkrachten/klassen en scholen, en om domeinspecifieke kenmerken van het onderwijsleerproces.

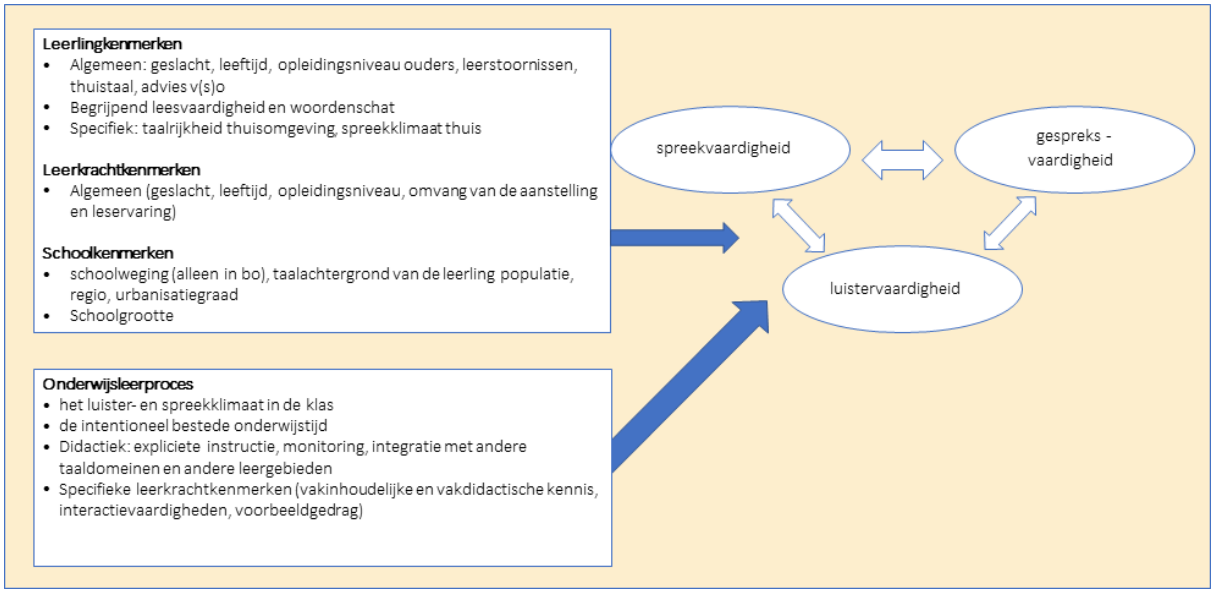
De informatie over de algemene en domeinspecifieke kenmerken is in dit peilingsonderzoek verzameld met de volgende instrumenten:

- Leerlinginformatielijst, in te vullen door de scholen (leerkrachten, IB'er, schooladministratie) over alle leerlingen in de getoetste klassen in bao, sbo en so;
- Quickscan, waarin de leerkracht enkele gegevens verstrekt over de leerlingen die een luister-, spreek- of gesprekstaak hebben uitgevoerd;

- Leerlingvragenlijst, in te vullen door de leerlingen zelf;
- Woordenschattoets, die de leerlingen in de getoetste klassen maakten als onderdeel van de leerlingvragenlijst;
- Leerkrachtvragenlijst voor de leerkracht(en) van de getoetste klassen;
- Schoolvragenlijst voor de schoolleider van de betreffende school/vestiging.

Voor de ontwikkeling van deze instrumenten is uitgegaan van het globale conceptueel model (Figuur 2.5.1) dat ook was opgenomen in ons onderzoeksvoorstel.

Model 2.5.1: Conceptueel model



De onderzoeksvragen (zie Hoofdstuk 1) met de daarin genoemde kenmerken zijn in het model als uitgangspunt genomen. De keuze voor met name de domeinspecifieke kenmerken is grotendeels gebaseerd op de adviezen uit de reviewstudie van Scheltinga (2022). De items en schalen die deze kenmerken representeren, zijn voor het merendeel afkomstig uit de vragenlijsten van de vorige peilingen voor mondelinge taalvaardigheid (o.a. Van Langen et al., 2018; Keuning et al., 2019) of van andere recente peilingen in het primair onderwijs waarbij de consortiumpartners betrokken waren (o.a. PIRLS & Peiling Leesvaardigheid 2021). Enkele vragen in de Quickscan zijn gebaseerd op het Zorgprofiel uit cohortonderzoek COOL$^{5\text{-}18}$ (zie o.a. Driessen et al., 2015[1]). Ook zijn enkele nieuwe vragen ontwikkeld. De achtergrondkenmerken van de deelnemende scholen (type, schoolweging, regio, schoolgrootte en urbanisatiegraad) zijn afkomstig uit in de onderwijsbestanden van DUO. De klassengrootte kon worden afgeleid uit het aantal opgegeven leerlingen op de Leerlinginformatielijst.

Tabellen 2.5.1 tot en met 2.5.5 geven een overzicht van de kenmerken die met de Leerlinginformatielijst, de Quickscan en de drie vragenlijsten in kaart zijn gebracht. Voor meer informatie over de herkomst en inhoud van de Woordenschattoets verwijzen we naar hoofdstuk 8 van dit rapport.

Tabel 2.5.1: Variabelen Leerlinginformatielijst peiling Mondelinge taalvaardigheid in het po (2023-2024)

Categorie	Kenmerk	Gemeten als:
Algemene leerlingkenmerken	Geslacht	Geslacht
	Leeftijd	Geboortejaar en maand
	Verwacht onderwijsniveau	Vo-advies of uitstroomprofiel

Tabel 2.5.2: – Variabelen Quickscan peiling Mondelinge taalvaardigheid in het po (2023-2024)

Categorie	**Kenmerk**	**Gemeten als:**
Beïnvloedende gedragskenmerken en stoornissen van leerlingen	Probleemgedrag (mate van)	Internaliserend probleemgedrag Externaliserend probleemgedrag Intro-/extraversie Spreekangst
	Communicatieproblemen (prevalentie)	Zwak taalbegrip Zwakke uitdrukkingsvaardigheid Zich niet kunnen verplaatsen in anderen Zwak auditief geheugen/verwerking
	Stimulerende en belemmerende factoren	Inhibitie Emotieregulatie Planning en organisatie Zelfstandigheid Motivatie Concentratie en aandacht

Tabel 2.5.3: Variabelen Leerlingvragenlijst peiling Mondelinge taalvaardigheid in het po (2023-2024)

Categorie	**Kenmerk**	**Gemeten als:**
Algemene leerlingkenmerken	SES (proxy)	Aantal boeken thuis
	Thuistaal	Frequentie thuis Nederlands gesproken Welke andere thuistaal Meest gesproken taal thuis Eerst geleerde taal
Domeinspecifieke leerlingkenmerken	Taalrijkheid thuisomgeving	Frequentie voorlezen vroeger Taal voorlezen vroeger Frequentie voorlezen nu Taal voorlezen nu
	Type taalactiviteiten thuis	Frequentie taalrijke activiteiten thuis Frequentie praten over dingen thuis Taal praten over dingen thuis Frequentie taalrijke activiteiten ouders
	Klasklimaat	Ervaren klasklimaat Ervaren spreekvrijheid

Tabel 2.5.4: Variabelen Leerkrachtvragenlijst peiling Mondelinge taalvaardigheid in het po (2023-2024)

Categorie	Kenmerk	Gemeten als:
Algemene leerkrachtkenmerken	Geslacht	Geslacht
	Leeftijd	Leeftijd in jaren
	Opleidingsniveau	Afgeronde lerarenopleiding
	Leservaring	Aantal jaren werkzaam als leerkracht
	Omvang aanstelling	Omvang huidig dienstverband
Domeinspecifieke leerkracht- en klaskenmerken	Houding en bekwaamheid mondelinge taalonderwijs	Belang Ervaren competentie
	Professionalisering mondelinge taalonderwijs	Professionaliseringsactiviteiten laatste 2 jaar
	Kenmerken onderwijsleerproces mondelinge taalvaardigheid in de klas	Absolute en relatieve onderwijstijd Methodegebruik Didactische aanpak Onderwijsaanbod luisteren, spreken en gesprekken

Tabel 2.5.5: Variabelen Schoolvragenlijst peiling Mondelinge taalvaardigheid in het po (2023-2024)

Categorie	Kenmerk	Gemeten als:
Algemene schoolkenmerken	Taalachtergrond leerlingpopulatie	Dominante thuistaal populatie Taalrijke/taalarme thuistaalsituatie populatie
Domeinspecifieke schoolkenmerken	Kenmerken onderwijsleerproces mondelinge taalvaardigheid in de school	Plaats in beleidsplan Grondslag doelen Methodegebruik Leerlijnen Monitoring Ondersteuning Percentage onderwijstijd Integratie met andere taaldomeinen & leergebieden Aandacht en visie

2.6 Het afnamedesign

In Tabel 2.6.1 is een overzicht van de taken, de afnamevorm en de -duur weergegeven.

Tabel 2.6.1: Overzicht van de taken, de afnamevorm en afnameduur.

Domein mondelinge taalvaardigheid	Taak	Afnamevorm	Duur
	Leerlingvragenlijst en Woordenschat (VL + WS)	Klassikaal	30 min
Luisteren	Luistertoets: 1, 2, 3 of 4	Klassikaal	30 – 45 min
Gesprekken	Samenwerkingsspel Raadsel	Groep: 3 à 4 leerlingen Groep: 3 leerlingen	30 – 45 min 15 min
Spreken	Betoog Naverteltaak	Individueel Individueel	15 min 10 min

Het afnemen van alle taken bij alle leerlingen was praktisch gezien onhaalbaar en onwenselijk, omdat het testen van één klas dan meerdere dagen in beslag zou nemen. Er is daarom besloten om een afnamedesign te gebruiken dat voorziet in een voldoende afname van (combinaties) van taken en dat daarnaast praktisch gezien haalbaar was voor scholen. Enkel de vragenlijst en woordenschattoets (zie hoofdstuk 8 voor een beschrijving) (VL+WT) is bij iedere leerling afgenomen. De basis van het design was dat in iedere klas 16 willekeurige leerlingen verdeeld werden over vier leerlinggroepjes. Indien een klas minder dan 16 leerlingen had, werden groepjes van drie leerlingen gevormd. Indien een klas te weinig leerlingen had om vier groepjes van drie leerlingen te maken, werden er minder groepjes gevormd. Vanwege de kleinere klassen in het speciaal onderwijs en speciaal basisonderwijs (respectievelijk so en sbo) vergeleken met het regulier basisonderwijs (bao), zijn op so/sbo-scholen willekeurig 16 leerlingen geselecteerd per school en niet per klas zoals in het bao. Iedere leerling uit een groepje voerde onder leiding van de toetsleider één van de volgende combinaties van twee taken uit:

- Beide gesprekstaken (GE-GE, combi 1)
- Beide spreektaken (SP-SP, combi 2)
- Een gesprekstaak en een spreektaak (GE-SP, combi 3- 6)
- Een gesprekstaak en een luistertoets (GE-LU, combi 7- 8)
- Een spreektaak en een luistertoets (SP – LU, combi 9- 10)

De leerlingen voerden dus een vooraf bepaalde selectie van taken uit, waardoor er sprake is van een structureel onvolledig design. Het afnamedesign wordt weergegeven in Tabel 2.6.2. Bij het toekennen van de combinaties van taken aan de leerlingen is erop gelet dat bepaalde combinaties vaker voorkwamen dan andere. In de kolom *Factor* is aangegeven hoe bij de toekenning is ingezet op bepaalde veelvouden van combinaties. In kolom *N* worden de gerealiseerde leerlingaantallen weergegeven.
Als aan een groepje een luistertoets was toegewezen, werd deze klassikaal afgenomen in de klas. De leerlingen in de klas die niet in een groepje waren ingedeeld, aangeduid als ‘combi 11’ in Tabel 2.6.2., maakten in dit geval ook de luistertoets. Indien alleen spreek- en gesprekstaken werden afgenomen in de klas, maakten deze overige leerlingen (combi 17) enkel de leerlingvragenlijst en woordenschattoets. Aangezien na de pilot is besloten dat het raadsel afgenomen wordt in een groepje van drie leerlingen in plaats van vier, zijn er ook leerlingen die één taak hebben gemaakt. Deze leerlingen, gecombineerd met leerlingen waarvan incidenteel een afname van een tweede taak niet gelukt is, zijn weergeven bij combi 12 tot en met 15.

Tabel 2.6.2: Gerealiseerd afnamedesign.

Populatie	Combi	Taken	Factor	GE1	GE2	SP1	SP2	LU4	LU3	LU2	LU1	VL+WT	N
Alle	1	GE1-GE2	8x										348
Alle	2	SP1-SP2	8x										366
Alle	3	GE1-SP1	6x										265
Alle	4	GE2-SP2	6x										226
Alle	5	GE1-SP2	6x										271
Alle	6	GE2-SP1	6x										218
Bao/sbo	7D	LU4-GE1	1x										41
Bao/sbo	8D	LU4-GE2	1x										40
Bao/sbo	9D	LU4-SP1	2x										82
Bao/sbo	10D	LU4-SP2	2x										77
Bao/so	7C	LU3-GE1	2x										76
Bao/so	8C	LU3-GE2	2x										76
Bao/so	9C	LU3-SP1	1x										33
Bao/so	10C	LU3-SP2	1x										38
Bao	7B	LU2-GE1	1x										49
Bao	8B	LU2-GE2	1x										46
Bao	9B	LU2-SP1	2x										110
Bao	10B	LU2-SP2	2x										105
Bao	7A	LU1-GE1	2x										121
Bao	8A	LU1-GE2	2x										101
Bao	9A	LU1-SP1	1x										57
Bao	10A	LU1-SP2	1x										54
Bao/sbo	11D	LU4	Rest										170
Bao/so	11C	LU3	Rest										131
Bao	11B	LU2	Rest										173
Bao	11A	LU1	Rest										123
Alle	12	GE1	Rest										156
Alle	13	GE2	Rest										42
Alle	14	SP1	Rest										107
Alle	15	SP2	Rest										71
Alle	16		Rest										67
Alle	17	VL+WT	Rest										553
Totaal													4393

Noot. GE1 = Gesprekstaak 1 (samenwerkingsspel); GE2 = Gesprekstaak 2 (raadsel); SP1 = Spreektaak 1 (naverteltaak); SP2 = Spreektaak 2 (betoog); LU1 = Luistertoets 1; VL+WT = Vragenlijst + Woordenschattoets; BAO = regulier basisonderwijs; SBO = speciaal basisonderwijs; SO = speciaal onderwijs

Na afname bleek - in aanvulling op het bovenstaande - dat 67 leerlingen drie taken hadden uitgevoerd (in wisselende samenstelling van taken), waarvan één leerling alle spreek- en gesprekstaken had gedaan. Deze leerlingen worden weergegeven in combi 16.

Dankzij dit afnamedesign zijn alle combinaties van taken voorgekomen binnen leerlingen en is er een link tussen alle domeinen en taken binnen de domeinen. Daarnaast is ervoor gekozen dat de combinaties van beide spreektaken en beide gesprekstaken het vaakst voor kwamen. Hierdoor kan de samenhang tussen domeinen worden geëvalueerd (paragraaf 4.4), en zijn er vele mogelijkheden ten behoeve van de schaling (paragraaf 4.5), waarbij domeinen los kunnen worden geschaald, en – indien daarvoor aanleiding is – eventueel ook gezamenlijk kunnen worden geschaald. Op de designs per domein (i.e., luistervaardigheid, spreekvaardigheid en gespreksvaardigheid) wordt nader ingegaan in de paragraaf 4.1.

3 Steekproef, werving, respons, dataverzameling

3.1 Inleiding

Het peilingsonderzoek Mondelinge taalvaardigheid 2023-2024 werd uitgevoerd aan het einde van het basisonderwijs. Meer specifiek gaat het hierbij om groep 8 van het regulier basisonderwijs (bao) en schoolverlaters van het speciaal basisonderwijs (sbo) en het speciaal onderwijs (so). Per type onderwijs is er een steekproef van scholen getrokken door de inspectie. Hierna wordt per schooltype toegelicht op basis van welke criteria deze steekproeven zijn getrokken.

Steekproef regulier basisonderwijs

De doelpopulatie bestond uit schoolvestigingen in het bekostigd basisonderwijs. Er zijn drie wervingstranches van ieder 155 basisscholen getrokken. Bij het trekken van de steekproef werd rekening gehouden met stratificatievariabelen regio[1] (expliciet) en schoolweging[2] (impliciet). Daarnaast werden scholen die in het jaar 2022-2023 minder dan zes leerlingen in groep 7 hadden, scholen voor speciaal basisonderwijs en particulier onderwijs, rijdende scholen, scholen geheel gericht op nieuwkomers en scholen in het buitenland en scholen die zich geheel richten op internationaal onderwijs uitgesloten. Per basisschool werden twee vervangende scholen getrokken als reservesteekproef. Deze reservescholen kwamen wat betreft de stratificatievariabelen zoveel mogelijk overeen met de school in de hoofdsteekproef. Scholen die voor andere stelselonderzoeken door de inspectie[3] worden benaderd, zijn zoveel als mogelijk gewisseld met hun gelijkwaardige vervangers in de laatste tranche(s).

Steekproef sbo

De steekproef van scholen voor speciaal basisonderwijs heeft de inspectie getrokken op basis van de stratificatiecriteria regio (expliciet) en uitstroomprofiel van de school[4] (impliciet). In totaal zijn er drie wervingstranches van ieder 70 scholen voor sbo getrokken, waarbij er voor iedere school in de eerste tranche (hoofdsteekproef) twee vervangende scholen zijn getrokken als reservesteekproef. Deze reservescholen kwamen wat betreft de stratificatievariabelen zoveel mogelijk overeen met de school in de hoofdsteekproef. Scholen zonder zicht op eindresultaten in relatie tot Nederlandse kerndoelen en/of waarbij opheffing in 2022-2023 is aangekondigd zijn uitgesloten voor de steekproeftrekking. Bij het trekken

1 De scholen horen bij de regio van de provincie waar in ze staan. De provincies zijn ingedeeld in vier regio's: Noord (Drenthe, Friesland, Groningen); Oost (Gelderland, Overijssel), Midden (Flevoland, Noord-, Zuid-Holland, Utrecht); Zuid (Limburg, Noord-Brabant, Zeeland).

2 Uitgegaan is van de schoolweging van de scholen in 2022-2023. De maat betreft een afgeleide van de berekening die CBS maakt t.b.v. het onderwijsachterstandenbeleid. Hoe hoger de schoolweging, des te complexer de leerlingpopulatie van de school.

3 Scholen die hebben deelgenomen aan Peil.Mens & maatschappij en Peil.Rekenen & Wiskunde in schooljaar 2022-2023, scholen die deel uitmaakten de eerste wervingstranche van Peil.Mens & Natuur of Peil.Bewegen & Sport in 2023-2024 en scholen die in 2023-2024 zijn onderzocht vanwege een landelijk representatief themaonderzoek van de inspectie.

4 Het uitstroomprofiel van de school is gebaseerd op schooladviezen aan leerlingen, meer specifiek een advies voor vso of pro zoals geregistreerd in DUO Open Onderwijsdata. Per school is het aandeel adviezen voor vso of pro in het totaal aantal afgegeven adviezen bepaald over de laatste drie schooljaren (2019-20, 2020-2021 en 2021-22).

van de steekproef is daarnaast rekening gehouden met de peiling Rekenen-Wiskunde 2022-2023 en een representatief themaonderzoek van de inspectie in 2023-2024. Deze scholen zijn zoveel als mogelijk gewisseld met hun gelijkwaardige vervangers in de laatste tranche(s).

Steekproef so

Voor het peilingsonderzoek mondelinge taalvaardigheid onder schoolverlaters in het speciaal onderwijs (so) heeft de inspectie een steekproef met twee wervingstranches van ieder 45 scholen getrokken. De doelpopulatie beperkt zich tot de schoolverlatende leerlingen met uitstroomprofiel vervolgonderwijs in voormalig clusters 3 en 4[5]. Scholen die minder dan vijf leerlingen behorende tot de doelpopulatie in het laatst bekende schooljaar hadden, worden uitgesloten van de steekproeftrekking. De gehanteerde stratificatiecriteria voor de steekproeftrekking betreffen cluster (expliciet) en uitstroompercentage vmbo-g/t of hoger[6] (impliciet). Elke school in de eerste tranche heeft een gelijkwaardige vervanger qua stratificatiecriteria in de tweede tranche. Scholen die reeds waren geworven voor de peiling Reken-Wiskunde in 2022-2023 en het representatief thema-onderzoek van de inspectie in 2023-2024 zijn zoveel als mogelijk verplaatst naar de tweede tranche.

3.2 Werving

Organisatie en verloop werving

Om voldoende deelnemers voor het onderzoek te verkrijgen is er een intensieve wervingsprocedure gehanteerd, die bestond uit een aankondiging van het onderzoek in informatie- en aanbevelingsbrieven (ondertekend door de Inspecteur-Generaal van de Inspectie van het Onderwijs), gevolgd door een telefonisch contact door een getraind belteam. Scholen die bereid waren om deel te nemen, konden zich aanmelden via een aanmeldlink.

De werving startte met een aankondiging van het onderzoek in digitale nieuwsbrieven en op sociale media. De schoolleiders van de scholen uit de steekproef ontvingen vervolgens een informatiebrief waarin het belang en de praktische organisatie van het onderzoek werden beschreven en hen gevraagd werd zich aan te melden voor het onderzoek. Als bijlage was een door de Inspecteur-Generaal van de Inspectie van het Onderwijs ondertekende aanbevelingsbrief toegevoegd die nogmaals het belang en de anonimiteit van het onderzoek benadrukte. De brieven zijn in het najaar van 2023 zowel op papier als digitaal verstuurd naar de schoolleiders van alle scholen uit de hoofdsteekproef voor het bao, sbo en so.

Enkele dagen na het versturen van de brieven nam een onderzoeksassistent van het belteam telefonisch contact op met de schoolleider. In dit telefoongesprek werd desgewenst meer informatie gegeven over het onderzoek en werd de school gevraagd om deel te nemen. De onderzoeksassistent volgde hierbij een

5 Onder uitstroomprofiel vervolgonderwijs wordt vmbo-basis/vmbo-kader of hoger verstaan. Praktijkonderwijs wordt hierin niet meegenomen. Bovendien behoren clusters 1 en 2 niet tot de doelpopulatie.

6 Het percentage leerlingen dat is uitgestroomd naar vmbo-g/t of hoger in de laatste drie schooljaren.

belscript zodat alle schoolleiders dezelfde informatie ontvingen. Schoolleiders konden in het telefoongesprek alle vragen stellen die ze hadden met betrekking tot het onderzoek.
Om deelname aan het onderzoek zo aantrekkelijk mogelijk te maken, namen de onderzoekers en onderzoeksassistenten de voorbereidingen van de onderzoeksafname zoveel als mogelijk voor hun rekening. Om scholen goed te informeren is bij de informatiebrief voor de scholen een stappenplan gemaakt waarin alle te nemen stappen voor de scholen staan beschreven. Voor de ouders zijn er ouderbrieven gemaakt in zeven verschillende talen. Ook hebben we een informatiefilmpje gemaakt voor ouders waarin we uitleggen wat het onderzoek inhoudt, wat we met de resultaten kunnen en hoe de privacy en toestemming voor audio- en video-opnames geregeld zijn. Met de ouderbrieven en het filmpje hebben we ervoor gezorgd dat alle ouders konden begrijpen wat het onderzoek inhield en goed geïnformeerd wel of geen toestemming konden geven voor audio- en video-opnames. Onderdeel van de brief was een invulstrook, waarop ouders konden aangeven of we wel of geen opnames van hun kind mochten maken. Dit strookje werd vervolgens weer op de school ingeleverd, die deze informatie anoniem terugkoppelde aan ons. De leerlingen kregen na hun deelname een kleinigheidje (speelgoed). Na hun deelname ontvingen scholen bovendien een terugkoppeling van de resultaten in de vorm van een schoolrapportage. Ook werden scholen uitgenodigd om deel te nemen aan een workshop over effectief onderwijs in de mondelinge taal, waarbij ze niet alleen praktische handvatten kregen over mondelinge taalonderwijs, maar ook met andere scholen hun ervaringen en ideeën konden uitwisselen.

Voor zowel het bao, sbo als so gold dat wanneer een school zich afmeldde voor het onderzoek, de vervangende school uit de reservesteekproef benaderd werd. Scholen die besloten deel te nemen aan het onderzoek werd gevraagd om hun aanmelding te bevestigen via een digitale aanmeldlink, waarin ze ook een contactpersoon aanwezen en administratieve gegevens verstrekten zoals het aantal groepen 8 of schoolverlatersgroepen en het aantal leerlingen per groep. De scholen is vervolgens gevraagd de ouders te informeren over het onderzoek. Hiervoor is door de onderzoekers een ouderbrief (in verschillende talen) aangeleverd die de scholen aan de ouders konden sturen. Indien ouders geen toestemming gaven voor deelname van hun kind aan het peilingsonderzoek, konden zij dat aan de onderzoekers of de school doorgeven. Deze kinderen werden uitgesloten van deelname aan het onderzoek.

3.3 Respons

Regulier basisonderwijs

In totaal hebben 243 basisscholen een uitnodigingsbrief ontvangen. Aanvankelijk waren er 139 basisscholen bereid gevonden om deel te nemen aan het onderzoek. Uiteindelijk hebben er 131 basisscholen daadwerkelijk deelgenomen aan het onderzoek en heeft er op 8 basisscholen die deelname hadden toegezegd om verschillende redenen toch geen afname plaatsgevonden. Het totale responspercentage bedraagt 85% met de deelname van 131 scholen uit een steekproef van 155 scholen. De meerderheid van de scholen (57%) is afkomstig uit de hoofdsteekproef (tranche 1, zie Tabel 3.3.1).

Tabel 3.3.1: Overzicht werving en respons bao, naar tranche.

	Tranche 1	Tranche 2	Tranche 3	Totaal
Benaderd	**155**	**63**	**25**	**243**
Waarvan:				
a. Deelname toegezegd én gerealiseerd	*88 (57%)*	*34*	*9*	***131***
b. Deelname toegezegd, niet gerealiseerd (uitval)	*5*	*2*	*1*	*8*
c. Deelname geweigerd	*62*	*27*	*12*	*101*
d. Deelname onbekend (veelvuldig benaderd zonder respons)	*0*	*0*	*3*	*3*
Niet benaderd	**0**	**92**	**130**	**232**
Totaal	**155**	**155**	**155**	**465**

In totaal hebben 101 benaderde basisscholen afgezien van deelname. De voornaamste redenen voor hun weigering worden in Tabel 3.3.2 gepresenteerd.

Tabel 3.3.2: Overzicht redenen weigering deelname bao

Voornaamste reden weigering deelname	N scholen	%
Te druk, kost te veel tijd / moeite	57	56%
Onrust in school / al genoeg problemen	19	19%
Zien het belang niet in / geen interesse in onderwerp	6	6%
Deelname andere onderzoeken	4	4%
Doen nooit mee aan onderzoek	1	1%
Grote veranderingen (zoals verhuizing) / fusie	1	1%
Weigering zonder duidelijke reden	13	13%
Totaal	101	100%

In Tabel 3.3.3 wordt uiteengezet hoeveel scholen, klassen, leerlingen, leerkrachten en schoolleiders uiteindelijk hebben deelgenomen aan het onderzoek. Uiteindelijk hebben in het regulier basisonderwijs 169 klassen deelgenomen, met daarin 2.926 leerlingen. De docentvragenlijst werd ingevuld door 133 leerkrachten, de vragenlijst voor schoolleiders werd ingevuld door 97 schoolleiders.

Tabel 3.3.3: Deelname scholen, klassen, leerlingen, leerkrachten en schoolleiders

	Deelgenomen (incl. gedeeltelijke deelname)
Scholen	131
Klassen	169
Leerlingen	2.926
Leerkrachten	133
Schoolleiders	97

Representativiteit

Om na te gaan of de responderende bao-scholen representatief zijn voor de totale basisschoolpopulatie worden de scholen in Tabel 3.3.4 op enkele achtergrondkenmerken vergeleken met de steekproef en de doelpopulatie. De kenmerken betreffen regio en schoolweging, de stratificatiecriteria voor de steekproeftrekking. In Tabel 3.3.4 wordt duidelijk dat het aandeel bao-scholen afkomstig uit de regio's Midden, Noord, Oost en Zuid, redelijk vergelijkbaar is met het de verdeling in de steekproef en de doelpopulatie. Een chi^2-toets laat dan ook zien dat de verdeling in de respons niet significant afwijkt van de landelijke verdeling (χ^2=0,879, df=3; N=131; p=0,831). Hetzelfde geldt voor het kenmerk schoolweging[7]: de responderende bao-scholen zijn redelijk gelijk verdeeld over de vier kwartielen van schoolweging en de verdeling wijkt dan ook niet significant af van de landelijke verdeling[8] (χ^2=0,389, df=3; N=131; p=0,942). Dit betekent dat de bao-scholen die hebben deelgenomen aan het onderzoek representatief zijn voor de totale basisschoolpopulatie.

Tabel 3.3.4: Achtergrondkenmerken van bao-scholen in de steekproef, de totale basisschoolpopulatie en deelnemende bao-scholen

	Steekproef				
	Tranche 1	**Tranche 2**	**Tranche 3**	**Doelpopulatie**	**Respons**
Regio (%)					
Midden	43,2%	43,2%	43,2%	43,2%	39,7%
Noord	13,5%	13,5%	13,5%	13,4%	13, 0%
Oost	21,3%	21,3%	21,3%	21,4%	22,9%
Zuid	21,9%	21,9%	21,9%	22,0%	24,4%
Schoolweging (%)					
1e kwartiel	25,8%	25,8%	25,2%	25,0%	22,9%
2e kwartiel	24,5%	25,2%	25,2%	25,0%	25,2%
3e kwartiel	25,2%	24,5%	25,2%	25,0%	26,7%
4e kwartiel	24,5%	24,5%	24,5%	25,0%	25,2%
Totaal (N)	**155**	**155**	**155**	**5891**	**131**

Speciaal basisonderwijs (sbo)

In totaal hebben 120 sbo-scholen een uitnodigingsbrief voor het onderzoek ontvangen. Aanvankelijk hebben 60 van deze scholen toegezegd deel te nemen aan het onderzoek. Op vijf scholen daarvan is de deelname uiteindelijk toch niet gerealiseerd, zo wordt duidelijk in Tabel 3.3.5. Dit betekent dat het onderzoek heeft plaatsgevonden op 55 sbo-scholen en er een responspercentage van 79% is behaald. Van deze 55 scholen zijn er 35 afkomstig uit de hoofdsteekproef (Tranche 1). De respons uit de hoofdsteekproef bedraagt daarmee 50%.

7 Gebaseerd op schoolweging van de scholen in schooljaar 2022-2023
8 Dat geldt (logischerwijs) ook voor de verdeling in de steekproef.

Tabel 3.3.5: Overzicht werving en respons sbo, naar tranche

	Tranche 1	Tranche 2	Tranche 3	Totaal
Benaderd	**70**	**34**	**16**	**120**
Waarvan:				
a. Deelname toegezegd én gerealiseerd	*35 (50%)*	*16*	*4*	***55***
b. Deelname toegezegd, niet gerealiseerd (uitval)	*1*	*2*	*2*	*5*
c. Deelname geweigerd	*33*	*16*	*6*	*55*
d. Deelname onbekend (vaak genoeg benaderd zonder respons)	*1*	*0*	*4*	*5*
Niet benaderd	**0**	**36**	**54**	**90**
Totaal	**70**	**70**	**70**	**210**

55 van de 120 sbo-scholen die een uitnodigingsbrief hebben ontvangen, hebben aangegeven niet deel te willen nemen aan het onderzoek. In Tabel 3.3.6 worden de voornaamste redenen voor geweigering gepresenteerd.

Tabel 3.3.6: Overzicht redenen weigering deelname sbo

Voornaamste reden weigering deelname	N scholen	%
Te druk/ kost te veel tijd/ moeite	25	
Onrust in school/ al genoeg problemen	18	33%
Zien het belang niet in/ geen interesse in onderwerp	4	7%
Deelname andere onderzoeken	3	5%
Weigering zonder duidelijke reden	5	9%
	55	99,00%

In Tabel 3.3.7 wordt uiteengezet hoeveel scholen, klassen, leerlingen, leerkrachten en schoolleiders uiteindelijk hebben deelgenomen aan het onderzoek. Uiteindelijk hebben in het speciaal basisonderwijs 108 klassen deelgenomen, met daarin 1053 leerlingen. De docentvragenlijst werd ingevuld door 82 leerkrachten, de vragenlijst voor schoolleiders werd ingevuld door 42 schoolleiders.

Tabel 3.3.7: Aanmeldingen en deelgenomen

	Deelgenomen (Incl. gedeeltelijke deelname)
Scholen	55
Klassen	108
Leerlingen	1053
Leerkrachten	82
Schoolleiders	42

Representativiteit

Om na te gaan of de responderende sbo-scholen representatief zijn voor de totale populatie van scholen voor speciaal basisonderwijs worden de scholen in Tabel 3.3.8 op enkele achtergrondkenmerken vergeleken met de steekproef en de doelpopulatie. De kenmerken betreffen de regio en het uitstroomprofiel (% advies vso/pro), de stratificatiecriteria voor de steekproeftrekking. Een chi^2-toets laat zien dat de verdeling naar regio in de respons niet significant afwijkt van de landelijke verdeling (χ^2=**0,995961**, df=3; N=55; p=0,802229). Bij het uitstroomprofiel zijn de scholen ingedeeld in drie categorieën, te weten ≤ 35% advies vso/pro, 36% - 59% advies vso/pro en ≥ 60% advies vso/pro. We bezitten geen informatie voor de landelijke populatie van sbo-scholen over het percentage leerlingen dat uitstroomt naar vso of pro. Om deze reden worden er dan ook geen gegevens getoond in de vierde kolom voor uitstroomprofiel en voor het bepalen van de representativiteit de verdeling in de respons afgezet tegen de verdeling in de steekproeven. De verdeling naar de categorieën van uitstroomprofiel in de respons wijkt niet significant af van de verdeling in Hoofdsteekproef (tranche 1; χ^2=**0,659**, df=2; N=55; p=0,719)[9]. Daarmee kunnen we concluderen dat sbo-scholen die hebben deelgenomen aan het onderzoek representatief zijn voor de totale populatie van scholen voor speciaal basisonderwijs.

Tabel 3.3.8: Achtergrondkenmerken van sbo-scholen in de steekproef, de totale basisschoolpopulatie en deelnemende sbo-scholen.

	Tranche 1	Tranche 2	Tranche 3	Doelpopulatie	Respons
Regio (%)					
Midden	48,6%	48,6 %	48,6%	48,1%	47%
Noord	11,4%	11,4%	11,4%	11,0%	7%
Oost	20,0%	20,0%	20,0%	20,1%	22%
Zuid	20,0%	20,0%	20,0%	20,8%	24%
Uitstroomprofiel					
% advies vso/pro (gem/sd)	51,3/14,9	51,4/15,1	50,9/15,1	51,4/15,1	49,6 / 15,6
≤ 35%	18,6%	15,7%	17,1%		20,0%
36% - 59%	47,1%	51,4%	51,4%		50,9%
≥ 60%	34,3%	32,9%	31,4%		29,1%
Totaal (N)	**70**	**70**	**70**	**264**	**55**

Speciaal onderwijs (so)

Voor het speciaal onderwijs zijn er twee wervingstranches met ieder 45 scholen getrokken. In totaal hebben 68 so-scholen een uitnodigingsbrief ontvangen. Daarvan hebben 28 so-scholen deelname toegezegd én daadwerkelijk deelgenomen. Het responspercentage bedraagt daarmee 62%. Van deze 28

9 Dat geldt ook voor de overige wervingstranches; de verdeling in de respons wijkt niet significant af van de verdelingen in Tranche 2 en Tranche 3).

deelnemende so-scholen, behoren er 19 so-scholen (42%) tot de hoofdsteekproef (tranche 1). Relatief veel so-scholen (n=33) hebben deelname geweigerd. In Tabel 3.3.10 worden de voornaamste redenen voor weigering door so-scholen getoond.

Tabel 3.3.9: Aantallen benaderde so-scholen respons per tranche

	Tranche 1	Tranche 2	Totaal
Benaderd	**45**	**23**	**68**
Waarvan:			
a. Deelname toegezegd én gerealiseerd	*19 (42%)*	*9*	***28***
b. Deelname toegezegd, niet gerealiseerd (uitval)	*2*	*2*	*4*
c. Deelname geweigerd	*24*	*9*	*33*
d. Deelname onbekend (vaak genoeg benaderd zonder respons)	*0*	*3*	*3*
Niet benaderd	**0**	**22**	**22**
Totaal	**45**	**45**	**90**

Tabel 3.3.10 Overzicht redenen weigering deelname bao

Voornaamste reden weigering deelname	N scholen	%
Te druk, kost te veel tijd/ moeite	16	48%
Onrust in school/ al genoeg problemen	9	27%
Weigering zonder duidelijke reden	5	15%
Zien het belang niet in/ geen interesse in onderwerp	3	9%
Totaal	**33**	**100%**

In Tabel 3.3.11 wordt uiteengezet hoeveel scholen, klassen, leerlingen, leerkrachten en schoolleiders uiteindelijk hebben deelgenomen aan het onderzoek. Uiteindelijk hebben in het speciaal onderwijs 52 klassen deelgenomen, met daarin 414 leerlingen. De docentvragenlijst werd ingevuld door 38 leerkrachten, de vragenlijst voor schoolleiders werd ingevuld voor 21 schoolleiders.

Tabel 3.3.11: Aantallen benaderde so-scholen respons per tranche

	Deelgenomen (Incl. gedeeltelijke deelname)
Scholen	28
Klassen	52
Leerlingen	414
Leerkrachten	38
Schoolleiders	21

Representativiteit

Ook voor het speciaal onderwijs is nagegaan in hoeverre de responderende so-scholen representatief zijn voor de totale doelpopulatie van (cluster 3 en 4) so-scholen. In Tabel 3.3.12 worden de kenmerken van de scholen in de wervingstranches, doelpopulatie en responsgroep getoond. Het gaat hierbij om de stratificatiecriteria voor de steekproeftrekking, cluster en percentage uitstroom naar vmbo-gt of hoger. Een chi^2-toets laat zien dat de verdeling naar cluster in de responsgroep niet significant afwijkt van de verdeling in de doelpopulatie (χ^2=0,196, df=1; N=28; p=0,658). Om na te gaan of de responderende so-scholen vergelijkbaar zijn met de populatie op het gebied van het percentage leerlingen dat uitstroomt naar vmbo-gt of hoger, onderscheiden we drie categorieën. We beschikken niet over deze informatie voor de doelpopulatie. Er blijkt geen significant verschil in verdeling over deze categorieën te zijn tussen de responsgroep en in de wervingstranches (χ^2= 0,367779, df=2; N28; p=0,832028). Daarmee kunnen we concluderen dat de so-scholen die hebben deelgenomen aan het onderzoek representatief zijn voor de totale populatie van (voormalig cluster 3 en 4) so-scholen.

Tabel 3.3.12: Achtergrondkenmerken van so-scholen in de steekproef, de totale basisschoolpopulatie en deelnemende so-scholen

	Tranche 1	Tranche 2	Doelpopulatie	Respons
Cluster				
Voormalig cluster 3	18%	18%	18%	21%
Voormalig cluster 4	82%	82%	82%	79%
Percentage uitstroom vmbo-gt of hoger (gem/sd)	38,0 (14,3)	37,9 (14,0)	37,7 (13,7)	37,7 (16,5)
0 - ≤ 34%	36%	36%		36%
35% - 45%	33%	33%		29%
≥ 45%	31%	31%		36%
Totaal (N)	**45**	**45**	**110**	**28**

4 Schaling luisteren, spreken en gesprekken voeren

In dit hoofdstuk wordt dieper ingegaan op de psychometrische analyses van de onderdelen luistervaardigheid, spreekvaardigheid en gespreksvaardigheid. Eerst worden de uiteindelijke observatieaantallen besproken en de ontbrekende waarden. Daarna worden de klassieke psychometrische gegevens weergegeven en wordt de beoordelaarsovereenstemming geëvalueerd. Vervolgens wordt dieper ingegaan op de samenhang tussen de items met verschillende kenmerken. De keuze voor het itemresponstheoriemodel (IRT-model) wordt toegelicht en een *differential item functioning*-analyse (DIF-analyse) ten aanzien van de verschillende schooltypes (bao, sbo, so) wordt geëvalueerd.

4.1 Aantallen en ontbrekende waarden

4.1.1 Luistervaardigheid

Voor het onderdeel luistervaardigheid zijn de 12 fragmenten (met in totaal 43 individuele items) verdeeld over vier verschillende toetsversies. Het afnamedesign is weergegeven in Tabel 4.1.1. Om de toetsversies te kunnen equivaleren waren zeven fragmenten in twee toetsversies opgenomen, zodat alle toetsversies verbonden waren.

Tabel 4.1.1: Afnamedesign luistervaardigheid.

Fragment	Inschatting niveau		Bron	N items	V4	V3	V2	V1
Ridders	1F		Anker 1F	5				
Bank	2F		Anker 2017	2				
Melkmeisje	2F		Anker 2F	4				
Gooi	2F		Anker 2017	2				
Tekenares	1F		LVS E6	4				
Zebrastrepen	1F		Anker 1F	3				
Goud	1F		Anker 2017	3				
Bacteriën	2F		PPON 2007	5				
Julia	1F		Anker 2017	3				
Contant geld	2F		Anker 2017	2				
Internet	2F		PPON 2007	5				
Pindakaas	1F/2F		LVS E7	5				
N items					11	15	20	20

Uit Tabel 4.1.1 is af te leiden dat de lengte van de toetsversies varieert van 11 tot 20 items. Daarnaast werd de ingeschatte moeilijkheid van de fragmenten in acht genomen bij het opstellen van de toetsversies. Bij de toekenning van de toetsversies aan de leerlingen was sprake van een *targeted* design, waarbij de toekenning van de leerlingen is aangepast aan het schooltype van de leerlingen. Toetsversie 4 was het

kortst en bestond uit overwegend de makkelijkste fragmenten. Deze versie werd gemaakt door alle sbo-leerlingen. De iets moeilijkere en langere toetsversie 3 werd gemaakt door alle so-leerlingen. De twee moeilijkste toetsversies 1 en 2 werden gemaakt door bao-leerlingen. Daarnaast maakten bao-leerlingen ook toetsversie 3 en 4, zodat alle verschillende items een gelijk aantal observaties zouden krijgen. Het aantal leerlingen per toetsversie per populatie is weergegeven in Tabel 4.1.2.

Tabel 4.1.2: Aantallen leerlingen per toetsversie per populatie.

Toetsversie	Populatie	Aantal
1	bao	466
2	bao	506
3	bao	244
	so	120
4	bao	45
	sbo	373
Totaal		1754

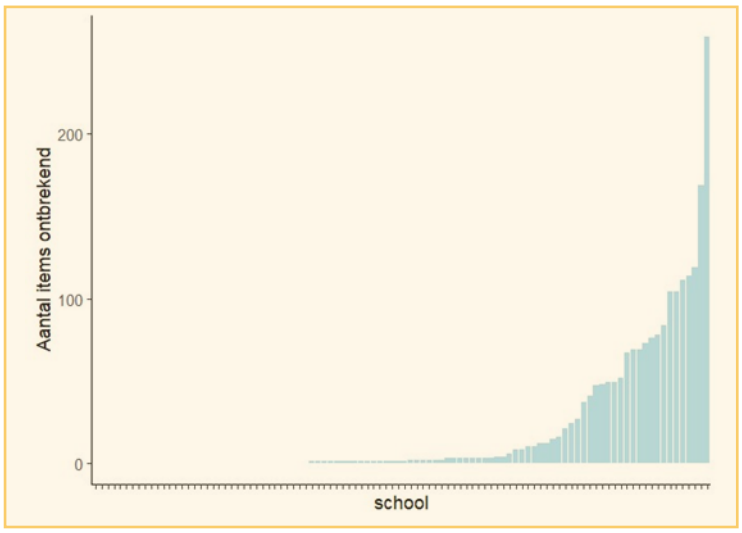
Van deze 1754 leerlingen is in Tabel 4.1.3 weergegeven hoeveel leerlingen een bepaald aantal items hebben gemaakt van de toetsversie die zij voorgelegd hadden gekregen.

Tabel 4.1.3: Aantal gemaakte items per leerling.

Toets versie	Totaal N items	Aantal gemaakte items																		
		Alle opgaven	**19**	**18**	**17**	**16**	**15**	**14**	**13**	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**1**
1	20	**327**	19	9	5	1	31	4	9	11	16	12	6	5	7	3	1	0	0	0
2	20	**375**	29	4	19	12	16	23	1	15	0	2	2	0	1	2	2	1	0	2
3	15	**265**						17	6	22	4	30	2	13	3	2	0	0	0	0
4	11	**384**										19	11	1	1	0	0	0	2	0

Als alle leerlingen alle items hadden gemaakt van hun toetsversie, zou het totale aantal 29.498 responses zijn geweest. In de praktijk bleken er 27.450 responses te zijn. Hiermee ontbreekt ongeveer 6,9% van de data. De ontbrekende waarden bleken redelijk geclusterd rondom scholen te zitten. In Figuur 4.1.1 zijn de scholen gesorteerd op het aantal ontbrekende waarden.

Figuur 4.1.1: Aantal ontbrekende waarden per school.



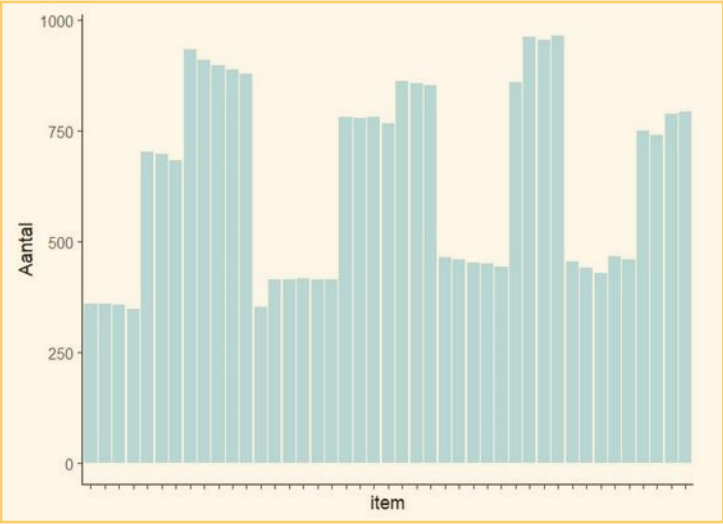
In figuur 4.1.1 is te zien dat er een aantal uitschieters zijn. Op deze scholen was volgens toetsleiders onvoldoende tijd beschikbaar om de afname van de toetsversie af te maken. De ontbrekende data leken hierdoor meer het resultaat te zijn van praktisch omstandigheden dan van de luistervaardigheid van de leerlingen. Uit de analyse bleek ook dat de luistervaardigheid niet samenhing met het aantal ontbrekende waarden (correlatie: -0,002). Om die reden is besloten om de ontbrekende data niet te imputeren met de minimale score (i.e., 0), maar om deze weg te laten uit de analyse. Van de 1754 leerlingen hadden drie leerlingen minder dan 75% van de items gemaakt. De respons van de deze leerlingen is hiermee als ongeldig beschouwd. Van de overige leerlingen met ontbrekende waarden, is een vaardigheidsschatting gemaakt op basis van de beschikbare items. De toets was daarmee per leerling zo lang als het aantal door de betreffende leerling gemaakte items. De analyses zijn vervolgens uitgevoerd met de overgebleven 1751 leerlingen uit 100 scholen en 155 klassen. De verdeling van deze leerlingen over het type school is weergegeven in Tabel 4.1.4.

Tabel 4.1.4: Overzicht van aantal deelnemende leerlingen per schooltype die deelnamen aan de luistertoets.

Onderwijstype	Aantal scholen	Aantal klassen	Aantal leerlingen
Basisonderwijs	63	88	1258
Speciaal basisonderwijs	24	47	373
Speciaal onderwijs	13	24	120

Het aantal (niet ontbrekende) observaties per item is weergegeven in Bijlage 1. In Figuur 4.1.2 worden deze visueel weergegeven.

Figuur4.1.2: Observatieaantallen per item.



In Figuur 4.1.2 is een duidelijk onderscheid zichtbaar tussen de items die in twee toetsversies voorkwamen en de items die in slechts één toetsversie voorkwamen. Het item met het laagste aantal valide responses heeft 348 responses. In peilingsonderzoek wordt 300 observaties vaak gezien als een ondergrens. Zodoende zijn in deze peiling geen items uitgesloten op basis van een laag observatie-aantal.

4.1.2 Spreekvaardigheid

Bij de naverteltaak en het betoog werd de spreekvaardigheid van de leerlingen op, respectievelijk, 15 en 16 verschillende criteria (items) gescoord. Ieder item werd gescoord op een driepuntsschaal (0 = *onvoldoende*, 1 = *voldoende*, 2 = *goed*) en zo konden voor de naverteltaak en het betoog dus maximaal respectievelijk 30 punten en 32 punten behaald worden.

Beoordelaars waren in de gelegenheid om bij de start van het beoordelen van een spreektaak aan te geven of de geproduceerde spraak te beoordelen was, i.e., of er werd voldaan aan de zogenaamde preconditie. Zowel bij de naverteltaak (*N* = 14) als bij het betoog (*N* = 65) kwam het voor dat beoordelaars aangaven dat de spraakproductie van de leerlingen niet toereikend was om beoordeeld te kunnen worden. Deze beoordelingen zijn niet meegenomen in verdere analyses. Net zoals in de vorige sbo-peiling zijn leerlingen niet meegenomen voor verdere analyses indien meer dan 20% van de data ontbrak. Zodoende zijn de resultaten van drie leerlingen ten aanzien van het betoog niet meegenomen. De resterende 2174 leerlingen voerden ofwel de naverteltaak, ofwel het betoog, ofwel beide taken uit (zie Tabel 4.1.5).

Tabel 4.1.5: Design spreektaken.

Groep	Naverteltaak	Betoog	*N* leerlingen	*N* items
A			915	15
B			388	31
C			871	16
	N = 1303	N = 1259		

De items binnen de taken zijn nagenoeg hetzelfde voor beide spreektaken. Bij het betoog wordt nog op een extra item gescoord, namelijk *argumenten*. Hierna worden de items opgesomd waarop wordt gescoord. De items zijn ingedeeld in zes subdomeinen.

1. Uitvoering
   a. Volledigheid
   b. Inhoudelijke kwaliteit
2. Samenhang
   a. Opbouw
   b. Signaalwoorden
3. Afstemming op doel
   a. Spreekdoel
   b. Argumenten (alleen bij betoog)
4. Afstemming op gesprekspartner
   a. Taalgebruik afstemmen
   b. Gesprekspartner boeien
   c. Non-verbale communicatie
5. Woordgebruik en woordenschat
   a. Woordgebruik
   b. Woordenschat
6. Vloeiendheid, verstaanbaarheid en grammaticale beheersing
   a. Vloeiendheid
   b. Uitspraak en intonatie
   c. Grammatica: vervoeging
   d. Grammatica: verbuiging
   e. Grammatica: zinsbouw

Leerlingen van de verschillende schooltypen (bao, sbo en so) werden random ingedeeld bij de verschillende taken. De verdeling spreektaken per schooltype wordt weergegeven in Tabel 4.1.6.

Tabel 4.1.6: Aantal leerlingen per spreektaak per schooltype.

Schooltype	Spreektaak	Aantal
Basisonderwijs	Naverteltaak	873
	Betoog	869
Speciaal basisonderwijs	Naverteltaak	312
	Betoog	287
Speciaal onderwijs	Naverteltaak	118
	Betoog	103

Van de 1303 leerlingen die de naverteltaak uitvoerden waren er zeven leerlingen die in totaal negen ontbrekende waarden (9/ (1303 * 15 scorepunten) = 0,046%) hadden op individuele items. Deze negen ontbrekende waarden zijn geïmputeerd met de minimale score (i.e., 0). Van de 1259 leerlingen die het betoog uitvoerden waren er tien leerlingen die in totaal 14 ontbrekende waarden (14/ (1259 * 16 scorepunten) = 0,069%) hadden op individuele items. Deze 14 ontbrekende waarden zijn eveneens geïmputeerd met de minimale score.

In Tabel 4.1.7 is weergegeven hoe de 2174 leerlingen uit 305 klassen van 212 scholen verdeeld zijn over de schooltypen.

Tabel 4.1.7: Overzicht van aantal deelnemende leerlingen per schooltype die deelnamen aan de naverteltaak en/of het betoog.

Schooltype	Aantal scholen	Aantal klassen	Aantal leerlingen
Basisonderwijs	130	167	1482
Speciaal basisonderwijs	55	97	508
Speciaal onderwijs	27	41	184

4.1.3 Gespreksvaardigheid

Bij zowel het samenwerkingsspel als het raadsel werd de gespreksvaardigheid van de leerlingen op 17 verschillende criteria (items) gescoord. Ieder item werd gescoord op een driepuntsschaal (0 = *onvoldoende*, 1 = *voldoende*, 2 = *goed*) en zo konden voor zowel het samenwerkingsspel als het raadsel dus maximaal 34 punten behaald worden.

Net als bij de taken voor spreekvaardigheid, zijn de resultaten niet meegenomen van leerlingen waarbij meer dan 20% van de data ontbrak. Voor het samenwerkingsspel was dit voor twee leerlingen het geval, bij het raadsel waren dit vijf leerlingen. De overgebleven 2147 leerlingen deden ofwel het samenwerkingsspel, ofwel het raadsel, ofwel beide taken (zie Tabel 4.1.8).

Tabel 4.1.8: Design gesprekstaken.

Groep	Samenwerkingsspel	Raadsel	N leerlingen	N items
A			1006	17
B			366	34
C			775	17
	N = 1372	N = 1141		

De items binnen de taken zijn hetzelfde bij beide taken. Hieronder worden de items opgesomd waarop wordt gescoord. De items zijn ingedeeld in zes subdomeinen:

1. Uitvoering
    a. Kwaliteit van de inhoud

2. Woord nemen en bijdragen aan samenhang
    a. Gesprek onderhouden
    b. Passendheid van het woord nemen
    c. Inbreng gesprekspartners integreren
3. Afstemming op doel
    a. Doelgerichtheid
    b. Begrijpelijk maken
4. Afstemming op gesprekspartners
    a. Taalgebruik afstemmen
    b. Non-verbale communicatie
    c. Actief luistergedrag
    d. Reageren op gesprekspartner
5. Woordgebruik en woordenschat
        a. Woordgebruik
        b. Woordenschat
6. Vloeiendheid, verstaanbaarheid, en grammaticale beheersing
    a. Vloeiendheid
    b. Uitspraak en intonatie
    c. Grammatica: vervoeging
    d. Grammatica: verbuiging
    e. Grammatica: zinsbouw

Leerlingen van de verschillende schooltypen (bao, sbo en so) werden random ingedeeld bij de verschillende taken. De verdeling gesprekstaken per schooltype wordt weergegeven in Tabel 4.1.9. Het raadsel is door minder leerlingen uitgevoerd dan het samenwerkingsspel. Dit kwam omdat het raadsel in kleiner groepsverband werd afgenomen. Bij het samenwerkingsspel namen drie à vier leerlingen deel, aan het raadsel namen uitsluitend drie leerlingen deel.

Tabel 4.1.9: Aantal leerlingen per gesprekstaak per schooltype.

Schooltype	Spreektaak	Aantal
Basisonderwijs	Samenwerkingsspel	933
	Raadsel	774
Speciaal basisonderwijs	Samenwerkingsspel	291
	Raadsel	241
Speciaal onderwijs	Samenwerkingsspel	148
	Raadsel	126

Van de 1372 leerlingen die het samenwerkingsspel deden, waren er bij 109 leerlingen ontbrekende waarden doordat het item niet was ingevuld, of doordat de beoordelaar had aangegeven dat het item niet te beoordelen was. Voor 32 leerlingen was er één item niet ingevuld (32 / (1372 * 17 scorepunten) = 0,137%). Het verschilde per leerling welk item ontbrak. Deze 32 ontbrekende waarden zijn geïmputeerd

met de minimale score (i.e., 0). Bij 77 leerlingen werd bij één of meer items aangegeven dat dit niet beoordeeld kon worden, dit werd dit in totaal 222 keer aangegeven (222 / (1372 * 17 scorepunten) = 0,951%). Ook hier werd de minimale score geïmputeerd.

Van de 1141 leerlingen die het raadsel deden, waren er voor 116 leerlingen ontbrekende waarden doordat het item niet was ingevuld, of doordat de beoordelaar had aangegeven dat het item niet te beoordelen was. Voor 21 leerlingen waren in totaal 24 ontbrekende waarden (24 / (1141 * 17 scorepunten) = 0,124%) op individuele items. Deze 24 ontbrekende waarden zijn geïmputeerd met de minimale score. Bij 95 leerlingen werd in totaal bij 489 items aangegeven dat het item niet te beoordelen was (489 / (1141 * 17 scorepunten) = 2,521%). Ook deze waarden zijn geïmputeerd met de minimale score.
In Tabel 4.1.10 is weergegeven hoe de 2147 leerlingen uit 301 klassen van 212 scholen verdeeld zijn over de schooltypen.

Tabel 4.1.10: Overzicht van aantal deelnemende leerlingen per schooltype die deelnamen aan het samenwerkingsspel en/of het raadsel.

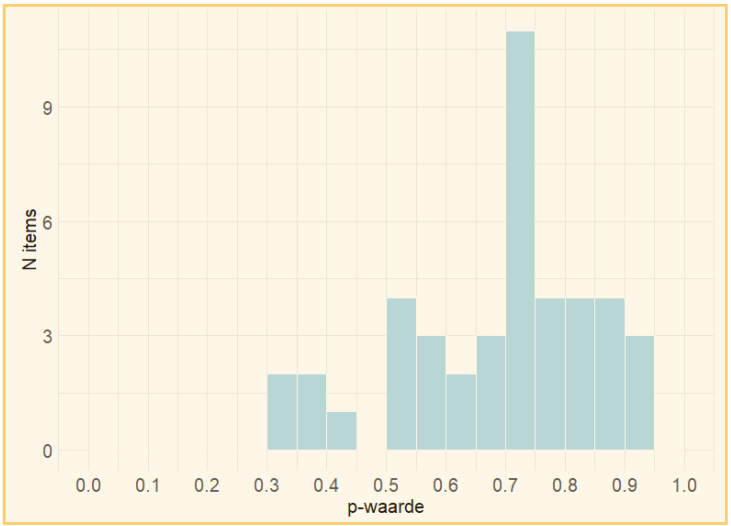
Schooltype	Aantal scholen	Aantal klassen	Aantal leerlingen
Basisonderwijs	129	166	1471
Speciaal basisonderwijs	55	90	446
Speciaal onderwijs	28	45	230

4.2 Klassieke analyses

4.2.1 Luistervaardigheid

Alle 43 items van de luistertoetsversies waren meerkeuze-items. Een item was goed of fout, en werd zodoende automatisch binair gescoord met, respectievelijk, 1 of 0. In Figuur 4.2.1 wordt de verdeling van p-waarden van de 43 items weergegeven.

Figuur 4.2.1: Verdeling p-waarden.



De gemiddelde p-waarde was 0,690 (*SD* = 0,161). De laagste p-waarde (0,342) werd gevonden voor een item uit fragment *Internet* uit toetsversie 1. De hoogste p-waarde (0,943) werd gevonden voor een item uit fragment *Goud* uit toetsversie 2.

Vervolgens is bij de verschillende toetsversies voor ieder item de item-restcorrelatie (rir) berekend om te zien of eventuele negatieve rir-waarden zouden kunnen duiden op sleutelfouten. Hiervoor zijn alleen de resultaten van de leerlingen gebruikt die geen ontbrekende waarden bevatten. De laagst gevonden rir-waarde bedroeg 0,031. De alternatief-restcorrelaties (rar) gaven geen aanleiding om een sleutelfout te vermoeden. Daarnaast zijn alle items na de pilot inhoudelijk gescreend op sleutelfouten.

In Tabel 4.2.1 worden tot slot nog de gemiddelde p-waarden per toetsversie weergegeven en in Tabel 4.2.2 wordt de gemiddelde p-waarde per fragment weergegeven. De p-waarden zijn niet onderling vergelijkbaar, aangezien de verschillende toetsversies gemaakt zijn door leerlingen van klassen van verschillende schooltypes. De p-waarden geven aan dat de moeilijkheid van de taken accuraat is ingeschat en daarmee passend zijn geweest voor het niveau van de leerling.

Tabel 4.2.1: Gemiddelde p-waarde per toetsversie.

Toetsversie	p-waarde
1	0,644
2	0,727
3	0,793
4	0,644

Tabel 4.2.2: Gemiddelde p-waarde per fragment.

Fragment	Inschatting moeilijkheid	p-waarde
Internet	2F	0,587
Julia	1F	0,589
Contant geld	2F	0,594
Ridders	1F	0,601
Bacteriën	2F	0,666
Pindakaas	1F/2F	0,708
Melkmeisje	2F	0,714
Goud	1F	0,755
Zebrastrepen	1F	0,787
Tekenares	1F	0,791
Bank	2F	0,811
Gooi	2F	0,818

De gemiddelde p-waarden in Tabel 4.2.2 laten zien dat de verschillende fragmenten en daarmee de verschillende toetsversies niet beslist te moeilijk zijn geweest. Sommige fragmenten zijn eerder wat aan de makkelijke kant geweest. Geen fragmenten waren dusdanig makkelijk of moeilijk dat het noodzakelijk werd geacht om deze te verwijderen voor analyse.

Op basis van deze initiële controle van klassieke psychometrische iteminformatie is geen item verwijderd voor verdere analyses, en zijn de analyses vervolgd met alle 43 items. De individuele p-waarden zijn te raadplegen in Bijlage 1.

4.2.2 Spreekvaardigheid

Naverteltaak

Zoals in de vorige paragraaf uiteengezet, namen 1303 leerlingen deel aan de naverteltaak. Alle 15 items uit de naverteltaak werden gescoord op een driepuntsschaal (0 = *onvoldoende*, 1 = *voldoende*, 2 = *goed*). In Tabel 4.2.3 worden de klassieke resultaten gepresenteerd van de naverteltaak.

Tabel 4.2.3: Klassieke itemanalyse – naverteltaak.

item_id	itembeschrijving	score (M)	score (SD)	p-waarde	rir
SP1_01	volledigheid	1,513	0,639	0,757	0,615
SP1_02	inhoudelijke kwaliteit	1,328	0,644	0,664	0,697
SP1_03	opbouw	1,208	0,643	0,604	0,581
SP1_04	signaalwoorden	1,736	0,530	0,868	0,634
SP1_05	spreekdoel	1,488	0,608	0,744	0,688
SP1_06	taalgebruik afstemmen	1,932	0,257	0,966	0,340
SP1_07	gesprekspartner boeien	0,843	0,640	0,421	0,388
SP1_08	non-verbale communicatie	1,002	0,469	0,501	0,285
SP1_09	woordgebruik	1,424	0,639	0,712	0,662
SP1_10	woordenschat	1,632	0,535	0,816	0,588
SP1_11	vloeiendheid	1,506	0,615	0,753	0,529
SP1_12	uitspraak en intonatie	1,822	0,406	0,911	0,335
SP1_13	grammatica: vervoeging	1,830	0,416	0,915	0,518
SP1_14	grammatica: verbuiging	1,845	0,398	0,922	0,462
SP1_15	grammatica: zinsbouw	1,705	0,522	0,852	0,660

In Tabel 4.2.3 is weergegeven dat op sommige items erg hoog werd gescoord (e.g., SP1_06), en dat op SP1_07, vergeleken met de andere items, lager werd gescoord. Dit patroon is vergelijkbaar met de items van de naverteltaak in de vorige so/sbo-peiling (Keuning et al., 2019). De gemiddelde p-waarde van de naverteltaak was 0,761. De items hebben allen een hoge rir-waarde, en de betrouwbaarheid van deze taak was hoog (Cronbach's alpha = 0,879).

Betoog

Zoals in de vorige paragraaf uiteengezet, namen 1259 leerlingen deel aan het betoog. Alle 16 items uit het betoog werden gescoord op een driepuntsschaal (0 = *onvoldoende*, 1 = *voldoende*, 2 = *goed*). In Tabel 4.2.4 worden de klassieke resultaten gepresenteerd van het betoog.

Tabel 4.2.4: Klassieke itemanalyse – betoog

item_id	itembeschrijving	score (M)	score (SD)	p-waarde	rir
SP2_01	volledigheid	1,592	0,613	0,796	0,586
SP2_02	inhoudelijke kwaliteit	1,310	0,650	0,655	0,700
SP2_03	opbouw	1,077	0,607	0,539	0,542
SP2_04	signaalwoorden	1,722	0,529	0,861	0,591
SP2_05	spreekdoel	1,232	0,664	0,616	0,705
SP2_06	argumenten	1,572	0,531	0,786	0,476
SP2_07	taalgebruik	1,890	0,337	0,945	0,398
SP2_08	gesprekspartner boeien	0,886	0,701	0,443	0,511
SP2_09	non-verbale communicatie	0,997	0,511	0,498	0,424
SP2_10	woordgebruik	1,421	0,639	0,710	0,685
SP2_11	woordenschat	1,733	0,479	0,867	0,534
SP2_12	vloeiendheid	1,554	0,587	0,777	0,442
SP2_13	uitspraak en intonatie	1,783	0,451	0,892	0,384
SP2_14	grammatica: vervoeging	1,843	0,400	0,921	0,436
SP2_15	grammatica: verbuiging	1,851	0,393	0,925	0,376
SP2_16	grammatica: zinsbouw	1,705	0,520	0,853	0,631

Net als bij de naverteltaak is er op sommige items erg hoog gescoord, en is op SP2_08 en SP2_09 relatief laag gescoord. De gemiddelde p-waarde van het betoog was 0,755, en is daarmee sterk vergelijkbaar met de naverteltaak. De items hebben alle een hoge rir-waarde, en de betrouwbaarheid van deze taak was – vergelijkbaar met de naverteltaak - hoog (Cronbach's alpha = 0,882).

Samenhang spreektaken

In totaal namen 388 leerlingen deel aan zowel de naverteltaak als het betoog. De correlatie tussen de resultaten op beide taken was 0,614. Deze correlatie is echter behept met meetfout, en wanneer gecorrigeerd hiervoor, is de latente correlatie 0,697. Daarmee lijkt de ene taak niet helemaal hetzelfde construct te meten als de andere taak. In peilingsonderzoek is het immers gebruikelijk om bij een latente correlatie vanaf 0,8-0,9 te concluderen dat twee deeldomeinen hetzelfde meten. In paragraaf 4.4 (Samenhang subdomeinen) wordt hier dieper op ingegaan.

4.2.3 Gespreksvaardigheid

Samenwerkingsspel

Zoals in de vorige paragraaf uiteengezet, namen 1372 leerlingen deel aan het samenwerkingsspel. Alle 17 items uit het samenwerkingsspel werden gescoord op een driepuntsschaal (0 = *onvoldoende*, 1 = *voldoende*, 2 = *goed*). In Tabel 4.2.5 worden de klassieke resultaten gepresenteerd van het samenwerkingsspel.

Tabel 4.2.5: Klassieke itemanalyse – samenwerkingsspel

item_id	itembeschrijving	score (M)	score (SD)	p-waarde	rir
GE1_01	kwaliteit van de inhoud	1,480	0,644	0,740	0,692
GE1_02	gesprek onderhouden	1,456	0,663	0,728	0,615
GE1_03	passendheid van het woord nemen	1,732	0,546	0,866	0,439
GE1_04	inbreng gesprekspartners integreren	1,670	0,602	0,835	0,589
GE1_05	doelgerichtheid	1,361	0,628	0,680	0,646
GE1_06	begrijpelijk maken	1,494	0,676	0,747	0,681
GE1_07	taalgebruik afstemmen	1,828	0,421	0,914	0,461
GE1_08	non-verbale communicatie	1,275	0,576	0,637	0,464
GE1_09	actief luistergedrag	1,618	0,585	0,809	0,587
GE1_10	reageren op gesprekspartner	1,754	0,519	0,877	0,457
GE1_11	woordgebruik	1,764	0,515	0,882	0,531
GE1_12	woordenschat	1,893	0,370	0,946	0,499
GE1_13	vloeiendheid	1,837	0,419	0,919	0,433
GE1_14	uitspraak en intonatie	1,872	0,362	0,936	0,327
GE1_15	grammatica: vervoeging	1,895	0,353	0,948	0,455
GE1_16	grammatica: verbuiging	1,917	0,313	0,958	0,422
GE1_17	grammatica: zinsbouw	1,829	0,429	0,915	0,559

Uit Tabel 4.2.5 kan worden afgelezen dat op de meeste items erg hoog werd gescoord en dat op GE1_08, vergeleken met de andere items, lager werd gescoord. Dit item over non-verbale communicatie werd ook het minst hoog gescoord bij de spreektaken. De gemiddelde p-waarde van het samenwerkingsspel was hoog: 0,843. De items hebben alle een hoge rir-waarde, en de betrouwbaarheid van deze taak was hoog (Cronbach's alpha = 0,884).

Raadsel

Aan het raadsel namen 1141 leerlingen deel. In Tabel 4.2.6 worden de klassieke resultaten gepresenteerd van het raadsel.

Tabel 4.2.6: Klassieke itemanalyse – raadsel

item_id	itembeschrijving	score (M)	score (SD)	p-waarde	rir
GE2_01	kwaliteit van de inhoud	1,414	0,678	0,707	0,672
GE2_02	gesprek onderhouden	1,479	0,684	0,740	0,625
GE2_03	passendheid van het woord nemen	1,721	0,546	0,861	0,534
GE2_04	inbreng gesprekspartners integreren	1,542	0,642	0,771	0,659
GE2_05	doelgerichtheid	1,397	0,664	0,699	0,637
GE2_06	begrijpelijk maken	1,521	0,664	0,761	0,690
GE2_07	taalgebruik afstemmen	1,801	0,462	0,901	0,597
GE2_08	non-verbale communicatie	1,075	0,541	0,538	0,416
GE2_09	actief luistergedrag	1,514	0,602	0,757	0,422
GE2_10	reageren op gesprekspartner	1,749	0,492	0,875	0,373
GE2_11	woordgebruik	1,736	0,547	0,868	0,677
GE2_12	woordenschat	1,833	0,477	0,916	0,648
GE2_13	vloeiendheid	1,772	0,500	0,886	0,573
GE2_14	uitspraak en intonatie	1,778	0,478	0,889	0,502
GE2_15	grammatica: vervoeging	1,779	0,515	0,890	0,575
GE2_16	grammatica: verbuiging	1,802	0,491	0,901	0,591
GE2_17	grammatica: zinsbouw	1,772	0,512	0,886	0,657

Uit de tabel kan worden afgelezen dat – net als bij het samenwerkingsspel - op de meeste items erg hoog werd gescoord en dat op GE2_08, wederom non-verbale communicatie, vergeleken met de andere items, lager werd gescoord. De gemiddelde p-waarde van het raadsel was hoog: 0,814. De items hebben alle een hoge rir-waarde, en de betrouwbaarheid van deze taak was hoog (Cronbach's alpha = 0,908).

Samenhang gesprekstaken

In totaal namen 366 leerlingen deel aan zowel het samenwerkingsspel als het raadsel. De correlatie tussen de resultaten op beide taken was 0,622. Deze correlatie is echter behept met meetfout, en wanneer gecorrigeerd hiervoor, is de latente correlatie 0,695. Daarmee lijkt de ene taak niet helemaal hetzelfde construct te meten als de andere taak. In peilingsonderzoek is het immers gebruikelijk om bij een latente correlatie vanaf 0,8-0,9 te concluderen dat twee deeldomeinen hetzelfde meten. In paragraaf 4.4 (Samenhang subdomeinen) wordt hier dieper op ingegaan.

Luisteren in een praktijksituatie

Het beoordelingsmodel van het samenwerkingsspel bevatte ook een aantal aspecten die betrekking hadden op luisteren in een praktijksituatie. Het ging om onderstaande beoordelingsaspecten die op een driepuntsschaal werden gescoord (0 = *onvoldoende*, 1 = *voldoende*, 2 = *goed*):

- Begrijpen
  - Instructie toepassen
  - Herhaling van instructie
- Samenvatten
  - Mondeling samenvatten
- Interpreteren

- Meningen herkennen

In Tabel 4.2.7 worden de klassieke resultaten gepresenteerd van het luisteren in een praktijksituatie.

Tabel 4.2.7: Klassieke itemanalyse – luisteren in een praktijksituatie.

item_id	itembeschrijving	score (M)	score (SD)	p-waarde	rir
LUp_01	instructie toepassen	1,771	0,489	0,885	0,563
LUp_02	herhaling van instructie	1,635	0,555	0,818	0,556
LUp_03	mondeling samenvatten	1,521	0,664	0,761	0,599
LUp_04	meningen herkennen	1,327	0,78	0,663	0,419

De p- en rir-waarden waren hoog. De betrouwbaarheid van deze kleine schaal was goed (Cronbach's alpha = 0,782).

4.3 Beoordelaarsonderzoek

De scoring van de spreek- en gesprekstaken is gebaseerd op het oordeel van beoordelaars (de toetsleiders). Om de consistentie tussen beoordelaars te evalueren, heeft een apart beoordelaarsonderzoek plaatsgevonden. Hiertoe is per spreektaak ongeveer 10% van de leerlingen door een tweede beoordelaar herbeoordeeld. De richtlijnen van Keuning en anderen (2019) zijn hiervoor gevolgd; dit betreft dezelfde procedure als bij de vorige peiling. Het design voor deze dubbele beoordelingen is weergegeven in Tabel 4.3.1.

Tabel 4.3.1: Design dubbele beoordelingen

			1e beoordelaar							
2e beoordelaar	leerling	niveau	1	2	3	4	5	6	..	*b*
1	1	p20	■							
2	2	p40	■							
3	3	p60	■							
4	4	p80	■							
2	5	p20		■						
1	6	p40		■						
4	7	p60		■						
3	8	p80		■						
4	9	p20			■					
3	10	p40			■					
1	11	p60			■					
2	12	p80			■					
3	13	p20				■				
4	14	p40				■				
2	15	p60				■				

1	16	p80								
…	…	…								
m	*j*	…								

Waarbij geldt:
- *b* = 1, … *B* 1e beoordelaars
- *m* = 2, … *M* 2e beoordelaars
- *j* = 1, … *J* leerlingen

Dit design is als volgt uitgevoerd. Vier beoordelaars werden uitgekozen om de herbeoordeling uit te voeren. Deze vier beoordelaars zijn geselecteerd op basis van een aantal criteria. Ten eerste moesten de beoordelaars in de hoofdmeting voldoende ervaring hebben opgedaan als toetsleider. Ten tweede is een inschatting gemaakt van de kwaliteit (inclusief nauwkeurigheid) van de toetsleiders. Een derde criterium was beschikbaarheid in de zomervakantie. In de toewijzing van de tweede beoordelaars aan taken, is ervoor gezorgd dat ze van alle eerste beoordelaars taken beoordeelden. Op deze manier is er een koppeling tussen alle eerste en tweede beoordelaars. De leerlingen van wie audiomateriaal beschikbaar was, werden opgedeeld in vier vaardigheidsgroepen (in het design aangegeven als p20, p40, p60 en p80). De vaardigheden zijn gebaseerd op de eerste beoordeling. Zoals in het design weergegeven, is er bij het toewijzen van de leerlingen aan de tweede beoordelaars zorg voor gedragen dat de leerlingen per tweede beoordelaar evenwichtig verdeeld zijn over de vier vaardigheidsgroepen.

In navolging van de richtlijnen van Keuning en anderen (2019) is voor iedere eerste beoordelaar de Mean Absolute Error (MAE_b) berekend:

$$MAE_b = \frac{\sum_j \sum_i \sum_m |s_{bji} - t_{jim}|}{N_j},$$

Waarbij geldt: *i* = item, *s* = score beoordelaar 1, *t* = score beoordelaar 2 *N* = aantal scorepunten per leerling. Voor iedere eerste beoordelaar is dus het volgende gedaan: van alle scorepunten die dubbel zijn beoordeeld wordt het absolute verschil genomen, en deze worden bij elkaar opgeteld. Deze somscore wordt vervolgens gedeeld door het aantal leerlingen dat door de tweede beoordelaar dubbel zijn beoordeeld. Dit resulteert in een gemiddelde afwijking voor iedere eerste beoordelaar. Vervolgens is de Median Absolute Deviation (*MAD*; Leys et al., 2013) berekend aan de hand van de *MAE*'s van alle beoordelaars.
Om optimaal rekening te kunnen houden met een mogelijke asymmetrische verdeling van *MAE* is de *MAD* van de mediaan gebaseerd op alle punten groter of gelijk aan de mediaan:

$$MAD = med(|\boldsymbol{Y} - med(\boldsymbol{MAE})|),$$

Waarin:

$$\boldsymbol{Y} = \{MAE_b \in \boldsymbol{MAE} : MAE_b \geq med(\boldsymbol{MAE})\}$$

Een afwijkende eerste beoordelaar is geïdentificeerd indien:

$$\frac{MAE_b - med(\boldsymbol{MAE})}{MAD} > 2{,}5,$$

(Leys et al., 2013).
Naast het evalueren van het gebrek aan overeenstemming zijn ook overall maten van interbeoordelaarsovereenstemming en -betrouwbaarheid berekend. Voor de interbeoordelaarsovereenstemming zijn de proportie overeenstemming en de Cohen's Kappa berekend en geïnterpreteerd (met inachtneming van de criteria van Landis en Koch (1977), zie Tabel 4.3.2).

Tabel 4.3.2: Criteria Landis en Koch interbeoordelaarsovereenstemming

	Mate van overeenstemming
<0,00	poor
0,00-0,20	slight
0,21-0,40	fair
0,41-0,60	moderate
0,61-0,80	substantial
0,81-1,00	almost perfect

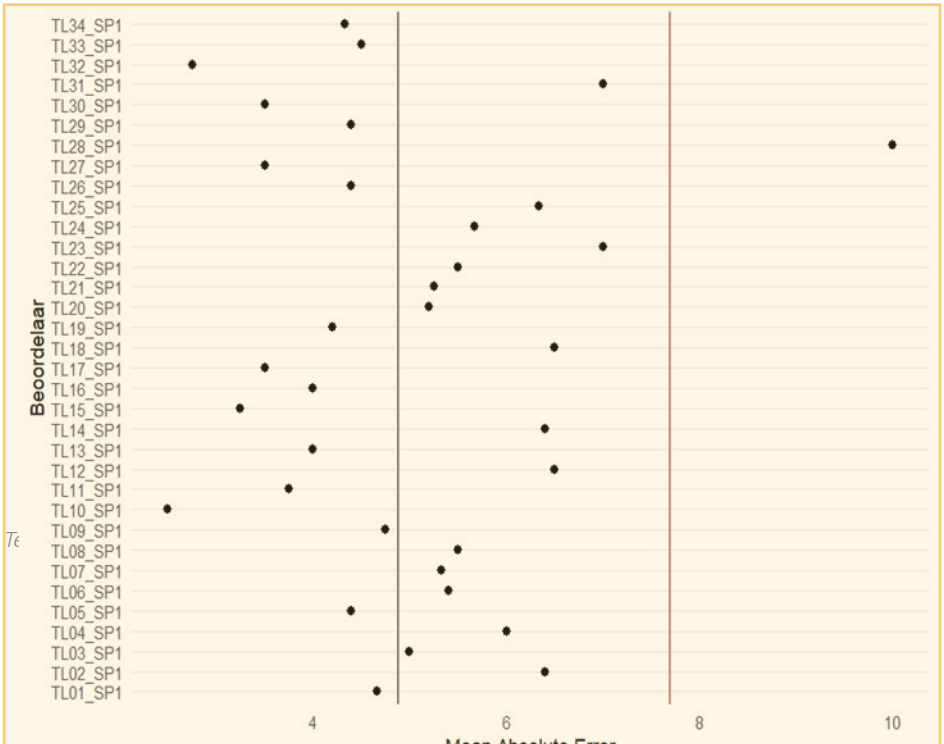
Voor de interbeoordelaarsbetrouwbaarheid is een Pearson-correlatie berekend op itemniveau.

4.3.1 Spreekvaardigheid

Naverteltaak

Bij de naverteltaak hadden in totaal 34 beoordelaars een eerste beoordeling gedaan. Van de 1303 leerlingen zijn 132 leerlingen voor een tweede maal beoordeeld door de vier tweede beoordelaars. Voor de tweede beoordeling is het item SP1_08 (non-verbale communicatie) niet beoordeeld, omdat non-verbale communicatie niet beoordeeld kan worden op basis van de audio-opname. In Figuur 4.3.1 is voor iedere eerste beoordelaar de MAE weergegeven. De zwarte lijn representeert de mediaan van de verzameling MAE's; de rode lijn is de grenswaarde voor het identificeren van afwijkende beoordelaars. In Figuur 4.3.1 is weergegeven dat één beoordelaar (TL28_SP1) geïdentificeerd kan worden als afwijkend.

Figuur 4.3.1: MAE per beoordelaar – naverteltaak



De betreffende beoordelaar heeft in totaal de spreekvaardigheid van 20 leerlingen beoordeeld op de naverteltaak. De resultaten van deze 20 leerlingen zijn voor verdere analyses buiten beschouwing gelaten. De interbeoordelaarsovereenstemmings- en betrouwbaarheidsmaten zijn weergegeven in Tabel 4.3.3.

Tabel 4.3.3: Overzicht van de interbeoordelaarsovereenstemming en interbeoordelaarsbetrouwbaarheid-maten op de naverteltaak.

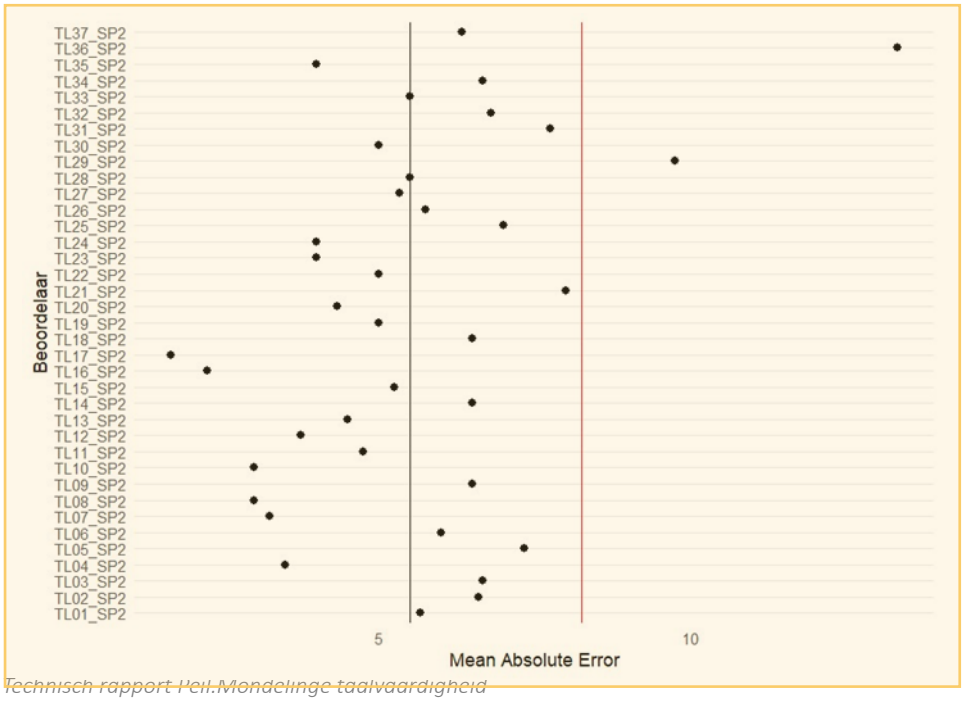
Maat	Waarde
Proportie overeenstemming	0,666
Gewogen Kappa	0,500
Correlatie op itemniveau	0,503
Correlatie op somscoreniveau per leerling	0,658

Volgens de criteria van Landis en Koch (1977) ten aanzien van de gewogen Kappa, is de interbeoordelaarsovereenstemming *moderate*, hetgeen kan worden opgevat als redelijk. Ondanks dat de correlatie op somscoreniveau van de leerling niet uitermate hoog is, wordt een betrouwbaarheid van 0,6 of hoger door COTAN beschouwd als voldoende betrouwbaar voor het doen van metingen op groepsniveau (Evers et al., 2009).

Betoog

Bij het betoog hebben in totaal 37 beoordelaars een eerste beoordeling gedaan. Van de 1259 leerlingen zijn 131 leerlingen voor een tweede maal beoordeeld door de vier tweede beoordelaars. Voor de tweede beoordeling is het item SP2_09 (non-verbale communicatie) niet beoordeeld, omdat non-verbale communicatie niet beoordeeld kan worden op basis van de audio opname. In Figuur 4.3.2 is voor iedere eerste beoordelaar de MAE weergegeven.

Figuur 4.3.2: MAE per beoordelaar – betoog



In Figuur 4.3.2 is weergegeven dat twee beoordelaars (TL36_SP2 en TL29_SP2) geïdentificeerd kunnen worden als afwijkend. De betreffende beoordelaars hebben samen in totaal 23 leerlingen beoordeeld op het betoog. De resultaten van deze 23 leerlingen zijn voor verdere analyses buiten beschouwing gelaten. Eén van deze twee beoordelaars was dezelfde afwijkende beoordelaar als bij de naverteltaak.
De interbeoordelaarsovereenstemmings- en interbeoordelaarsbetrouwbaarheidsmaten zijn hieronder weergegeven in Tabel 4.3.4.;

Tabel 4.3.4: Overzicht van de interbeoordelaarsovereenstemming en interbeoordelaarsbetrouwbaarheidsmaten op het betoog.

	Waarde
Proportie overeenstemming	0,663
Gewogen Kappa	0,472
Correlatie op itemniveau	0,479
Correlatie op somscoreniveau per leerling	0,635

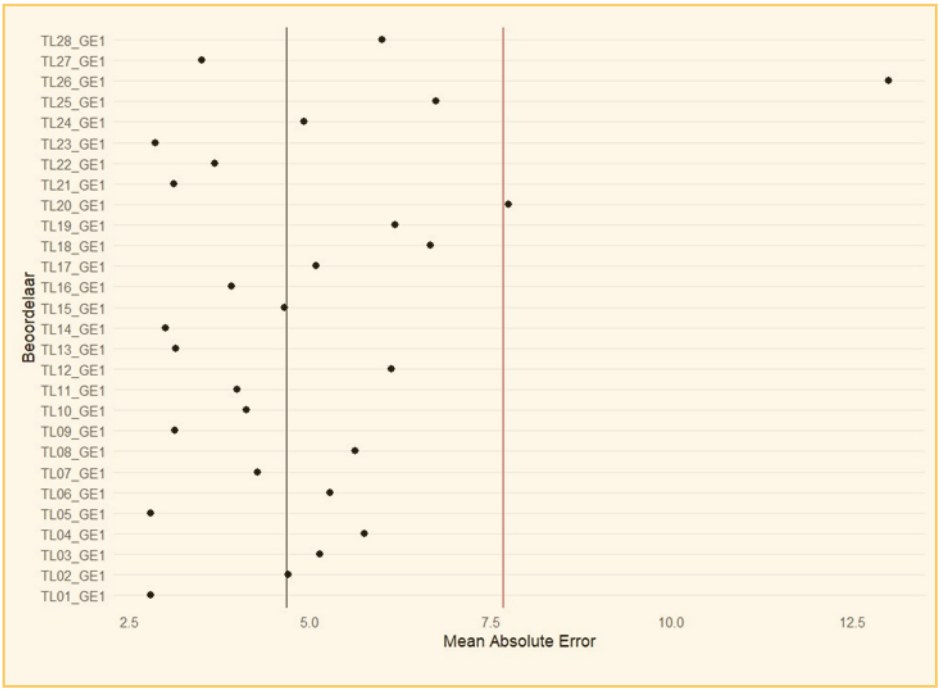
Volgens de criteria van Landis en Koch (1977) ten aanzien van de gewogen Kappa, is de interbeoordelaarsovereenstemming *moderate*, hetgeen kan worden opgevat als redelijk. Ondanks dat de correlatie op somscoreniveau van de leerling niet uitermate hoog is, wordt een betrouwbaarheid van 0,6 of hoger door COTAN beschouwd als voldoende betrouwbaar voor het doen van metingen op groepsniveau (Evers et al., 2009).

4.3.2 Gespreksvaardigheid

Samenwerkingsspel

Bij het samenwerkingsspel hadden in totaal 28 beoordelaars een eerste beoordeling gedaan. Van de 1372 leerlingen zijn 308 leerlingen voor een tweede maal beoordeeld door de vier tweede beoordelaars. In Figuur 4.3.3 is voor iedere eerste beoordelaar de MAE weergegeven. De zwarte lijn representeert de mediaan van de verzameling MAE's, de rode lijn is de grenswaarde voor het identificeren van afwijkende beoordelaars. In Figuur 4.3.3 is zichtbaar dat één beoordelaar (TL26_GE1) stevig afwijkt, en dat één beoordelaar (TL20_GE1) net de grenswaarde overschrijdt. Voor TL26_GE1 is besloten om de door deze beoordelaar beoordeelde leerlingen (N=8) buiten beschouwing te laten in de verdere analyses. Voor TL20_GE1 is besloten om de beoordelingen van de toetsleider op te nemen in de analyses. Het doel van het onderzoek is om afwijkende beoordelaars te identificeren. Van TL20_GE1 kan dat niet stellig worden beweerd.

Figuur 4.3.3: MAE per beoordelaar – samenwerkingsspel



De interbeoordelaarsovereenstemmings- en interbeoordelaarsbetrouwbaarheidsmaten zijn weergegeven in Tabel 4.3.5.

Tabel 4.3.5: Overzicht van de interbeoordelaarsovereenstemming en interbeoordelaarsbetrouwbaarheidsmaten op het samenwerkingsspel.

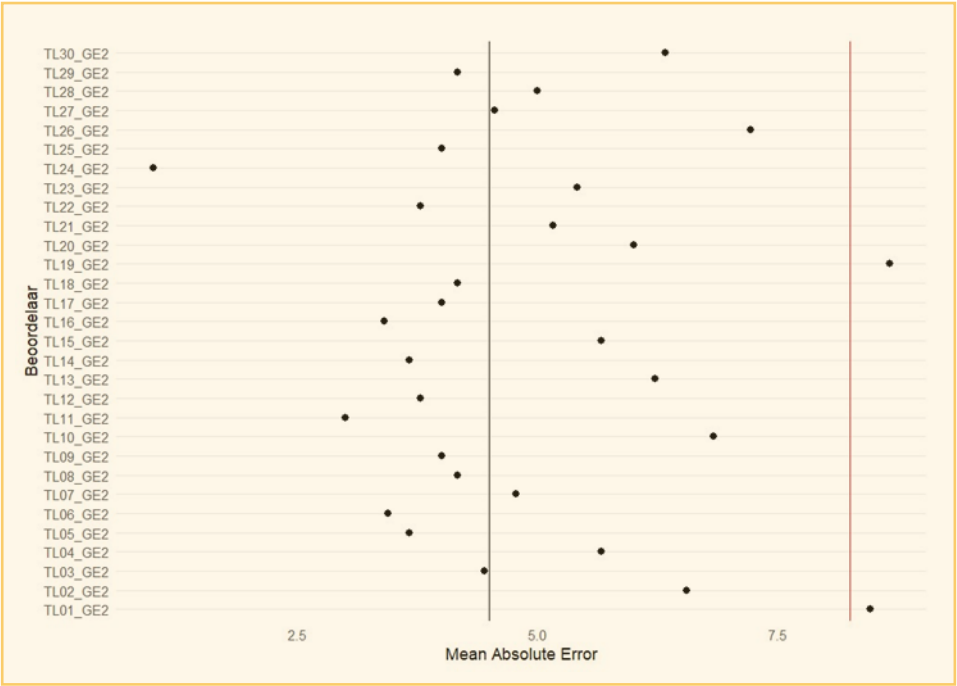
Maat	Waarde
Proportie overeenstemming	0,758
Gewogen Kappa	0,383
Correlatie op itemniveau	0,384
Correlatie op somscoreniveau per leerling	0,538

In vergelijking met de spreektaken valt op dat de proportie overeenstemming hoger is bij het samenwerkingsspel, maar dat tegelijkertijd de Kappa en de Pearson correlatie lager zijn. Dit kan verklaard worden doordat de score range bij het samenwerkingsspel beperkter is dan bij de spreektaken.

Raadsel

Bij het raadsel hebben in totaal 30 beoordelaars een eerste beoordeling gedaan. Van de 1141 leerlingen zijn 264 leerlingen voor een tweede maal beoordeeld door de vier tweede beoordelaars. In Figuur 4.3.4 is voor iedere eerste beoordelaar de MAE weergegeven. In Figuur 4.3.4 is weergegeven dat twee beoordelaars kunnen worden geïdentificeerd als afwijkend. Twee beoordelaars overschrijden net de grenswaarde. In lijn met de overwegingen die zijn genoemd bij het samenwerkingsspel, is besloten om de door deze twee beoordelaars beoordeelde leerlingen mee te nemen in de verdere analyses.

Figuur 4.3.4: MAE per beoordelaar – raadsel



De interbeoordelaarsovereenstemmings- en interbeoordelaarsbetrouwbaarheidsmaten zijn hieronder weergegeven in Tabel 4.3.6.

Tabel 4.3.6: Overzicht van de interbeoordelaarsovereenstemming en interbeoordelaarsbetrouwbaarheidmaten op het raadsel.

Maat	Waarde
Proportie overeenstemming	0,735
Gewogen Kappa	0,365
Correlatie op itemniveau	0,370
Correlatie op somscoreniveau per leerling	0,515

Deze maten laten een sterke gelijkenis zien met de resultaten van het samenwerkingsspel.

4.4 Samenhang subdomeinen

De domeinen van de mondelinge taalvaardigheid (i.e., luistervaardigheid, spreekvaardigheid en gespreksvaardigheid) zijn alle onder te verdelen in kleinere subdomeinen. In deze paragraaf onderzoeken we de samenhang tussen deze subdomeinen. Hiervoor wordt itemresponstheorie (IRT; zie paragraaf 4.5) gebruikt, en wordt bepaald of de subdomeinen verschillende vaardigheden (multidimensionaliteit) meten of één onderliggende vaardigheid. Om de dimensionaliteit te evalueren zijn latente correlaties berekend.

Daarvoor is – per domein - een afzonderlijke IRT-schaal geschat voor elk van de betreffende subdomeinen waarna *plausible values* zijn getrokken met een multivariaat normaalverdeelde prior. De latente correlaties zijn verkregen uit de updates van de covariantiematrix van de prior.

4.4.1 Luistervaardigheid

De 43 items zijn op vier manieren onderverdeeld in verschillende subdomeinen, namelijk op basis van kenmerken van de taakuitvoering, aard van de activiteit, gespreksvorm en referentieniveau (zie Tabel 2.2.1).

- Kenmerken van de taakuitvoering
  - Begrijpen (*N* = 19)
  - Interpreteren (*N* = 23)
  - Samenvatten (*N* = 1)
- Aard van de activiteit
  - Kijken en luisteren (*N* = 29)
  - Kritisch kijken en luisteren (*N* = 2)
  - Luisteren (*N* = 9)
  - Kritisch luisteren (*N* = 3)
- Gespreksvorm:
  - Monoloog (*N* = 28)
  - Dialoog (*N* = 15)
- Referentieniveau
  - 1F (*N* = 18)
  - 1F/2F (*N* = 5)
  - 2F (*N* = 20)

In Tabel 4.4.1 worden de correlaties weergegeven voor de verschillende 'subdomeinen' van luistervaardigheid. Voor het subdomein samenvatten is geen losse schaal geschat, aangezien deze slechts uit één item bestond. Kritisch luisteren en kritisch kijken en luisteren zijn samengenomen tot een kleine schaal.

Tabel 4.4.1: Latente correlaties

subdomein	vs. subdomein	latente correlatie
begrijpen	interpreteren	0,944
kijken & luisteren	luisteren	0,976
kijken & luisteren	kritisch (kijken en) luisteren	0,904
luisteren	kritisch (kijken en) luisteren	0,911
monoloog	dialoog	0,942
1F	1F/2F	0,963
1F/2F	2F	0,966
1F	2F	0,947

De samenhang tussen de subdomeinen is zeer sterk. Daarom is besloten om een unidimensionele schaal te gebruiken, en dus geen specifieke vaardigheden te schatten voor de afzonderlijke subdomeinen. Dit betekent dat er één luistervaardigheidsschaal is geschat. Tot slot is geëvalueerd of de items van Luisteren in een praktijksituatie (onderdeel van het samenwerkingsspel, zie paragraaf 4.2) samenhang vertoonden met de items uit de luistertoetsen. De latente correlatie was 0,406. Duidelijk is dat de items niet dezelfde constructen meten. In de vorige so/sbo-peiling was deze latente correlatie ook te laag om de items van luisteren in een praktijksituatie samen te nemen met de items van de luistertoetsen. De items van luisteren in een praktijksituatie worden daarom in deze peiling niet meegenomen in verdere analyses.

4.4.2 Spreekvaardigheid

De in totaal 31 items van de twee spreektaken samen zijn te verdelen onder de volgende kenmerken van de taakuitvoering:

- Uitvoering (*N* = 4)
- Samenhang (*N* = 4)
- Afstemming op doel (*N* = 3)
- Afstemming op publiek (*N* = 6)
- Woordgebruik en woordenschat (*N* = 4)
- Vloeiendheid, verstaanbaarheid en grammaticale beheersing (*N* = 10)

De latente correlaties tussen deze subdomeinen worden weergegeven in Tabel 4.4.2.

Tabel 4.4.2 Latente correlaties subdomeinen spreekvaardigheid.

	1	2	3	4	5	6
1 Uitvoering						
2 Samenhang	0,951					
3 Afstemming op doel	0,991	0,973				
4 Afstemming op publiek	0,633	0,756	0,829			
5 Woordgebruik	0,898	0,938	0,975	0,671		
6 Vloeiendheid	0,676	0,808	0,708	0,459	0,923	

In Tabel 4.4.2 is te zien dat de latente correlaties grotendeels hoog zijn (> 0,8). Sommige correlaties zijn niet overtuigend. Hier dient de nuance toegevoegd te worden dat ieder subdomein uit slechts enkele items bestaat.

In de vorige so/sbo-peiling zijn subdomeinen *uitvoering, samenhang, afstemming op doel* en *afstemming op gesprekspartner* samengenomen tot een betekenisvol subdomein, i.e., de *inhoudelijke aspecten*. De subdomeinen *woordgebruik* en *vloeiendheid* waren samengenomen en werden geduid als de *technische aspecten*. Tussen deze twee subdomeinen is de latente correlatie berekend en deze was 0,790. Tot slot is de latente correlatie berekend tussen de naverteltaak en het betoog, deze was hier 0,691.

Aan de hand van de berekende latente correlaties is het lastig te verdedigen dat hier sprake is van een duidelijk unidimensioneel construct spreekvaardigheid. In de vorige so/sbo-peiling werd deze conclusie eveneens getrokken. Desondanks werd in de vorige peiling besloten om een enkel construct te definiëren en is een standaardbepaling gedaan op het domein spreekvaardigheid als geheel. Deze lijn is in de huidige peiling ook gevolgd. Spreekvaardigheid is dus beschouwd als een enkel construct, en de referentieniveaus zijn overgezet van de naverteltaak op het betoog (zie paragraaf 4.6).

4.4.3 Gespreksvaardigheid

De in totaal 34 items van de twee gesprekstaken gecombineerd zijn te verdelen onder de kenmerken van de taakuitvoering:

- Uitvoering (*N* = 2)
- Samenhang (*N* = 6)
- Afstemming op doel (*N* = 4)
- Afstemming op publiek (*N* = 8)
- Woordgebruik en woordenschat (*N* = 4)
- Vloeiendheid, verstaanbaarheid en grammaticale beheersing (*N* = 10)

De latente correlaties tussen deze subdomeinen worden weergegeven in Tabel 4.4.3.

Tabel 4.4.3: Latente correlaties subdomeinen gespreksvaardigheid.

	1	2	3	4	5	6
1 Uitvoering						
2 Samenhang	0,982					
3 Afstemming op doel	0,988	0,990				
4 Afstemming op publiek	0,879	0,899	0,807			
5 Woordgebruik	0,879	0,782	0,804	0,556		
6 Vloeiendheid	0,814	0,698	0,682	0,562	0,881	

De meeste correlaties waren hoog (> 0,8). Sommige vielen wat lager uit, e.g., een aantal correlaties met afstemming op gesprekspartner. In de vorige so/sbo-peiling uit 2018 zijn subdomeinen *uitvoering, samenhang, afstemming op doel* en *afstemming op gesprekspartner* samengenomen tot een betekenisvol subdomein, i.e., de *inhoudelijke aspecten*. De subdomeinen *woordgebruik* en *vloeiendheid* waren samengenomen en werden geduid als de *technische aspecten*. Tussen deze twee subdomeinen is de latente correlatie berekend en deze was 0,715. Tot slot is de latente correlatie berekend tussen het samenwerkingsspel en het raadsel. De latente correlatie was hier 0,751.

Ondanks dat de correlaties niet voldoende lijken om te concluderen dat gespreksvaardigheid een unidimensioneel construct is, is dezelfde lijn gevolgd als bij spreekvaardigheid. Gespreksvaardigheid is dus

beschouwd als een enkel construct, en de referentieniveaus zijn overgezet van het samenwerkingsspel op het raadsel (zie paragraaf 4.6).

4.4.4 Samenhang hoofddomeinen

Na het evalueren van de samenhang binnen de subdomeinen van luistervaardigheid, spreekvaardigheid en gespreksvaardigheid, is er aanvullend gekeken of deze drie domeinen dusdanig samenhangen met elkaar dat zij samengenomen kunnen worden tot een overkoepelende mondelinge taalvaardigheidsschaal. De latente correlaties zijn weergegeven in Tabel 4.4.4.

Tabel 4.4.4: Latente correlaties tussen de domeinen van mondelinge taalvaardigheid.

	luisteren	spreken	gesprekken
luisteren			
spreken	0,381		
gesprekken	0,420	0,460	

De drie domeinen hangen in vergelijkbare mate met elkaar samen, echter zijn ze niet hoog genoeg om de drie domeinen te kunnen samenvoegen tot een enkel domein. In peilingsonderzoek is het gebruikelijk vanaf een latente correlaties van 0,8 à 0,9 te concluderen dat twee subdomeinen hetzelfde construct meten. In gevallen zoals hierboven omschreven, waarbij latente correlaties zijn gevonden van ongeveer 0,7, worden de deeldomeinen alsnog samengevoegd, omdat er inhoudelijk reden toe was en dan is dit ook duidelijk vermeld. De correlaties zoals gevonden in bovenstaande tabel zijn sterk lager dan 0,7-0,9, de rest van de analyses zal daarom apart voor ieder domein worden uitgevoerd en beschreven.

4.5 Schaling en DIF

4.5.1 Luistervaardigheid

Zoals reeds in paragraaf 4.1 is beschreven, hebben leerlingen verschillende toetsversies gemaakt, waardoor de resultaten van leerlingen niet direct met elkaar vergelijkbaar zijn. Itemresponstheorie (IRT; zie bijvoorbeeld Hambleton et al., 1991) is gebruikt om de items en leerlingen op een enkele onderliggende meetschaal te plaatsen. IRT heeft een aantal voordelen ten opzichte van klassieke testtheorie, waaronder dat bij IRT de toetsscore van een persoon onafhankelijk is van de set items die hij of zij heeft gemaakt. Daarnaast zijn itemkenmerken zoals moeilijkheid (en onderscheidend vermogen) onafhankelijk van de groep personen die de items hebben gemaakt. Hierdoor zijn de itemkenmerken van een item gelijk, ongeacht voor welke steekproef van personen deze berekend zijn.

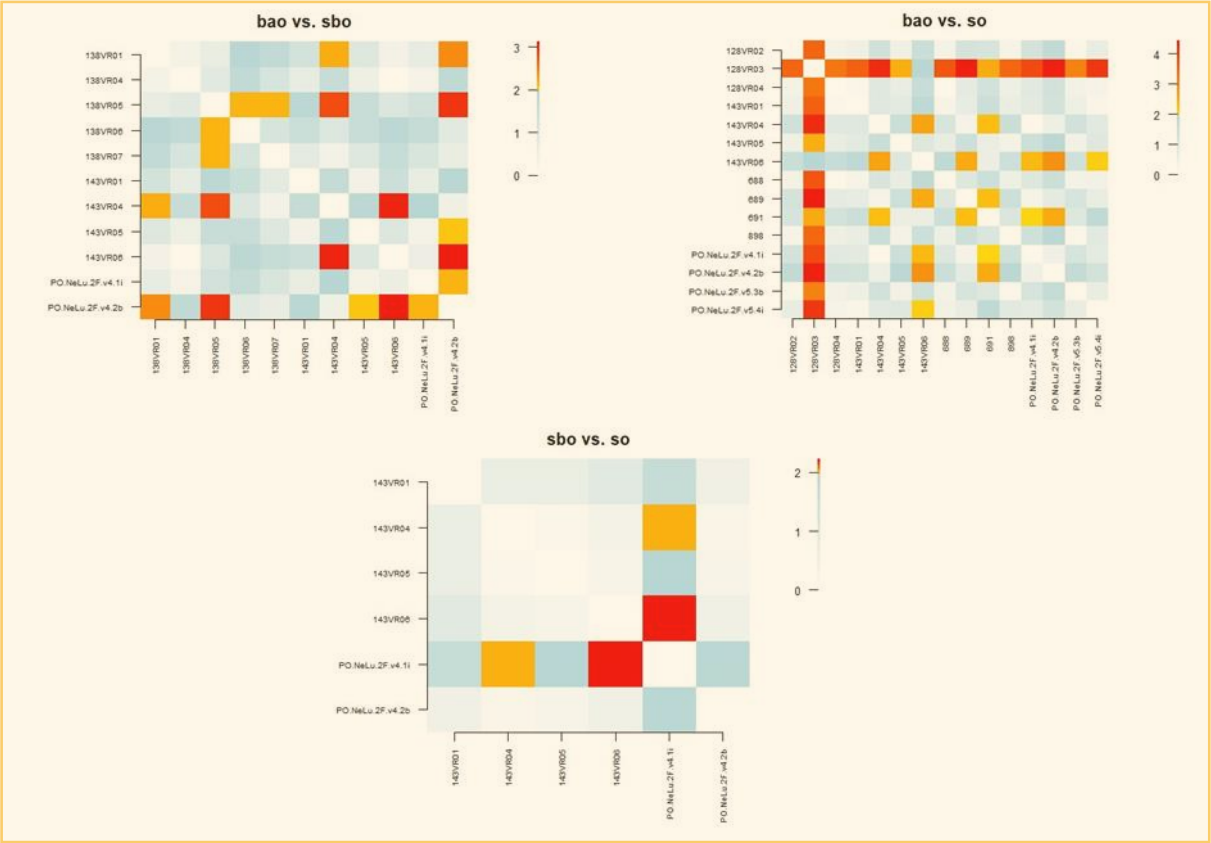
Met behulp van het Dexter (Maris et al., 2024) in R (R Core Team, 2023) zijn een aantal IRT-modellen gepast op de data om de modelpassing te evalueren. De 43 items zijn als binair gescoorde items meegenomen in de schaling.

Tweemaal is een Rasch-model geschat: eenmaal met Conditional Maximum Likelihood (CML) en eenmaal met Marginal Maximum Likelihood (MML) - waarbij het schooltype als marginaal is meegenomen. Bij het Rasch-model wordt enkel een moeilijkheidsparameter geschat per item. Daarnaast zijn het One Parameter Logistic Model (OPLM; Verhelst & Glas, 1995) en een 2PL-model gepast op de data. Bij deze twee IRT-modellen wordt – naast een moeilijkheidsparameter – eveneens een discriminatieparameter geschat. Door het schatten van de discriminatieparameter is het mogelijk om items die goed discrimineren zwaarder te laten wegen in het model. Na de modelschatting zijn de vaardigheden geschat met Weighted Maximum Likelihood (WML; Warm, 1989).

Voorafgaand aan de verschillende modelschattingen is eerst geëvalueerd of er mogelijk sprake is van *differential item functioning* (DIF) tussen de verschillende schooltypes (bao, sbo en so). Hierbij is een Rasch-model geschat en wordt beschouwd of items zich verschillend gedragen tussen de schooltypes. Als DIF op een item gevonden zou worden, dan zou bijvoorbeeld een leerling uit het so een hogere kans hebben een item goed te maken dan een leerling met een gelijke vaardigheid uit het sbo. Dit is niet de bedoeling. Voor dit onderzoek wordt *differential item pair functioning* (Bechger & Maris, 2014) beschouwd. Hierbij is een Rasch-model apart geschat in twee groepen. Vervolgens is het verschil in de relatieve moeilijkheid van alle paren van items vergeleken tussen groepen. De reden om een paarsgewijze vergelijking te gebruiken is dat de moeilijkheid van een item niet wordt geïdentificeerd op grond van de observaties, maar de verschillen in moeilijkheden (relatieve moeilijkheden) wel. De methode bestaat uit twee delen. Ze begint met een algehele statistische toets waarbij de twee matrices van gestandaardiseerde verschillen in relatieve moeilijkheden worden vergeleken. De gekwadrateerde Mahalanobis-afstand volgt een Chi-kwadraatverdeling met het aantal items minus 1 vrijheidsgraden. Deze toets geeft vrijwel altijd een significant resultaat wanneer de groepen worden gebaseerd op een wezenlijk verschil in achtergrondvariabelen. De statistische toets op zichzelf is daarom minder interessant. Des te interessant is het of, gegeven dat er vrijwel altijd DIF bestaat, het gekozen psychometrische model desondanks bruikbaar is voor het doel waarvoor het ingezet wordt. Hierbij helpt het om te kijken naar het patroon van de afwijkingen tussen itemparen. Aan de hand van dit patroon kunnen soms problematische items worden geïdentificeerd of kunnen groepen items worden herkend waarbinnen zeer weinig DIF optreedt.

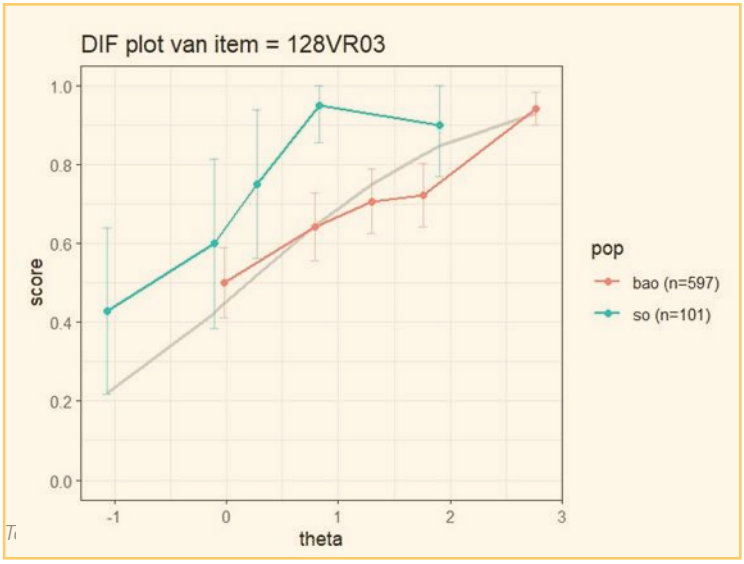
Als resultaat van de DIF-toets wordt telkens enkel de grafische weergave van de verschillen tussen itemparen gepresenteerd, deze zijn weergegeven in Figuur 4.5.1. Deze weergave is gebaseerd op absolute gestandaardiseerde verschillen. De kleurovergang tussen wit/blauw en geel/rood vindt plaats bij een absolute z-score van 1.96, oftewel een significantieniveau van 5%. De lengte van de kleurschaal verschilt echter per grafiek. In alle gevallen waar een grafiek wordt weergegeven is de DIF tussen de groepen significant bij een significantieniveau van 0,001.

Figuur 4.5.1: Differential item pair functioning



In Figuur 4.5.1 is in de rechterbovenplot op te merken dat het item 128VR03 (fragment *Zebrastrepen*) rood kleurt in combinatie met ongeveer alle andere items. Dit geeft aan dat dit item – in combinatie met nagenoeg ieder ander item – anders lijkt te functioneren bij de leerlingen in het bao dan bij de leerlingen in het so. Een dergelijke observatie geeft aanleiding om het item nader te bekijken. Bij het inspecteren van DIF-plot van het betreffende item (Figuur 4.5.2) wordt duidelijk dat de so-leerlingen – gegeven eenzelfde vaardigheid als bao-leerlingen - een hogere kans hadden om het item correct te beantwoorden.

Figuur 4.5.2 DIF-itemplot

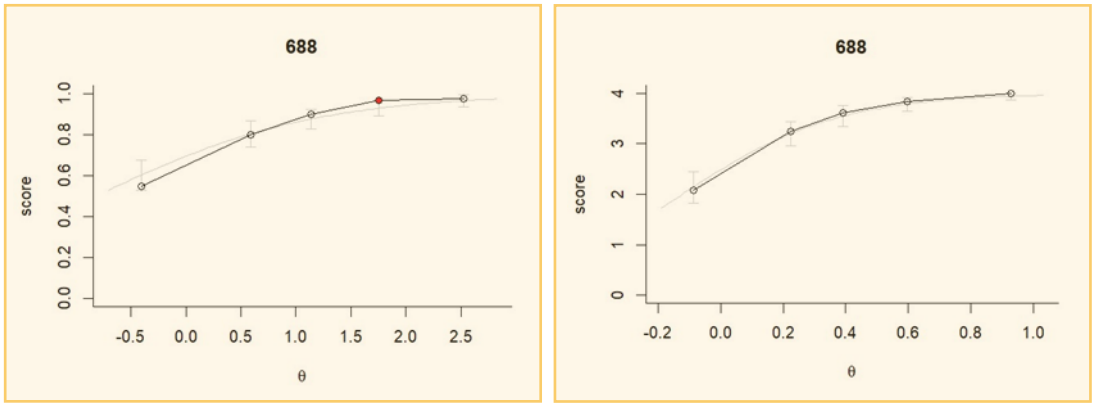


Een verklaring hiervoor is niet gevonden. Het was het op een na laatste item in toetsversie 2 en 3, dus in eerste instantie werd een verklaring gezocht bij het verschil in de lengte van beide toetsversies. Echter bleek dat er geen sprake was van DIF bij item 128VR04, het laatste item in beide toetsversies, en uit hetzelfde fragment als het DIF-item. Uiteindelijk is besloten om het item los te koppelen, i.e., het item te beschouwen als twee verschillende items: de bao-variant en de so-variant. Zodoende vervolgde de schaling met 44 items.

Bij de schaling lieten de twee Rasch-modellen een goede passing zien. De correlatie tussen de moeilijkheidsparameters van het Rasch-model met CML en de moeilijkheidsparameters van het Rasch-model met MML was hoger dan 0,999.

Bij het inspecteren van de itemkarakteristieke curves (*item characteristic curves* (ICC's); Figuur 4.5.3) werd geconcludeerd dat het Rasch-model goed past bij de data. Bij sommige items ontstond de indruk dat de items met een andere discriminatieparameter net wat beter zouden passen, zoals in onderstaand voorbeeld. In de linkerplot is te zien dat de geobserveerde scores (de zwarte lijn) bijna niet te onderscheiden zijn van de posterior verwachte scores (de grijze lijn). Desalniettemin zou de itemfit bevorderd kunnen worden met een hogere discriminatieparameter, zoals, e.g., in het OPLM is toegestaan (rechterplot). Door de hogere discriminatieparameter wordt de grijze modelcurve steiler, en ligt deze meer tegen de zwarte geobserveerde curve aan.

Figuur 4.5.3: Itemfit item 688 Rasch (links) en OPLM (rechts)



Voor het evalueren van de modelpassing zijn - in aanvulling op de visuele inspectie - een aantal statistische toetsen uitgevoerd die bepalen of een model voor een betreffende schaal een adequate beschrijving geeft van de data. Belangrijk zijn de zogenaamde itemgeoriënteerde S-toets en de R1c-toets voor het gehele model. De S-toets is gebaseerd op de verschillen tussen de geobserveerde en verwachte proporties antwoorden in homogene scoregroepen. Als de p-waarden (overschrijdingskansen) voor de S-toetsen uniform verdeeld zijn in het interval [0,1] kan geconcludeerd worden dat het model past (Verhelst, Glas & Verstralen, 1993). De R1c-toets voor het gehele model heeft dezelfde onderliggende rationale als de S-toets, waar een niet-significante R1c-toets een goede passing weergeeft. De waarde van de R1c-toets is echter zeer gevoelig voor afwijkingen van het model, waardoor het ook bij relatief weinig observaties vaak

al snel aangeeft dat het IRT-model niet past bij de data. Daarom wordt als effectmaat vaak de Phi gebruikt, die de wortel van de fractie R1c/vrijheidsgraden aanduidt. Een Phi < 1,225 geeft een indicatie van een goede modelpassing.

In Tabel 4.5.1 zijn de resultaten van de modelfitschattingen weergegeven. Kolommen P01, P05, P10, %<.01, %<.05 en %<.1 hebben betrekking op de S-statistiek. P01, P05 en P10 geven respectievelijk aan bij hoeveel items de p-waarde van de S-statistiek lager was dan 0,01, tussen 0,01 en 0,05 en tussen 0,05 en 0,10. Vervolgens wordt gekeken hoeveel procent van de items per schaal een misfit met een p<.01 (%p<.01), p<.05 (%p<.05) of p<.10 (%p<.10) heeft.

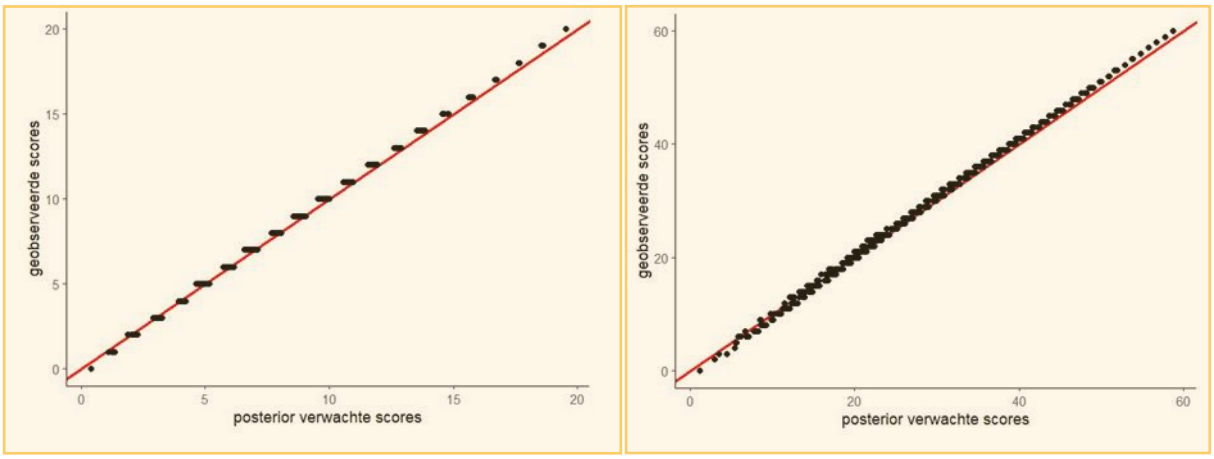
Tabel 4.5.1: Modelfit

model	N	P01	P05	P10	%<.01	%<.05	%<.1	R1c	df	p	Phi
Rasch (CML)	44	5	9	5	11,4	31,8	43,2	434	270	<0,001	1,27
Rasch (MML)	44	5	8	6	11,4	29,5	43,2	436	270	<0,001	1,27
OPLM	44	2	4	7	4,5	13,6	29,5	351	270	<0,001	1,14
2PL	44	3	3	7	6,8	13,6	29,5	345	270	<0,001	1,13

De modelpassing van de twee modellen met vrij geschatte discriminatieparameters laten een relatief betere modelpassing zien dan de Rasch-modellen. Het is hier vooral de vraag in welke mate deze modellen beter passen dan de Rasch-modellen.

Om de accuraatheid van het Rasch-model te vergelijken met het OPLM, is een robuustheidsanalyse gedaan. Hierbij is voor iedere leerling de somscore van alle door de leerling gemaakte items afgezet tegen de verwachte somscore voor deze leerling op deze betreffende items – gegeven dan wel het Rasch-model, dan wel het OPLM. De leerlingscores die op grond van de modelparameters worden verwacht worden dus vergeleken met de geobserveerde waarden. Onder de verwachte leerlingscore worden hier de *IRT True Score* gebaseerd op de item- en persoonsparameters (zie e.g., Hambleton, Swaminathan & Rogers, 1991) verstaan. Het evalueren van de robuustheid van het IRT-model is ook op een soortgelijke manier uitgevoerd in de so/sbo-peiling in 2018. De resultaten worden gevisualiseerd in Figuur 4.5.4.

Figuur 4.5.4: Geobserveerde vs. geschatte somscores (links Rasch; rechts OPLM).



Uit bovenstaande plots wordt helder dat beide modellen de scores van de leerlingen uitermate goed 'terugschatten'. De gemiddelde absolute afwijking tussen geobserveerde en geschatte score is 0,191 scorepunt (*SD* = 0,124) bij het Rasch-model en de gemiddelde absolute afwijking bij het OPLM, na correctie voor score-range, is 0,24 scorepunt (*SD* = 0,129). Hiervan uitgaand lijkt het weinig verschil te maken of nu voor een model wordt gekozen waarbij enkel een moeilijkheidsparameter wordt geschat of een model waarbij eveneens een discriminatieparameter wordt geschat.

Op basis van alle genoemde modellen die zijn gepast op de data, zijn WML's geschat. De samenhang tussen deze WML's worden in onderstaande correlatiematrix weergegeven.

Tabel 4.5.2: Correlatiematrix WML-schattingen IRT-modellen.

	Rasch (CML)	Rasch (MML)	OPLM	2PL
Rasch(CML)				
Rasch(MML)	1,000			
OPLM	0,985	0,985		
2PL	1,000	1,000	0,985	

Ook hieruit volgt dat de modelkeuze weinig verschil lijkt te maken ten aanzien van de vaardigheidsschatting.
Tot slot is stilgestaan bij de betrouwbaarheid van de vaardigheidsschattingen aan de hand van de verschillende modellen. In de peiling worden conclusies over de vaardigheid gebaseerd op het IRT-model. De voor een peiling relevante betrouwbaarheid is daarom de betrouwbaarheid van de vaardigheidsschatting. Deze betrouwbaarheid is gedefinieerd als de proportie ware variantie in de populatie als onderdeel van de variantie in de schatting. Deze kan worden bepaald door de correlatie te nemen tussen uitwisselbare replicaties van de schatting van de vaardigheid (Bechger, et al., 2003). Hiervoor wordt de gemiddelde correlatie tussen een set van 100 plausible values genomen. De verschillen in betrouwbaarheid van de vaardigheidsschatting tussen de verschillende modellen lijken erg klein (Tabel 4.5.3).

Tabel 4.5.3: Betrouwbaarheid van de vaardigheidsschatting.

	Betrouwbaarheid
Rasch (CML)	0,705
Rasch (MML)	0,711
OPLM	0,719
2PL	0,722

Concluderend, de modellen laten alle een goede passing zien, ze schatten de geobserveerde scores goed terug en de betrouwbaarheid van de vaardigheidsschatting is vergelijkbaar. Zodoende kan worden gesteld dat elk van de genoemde modellen gekozen zouden kunnen worden. Uiteindelijk is besloten om het Rasch-

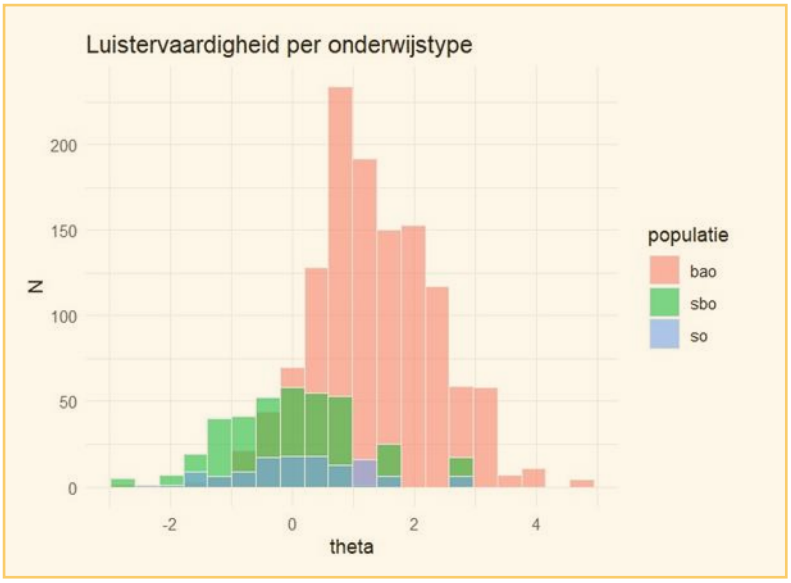
model (met CML) te gebruiken. Hier is voor gekozen omdat de noodzakelijke aannames bij gebruik van een CML-procedure zeer minimaal zijn. Daarnaast is het voor het gebruik van een 2PL van belang om een correcte specificatie van de achtergrondvariabelen te geven. Verder wordt formeel een random steekproef en een gekende vorm van de populatieverdeling vereist. De aanname van een normaal verdeelde populatie per niveau van de achtergrondvariabele lijkt weliswaar zeer redelijk en een MML-procedure is over het algemeen robuust voor kleine schendingen van normaliteit, mits de achtergrondvariabelen correct gespecificeerd zijn. Dit is echter geen zekerheid en de betreffende aannames zijn niet te toetsen.

In Tabel 4.5.4 wordt de verdeling van de WML-schattingen (gemiddelden en standaarddeviaties) weergegeven per schooltype. In Figuur 4.5.5 wordt dit visueel weergegeven.

Tabel 4.5.4: WML-verdeling per onderwijstype

Schooltype	WML (M)	WML (SD)
bao	1,332	1,002
sbo	-0,068	1,067
so	0,189	1,095

Figuur 4.5.5: Vaardigheidsverdeling per schooltype



Om te evalueren of de onderwijstypes significant verschillen in vaardigheid, is een ANOVA uitgevoerd. Bij een groot aantal deelnemers in een ANOVA wordt bij het kleinste verschil vaak al als significant geduid. Om enige nuance aan te brengen is de *-eta-squared* (η^2) berekend, die robuust is tegen grote groepsgrootten. Een door Cohen (1988) voorgestelde vuistregel om de η^2 te interpreteren is:

- Klein effect: 0,01
- Middelgroot effect: 0,06
- Groot effect: > 0,14

De schooltypes verschillen significant in vaardigheid, en het effect kan worden beschouwd als groot ($F(2,1748) = 306,1$; $p < 0,001$; $\eta^2 = 0,26$).

Tot slot zijn in Tabel 4.5.5 de paarsgewijze verschillen in vaardigheid tussen de schooltypes geduid aan de hand van Cohen's d (Cohen, 1988). De vuistregels worden hieronder opgesomd.

- Klein effect: 0,2
- Middelgroot effect: 0,5
- Groot effect: > 0,8

Tabel 4.5.5: Cohen's d voor paarsgewijze vergelijkingen

	bao	sbo	so
bao			
sbo	1,38		
so	1,13	0,24	

Concluderend, er werd DIF gevonden bij een item tussen bao en so. Dit item is losgekoppeld en is beschouwd als twee verschillende items. Op basis van de resultaten kan gesteld worden dat de bao-leerlingen een hogere luistervaardigheid hebben dan de leerlingen van het sbo en het so. De verschillen in vaardigheid tussen bao vs. sbo en bao vs. so zijn te omschrijven als groot. Daarnaast is een klein verschil gevonden in luistervaardigheid tussen de leerlingen van het sbo en so, waarbij de leerlingen in het so net wat beter hebben gescoord dan de leerlingen in het sbo.

4.5.2 Spreekvaardigheid

Zoals weergegeven in Tabel 4.1.4 hebben niet alle leerlingen alle spreektaken gedaan. Zo zijn er leerlingen die enkel de naverteltaak deden, leerlingen die enkel het betoog deden en leerlingen die beide taken uitvoerden. De resultaten van deze verschillende groepen zijn daardoor niet zonder meer vergelijkbaar. Net als bij luistervaardigheid is gebruik gemaakt van IRT. Voordat modellen geschat konden worden diende er nog een aanpassing in de scoring plaats te vinden. In de vorige so/sbo-peiling werden de technische aspecten minder zwaar gewogen dan de inhoudelijke aspecten bij de naverteltaak. Hier werd bij de totaalscore voor een leerling uitgegaan van een verhouding van 20% technische aspecten, en 80% inhoudelijke aspecten. Er waren echter zeven technische items en acht inhoudelijke items. Om tot een 20%/80 te komen werden de individuele scores op de technische items vermenigvuldigd met 0,429 en werden de individuele scores op de inhoudelijke items vermenigvuldigd met 1,500. Hierdoor kon een leerling maximaal 6 scorepunten (7 items * 2 scorepunten * 0,429) halen voor de technische aspecten, en maximaal 24 scorepunten (8 items * 2 scorepunten * 1,500) voor de inhoudelijke aspecten. Door deze weging ontstond een 20%/80%-verhouding van, respectievelijk, technische en inhoudelijke aspecten.

In de huidige peiling heeft iets soortgelijks plaatsgevonden. Doordat – anders dan in de vorige so/sbo-peiling – IRT is gebruikt voor de schaling, dienen de individuele scores op de items discrete waarden te zijn. Daartoe is een weging zoals hierboven beschreven niet toepasbaar in de huidige peiling. Door de individuele scores van de leerlingen op de inhoudelijke items te vermenigvuldigen met drie, ontstaat een

verhouding inhoudelijk/technisch die zeer dicht in de buurt komt van 20%/80% en is een discrete scoring behouden, waardoor het mogelijk blijft om er een IRT-model op te passen. De ongewogen en gewogen scoring voor de naverteltaak en het betoog worden weergegeven in Tabel 4.5.6. De technische items behouden gewicht 1, de inhoudelijke items krijgen gewicht 3.

Tabel 4.5.6. – Transformatie scoring spreektaken

		N items	scoring	verhouding	nieuwe scoring	nieuwe verhouding
naverteltaak	inhoudelijk	8	0-1-2	53,3%	0-3-6	77,4%
	technisch	7	0-1-2	46,7%	0-1-2	22,6%
betoog	inhoudelijk	9	0-1-2	56,25%	0-3-6	79,4%
	technisch	7	0-1-2	43,75%	0-1-2	20,6%

Na transformatie van de scores, is een One Parameter Logistic Model (OPLM; Verhelst & Glas, 1995) gepast op de data. Anders dan gebruikelijk bij een OPLM stond op voorhand dus al vast welke items welke discriminatieparameter zouden krijgen. Na modelschatting zijn de vaardigheden geschat met Weighted Maximum Likelihood (WML; Warm, 1989).

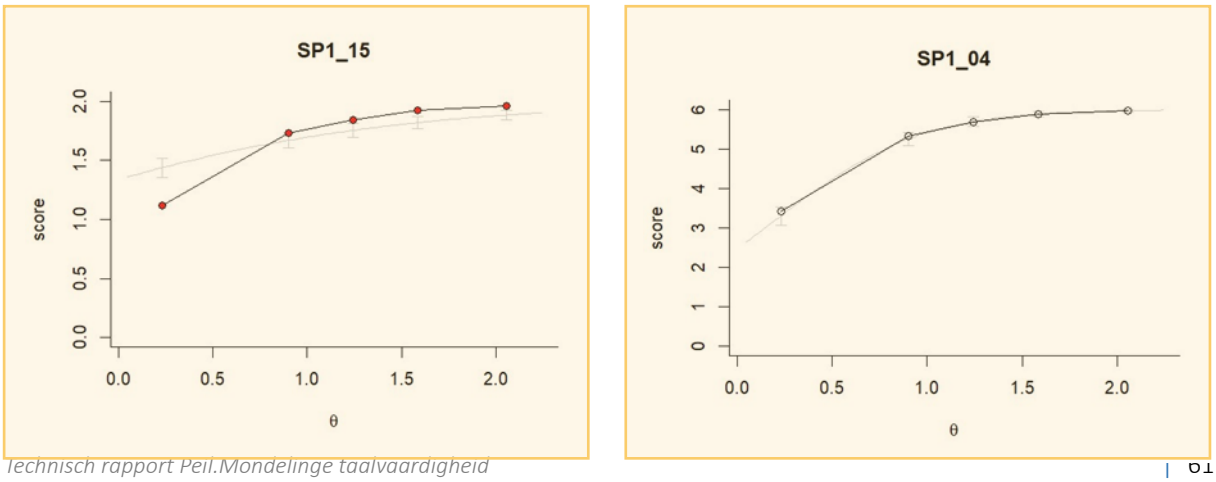
In Tabel 4.5.7 wordt de passing van het OPLM gepresenteerd. De passing van het model is niet overtuigend. De Phi is met 1,65 hoger dan de vaak in gebruik genomen grenswaarde van 1,225, maar laat geen extreem beeld zien. Doordat de discriminatieparameters op voorhand al waren bepaald, is enige misfit in de lijn der verwachting.

Tabel 4.5.7: Modelpassing OPLM

Model	N	P01	P05	P10	%<.01	%<.05	%<.1	R1c	df	p	Phi
OPLM	31	12	4	3	38,7	51,6	61,3	564	206	<0,001	1,65

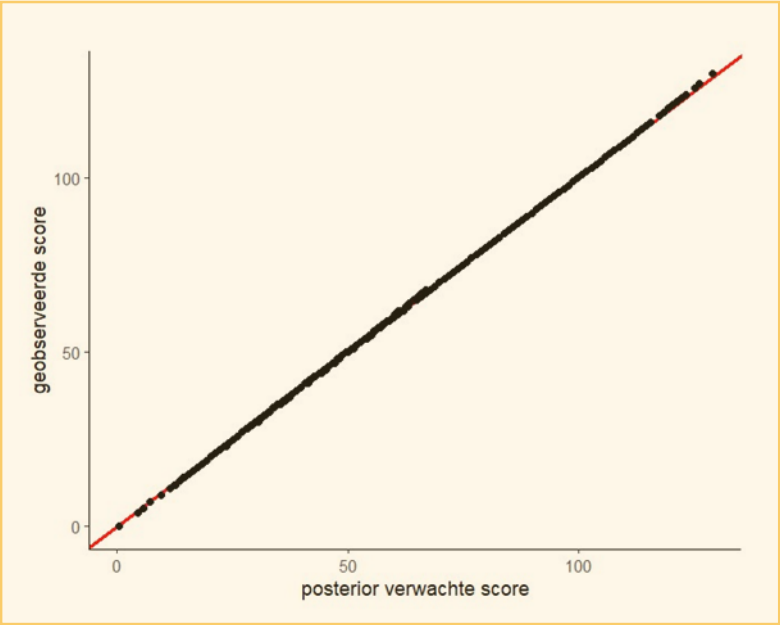
In Figuur 4.5.6 worden twee fit plots van individuele items gepresenteerd. De linker plot laat een misfit zien, i.e., de modelcurve (grijze curve) heeft een ander verloop dan de geobserveerde (zwarte) curve. Te zien is dat het item beter had gepast indien een discriminatieparameter vrij was geschat voor dit item. Bij item SP1_04 in de rechterplot kan bijvoorbeeld geconstateerd worden dat het item erg goed past.

Figuur 4.5.6: Misfit (links); goede fit (rechts)



Om de accuraatheid van het OPLM te evalueren, is een robuustheidsanalyse gedaan. Hierbij is voor iedere leerling de somscore van alle door de leerling gemaakte items afgezet tegen de verwachte somscore voor deze leerling op deze betreffende items – gegeven het OPLM. De resultaten worden gevisualiseerd in Figuur 4.5.7.

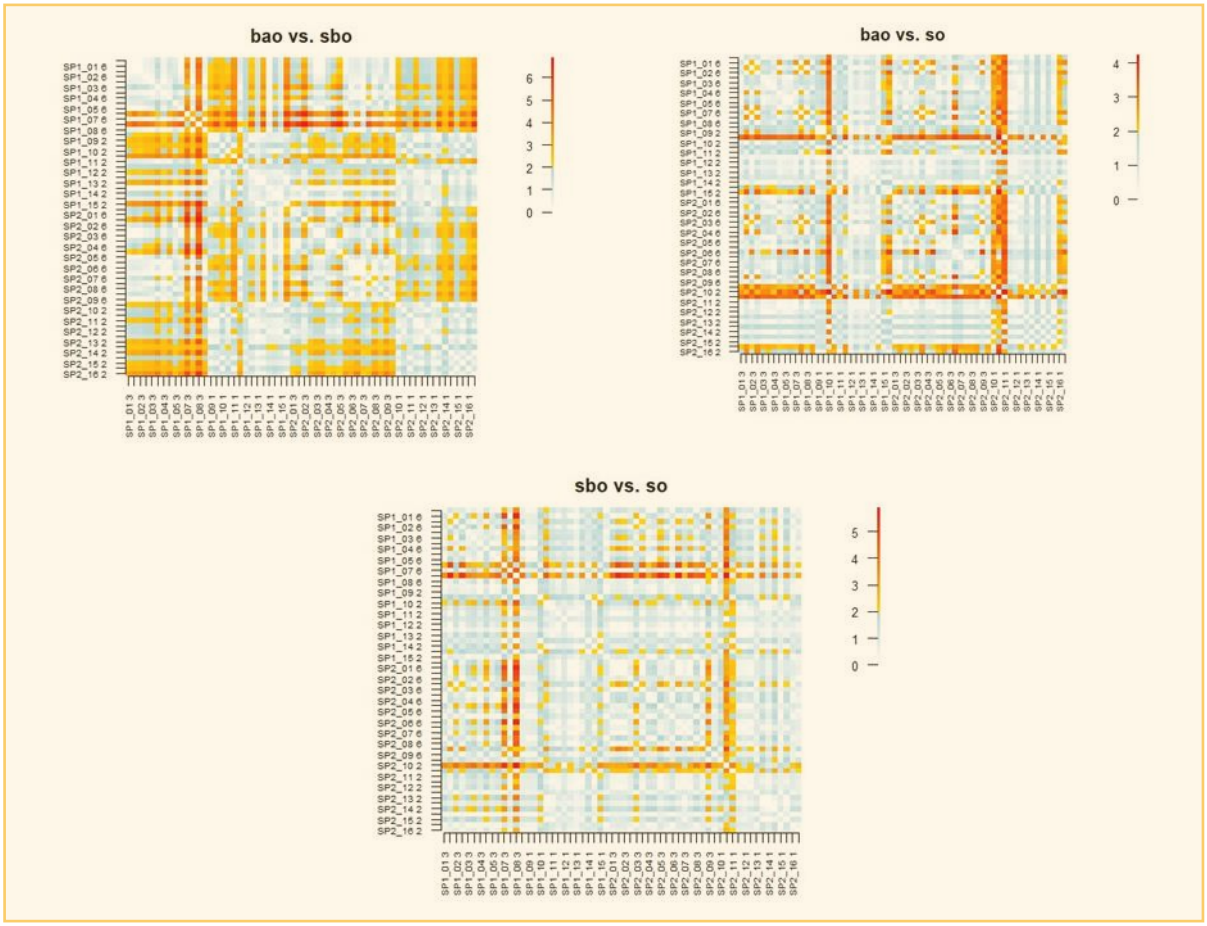
Figuur 4.5.7: Geobserveerde vs. geschatte somscores OPLM



Uit Figuur 4.5.7 wordt duidelijk dat het model de geobserveerde scores bijzonder accuraat 'terugschat'. De gemiddelde absolute afwijking tussen de geobserveerde score en verwachte score per persoon is 0,307 scorepunt (*SD* = 0,201). Concluderend, er is met vertrouwen besloten om verder te analyseren met dit model.

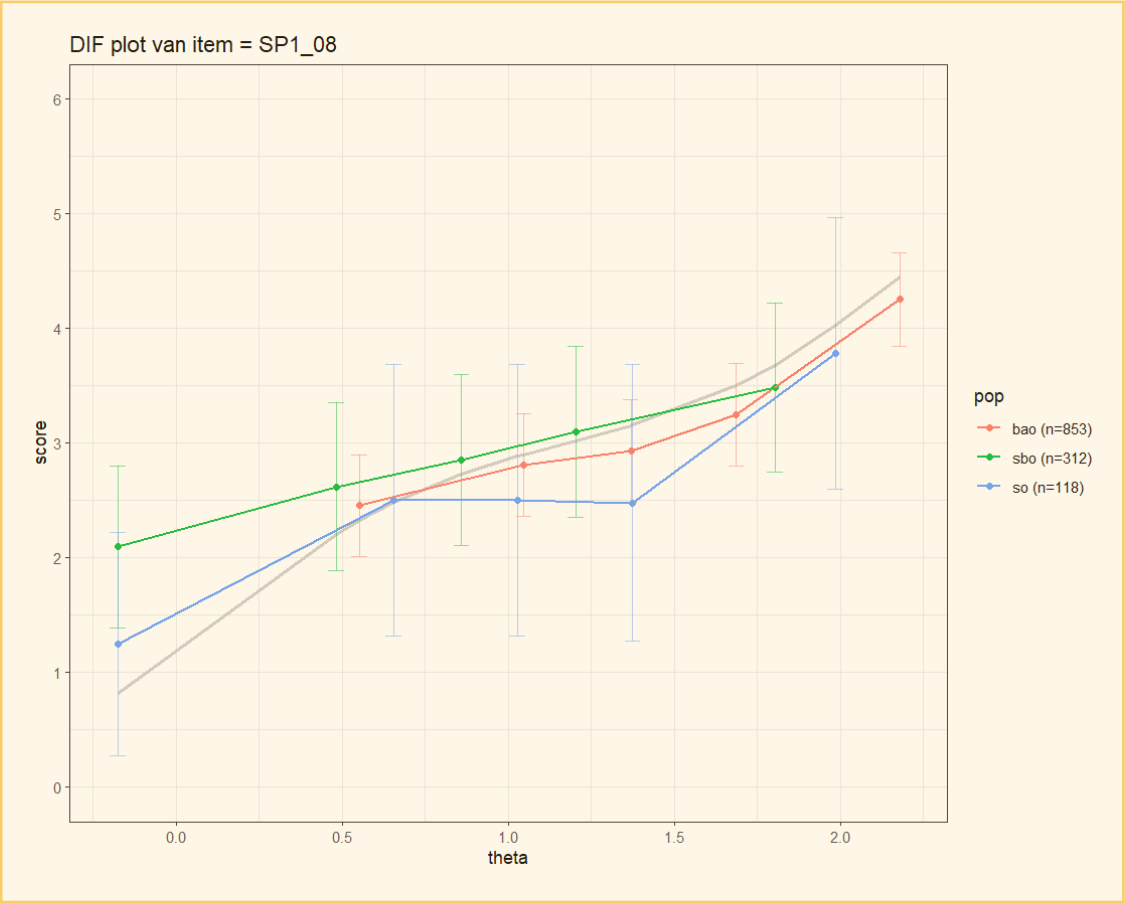
In aanvulling op de evaluatie van de passing van het model, is - net als bij luistervaardigheid - ook gekeken naar DIF tussen de onderwijstypes (bao, sbo, so). De differential item pair functioning (DIF) wordt visueel weergegeven in Figuur 4.5.8. Er leek enige DIF te zitten tussen bao en sbo. Op enkele items na, leek er weinig DIF te zijn tussen bao en so, evenals tussen sbo en so.

Figuur 4.5.8: Differential item pair functioning



Eén van de items waarbij de meeste DIF werd geïdentificeerd is weergegeven in Figuur 4.5.9. Bij dit item werd DIF gevonden tussen bao en sbo enerzijds en tussen sbo en so anderzijds. Echter, deze plot beschouwende, lijkt de DIF niet extreem te zijn. Er is met name sprake van algemene misfit. Uiteindelijk is besloten om alle items mee te nemen in de verdere analyses.

Figuur 4.5.9: Item aangemerkt als DIF-item



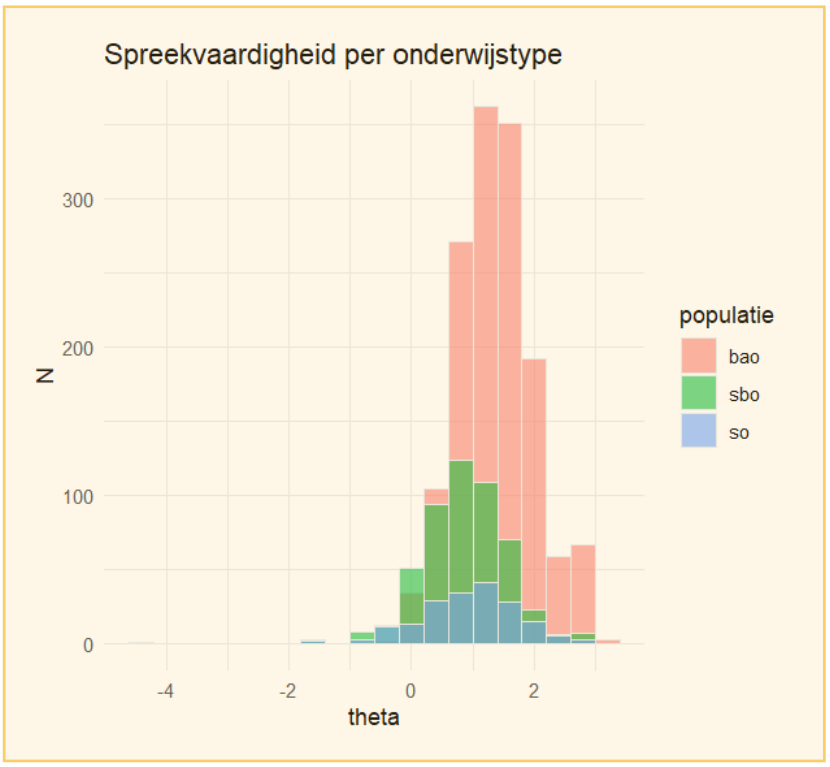
Tot slot is stilgestaan bij de betrouwbaarheid van de vaardigheidsschattingen. In de peiling worden conclusies over de vaardigheid gebaseerd op het IRT-model. De voor een peiling relevante betrouwbaarheid is daarom de betrouwbaarheid van de vaardigheidsschatting. Deze betrouwbaarheid is gedefinieerd als de proportie ware variantie in de populatie als onderdeel van de variantie in de schatting. Deze kan worden bepaald door de correlatie te nemen tussen uitwisselbare replicaties van de schatting van de vaardigheid (Bechger, et al., 2003). Hiervoor wordt de gemiddelde correlatie tussen een set van 100 plausible values genomen. De betrouwbaarheid van 0,869 wordt beschouwd als hoog.

In Tabel 4.5.8 wordt de verdeling van de WML-schattingen (gemiddelden en standaarddeviaties) weergegeven per schooltype. In Figuur 4.5.10 wordt dit visueel weergegeven.

Tabel 4.5.8: WML-verdeling per onderwijstype

Schooltype	WML (M)	WML (SD)
bao	1,375	0,627
sbo	0,867	0,723
so	0,941	0,783

Figuur 4.5.10: Verdeling spreekvaardigheid per onderwijstype



Om te evalueren of de onderwijstypes significant verschillen in vaardigheid, is een ANOVA uitgevoerd. Echter, bij een groot aantal deelnemers in een ANOVA wordt bij het kleinste verschil vaak al als significant geduid. Om enige nuance aan te brengen is de *-eta-squared* (η^2) gebruikt, die robuust is tegen grote groepsgrootten. Een door Cohen (1988) voorgestelde vuistregel om de η^2 te interpreteren is:

- Klein effect: 0,01
- Middelgroot effect: 0,06
- Groot effect: > 0,14

De schooltypes verschillen significant in vaardigheid, en het effect kan worden beschouwd als middelgroot tot groot ($F(2,2135) = 127$; $p < 0{,}001$; $\eta^2 = 0{,}11$). De spreekvaardigheid van de bao-leerlingen is hoger dan die van de sbo- en so-leerlingen.

Tot slot zijn in Tabel 4.5.9 de paarsgewijze verschillen in vaardigheid tussen de schooltypes geduid aan de hand van Cohen's d (Cohen, 1988). De vuistregels worden hieronder opgesomd.

– Klein effect: 0,2
– Middelgroot effect: 0,5
– Groot effect: > 0,8

Tabel 4.5.9: Cohen's d voor paarsgewijze vergelijkingen

	bao	sbo	so
bao			
sbo	0,78		

so	0,67	0,10	

Concluderend, er leek wel enige sprake van DIF te zijn tussen bao en sbo, maar deze bleek verwaarloosbaar te zijn. Op basis van de resultaten kan gesteld worden dat de bao-leerlingen een hogere spreekvaardigheid hebben dan de leerlingen van het sbo en het so. Het verschil in vaardigheid tussen bao en sbo is te omschrijven als groot. Het verschil in vaardigheid tussen bao en so is te omschrijven als middelgroot tot groot. Daarnaast is er een verwaarloosbaar verschil gevonden in vaardigheid tussen de leerlingen van het sbo en so.

4.5.3 Gespreksvaardigheid

Zoals weergegeven in Tabel 4.1.6 hebben niet alle leerlingen alle gesprekstaken gedaan. Zo zijn er leerlingen die enkel het samenwerkingsspel deden, leerlingen die enkel het raadsel deden en leerlingen die beide taken deden. De resultaten van deze verschillende groepen zijn daardoor niet zonder meer vergelijkbaar. Net als bij luistervaardigheid en spreekvaardigheid is hier gebruik gemaakt van IRT. Voordat modellen geschat konden worden diende er nog een aanpassing in de scoring plaats te vinden, net zoals bij spreekvaardigheid. In de vorige so/sbo-peiling werden de technische aspecten minder zwaar gewogen dan de inhoudelijke aspecten bij het samenwerkingsspel. Hier werd bij de totaalscore voor een leerling uitgegaan van een verhouding van 20% technische aspecten, en 80% inhoudelijke aspecten. Er waren echter zeven technische items en tien inhoudelijke items. Om tot een 20%/80% te komen werden de individuele scores op de technische items vermenigvuldigd met 0,486 en werden de individuele scores op de inhoudelijke items vermenigvuldigd met 1,360. Hierdoor kon een leerling maximaal 6,8 scorepunten (7 items * 2 scorepunten * 0,486) halen voor de technische aspecten, en maximaal 27,2 scorepunten (10 items * 2 scorepunten * 1,360) voor de inhoudelijke aspecten. Door deze weging ontstond een 20%/80%-verhouding van, respectievelijk, technische en inhoudelijke aspecten.

In de huidige peiling heeft iets soortgelijks plaatsgevonden, vergelijkbaar met de scoring van de spreektaken. Doordat – anders dan in de vorige so/sbo-peiling – IRT is gebruikt voor de schaling, dienen de individuele scores op de items discrete waarden te zijn. Daartoe is een weging zoals hierboven beschreven niet toepasbaar in de huidige peiling. Door de individuele scores van de leerlingen op de inhoudelijke items te vermenigvuldigen met drie, ontstaat een verhouding inhoudelijk/technisch die zeer dicht in de buurt komt van 20%/80% en is een discrete scoring behouden, waardoor het mogelijk blijft om een IRT-model op te passen. De ongewogen en gewogen scoring voor de naverteltaak en het betoog worden getoond in Tabel 4.5.10. De technische items behouden gewicht 1, de inhoudelijke items krijgen gewicht 3.

Tabel 4.5.10: Transformatie scoring spreektaken

		N items	scoring	verhouding	nieuwe scoring	nieuwe verhouding
samenwerkingsspel	inhoudelijk	10	0-1-2	58,8%	0-3-6	81,1%
	technisch	7	0-1-2	41,2%	0-1-2	18,9%
raadsel	inhoudelijk	10	0-1-2	58,8%	0-3-6	81,1%
	technisch	7	0-1-2	41,2%	0-1-2	18,9%

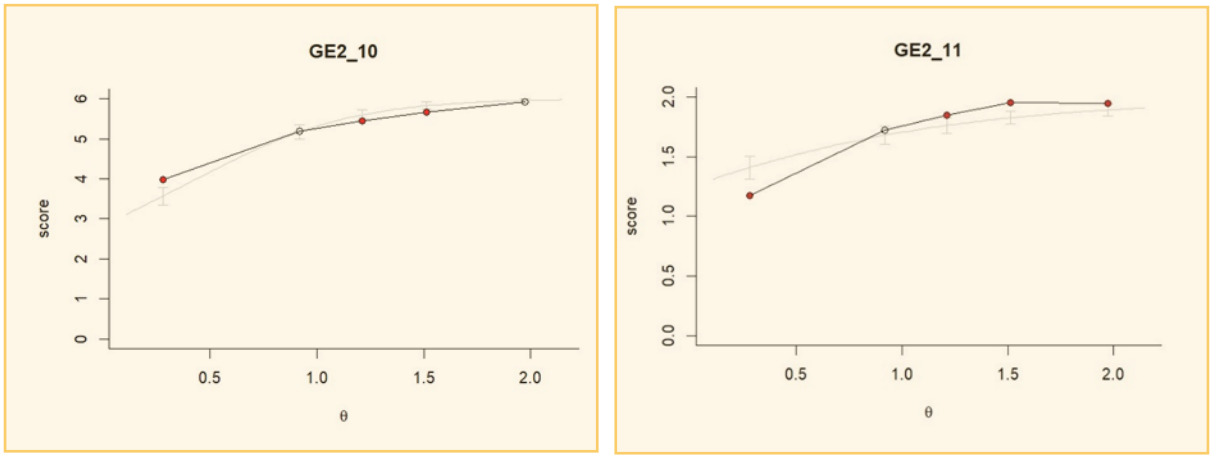
Na transformatie van de scores, is een One Parameter Logistic Model (OPLM; Verhelst & Glas, 1995) gepast op de data. Anders dan gebruikelijk bij een OPLM stond op voorhand dus al vast welke items welke discriminatieparameter zouden krijgen. Na modelschatting zijn de vaardigheden geschat met Weighted Maximum Likelihood (WML; Warm, 1989).
In Tabel 4.5.11 wordt de passing van het OPLM gepresenteerd. De passing van het model is niet overtuigend. De Phi is met 1,61 hoger dan de vaak in gebruik genomen grenswaarde van 1,225, maar laat geen extreem beeld zien.

Tabel 4.5.11: Modelpassing OPLM

Model	N	P01	P05	P10	%<.01	%<.05	%<.1	R1c	df	p	Phi
OPLM	34	14	3	4	41,2	50	61,8	586	227	<0,001	1,61

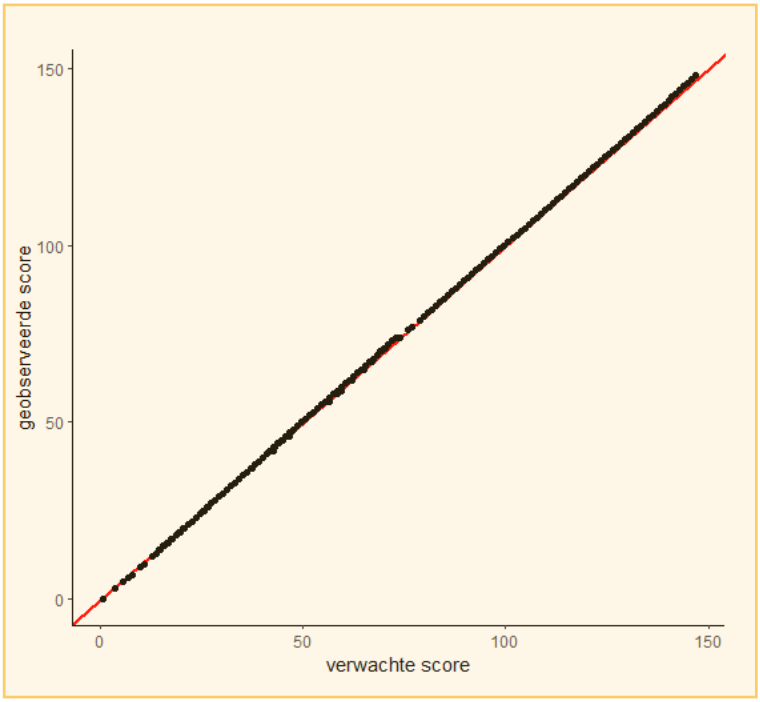
De modelpassing vertoont sterke gelijkenissen met de modelpassing van het OPLM bij spreekvaardigheid. Doordat de discriminatieparameters op voorhand al waren bepaald, is enige misfit in de lijn der verwachting. Dit wordt geïllustreerd in Figuur 4.5.11. Hier worden de ICC's weergegeven van twee items met enige misfit. Het item in de linkerplot is een hoge discriminatieparameter toegekend, terwijl de passing van het item gebaat zou zijn bij een minder hoge discriminatieparameter. De grijze modellijn zou dan vlakker liggen, en beter passen bij de zwarte geobserveerde data (zwarte lijn). Het tegenovergestelde geldt voor het item in de rechterplot. Bij dit item zou een hogere discriminatieparameter beter passen. De grijze modellijn zou dan steiler lopen, en beter passen bij de geobserveerde data.

Figuur 4.5.11: Passing items



Om de accuraatheid van het OPLM te evalueren, is een robuustheidsanalyse gedaan. Hierbij is voor iedere leerling de somscore van alle door de leerling gemaakte items afgezet tegen de verwachte somscore voor deze leerling op deze betreffende items – gegeven het OPLM. De resultaten worden gevisualiseerd in Figuur 4.5.12.

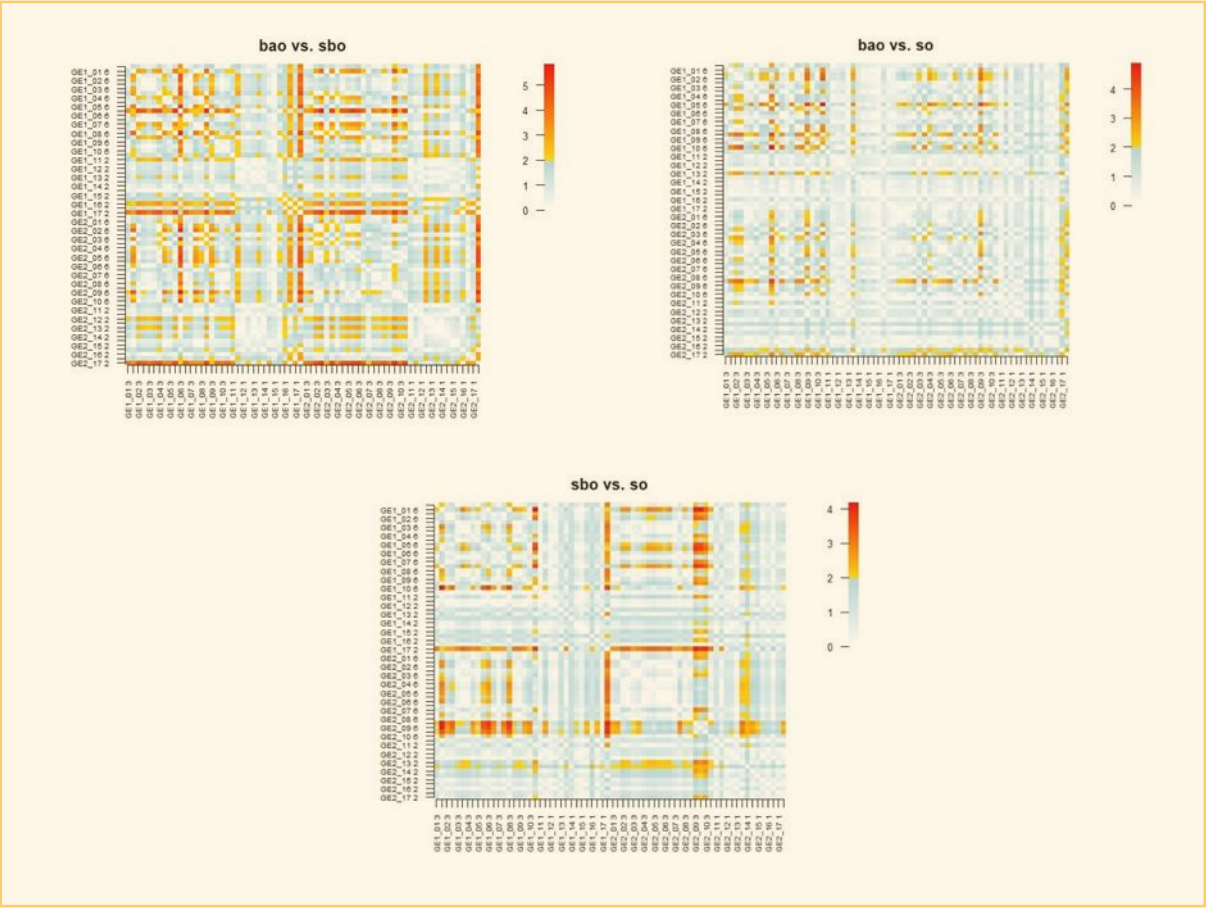
Figuur 4.5.12: Geobserveerde vs. geschatte somscores OPLM



Uit Figuur 4.5.12 wordt duidelijk dat het model de geobserveerde scores bijzonder accuraat 'terugschat'. De gemiddelde absolute afwijking tussen de geobserveerde score en verwachte score per persoon is slechts 0,628 scorepunt (*SD* = 0,301). Concluderend, met vertrouwen is besloten om verder te analyseren met dit model.

In aanvulling op de evaluatie van de passing van het model, is - net als bij luistervaardigheid en spreekvaardigheid - ook gekeken naar DIF tussen de onderwijstypes (bao, sbo, so). De differential item pair functioning (DIF) wordt visueel weergegeven in Figuur 4.5.13.

Figuur 4.5.13: Differential item pair functioning



Relatief gezien leek de meeste DIF te zitten tussen de bao- en sbo-leerlingen. Echter is vooral op te merken dat er van DIF weinig sprake is.

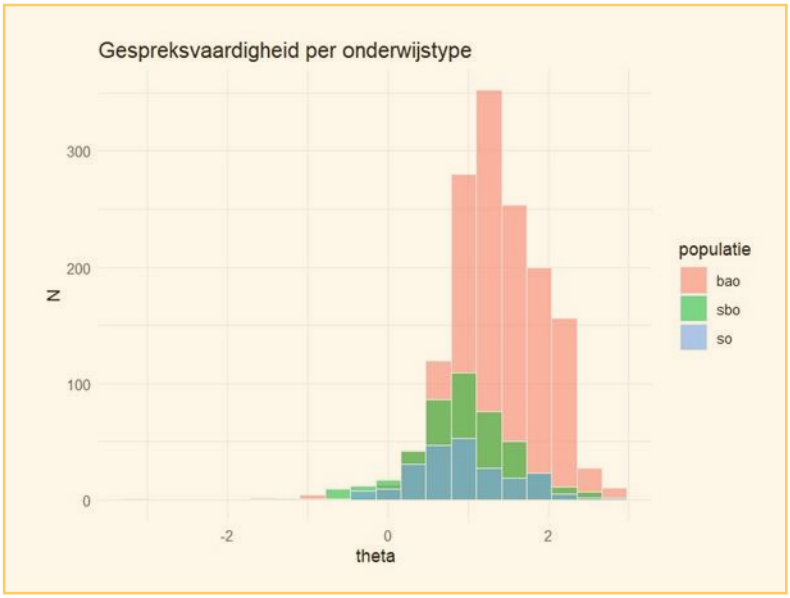
Tot slot is stilgestaan bij de betrouwbaarheid van de vaardigheidsschattingen. In de peiling worden conclusies over de vaardigheid gebaseerd op het IRT-model. De voor een peiling relevante betrouwbaarheid is daarom de betrouwbaarheid van de vaardigheidsschatting. Deze betrouwbaarheid is gedefinieerd als de proportie ware variantie in de populatie als onderdeel van de variantie in de schatting. Deze kan worden bepaald door de correlatie te nemen tussen uitwisselbare replicaties van de schatting van de vaardigheid (Bechger, et al., 2003). Hiervoor wordt de gemiddelde correlatie tussen een set van 100 plausible values genomen. De betrouwbaarheid, met een waarde van 0,850, kan beschouwd worden als hoog.

In Tabel 4.5.12 wordt de verdeling van de WML-schattingen (gemiddelden, standaarddeviaties) weergegeven per schooltype. In Figuur 4.5.14 wordt dit visueel weergegeven.

Tabel 4.5.12: WML-verdeling per onderwijstype

Schooltype	WML (M)	WML (SD)
bao	1,357	0,573
sbo	0,928	0,650
so	0,912	0,661

Figuur 4.5.14: Verdeling gespreksvaardigheid per onderwijstype



Om te evalueren of de onderwijstypes significant verschillen in vaardigheid, is een ANOVA uitgevoerd. Echter, bij een groot aantal deelnemers in een ANOVA wordt het kleinste verschil vaak al als significant geduid. Om enige nuance aan te brengen is de -*eta-squared* (η^2) gebruikt, die robuust is tegen grote groepsgrootten. Een door Cohen (1988) voorgestelde vuistregel om de η^2 te interpreteren is:

- Klein effect: 0,01
- Middelgroot effect: 0,06
- Groot effect: > 0,14

De schooltypes verschillen significant in vaardigheid, en het effect kan worden beschouwd als middelgroot tot groot ($F(2,2136) = 121,7$; $p < 0,001$; $\eta^2 = 0,10$). De gespreksvaardigheid van de bao-leerlingen is hoger dan die van de sbo- en so-leerlingen.

Tot slot zijn in Tabel 4.5.13 de paarsgewijze verschillen in vaardigheid tussen de schooltypen geduid aan de hand van Cohen's d (Cohen, 1988). De vuistregels worden hieronder opgesomd.

- Klein effect: 0,2
- Middelgroot effect: 0,5
- Groot effect: > 0,8

Tabel 4.5.13: Cohen's d voor paarsgewijze vergelijkingen

	bao	sbo	so
bao			
sbo	0,73		
so	0,76	0,02	

Concluderend, er werd nagenoeg geen DIF gevonden tussen de schooltypen. Op basis van deze resultaten kan gesteld worden dat de bao-leerlingen een hogere gespreksvaardigheid hebben dan de leerlingen van het sbo en het so. Het verschil in vaardigheid tussen bao en sbo is te omschrijven als middelgroot tot groot. Dezelfde conclusie kan getrokken worden ten aanzien van het verschil in vaardigheid tussen bao en so.

Daarnaast is er een verwaarloosbaar verschil gevonden in vaardigheid tussen de leerlingen van het sbo en so.

5 Schaling vragenlijsten

5.1 Inleiding

Voor het achterhalen van achtergrondkenmerken en domeinspecifieke kenmerken van leerlingen, leerkrachten en scholen en het achterhalen van kenmerken van het onderwijsleerproces, zijn vragenlijsten voor leerlingen, leerkrachten en schoolleiders ontwikkeld (zie Hoofdstuk 2). Bij het samenstellen van de vragenlijsten is zoveel mogelijk gebruikgemaakt van bestaande schalen. Om na te gaan of de veronderstelde schalen ook van toepassing zijn op de data uit het peilingsonderzoek is voor iedere veronderstelde schaal uit de leerling-, leerkracht- en schoolleidersvragenlijst een Principale Componentenanalyse (PCA) uitgevoerd. In dit hoofdstuk wordt verslag gedaan van de schaling die heeft plaatsgevonden voor deze vragenlijsten. Resultaten (op item- en schaalniveau) zijn terug te vinden in Hoofdstuk 7.

Allereerst is een initiële analyse uitgevoerd om de eigenwaarden van iedere component per variabele te verkrijgen. Eigenwaarden geven het inhoudelijke belang van een component aan waardoor het logisch is om alleen factoren te behouden met 'grote' eigenwaarden. Componenten met een eigenwaarde groter dan 1 (Kaiser's criterium; Kaiser, 1960) werden in principe behouden en gezien als schalen binnen de variabele. Om niet alleen af te gaan op de eigenwaarden, zijn ook de screeplots bekeken. Op basis van de eigenwaarden en screeplots zijn besluiten genomen over de schaling. Binnen iedere variabele werd verondersteld dat wanneer er meerdere schalen te onderscheiden waren, deze met elkaar zouden mogen correleren. Daarom is gekozen voor een oblique rotatie methode. Geschiktheid van de steekproeven werd geverifieerd aan de hand van de Kaiser-Meyer-Olkin maat (KMO), waarbij als criterium een waarde van groter dan .5 werd gehanteerd (Kaiser, 1970). Aan de hand van Bartlett's test of sphericity werd bepaald of de items onderling voldoende correleerden. In alle gevallen bleek de KMO groter dan 0.5 te zijn en was er voldoende onderlinge correlatie tussen de items volgens de Bartlett's test.

Na het uitvoeren van deze eerste analyse werd voor iedere verkregen schaal de betrouwbaarheid bepaald aan de hand van Cronbach's alpha. Voor deze indicator geldt dat een score van .70 of groter een indicatie is voor een goede betrouwbaarheid van de schaalscores. Een score van .60 - .70 wordt gezien als voldoende (Kline, 1999). Voor schalen met een betrouwbaarheid < .70 werd nagegaan of het verwijderen van een of meerdere items uit de schaal zou leiden tot een betere betrouwbaarheid.

Voor het berekenen van de schaalscores is een gemiddelde berekend van de antwoorden op de items van de schaal. Op deze manier blijft de schaalscore op de oorspronkelijke schaal, wat de interpreteerbaarheid van de schaalscore ten goede komt. Om een schaalscore te krijgen, moesten leerlingen, leerkrachten en schoolleiders minimaal de helft van het aantal items beantwoord hebben. Voor schalen met oneven items is naar boven afgerond: wanneer een schaal uit 7 items bestond moesten leerlingen, leerkrachten en schoolleiders minimaal 4 items beantwoord hebben.

De component- en betrouwbaarheidsanalyses zijn uitgevoerd op de totale respons van bao-, sbo- en so-leerlingen, leerkrachten en schoolleiders. Ter controle zijn de analyses ook separaat uitgevoerd voor de verschillende schooltypen. Op basis van deze analyses was er geen aanleiding om reeds genomen besluiten te veranderen en is er voor gelijke schalen voor de verschillende schooltypen gekozen. In de volgende paragrafen worden de eigenwaarden, componentladingen en Cronbach's Alpha's voor het primair onderwijs als geheel (bao, sbo en so) getoond.

5.2 Leerlingvragenlijst

De volgende schalen zijn uit de leerlingvragenlijst herleid: Thuis praten over, frequentie taalrijke activiteiten ouders, ervaren klasklimaat en preekvrijheid in de klas. In de tabellen hierna zijn de componentladingen en betrouwbaarheidsstatistieken per schaal weergegeven.

Thuis praten over..	
Resultaten PCA: één component met eigenwaarde=1,895 Cronbach's Alpha= 0,623	
Hoe vaak praat jij thuis over deze dingen?	**Componentlading**
1. Praten over een boek dat ik heb gelezen	0,737
2. Praten over een televisieprogramma of film waarnaar we samen hebben gekeken	0,723
3. Praten over wat ik op of na school heb gedaan	0,560
4. Praten over het nieuws (bijvoorbeeld over het klimaat of de politiek)	0,718
Besluit: er wordt één schaal gevormd op basis van de vier items: Schaal *Thuis praten over*	

Frequentie taalrijke activiteiten ouders	
Resultaten PCA: één component met eigenwaarde = 2,353 Cronbach's Alpha= 0,717	
Hoe vaak zie of hoor jij jouw ouders thuis deze dingen doen?	**Componentlading**
1. Lezen van een boek voor hun plezier	0,658
2. Brieven of mails schrijven of typen	0,653
3. De krant of een tijdschrift lezen (op papier of digitaal)	0,753
4. Kijken naar het journaal of een nieuwsprogramma	0,630
5. Praten over het nieuws (bijv. klimaat of politiek)	0,727
Besluit: er wordt één schaal gevormd op basis van de vijf items: Schaal *Frequentie taalrijke activiteiten ouders* Schaal loopt van 1 -4, waarbij 1= (bijna) nooit 2= 1 of 2 keer per maand 3= 1 of 2 keer per week 4=(Bijna) elke dag	

Ervaren Klasklimaat	
Resultaten PCA: twee componenten met eigenwaarden = 2,250 en 1,194 Ook te schalen als één component Cronbach's Alpha= 0,693	
Hoe vaak gebeuren deze dingen in jouw klas?	**Componentlading**
Er is te veel lawaai om goed te kunnen horen wat de juf/meester zegt	0,752
De juf/meester moet lang wachten voordat de kinderen stil zijn	0,757
Kinderen luisteren niet naar wat de juf/meester zegt	0,741
Mijn juf/meester praat te zacht of te snel waardoor ik het niet kan volgen	0,555
De juf/meester gebruikt (moeilijke) woorden die ik niet begrijp	0,506
Besluit: er wordt één schaal gevormd op basis van de vijf items: Schaal Ervaren klasklimaat Schaal loopt van 1 -4, waarbij 1= (bijna) nooit 2= 1 of 2 keer per maand 3= 1 of 2 keer per week 4=(Bijna) elke dag	

Spreekvrijheid in de klas	
Resultaten PCA: één component met eigenwaarde= 2,054 Cronbach's Alpha= 0,632	
Hoe vaak gebeuren deze dingen in jouw klas?	**Componentlading**
In mijn klas luisteren de kinderen goed naar elkaar	0,674
In mijn klas durven de kinderen vragen te stellen	0,619
In mijn klas word je geplaagd, gepest of uitgelachen als je een andere mening hebt dan de meesten (omgedraaid)	0,470
In mijn klas kan iedereen zeggen wat hij/zij van iets vindt	0,711
In mijn klas durven de kinderen over hun gevoelens te praten	0,700
Besluit: er wordt één schaal gevormd op basis van de vijf items: Schaal Spreekvrijheid in de klas Schaal loopt van 1 -4, waarbij 1= nooit 2= meestal 3=soms 4=Altijd	

5.3 Quickscan

In de quickscan hebben leerkrachten enkele gegevens over beïnvloedende gedragskenmerken en -stoornissen van leerlingen verstrekt. Middels een componentenanalyse is nagegaan of er schalen uit de quickscan te herleiden zijn. Dat bleek het geval te zijn voor communicatieproblemen en stimulerende en belemmerende factoren. Enkele gedrags- en persoonlijkheidskenmerken werden gemeten aan de hand van één of enkele items en zijn daarmee niet schaalbaar. Deze items worden wel meegenomen in de verdere analyses. Om deze reden presenteren we hieronder ook de resultaten van de componentenanalyse voor deze items.

In de tabellen hieronder zijn de componentladingen en betrouwbaarheidsstatistieken per schaal weergegeven.

Gedrags- en persoonlijkheidskenmerken	
Resultaten PCA: één component met eigenwaarde =1,029 Cronbach's Alpha= 0,345	
	Componentlading
Probleemgedrag Score 1-5 van geen probleemgedrag tot (sterk) probleemgedrag	
Externaliserend probleemgedrag	0,778
Internaliserend probleemgedrag	0,778
Besluit: Op basis van de componentenanalyse zou er één component gevormd kunnen worden. Het zijn echter maar twee items en de aard van het probleemgedrag is zeer uiteenlopend. Bovendien is de betrouwbaarheid van de schaal ook zeer laag. De samenhang (Pearson R) tussen deze twee items is minder sterk (0.209). Om deze reden worden de items als losse items meegenomen in de verdere analyses.	

Gedrags- en persoonlijkheidskenmerken	
Resultaten PCA: één component met eigenwaarde =1,732 Cronbach's Alpha= 0,845	
	Componentlading
Introvert- extravert (1 -(sterk) introvert – 5 (sterk) extravert)	0,931
Spreekangst – geen spreekangst (1- (sterk) spreekangstig– 5 niet spreekangstig)	0,931
Besluit: Op basis van de factoranalyse zou er één component gevormd kunnen worden. Het zijn echter maar twee items en interpretatie van de schaal wordt erg lastig. Correlatie tussen de items is hoog (0.732). Besluit om niet te schalen, maar als losse items mee te nemen. In de multilevel analyse wordt één van de twee items (Spreekangst) meegenomen, vanwege de hoge correlatie.	

Communicatieproblemen	
Resultaten PCA: één component met eigenwaarde = 2,685 Cronbach's Alpha= 0,827	
Communicatieproblemen (1=niet aanwezig, 2= in lichte mate, 3 = in behoorlijke mate, 4 = in ernstige mate)	**Componentlading**
Zwak taalbegrip	0,890
Zwakke uitdrukkingsvaardigheid	0,893
Zich niet kunnen verplaatsen in een ander	0,584
Zwak auditief geheugen/auditieve verwerking	0,869
Besluit: Ondanks dat de betrouwbaarheid toeneemt tot 0,885 bij verwijdering van het derde item (zich niet kunnen verplaatsen in een ander) uit de schaal, wordt er vanwege een voldoende hoge betrouwbaarheid en theoretische overwegingen één schaal gevormd op basis van de vier items: Schaal Communicatieproblemen Schaal loopt van 1 -4, waarbij 1= niet aanwezig, 2= in lichte mate, 3 = in behoorlijke mate, 4 = in ernstige mate	

Stimulerende en belemmerende factoren	
Resultaten PCA: één component met eigenwaarde =3,528 Cronbach's Alpha= 0,858	
Stimulerende en belemmerende factoren (1=belemmerend, 2= neutraal, 3 = stimulerend)	Componentlading
Inhibitie	0,731
Emotieregulatie	0,664
Planning en organisatie	0,788
Zelfstandigheid	0,793
Motivatie	0,789
Concentratie en aandacht	0,825
Besluit: er wordt één schaal gevormd op basis van de zes items: Schaal Stimulerende en belemmerende factoren Schaal loopt van 1-3, waarbij 1= belemmerend 2= neutraal 3 =stimulerend.	

5.4 Leerkrachtvragenlijst

De volgende schalen zijn uit de leerkrachtvragenlijst herleid:

- Relatieve belang mondelinge taalvaardigheid (Tabel 5.4.1);
- Ervaren bekwaamheid mondelinge taalvaardigheid (Tabel 5.4.2);
- Frequentie gesprekstaken (Tabel 5.4.3);
- Gepercipieerde toepassing van effectieve instructie in gesprekken (Tabel 5.4.4);
- Frequentie luistertaken (Tabel 5.4.5);
- Gepercipieerde toepassing van effectieve instructie in luisteren (Tabel 5.4.6);
- Frequentie spreektaken (Tabel 5.4.7);
- Gepercipieerde toepassing van effectieve instructie in gesprekken (Tabel 5.4.8).

De tabellen hierna zijn de componentladingen en betrouwbaarheidsstatistieken per schaal weergegeven.

Tabel 5.4.1: Relatieve belang mondelinge taalvaardigheid

Belang mondelinge taalvaardigheid		
Resultaten PCA: twee componenten met eigenwaarde = 1,609 en 1,583 Component 1: items 4 en 5 | component 2: items 1 t/m 3 en 6. Cronbach's Alpha component 1= 0,717 Cronbach's Alpha component 2= 0,468		
In hoeverre bent u het eens met de volgende stellingen?	**Componentlading Component 1**	**Componentlading Component 2**
Mondelinge taalvaardigheid moet van groep 1 t/m eind groep 8 expliciet worden onderwezen		0,744
Mondelinge taalvaardigheid is essentieel voor de lees- en schrijfontwikkeling van leerlingen		0,746
Ik besteed voldoende aandacht aan onderwijs in mondelinge taalvaardigheid		0,464
Een goede leesvaardigheid is belangrijker dan een goede mondelinge taalvaardigheid	0,891	
Een goede schrijfvaardigheid is belangrijker dan een goede mondelinge taalvaardigheid	0,871	
Onderwijs in mondelinge taalvaardigheid is voor alle leerlingen in mijn klas belangrijk		0,497
Besluit: er wordt één schaal gevormd op basis van item vier en vijf: Schaal *Relatieve belang mondelinge taalvaardigheid* Schaal loopt van 1 -4, waarbij 1= zeer mee oneens 2= mee oneens 3= mee eens 4=zeer mee eens Overige stellingen (1, 2, 3 en 6) zijn gezamenlijk als schaal onvoldoende betrouwbaar en worden als losse items geanalyseerd (in hoofdstuk 7).		

Tabel 5.4.2: Ervaren bekwaamheid

Ervaren bekwaamheid	
Resultaten PCA: drie component met eigenwaarden = 2,455 ; 1,145 en 1,027 Component 1: 1 t/m 3 | component 2: 4 en 5 | component 3: 6 Ook te schalen op één component: item 1 t/m 5, 6 te lage componentlading Cronbach's Alpha (schaal items 1 t/m 5)= 0,734	
In hoeverre bent u het eens met de volgende stellingen?	**Componentlading**
• Ik weet in grote lijnen hoe de ontwikkeling van de mondelinge taalvaardigheid in groep 1 t/m 8 verloopt	0,777
• Ik ben bekend met de beheersingsdoelen in het referentiekader voor mondelinge taalvaardigheid	0,742
• Ik weet hoe ik de mondelinge taalvaardigheid van mijn leerlingen kan bevorderen	0,675
• Ik vind het lastig om leerlingen te ondersteunen bij de ontwikkeling van hun mondelinge taalvaardigheid (omgedraaid)	0,616
• Ik vind het moeilijk om de lessen in mondelinge taalvaardigheid goed vorm te geven (omgedraaid)	0,679
• De wijze waarop ik gesprekken voer, luister en spreek, heeft invloed op hoe mijn leerlingen deze vaardigheden ontwikkelen	0,072
Besluit: er wordt één schaal gevormd op basis van item 1 t/m 3, 4 en 5 (beide omgedraaid): Schaal Ervaren bekwaamheid Schaal loopt van 1-4, waarbij 1= zeer mee oneens 2= mee oneens 3= mee eens 4=zeer mee eens	

Tabel 5.4.3: Frequentie gesprekstaken

Frequentie gesprekstaken	
Resultaten PCA: twee componenten met eigenwaarden = 2,455 en 1,132 Component 1: 3 t/m 5 en 7 | component 2: 1 en 2 Ook te schalen als één component: items 1 t/m 7 Cronbach's Alpha= 0,682	
Hoe vaak heeft u onderstaande gesprekstaken intentioneel geoefend met de leerlingen in de afgelopen 12 weken?	**Componentlading**
Discussies voeren (bijv. uitwisselen wat ze van een boek of film vinden of discussiëren over een omstreden of actueel onderwerp)	0,569
Debatteren (bijv. een wedstrijddebat waarin een stelling wordt verdedigd of bestreden)	0,462
Overleggen (bijv. over hoe ze samen een taak kunnen uitvoeren)	0,720
Brainstormen (bijv. ideeën over de aanpak van een taak uitwisselen)	0,748
Interviews afnemen en geven	0,554
Leergesprekken (bijv. een gesprek met het doel kennis over een geschiedenisonderwerp te vergaren of vragen te formuleren voor een onderzoek)	0,600
Vrije gesprekken (bijv. een kringgesprek om ervaringen en gevoelens uit te wisselen)	0,415
Besluit: er wordt één schaal gevormd op basis van item 1 t/m 7: Schaal Frequentie gesprekstaken Schaal loopt van 1 -5, waarbij 1= nooit 2= 1 of 2 keer 3= 3 tot 6 keer 4= 7 tot 12 keer 5= meer dan 12 keer	

Tabel 5.4.4: Stellingen over gesprekken

Stellingen over gesprekken	
Resultaten PCA: twee componenten met eigenwaarden =2,198 en 1,062 Component 1: 1 t/m 5 | component 2: 6 Cronbach's Alpha component 1= 0,673	
In hoeverre bent u het eens met de volgende stellingen?	**Componentlading**
Ik leer leerlingen expliciet aan hoe ze goed een gesprek kunnen voeren (bijv. beurt nemen, elkaars perspectief begrijpen, redeneren, argumenteren)	0,703
Ik laat leerlingen oefenen met verschillende gespreksdoelen, zoals uitleggen, discussiëren en argumenteren	0,572
Ik stimuleer een veilig gesprekskliimaat in de klas (bijvoorbeeld door leerlingen de ruimte te geven om hun mening te geven, aandachtig naar ze te luisteren, ze serieus te nemen en alle leerlingen aan te moedigen dit ook onderling te doen)	0,587
Ik laat leerlingen gesprekken voeren in kleine groepjes	0,678
Ik stimuleer alle leerlingen om tijdens gesprekken mee te denken en te praten	0,753
Ik sta toe dat leerlingen in een andere taal dan het Nederlands overleggen	-
Besluit: er wordt één schaal gevormd op basis van item 1 t/m 5: Schaal Gepercipieerde toepassing van effectieve instructie in gesprekken Schaal loopt van 1-4, waarbij 1= helemaal niet van toepassing 2= niet van toepassing 3 = van toepassing 4= helemaal van toepassing	

Tabel 5.4.5: Frequentie Luistertaken

Frequentie Luistertaken	
Resultaten PCA: één component met eigenwaarde =1,728 Cronbach's Alpha= 0,622	
Hoe vaak heeft u onderstaande luistertaken intentioneel geoefend met de leerlingen in de afgelopen 12 weken?	**Componentlading**
Luisteren naar een verhaal verteld door iemand die fysiek aanwezig is in het klaslokaal	0,835
Luisteren naar een gesprek tussen mensen die fysiek aanwezig zijn in het klaslokaal	0,775
Luisteren naar gesproken tekst op een digitaal apparaat die leerlingen kunnen terugspoelen (bijv. filmpjes of podcasts)	0,655
Besluit: er wordt één schaal gevormd op basis van item 1 t/m 3: Schaal Frequentie luistertaken Schaal loopt van 1 -5, waarbij 1= nooit 2= 1 of 2 keer 3= 3 tot 6 keer 4= 7 tot 12 keer 5= meer dan 12 keer.	

Tabel 5.4.6: Stellingen over Luisteren

Stellingen over Luisteren	
Resultaten PCA: drie componenten met eigenwaarden = 2,539 ; 1,221 en 1,009 Component 1: 5, 7 en 8 | component 2: 1, 3 | component 3: 6 Item 2 laadt dubbel (op c2 en c3), item 4 te lage componentladingen. Ook te schalen op één component Cronbach's Alpha= 0,689	
In hoeverre bent u het eens met de volgende stellingen?	**Componentlading**
Ik leer leerlingen expliciet luisterstrategieën aan (bijv. luisterdoel bepalen, onbekende woorden verduidelijken, voorspellingen doen, samenvatten)	0,608
Ik stimuleer een goed luisterklimaat in de klas (bijv. weinig afleiding en ruis)	0,392
Ik laat leerlingen aantekeningen maken tijdens het luisteren	0,458
Ik bied visuele ondersteuning aan bij het luisteren, bijvoorbeeld door middel van plaatjes, een woordweb, stroomdiagrammen, etc.	0,486
Ik bied luisterteksten herhaald aan bij het oefenen van begrijpend luisteren	0,586
Ik leg uit wat de moeilijke woorden die leerlingen horen betekenen	0,435
Ik deel luistertaken op in korte fragmenten voor leerlingen die luisteren lastig vinden	0,758
Ik geef leerlingen die extra ondersteuning nodig hebben bij luisteren instructie in kleine groepen of één-op-één	0,681
Besluit: er wordt één schaal gevormd op basis van item 1 t/m 8: Schaal Gepercipieerde toepassing van effectieve instructie in Luisteren Schaal loopt van 1-4, waarbij 1= helemaal niet van toepassing 2= niet van toepassing 3 = van toepassing 4= helemaal van toepassing Hoewel het item 2 een (net) te lage componentlading heeft, wordt het toch meegenomen in de schaal. De betrouwbaarheid van de schaal is namelijk iets lager zonder het item (0,684).	

Tabel 5.4.7: Frequentie Spreektaken

Frequentie Spreektaken	
Resultaten PCA: één component met eigenwaarde =2,340 Item 4 heeft een te lage communaliteitlading. Cronbach's Alpha= 0,733	
Hoe vaak heeft u onderstaande spreektaken intentioneel geoefend met de leerlingen in de afgelopen 12 weken?	**Componentlading**
Een beschrijving geven (bijv. van een persoon, plaats, zaak of voorwerp)	0,729
Verslag van een gebeurtenis uitbrengen	0,752
Een verhaal vertellen of navertellen	0,787
Presenteren (bijv. een spreekbeurt)	0,416
Leerlingen uitleg of instructie laten geven (bijv. de uitleg over een apparaat, instructie bij een spel of een routebeschrijving geven)	0,670
Besluit: er wordt één schaal gevormd op basis van item 1 t/m 3 en 5: Schaal Frequentie spreektaken Schaal loopt van 1 -5, waarbij 1= nooit 2= 1 of 2 keer 3= 3 tot 6 keer 4= 7 tot 12 keer 5= meer dan 12 keer	

Tabel 5.4.8: Stellingen over spreken

Stellingen over spreken	
Resultaten PCA: één component met eigenwaarde =3,002 Cronbach's Alpha= 0,832	
In hoeverre bent u het eens met de volgende stellingen?	**Componentlading**
Ik zorg bij spreektaken voor opbouw in taakeisen en complexiteit	0,712
Ik reflecteer na de lesactiviteit met leerlingen op de spreektaak en of de doelen zijn behaald	0,862
Ik laat leerlingen elkaar feedback geven op de uitvoering van een spreektaak	0,781
Ik modelleer (=hardop voordoen) hoe leerlingen spreektaken goed kunnen uitvoeren	0,759
Ik laat leerlingen spreektaken oefenen in kleine groepjes	0,753
Besluit: er wordt één schaal gevormd op basis van item 1 t/m 5: Schaal Gepercipieerde toepassing van effectieve instructie in Spreken Schaal loopt van 1-4, waarbij 1= helemaal niet van toepassing 2= niet van toepassing 3 = van toepassing 4= helemaal van toepassing	

5.5 Schoolleidersvragenlijst

De volgende schalen zijn uit de leerkrachtvragenlijst herleid:

- aandacht voor mondelinge taalvaardigheid in vergaderingen (Tabel 5.5.1);
- aandacht voor mondelinge taalvaardigheid in visie school en bestuur (Tabel 5.5.2).

De tabellen hierna zijn de componentladingen en betrouwbaarheidsstatistieken per schaal weergegeven.

Tabel 5.5.1: Mondelinge taalvaardigheid in vergaderingen

Mondelinge taalvaardigheid in vergaderingen	
Resultaten PCA: één component met eigenwaarde =2,614 Cronbach's Alpha= 0,818 ; verhoogd tot 0,850 bij verwijdering item 1	
Stel alle teamvergaderingen op uw school in een gemiddeld schooljaar op 100%. Bij hoeveel procent van deze vergaderingen komen de onderstaande thema's aan de orde?	**Componentlading**
(De veiligheid van) het spreekklimaat voor leerlingen in onze school	0,644
De prestaties voor mondelinge taalvaardigheid op groepsniveau en/of schoolniveau	0,847
De kwaliteit van het mondelinge taalvaardigheidsonderwijs (leerstof aanbod, leertijd, didactiek)	0,871
De visie van onze school op het onderwijs in mondelinge taalvaardigheid	0,842
Besluit: Vanwege inhoud en betrouwbaarheid wordt er één schaal gevormd op basis van item 2 t/m 4: Schaal Aandacht voor mondelinge taalvaardigheid in vergaderingen Schaal loopt van 1-4, waarbij 1= minder dan 25% 2= 25-50% 3 = 50-75% 4 = >75%	

Tabel 5.5.2: Stellingen over mondelinge taalvaardigheid: aandacht in visie

Stellingen over mondelinge taalvaardigheid: aandacht in visie	
Resultaten PCA: één component met eigenwaarde =2,431 Cronbach's Alpha= 0,781	
In hoeverre bent u het eens met de volgende uitspraken over het onderwijs in mondelinge taalvaardigheid op uw school?	**Componentlading**
Onze school heeft een duidelijke visie over de plaats van mondelinge taalvaardigheid in ons curriculum	0,792
Het belang van (meer) aandacht voor mondelinge taalvaardigheid in het onderwijs wordt door de leerkrachten op onze school breed gedragen	0,640
Het schoolbestuur heeft zicht op de kwaliteit van het mondelinge taalvaardigheidsonderwijs van onze school	0,872
Het schoolbestuur heeft in beeld welke leerlingen van onze school in het leergebied mondelinge taalvaardigheid een bepaald referentieniveau behalen	0,797
Besluit: er wordt één schaal gevormd op basis van item 1 t/m 4: Schaal Aandacht voor mondelinge taalvaardigheid in visie school en bestuur Schaal loopt van 1-4, waarbij 1= Zeer mee oneens 2= enigszins mee oneens 3 = enigszins mee eens 4= zeer mee eens	

6 Resultaten luisteren, spreken en gesprekken voeren

In dit hoofdstuk wordt het overzetten van de prestatiestandaarden uit vorige peilingen beschreven en vervolgens de leerlingprestaties. Dit wordt vervolgd door een trendanalyse, waarin de vaardigheid van de leerlingen uit de huidige peiling wordt vergeleken met de vaardigheid uit vorige peilingen: de bao-peiling uit 2017 en de so/sbo-peiling uit 2018 (resp., van Langen et al., 2017; Keuning et al., 2019). Hier wordt een DIF-analyse ten aanzien van de verschillende jaargangen geëvalueerd. Tot slot wordt de beheersing van de vaardigheid op item- en domeinniveau besproken.

Hieronder wordt eerst het overzicht gegeven van het aantal responses waarop de resultaten zijn gebaseerd.

Taak	Aanvankelijk	Ongeldige respons	Preconditie = 0	Verwijderd na beoordelaarsonderzoek	Uiteindelijk
Luistertoets 1	466	0	n.v.t.	n.v.t.	466
Luistertoets 2	506	3	n.v.t.	n.v.t.	503
Luistertoets 3	364	0	n.v.t.	n.v.t.	364
Luistertoets 4	418	0	n.v.t.	n.v.t.	418
Naverteltaak	1317	0	14	20	1283
Betoog	1327	3	65	23	1236
Samenwerkingsspel	1374	2	0	8	1364
Raadsel	1146	5	0	0	1141

6.1 Prestatiestandaarden

6.1.1 Luistervaardigheid

Voor het bepalen van de referentieniveaus van de leerlingen in de huidige peiling, wordt gebruikgemaakt van informatie uit vorige peilingen mondelinge taalvaardigheid: de bao-peiling uit 2017 en de so/sbo-peiling uit 2018. In beide peilingen heeft een standaardbepaling plaatsgevonden, waarbij de referentieniveaus van de leerlingen werden geduid. In de huidige peiling vindt geen standaardbepaling plaats, maar worden de standaarden uit deze vorige peilingen overgezet naar de huidige peiling. Dit is mogelijk doordat fragmenten en items uit de huidige peiling onderdeel waren van de vorige peilingen (weergegeven in Tabel 6.1.1).

Tabel 6.1.1: Verdeling fragmenten over peilingen (enkel selectie fragmenten huidige peiling)

			Peiling		
Fragment	**Bron**	**N items**	**bao/s(b)o2024**	**bao2017**	**s(b)o2018**
Ridders (1F)	Anker 1F	5			
bank (2F)	Anker 2017	2			
Melkmeisje (2F)	Anker 2F	4			
Gooi (2F)	Anker 2017	2			
Tekenares (1F)	LVS E6	4			
Zebrastrepen (1F)	Anker 1F	3			
goud (1F)	Anker 2017	3			
Bacteriën (2F)	PPON 2007	5			
julia (1F)	Anker 2017	3			
contant geld (2F)	Anker 2017	2			
Internet (2F)	PPON 2007	5			
Pindakaas (1F/2F)	LVS E7	5			
N items			43	12	26

Van de 43 items in de huidige peiling waren 12 items onderdeel van de bao-peiling en 26 items van de so/sbo-peiling. Van beide peilingen is bekend welke luistervaardigheid nodig was om 1F en 2F te behalen. Aan de hand van deze vaardigheidscesuren is het mogelijk om aan de hand van de destijds geschatte IRT-modellen de posterior verwachte somscore te berekenen op het setje items per peiling: de True Score (*True Score Equating* (TSE); Kolen & Brennan, 2014). In Tabel 6.1.2 worden de True Scores per peiling per referentieniveau gepresenteerd.

Tabel 6.1.2: True scores per peiling per referentieniveau

peiling	N items	True Score 1F	True Score 2F
bao2017	12	5,146	8,901
s(b)o2018	26	14,532	21,002
samen	38	19,678	29,902

Een posterior verwachte somscore van 5,146 op de 12 bao-items - gegeven het IRT-model uit de bao-peiling – zit precies op het kantelpunt <1F/1F. De True Scores uit de twee bronnen zijn bij elkaar opgeteld en het is duidelijk welke posterior verwachte somscore op de 38 items vereist is voor het behalen van 1F, dan wel 2F. Aan de hand van de geschatte itemparameters uit het huidige model kan dan andersom weer bepaald worden welke vaardigheidsgrenzen horen bij de gegeven True Scores. De vaardigheidsgrenzen horend bij de True Scores op de 38 items zijn bepaald op:

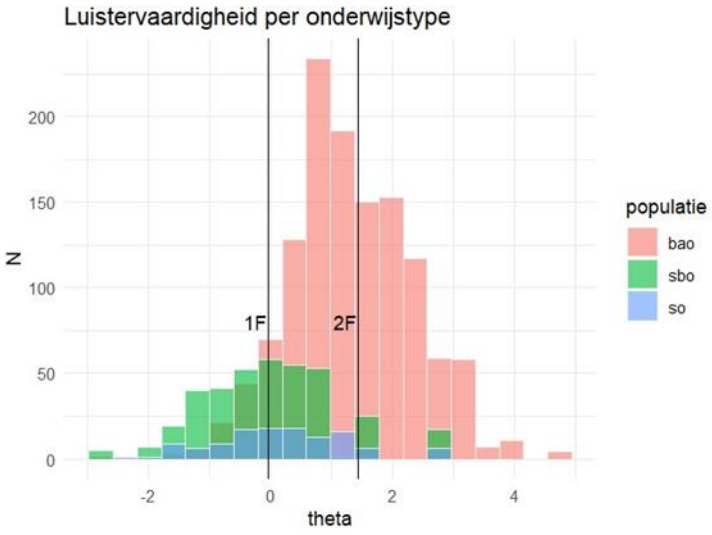
- 1F:-0,02899204
- 2F: 1,435654

Aan de hand van deze cesuren zijn de behaalde referentieniveaus bepaald per leerling. De proporties behaalde referentieniveaus worden weergegeven in Tabel 6.1.3. In Figuur 6.1.1 wordt dit visueel weergegeven.

Tabel 6.1.3: Behaalde referentieniveaus luistervaardigheid per schooltype

Schooltype	Basisonderwijs		Speciaal basisonderwijs		Speciaal onderwijs	
ref. niveau	aantal	%	aantal	%	aantal	%
<1F	101	8,0	219	58,7	48	40,0
1F	600	47,7	112	30,0	60	50,0
2F	557	44,3	42	11,3	12	10,0

Figuur 6.1.1. - Vaardigheid en behaald referentieniveaus



Zoals al werd geconcludeerd in de paragraaf over de schaling, hebben bao-leerlingen een hogere vaardigheid dan leerlingen van de overige schooltypes.
De cesuren bleken uiteindelijk niet zonder meer over te zetten te zijn. Een voorwaarde voor betrouwbare True Score Equating is dat er geen DIF is tussen de jaargangen. Dit werd niet gevonden bij de items tussen de huidige peiling en de so/sbo-peiling, maar wel in vrij grote mate tussen de items in de huidige peiling en de items in de bao-peiling. In paragraaf 6.2 over de trendbepaling wordt dit besproken. Daarom is ter aanvulling nog een keer de True Score Equating uitgevoerd, maar dan enkel op basis van de 26 so/sbo-items. Op dezelfde manier zijn via deze route de behaalde referentieniveaus bepaald voor de leerlingen in de huidige peiling. In Tabel 6.1.4 is een verwarringsmatrix weergegeven, die de resultaten weergeven waarbij True Scores zijn berekend aan de hand van alle 38 items, en waarbij True Scores zijn berekend op basis van de 26 so/sbo-items.

Tabel 6.1.4: Verwarringsmatrix behaalde referentieniveaus TSE a.d.h.v. alle 38 items vs. 26 s(b)o-items

		a.d.h.v. 38 items		
		<1F	1F	2F
a.d.h.v. 26 items	<1F	368	39	0
	1F	0	733	43
	2F	0	0	568

Hieruit volgt dat de misclassificatie ongeveer 4,7% bedraagt, waarbij in alle gevallen de leerling op basis van de 38 items hoger beoordeeld zijn dan aan de hand van de 26 items. Aangezien het percentage misclassificatie geen extreem beeld laat zien, is uiteindelijk besloten om toch uit te gaan van alle 38 items (dus inclusief de bao-items) bij het berekenen van de True Scores en de resultaten uit Tabel 6.1.3 te handhaven. Het toestaan van de bao-items in de procedure, rechtvaardigt het om in de trendvergelijking iets – in plaats van niets - te kunnen zeggen over de verschillen in prestaties van bao-leerlingen tussen de huidige peiling en de bao-peiling uit 2017.

6.1.2 Spreekvaardigheid

Voor het bepalen van de referentieniveaus van de leerlingen uit de huidige peiling wordt gebruikgemaakt van informatie uit de vorige so/sbo-peiling. Zoals eerder vermeld, heeft indertijd een standaardbepaling plaatsgevonden op de spreektaken die toen werden afgenomen (zie Hemker, Van Ingen & Linthorst, 2018). Eén van deze taken was de naverteltaak. Voor deze taak is bepaald:

- Leerlingen die 0-19 punten scoren op de naverteltaak behalen <1F
- Leerlingen die 20-26 punten scoren op de naverteltaak behalen 1F
- Leerlingen die 27-30 punten scoren op de naverteltaak behalen 2F

Dit betreffen gewogen scores zoals die indertijd zijn bepaald. Deze gewogen scores zijn in de huidige peiling exact zo berekend voor de leerlingen die enkel de naverteltaak hadden gedaan. Bij leerlingen die op de grens zaten van <1F/1F en 1F/2F is de bijbehorende theta gezocht aan de hand van het gebruikte IRT-model zoals beschreven in paragraaf 4.5. Omdat alle leerlingen op eenzelfde schaal staan in het IRT-model, was het aan de hand van de cesuurtheta's mogelijk om voor iedere leerling het behaalde referentieniveau te bepalen. De theta's die werden gevonden bij leerlingen die op de grens scoorden waren:

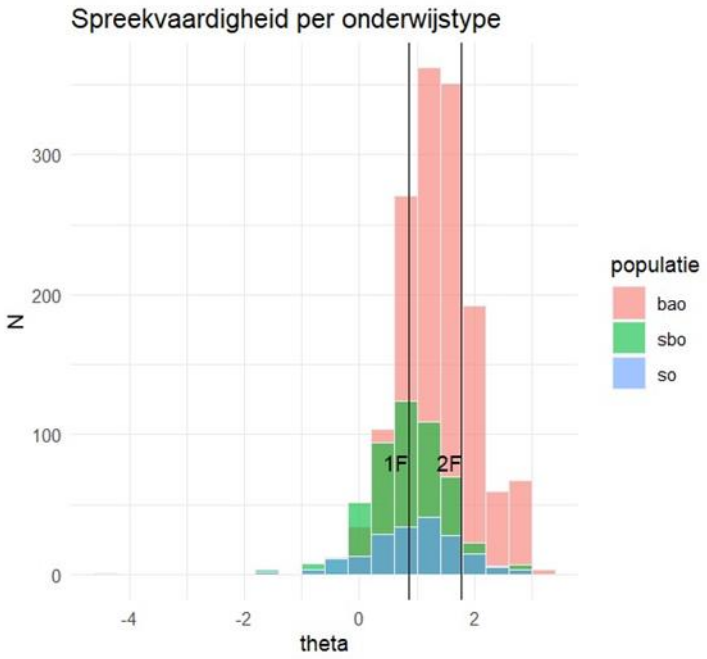
- <1F/1F: 0,851
- 1F/2F: 1,772

In Tabel 6.1.5 zijn de behaalde referentieniveaus weergegeven voor alle leerlingen. In Figuur 6.1.2 is wederom de verdeling van de spreekvaardigheid weergegeven, ditmaal met de cesuren erbij.

Tabel 6.1.5: Verdeling referentieniveaus spreekvaardigheid per onderwijstype

Schooltype	Basisonderwijs		Speciaal basisonderwijs		Speciaal onderwijs	
ref. niveau	aantal	%	aantal	%	aantal	%
<1F	266	18,4	236	46,5	78	42,4
1F	808	55,9	229	45,1	80	43,5
2F	372	25,7	43	8,5	26	14,1

Figuur 6.1.2: Vaardigheid en behaalde referentieniveaus



Zoals al geconcludeerd in paragraaf 4.5, scoort het bao overtuigend hoger dan het sbo en so. Bij het bao scoort het merendeel 1F of 2F, terwijl voor sbo en so geldt dat zij vooral <1F en 1F scoren.

6.1.3 Gespreksvaardigheid

Voor het bepalen van de referentieniveaus van de leerlingen uit de huidige peiling wordt gebruikgemaakt van informatie uit de vorige so/sbo-peiling. Zoals eerder vermeld, heeft indertijd een standaardbepaling plaatsgevonden op het samenwerkingsspel dat toen werden afgenomen (zie Hemker, Van Ingen & Linthorst, 2018). Hiervoor is bepaald:

- Leerlingen die 0-27 punten scoren op het samenwerkingsspel behalen <1F
- Leerlingen die 28-30 punten scoren op het samenwerkingsspel behalen 1F
- Leerlingen die 31-34 punten scoren op het samenwerkingsspel behalen 2F

Dit betreffen gewogen scores zoals die indertijd zijn bepaald. Deze gewogen scores zijn in de huidige peiling exact zo berekend voor de leerlingen die enkel het samenwerkingsspel hadden gedaan. Bij leerlingen die op

de grens zaten van <1F/1F en 1F/2F is de bijbehorende theta gezocht aan de hand van het gebruikte IRT-model zoals beschreven in paragraaf 6.1. Omdat alle leerlingen op eenzelfde schaal staan in het IRT-model, was het aan de hand van de cesuurtheta's mogelijk om voor iedere leerling het behaalde referentieniveau te bepalen.
De theta's die werden gevonden bij leerlingen die op de grens scoorden waren:

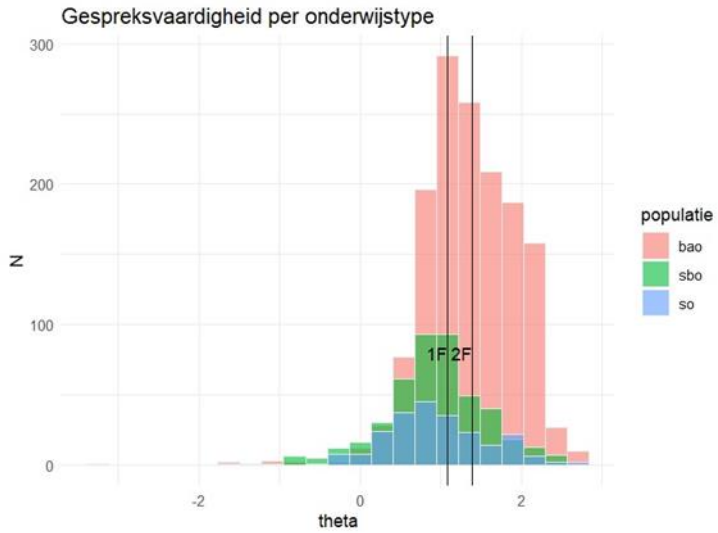
- <1F/1F: 1,080
- 1F/2F: 1,390

In Tabel 6.1.6 zijn de behaalde referentieniveaus weergegeven voor alle leerlingen. In Figuur 6.1.3 is wederom de verdeling van de gespreksvaardigheid weergegeven, ditmaal met de cesuren erbij.

Tabel 6.1.6: Verdeling referentieniveaus gespreksvaardigheid per onderwijstype

Schooltype	Basisonderwijs		Speciaal basisonderwijs		Speciaal onderwijs	
ref. niveau	**aantal**	**%**	**aantal**	**%**	**aantal**	**%**
<1F	458	31,3	277	62,1	150	65,2
1F	358	24,5	77	17,3	29	12,6
2F	647	44,2	92	20,6	51	22,2

Figuur 6.1.3. - *Vaardigheid en behaalde referentieniveaus*



Zoals al geconcludeerd in paragraaf 4.5, scoort het ba overtuigend hoger dan het sbo en so. Bij het bao wordt 2F het vaakst gescoord, terwijl voor sbo en so geldt dat zij vooral <1F scoren.

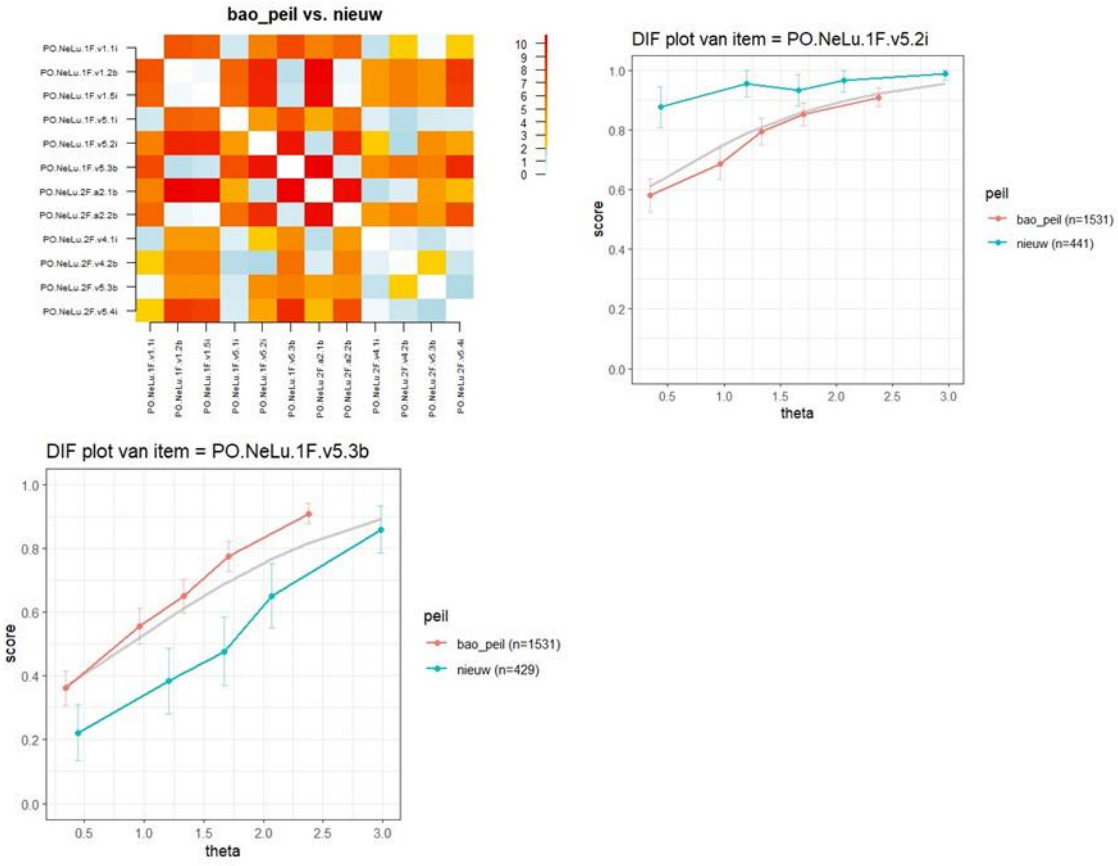
6.2 Trendanalyses

6.2.1 Luistervaardigheid

In paragraaf 6.1 is uiteengezet hoe de data uit de bao-peiling uit 2017 en uit de so/sbo-peiling uit 2018 zijn gebruikt om de prestatiestandaarden over te zetten. Naast het overzetten van de standaarden zijn deze data eveneens gebruikt om te evalueren of er prestatieverschillen tussen de jaargangen zijn. Hiertoe zijn onderstaande leerlinggroepen vergeleken met elkaar:
bao-leerlingen uit de huidige peiling vs. leerlingen uit de bao-peiling 2017
sbo-leerlingen uit de huidige peiling vs. sbo-leerlingen uit de so/sbo-peiling 2018
so-leerlingen uit de huidige peiling vs. so-leerlingen uit de so/sbo-peiling 2018

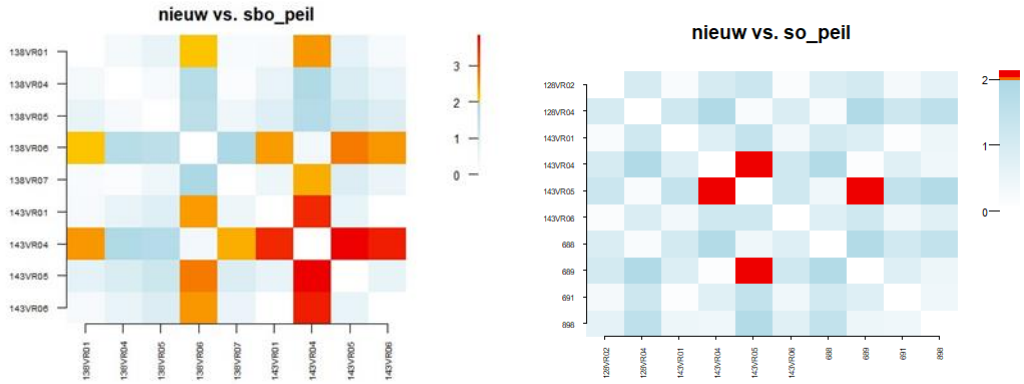
Middels concurrent calibration zijn alle leerlingen uit de verschillende peilingen op een schaal gezet. De data uit de drie jaargangen zijn samengevoegd, en op deze data is een Rasch-model gepast. De kalibratie is uitgevoerd met Conditional Maximum Likelihood (CML) en de vaardigheden zijn geschat met Weighted Maximum Likelihood (WML).
Eerst is DIF geëvalueerd tussen de verschillende peilingen. Zoals in paragraaf 4.5 al werd vermeld, werd veel DIF gevonden tussen de items uit de bao-peiling en dezelfde items uit de huidige peiling (zie Figuur 6.2.1).

Figuur 6.2.1: Differential item pair functioning bao en twee DIF-plots



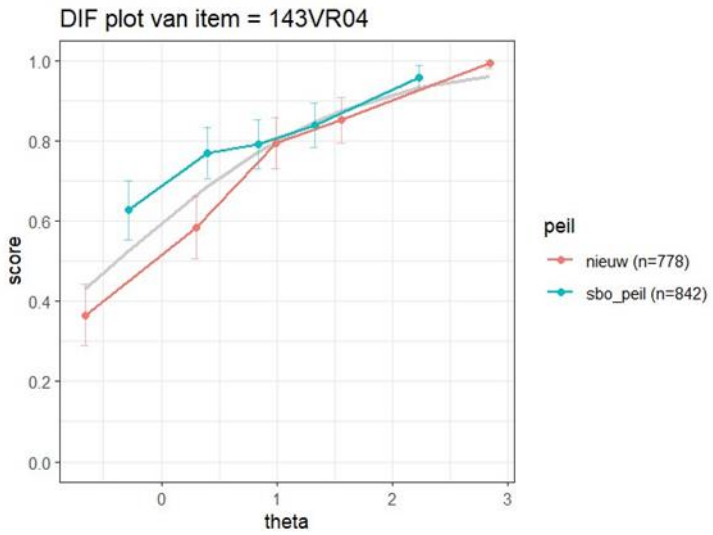
Voor de DIF kon geen verklaring worden gevonden. Er is gecontroleerd of er wellicht coderingsfouten zijn gemaakt, of de items anders gelabeld waren, of dat er wellicht sprake kon zijn van volgorde-effecten. Niets hiervan leek echter aan de orde te zijn.
Voor de so/sbo-leerlingen uit de huidige peiling in vergelijking met de leerlingen uit de vorige so/sbo-peiling leek er geen sterke indicatie van DIF te zijn. De plots voor differential item pair functioning worden weergegeven in Figuur 6.2.2.

Figuur 6.2.2: Differential item pair functioning sbo en so



Het item 143VR04, met de hoogste indicatie van DIF, werd visueel geïnspecteerd, maar liet geen extreme DIF zien (zie Figuur 6.2.3).

Figuur 6.2.3: Item 143VR04



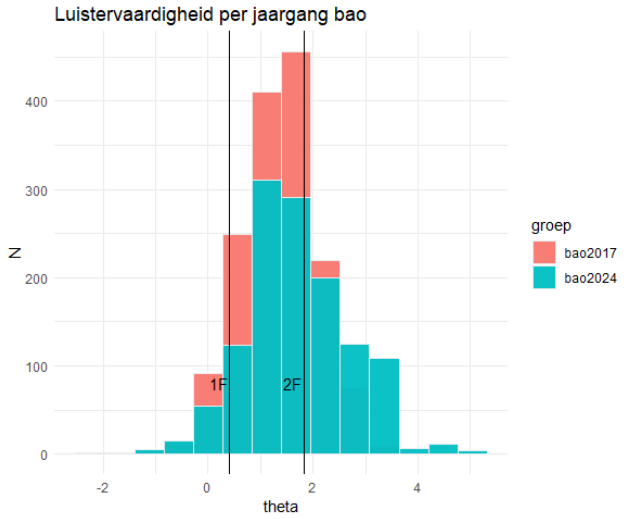
Na de evaluatie van DIF is besloten om geen items los te koppelen of weg te laten. Voor de bao-items zou bijna ieder item losgekoppeld moeten worden. Dit zou een trendvergelijking tussen bao-leerlingen uit de huidige peiling en de leerlingen uit de bao-peiling weinig tot niet betrouwbaar maken. Zodoende worden de bao-items meegenomen, maar dienen de resultaten van de ANOVA's ten aanzien van bao met zorg geïnterpreteerd te worden.
In Tabel 6.2.1 worden de resultaten van de ANOVA's gepresenteerd.

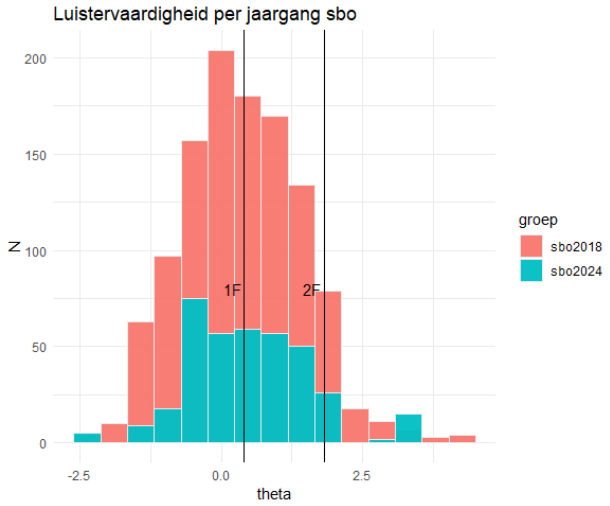
Tabel 6.2.1: Resultaten ANOVA's per schooltype

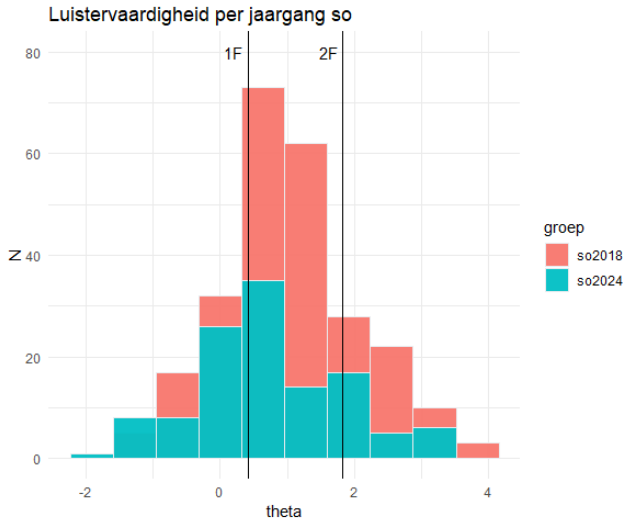
	Vorige peiling			Huidige peiling						
Schooltype	*N*	*M*	*SD*	*N*	*M*	*SD*	*df*	*F*	*p*	*eta2*
bao	1531	1,346	0,733	1258	1,733	0,968	1,2787	144,4	<0,001	0,05
sbo	1141	0,368	1,053	373	0,423	1,056	1,1512	0,779	0,378	<0,001
so	252	1,016	1,019	120	0,732	1,081	1,370	6,084	0,014	0,02

Uit Tabel 6.2.1 blijkt dat de luistervaardigheid van de bao-leerlingen vooruit is gegaan ten opzichte van de peiling uit 2017. Hier is een middelgroot effect gevonden. Bij de so-leerlingen is een klein effect gevonden: de leerlingen scoorden in de huidige peiling lager dan in de peiling van 2018. Tot slot is er geen verschil gevonden tussen de prestaties van de sbo-leerlingen. In Figuur 6.2.4 wordt de vaardigheidsverdeling van alle groepen weergegeven.

Figuur 6.2.4: Verdeling luistervaardigheid







In aanvulling op bovenstaande is de *concurrent calibration* nogmaals herhaald. Ditmaal zonder de data uit de bao-peiling. Hierdoor is geen trendvergelijking mogelijk tussen bao-leerlingen uit de verschillende

jaargangen, maar door het weglaten van de data wordt de schaal (waarin immers alle groepen zijn meegenomen) niet beïnvloed door de DIF. De resultaten worden gepresenteerd in Tabel 6.2.2.

Tabel 6.2.2: Resultaten ANOVA's per schooltype

	Vorige peiling			Huidige peiling						
schooltype	*N*	*M*	*SD*	*N*	*M*	*SD*	*df*	*F*	*p*	*eta2*
sbo	1141	0,394	1,056	373	0,345	1,068	1,1512	0,633	0,427	<0,001
so	252	1,047	1,025	120	0,672	1,085	1,370	6,084	0,001	0,03

De conclusies blijven onveranderd. Bij de sbo-leerlingen zijn geen verschillen gevonden tussen de jaargangen. Bij de so-leerlingen is een klein (tot medium) effect gevonden: so-leerlingen scoorden hoger bij de vorige peiling.
Tot slot is opnieuw een volledige kalibratie uitgevoerd, ditmaal met de MML-methode, in plaats van de CML-methode. Jaargang en schooltype zijn hierbij als marginaal meegenomen in de kalibratie. Daarna zijn opnieuw ANOVA's uitgevoerd om te zien of deze andere kalibratie heeft geleid tot andere conclusies dan hierboven gepresenteerd. De resultaten zijn vergelijkbaar met de resultaten uit de voorgaande ANOVA's.

Tabel 6.2.3: Resultaten ANOVA's per schooltype na MML-kalibratie

	vorige peiling			Huidige peiling						
Schooltype	*N*	*M*	*SD*	*N*	*M*	*SD*	*df*	*F*	*p*	*eta2*
bao	1531	-0,004	0,734	1258	0,378	0,968	1,2787	140,2	<0,001	0,05
sbo	1141	-0,807	0,967	373	-1,017	1,052	1,1512	12,71	<0,001	0,001
so	252	-0,254	0,932	120	-0,612	1,087	1,370	10,77	0,001	0,03

Concluderend, de luistervaardigheid van de sbo-leerlingen is hetzelfde gebleven, de vaardigheid van de so-leerlingen is wat achteruitgegaan, en de vaardigheid van de bao-leerlingen is redelijk stevig vooruitgegaan.

6.2.2 Spreekvaardigheid

De data van de vorige so/sbo-peiling zijn gebruikt om te evalueren of er prestatieverschillen zijn tussen de jaargangen. Hiertoe zijn onderstaande leerlinggroepen vergeleken met elkaar:

- sbo-leerlingen uit de huidige peiling vs. sbo-leerlingen uit de so/sbo-peiling 2018
- so-leerlingen uit de huidige peiling vs. so-leerlingen uit de so/sbo-peiling 2018

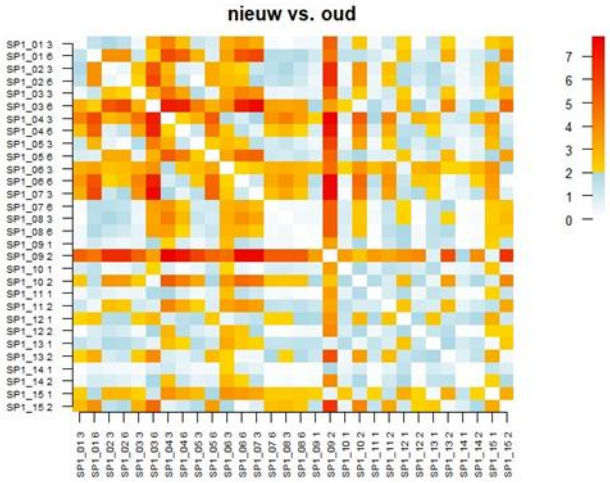
Middels *concurrent calibration* zijn de leerlingen uit de verschillende peilingen op een schaal gezet. De data uit de twee jaargangen zijn samengevoegd, en op deze data is een OPLM gepast, op dezelfde manier als is uitgevoerd bij de vaardigheidsschatting. Dit is uiteengezet in paragraaf 4.5. In de vorige peiling werd geen IRT toegepast op de spreektaken, aangezien de leerlingen – anders dan in de huidige peiling - de spreektaken volgens een volledig design uitvoerden (i.e., alle leerlingen deden alle taken). Door de data uit de vorige peiling op eenzelfde manier als de data uit de huidige peiling in een OPLM te zetten, is het mogelijk geweest om een vergelijking te maken.

Allereerst is geëvalueerd of er mogelijk sprake is van DIF tussen de jaargangen ten aanzien van de naverteltaak. De differential item pair functioning plots in Figuur 6.2.5 laten met name DIF zien bij een item: SP1_09. In mindere mate is ook DIF geconstateerd bij SP1_01, SP1_03 en SP1_05. Bij het inspecteren van de item characteristic curve (ICC; zie Figuur 6.2.6) van item SP1_09 leek het er inderdaad op dat – gegeven dezelfde vaardigheid in de hogere vaardigheidsregionen – de posterior verwachte score op het item iets hoger was voor de leerlingen uit de vorige peiling dan voor de leerlingen uit de huidige peiling. Bij nadere inspectie van de aantallen per antwoordcategorie bleek dat het betreffende item per jaargang sterk varieerde in verdeling. In Tabel 6.2.4 kan worden opgemerkt dat bij de vorige peiling antwoordcategorie 2 met afstand het meest werd gekozen. In de huidige peiling werd categorie 1 het meest gekozen. Dit werd in mindere mate ook gezien bij items SP1_01, SP1_03 en SP1_05. De DIF is hier per antwoordcategorie beschouwd. De stap van 1 punt verdienen naar 2 punten verdienen, leek – gegeven een gelijke vaardigheid – lastiger in de vorige peiling dan in de huidige peiling.

Tabel 6.2.4: Scoreverdeling item SP1_09 per peilingsjaargang

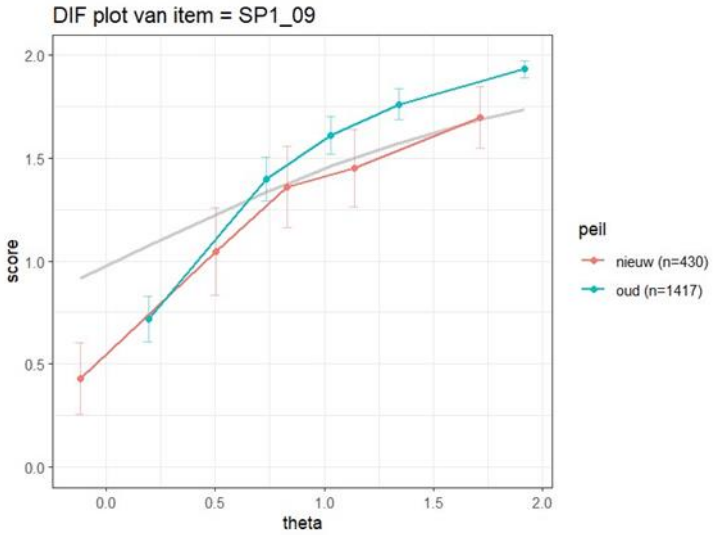
item_id	item_score	N oud	N nieuw
SP1_09	0	122	65
SP1_09	1	487	215
SP1_09	2	808	150

Figuur 6.2.5: Differential item pair functioning plots vorige peiling vs. huidige peiling



In Figuur 6.2.6 is het verschil in functioneren van het hele item (in plaats van per antwoordcategorie) per subgroep afgebeeld. Onderaan de streep lijkt de DIF voor het item niet extreem.

Figuur 6.2.6: Groepsverschillen item SP1_09



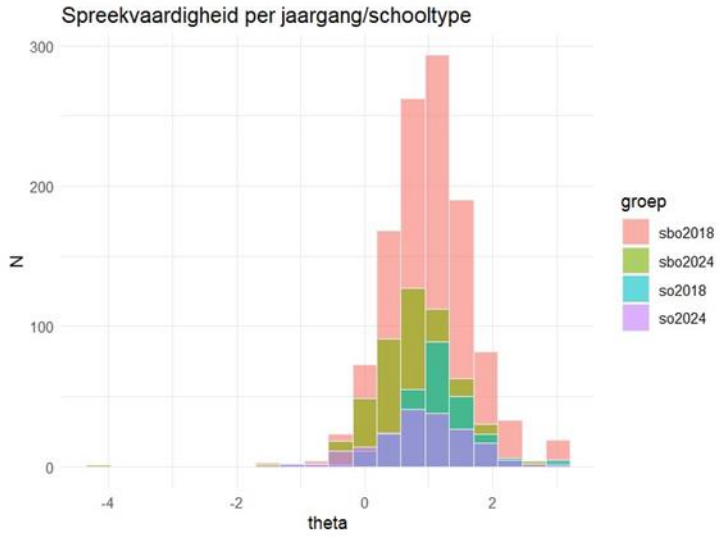
Desalniettemin is de IRT-kalibratie op twee manieren uitgevoerd. Bij de eerste manier is geen acht op de gevonden DIF geslagen, en gaat het item SP1_09 zijnde hetzelfde item in beide jaargangen mee in de IRT-kalibratie. Bij de tweede manier wordt item SP1_09 losgekoppeld. Bij deze kalibratie worden SP1_09 in 2018 en SP1_09 in 2024 beschouwd als twee verschillende items.
Op basis van de WLE-vaardigheidsschattingen die voortvloeien uit beide modellen zijn ANOVA's uitgevoerd. De resultaten van de ANOVA's worden gepresenteerd in Tabel 6.2.5.

Tabel 6.2.5: Resultaten ANOVA's per schooltype

	Vorige peiling			Huidige peiling						
schooltype	*N*	*M*	*SD*	*N*	*M*	*SD*	*df*	*F*	*p*	*eta2*
Manier 1										
Sbo	1153	1,022	0,644	508	0,832	0,665	1,1659	25,79	<0,001	0,02
So	264	1,141	0,568	184	0,902	0,710	1,446	15,49	<0,001	0,03
Manier 2										
Sbo	1153	1,011	0,643	508	0,834	0,667	1,1659	29,89	<0,001	0,02
So	264	1,129	0,567	184	0,905	0,713	1,446	13,74	<0,001	0,03

In Tabel 6.2.5 is weergegeven dat leerlingen in zowel het sbo als het so achteruit zijn gegaan ten aanzien van de spreekvaardigheid, waarbij het voor de resultaten van de ANOVA's niet uit maakt of item SP1_09 losgekoppeld werd of niet. Voor de helderheid is de ANOVA beschouwd waarbij het item niet is losgekoppeld (manier 1). Uit de resultaten blijkt dat er een klein (tot middelgroot) effect is voor zowel sbo als so tussen de jaargangen. De verdeling van de spreekvaardigheid over de twee peilingen en de twee populaties is visueel weergegeven in Figuur 6.2.7.

Figuur 6.2.7: Verdeling spreekvaardigheid sbo/so 2018-2024



Concluderend, de sbo- en so-leerlingen hebben een kleine achteruitgang laten zien ten aanzien van de spreekvaardigheid.

Trendvergelijking bao

De spreekvaardigheid van de so/sbo-populaties uit beide jaargangen konden direct met elkaar worden vergeleken, aangezien een spreektaak voorkwam in beide jaargangen. Deze vergelijking kon niet worden gedaan voor de bao-populatie. De leerlingen uit de vorige bao-peiling (Langen et al., 2017) deden andere spreektaken dan de bao-leerlingen uit de huidige peiling. Zodoende is een directe vergelijking zoals bij de so/sbo-populaties niet mogelijk. De informatie die wel met elkaar vergeleken kan worden, is de verdeling behaalde referentieniveaus. Uit beide bao-jaargangen is de verdeling behaalde referentieniveaus bekend. Deze worden weergegeven in Tabel 6.2.6.

Tabel 6.2.6: Behaalde referentieniveaus bao

	peiling2017 (N)	peiling2017 (%)	peiling2024 (N)	peiling2024 (%)
<1F	110	8,1	266	18,4
1F	406	29,7	808	55,9
2F	849	62,2	372	25,7

Op het eerste gezicht is het duidelijk dat het merendeel van de bao-leerlingen 2F behaalde in de peiling uit 2017. Daarentegen behaalde het merendeel 1F in de huidige peiling. Een Chi-kwadraattoets wijst uit dat het verschil in de verdeling van behaalde referentieniveaus significant is ($\chi^2(2) = 382{,}17$; $p<0{,}001$). Aan de hand van effectgrootte Phi werd de grootte bepaald van het effect. Een Phi > 0,5 wordt geïnterpreteerd als een groot effect (Cohen, 1988). De Phi was 0,530 in bovenstaande verdeling. Daarmee wordt de indruk gewekt dat de bao-leerlingen volgens bovenstaande sterk achteruitgegaan zijn ten aanzien van de spreekvaardigheid. Bij bovenstaande stelling is echter enige nuance geboden. Bij deze poging tot vergelijking van de vaardigheid tussen de twee peilingsjaargangen is het alleen mogelijk geweest om de referentieniveaus te vergelijken. De verdeling behaalde referentieniveaus is een extreem beknopte samenvatting van de vaardigheid, waarbij erg veel nuance verloren gaat. Daarnaast zijn de behaalde referentieniveaus uit beide jaargangen gebaseerd op verschillende taken. Er is niet gevalideerd of de taken uit de huidige peiling tot dezelfde resultaten komen als bij de vorige peiling. Eventuele beoordelaarseffecten kunnen niet worden geëvalueerd. In de vorige peiling is enkel de spreekvaardigheid bij de bao-populatie gemeten, in de huidige peiling hebben beoordelaars afwisselend leerlingen uit zowel de bao-, sbo- als so-populatie beoordeeld. Het is niet te achterhalen of beoordelaars in de huidige peiling bijvoorbeeld extra streng zijn geweest in vergelijking met de vorige peiling. Zodoende kan bovenstaande conclusie ten aanzien van de trendverschuiving niet stellig worden getrokken.

6.2.3 Gespreksvaardigheid

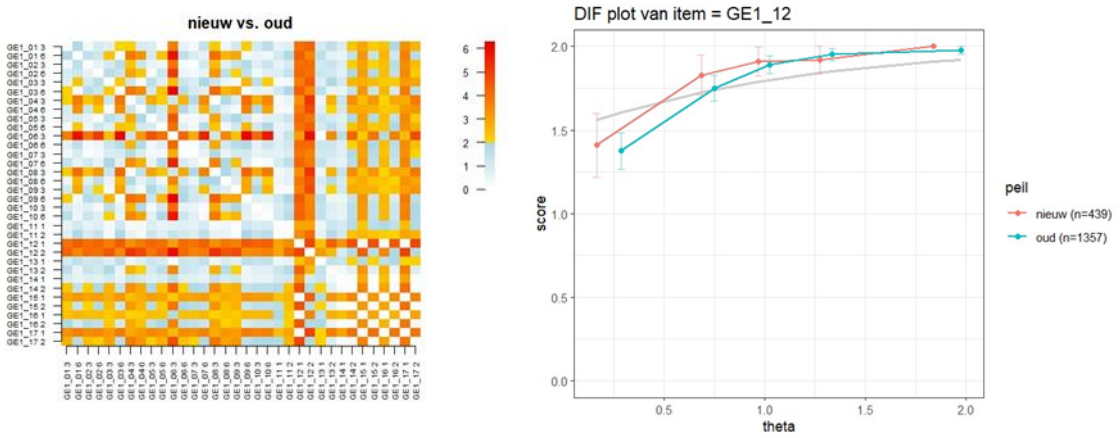
De data van de vorige so/sbo-peiling zijn gebruikt om te evalueren of er prestatieverschillen zijn tussen de jaargangen. Hiertoe zijn onderstaande leerlinggroepen vergeleken met elkaar:

- sbo-leerlingen uit de huidige peiling vs. sbo-leerlingen uit de so/sbo-peiling 2018
- so-leerlingen uit de huidige peiling vs. so-leerlingen uit de so/sbo-peiling 2018

Middels *concurrent calibration* zijn de leerlingen uit de verschillende peilingen op een schaal gezet. De data uit de twee jaargangen zijn samengevoegd, en op deze data is een OPLM gepast, op dezelfde manier als is uitgevoerd bij de vaardigheidsschatting, uiteengezet in paragraaf 4.5. In de vorige peiling is geen IRT uitgevoerd op de gesprekstaken, aangezien de leerlingen – anders dan in de huidige peiling - de gesprekstaken volgens een volledig design uitvoerden (i.e., alle leerlingen deden alle taken). Door de data uit de vorige peiling op eenzelfde manier als de data uit de huidige peiling in een OPLM te zetten, is het mogelijk geweest om een vergelijking te maken.
Allereerst is geëvalueerd of er mogelijk sprake is van DIF tussen de jaargangen ten aanzien van het samenwerkingsspel. De differential item pair functioning plots in Figuur 6.2.8 laten met name DIF zien bij een item: GE1_12. Bij het inspecteren van de ICC leek het er inderdaad op dat de posterior verwachte score op het item hoger was voor de leerlingen uit de huidige peiling dan voor de leerlingen uit de vorige peiling.

Echter zijn de verschillen tussen de twee curves niet te classificeren als extreem, en is besloten om het item mee te nemen in de trendanalyse.

Figuur 6.2.8: Differential item pair functioning en ICC item met meeste DIF



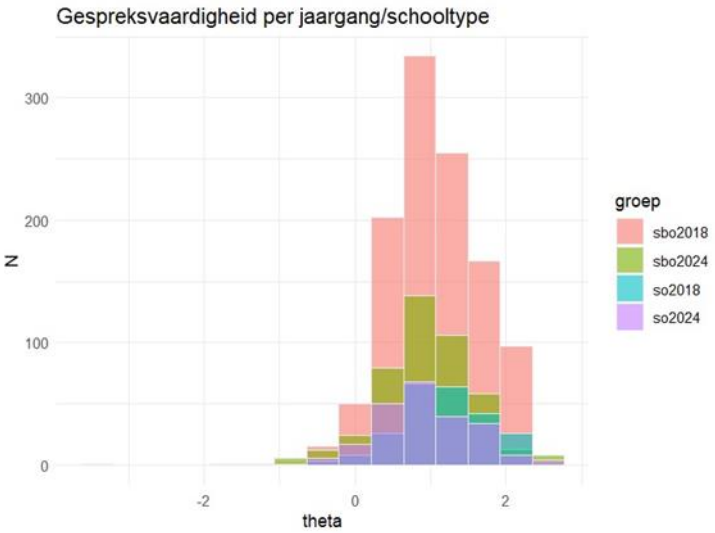
Op basis van de WLE-vaardigheidsschattingen die voortvloeien uit de OPLM zijn ANOVA's uitgevoerd. De resultaten van de ANOVA's worden gepresenteerd in Tabel 6.2.7.

Tabel 6.2.7: Resultaten ANOVA's per onderwijstype

	Vorige peiling			Huidige peiling						
Schooltype	*N*	*M*	*SD*	*N*	*M*	*SD*	*df*	*F*	*p*	*eta2*
Sbo	1122	1,052	0,598	446	0,935	0,644	1,1566	11,78	<0,001	<0,001
So	235	1,172	0,591	230	0,916	0,661	1,446	15,49	<0,001	0,04

Uit de resultaten blijkt dat de sbo-leerlingen uit de vorige peiling en de huidige peiling niet verschillen van elkaar in gespreksvaardigheid (eta-squared <0,001). Voor het so is gevonden leerlingen een hogere gespreksvaardigheid hadden in de vorige peiling dan in de huidige peiling. De effectgrootte is klein tot middelgroot.
De verdeling van de gespreksvaardigheid voor de verschillende populaties is weergegeven in Figuur 6.2.9.

Figuur 6.2.9: Verdeling gespreksvaardigheid sbo/so 2018-2024



Concluderend, de so-leerlingen zijn hetzelfde blijven scoren als bij de vorige peiling, de sbo-leerlingen lieten een achteruitgang zien.

Trendvergelijking bao

Tot slot is er een trendvergelijking gedaan voor de bao-leerlingen. Hier deed zich echter hetzelfde obstakel voor als bij de spreektaken. Anders dan bij de vorige so/sbo-peiling zat er namelijk geen gesprekstaak in de vorige bao-peiling (van Langen et al., 2017) die in wel in de huidige peiling zat. Hierdoor is het niet mogelijk om een trendvergelijking te doen tussen de populatie van de vorige bao-peiling en de bao-populatie van de huidige peiling, zoals dat hierboven is gedaan voor de sbo- en so-populaties.
Van de vorige peiling is de proportie behaalde referentieniveaus bekend, net zoals die bekend zijn in de huidige peiling. Deze worden weergegeven in Tabel 6.2.8.

Tabel 6.2.8: Behaalde referentieniveaus bao

	peiling2017 (N)	peiling2017 (%)	peiling2024 (N)	peiling2024 (%)
<1F	183	13,3%	458	31,3
1F	521	37,8%	358	24,5
2F	676	48,9%	647	44,2

Het aantal bao-leerlingen dat 2F behaalde was redelijk vergelijkbaar tussen de jaargangen. Echter is duidelijk op te merken dat leerlingen relatief gezien vaker <1F scoorden in de huidige peiling, vergeleken met de vorige peiling. Een Chi-kwadraattoets wijst uit dat het verschil in de verdeling van behaalde referentieniveaus significant is ($\chi^2(2) = 146{,}54$; $p<0{,}001$). Aan de hand van effectsize Phi werd de grootte van het effect bepaald. De Phi was 0,227 en was daarmee een klein tot middelgroot effect. Daarmee wordt de indruk gewekt dat de bao-leerlingen ten aanzien van de gespreksvaardigheid achteruit zijn gegaan. Dezelfde nuance dient hier echter gemaakt te worden als bij de spreektaken. De vergelijking kan niet zonder meer gedaan worden.

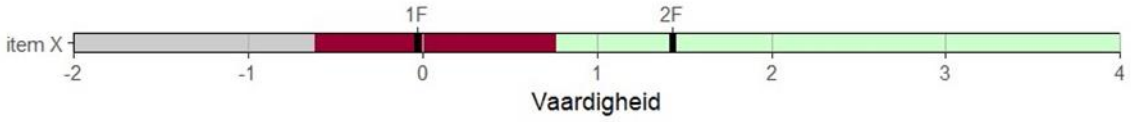
6.3 Beheersing vaardigheid op itemniveau

Aan de hand van de geschatte IRT-modellen kan worden voorspeld wat de kans is op de maximaal te behalen score van een item, gegeven de vaardigheid van een leerling. Er kan ook worden nagegaan welk niveau van vaardigheid benodigd is om een bepaalde kans te hebben op het correct maken van een item. Beheersing wordt hier als volgt gedefinieerd:

- Geen beheersing: kans op maximaal te behalen punten kleiner dan 50 procent
- Matige beheersing: kans op maximaal te behalen punten ligt tussen de 50 en 80 procent
- Goede beheersing: kans op maximaal te behalen punten is groter dan 80 procent

Voor elk item wordt dus uitgerekend hoeveel vaardigheid nodig is om 50 of 80 procent kans te hebben op het behalen van de maximale score van het item. Vervolgens wordt de beheersing per item grafisch weergegeven zoals in Figuur 6.3.1.

Figuur 6.3.1: Moeilijkheidsgraad van een item in relatie tot referentieniveaus 1F en 2F op de latente vaardigheidsschaal



Het groen gemarkeerde gedeelte in Figuur 6.3.1 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. Vervolgens worden de prestaties van leerlingen gerelateerd aan de kenmerken van de taakuitvoering. Daartoe wordt op itemniveau in kaart gebracht wat leerlingen op de grenzen <1F/1F en 1F/2F goed, matig en niet beheersen. In het gegeven voorbeeld beheersen leerlingen met een vaardigheid rondom de 1F-cesuur het item matig. Leerlingen met een vaardigheid rondom de 2F-cesuur of hoger, beheersen dit item goed. In aanvulling op de cesuren <1F/1F en 1F/2F wordt ook stilgestaan bij de beheersing van leerlingen die tot het 10^{e}, 50^{e} en 90^{e} percentiel (hierna, resp.: P10-, P50-, P90-leerling) behoren van alle leerlingen wiens vaardigheden zijn geschat in deze peiling.
Gegeven de itemparameters en de vaardigheidsschatting voor een leerling, is het mogelijk om voor deze leerling de posterior kans te berekenen op het behalen van de maximale score op alle items uit de itembank. Dit geldt dus ook voor de items die de leerling niet gemaakt heeft. De maximale scores van alle items van de itembank bij elkaar opgeteld geven een maximaal te behalen score op het domein. De posterior kansen op de maximale score van alle items – gegeven de vaardigheid van de leerling – bij elkaar opgeteld geven de verwachte score van de leerling op het domein. De verwachte domeinscore/maximale domeinscore geeft dan de bankscore. Als een leerling een bankscore van 75% behaalt, betekent dat dat de betreffende leerling – gegeven diens vaardigheid en gegeven de itemparameters – naar verwachting een score behaalt die 75% is van de maximaal te behalen score, had deze leerling alle opgaven uit de itembank gemaakt.

6.3.1 Luistervaardigheid

In Tabel 6.3.1 worden de bankscores – ten aanzien van luistervaardigheid - weergegeven voor P10-, P50- en P90-leerlingen, evenals voor leerlingen die precies op de cesuren van de referentieniveaus zitten. De P10-

leerling heeft een bankscore van 40,8% (i.e., de P10-leerling behaalt naar verwachting een score die 40,8% is van de maximale te behalen score in de itembank, had deze leerling alle opgaven gemaakt). De P90-leerling heeft een bankscore van bijna 90%. De mediane leerling (P50) zit tussen de 1F- en 2F-cesuur in, en behaalt daarmee op basis van kans wel 1F, maar geen 2F.

Tabel 6.3.1: Bankscores luistervaardigheid

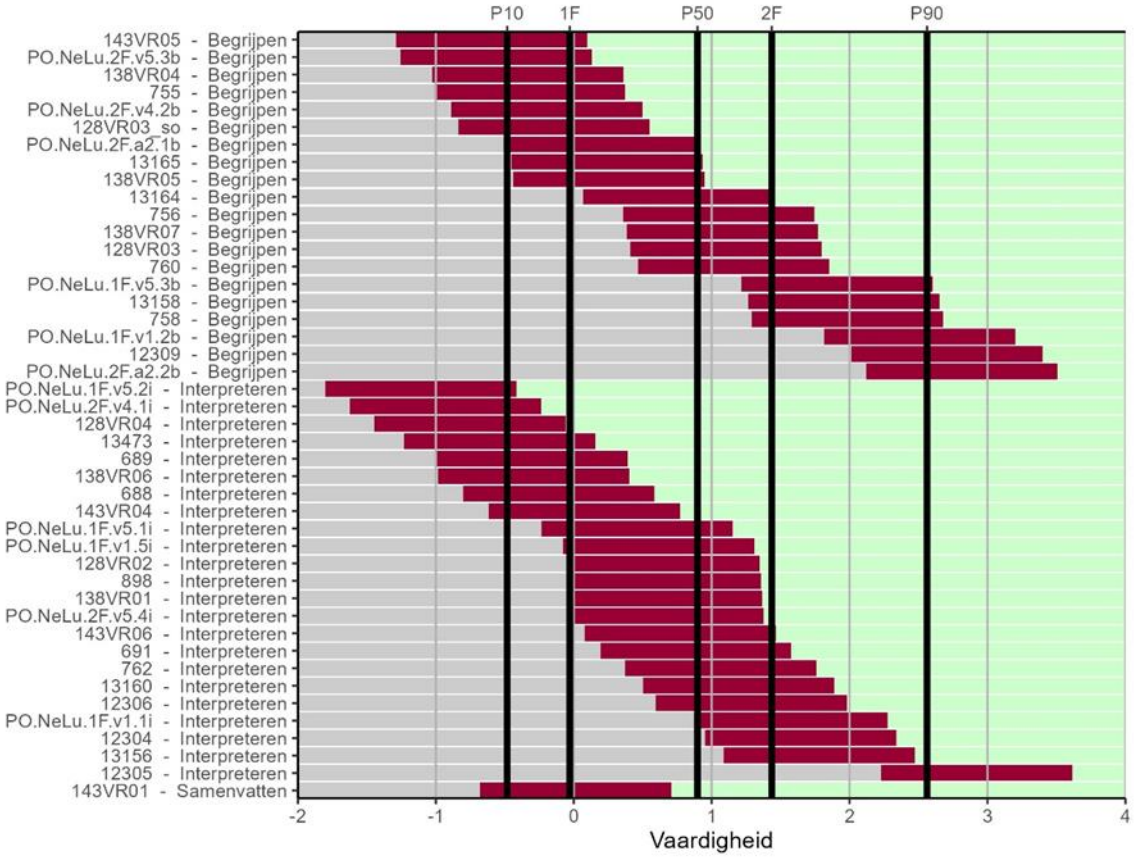
vaardigheidsniveau	bankscore%
P10	40,8
1F	49,9
P50	67,9
2F	76,7
P90	89,6

In Figuur 6.3.2 t/m 6.3.5 wordt de beheersing op itemniveau gepresenteerd. De items worden viermaal op verschillende subdomeinen gerangschikt.

Kenmerken van de taakuitvoering

De items van luistervaardigheid zijn voor het duiden van de kenmerken van de taakuitvoering onderverdeeld in *begrijpen, interpreteren* en *samenvatten*.

Figuur 6.3.2: Relatie tussen de kenmerken van de taakuitvoering en referentieniveaus 1F en 2F



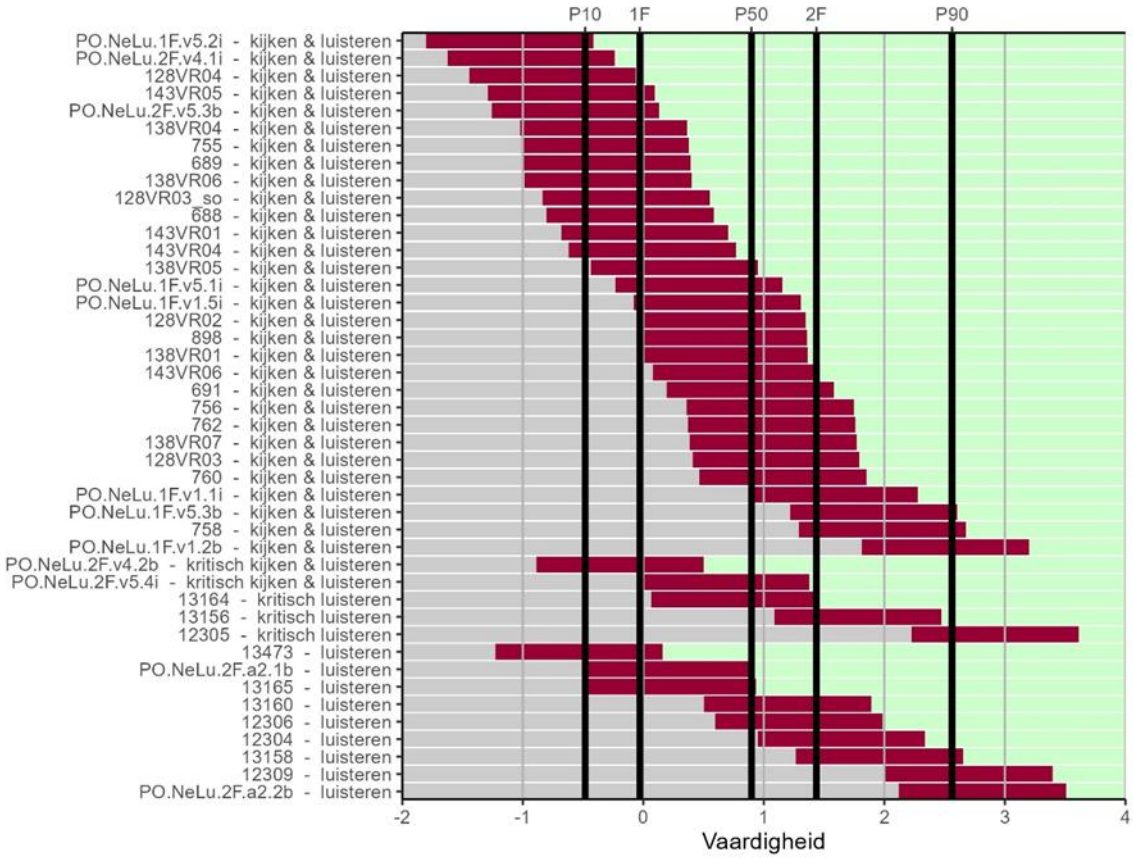
Figuur 6.3.2 laat zien dat er weinig verschil in beheersing is tussen begrijpen en interpreteren als kenmerk van de taakuitvoering. De leerlingen beheersen de interpreteren-items iets beter dan de begrijpen-items:

- De bankscore voor de P50-leerlingen op de begrijpen-items is 62,9%
- De bankscore voor de P50-leerlingen op de interpreteren-items is 70,6%

Aard van de activiteit

De items van luistervaardigheid zijn voor het duiden van de aard van de activiteit onderverdeeld in *kijken & luisteren*, *kritisch luisteren*, *kritisch kijken en luisteren* en *luisteren*.

Figuur 6.3.3: Relatie tussen de aard van de activiteit en referentieniveaus 1F en 2F



De 'kijken & luisteren'-items zijn oververtegenwoordigd wat betreft de aard van de activiteit. Relatief gezien worden de 'kijken & luisteren'-items wat beter beheerst dan de luisteren-items:

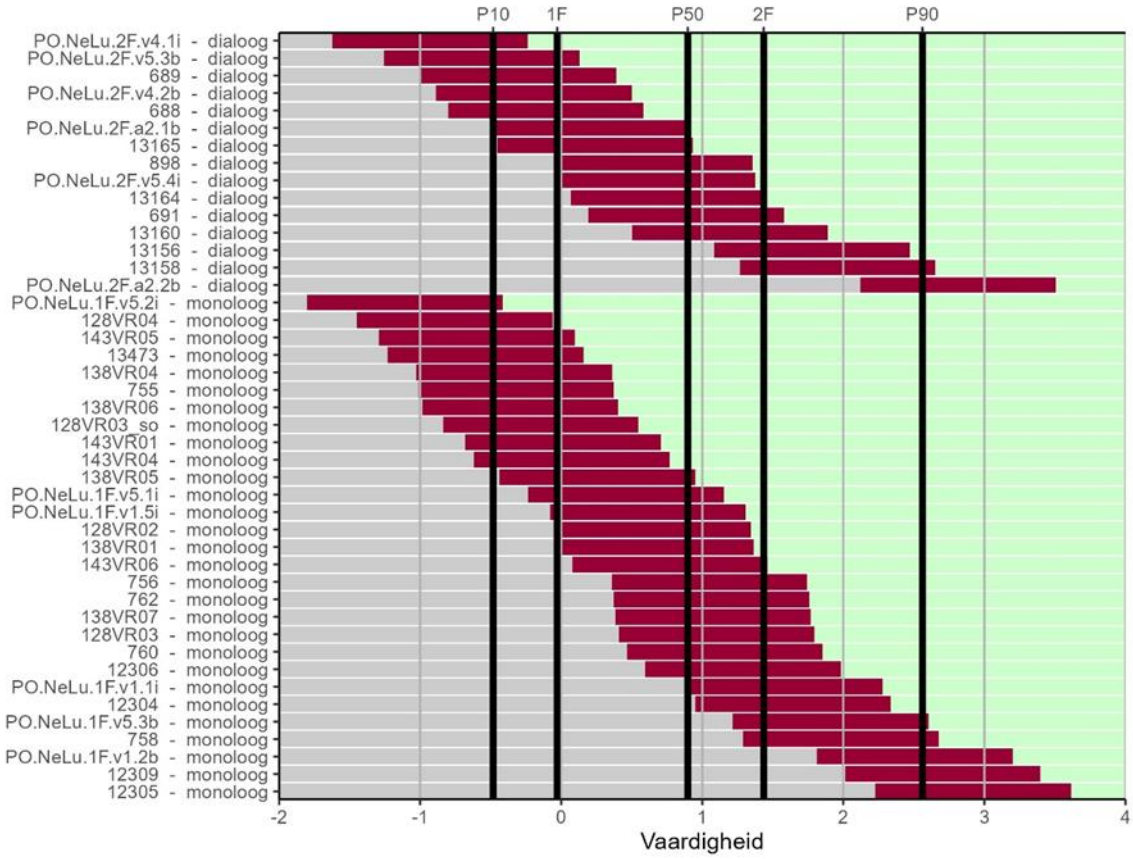
- De bankscore voor de P50-leerlingen op de kijken & luisteren-items is 72,6%
- De bankscore voor de P50-leerlingen op de luisteren-items is 55,9%
- De bankscore voor de P50-leerlingen op de kritisch luisteren-items is 45,3%
- De bankscore voor de P50-leerlingen op de kritisch kijken & luisteren-items is 78,4%

Let wel, de laatste twee categorieën bestaan uit respectievelijk drie en twee items, dus is het lastig om op basis hiervan generalisaties te doen.

Gespreksvorm

De items van luistervaardigheid zijn voor het duiden van de gespreksvorm onderverdeeld in *dialoog* en *monoloog*.

Figuur 6.3.4: Relatie tussen de gespreksvorm en referentieniveaus 1F en 2F



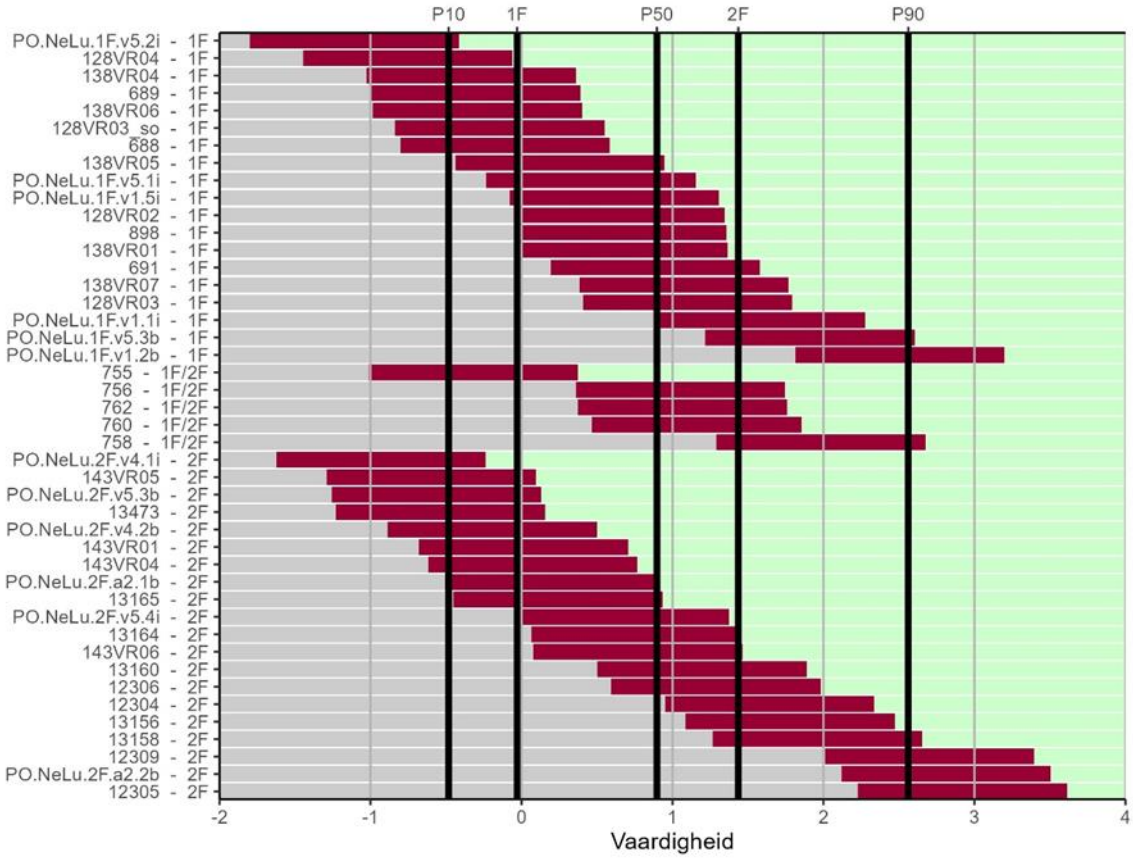
De dialoog- en monoloog-items worden in vergelijkbare mate beheerst:

- De bankscore voor de P50-leerlingen op de dialoog-items is 69,8%
- De bankscore voor de P50-leerlingen op de monoloog-items is 66,2%

Tekstkenmerken

De items van luistervaardigheid zijn voor het duiden van de tekstkenmerken onderverdeeld in 1F, 1F/2F en 2F.

Figuur 6.3.5: Relatie tussen de tekstkenmerken en referentieniveaus 1F en 2F



De leerlingen beheersen de 1F-items iets beter dan de 1F/2F- en 2F-items. De 1F/2F-items en de 2F-items worden in meer vergelijkbare mate beheerst:

- De bankscore voor de P50-leerlingen op de 1F-items is 71,4%
- De bankscore voor de P50-leerlingen op de 1F/2F-items is 62,8%
- De bankscore voor de P50-leerlingen op de 2F-items is 65,1%

6.3.2 Spreekvaardigheid

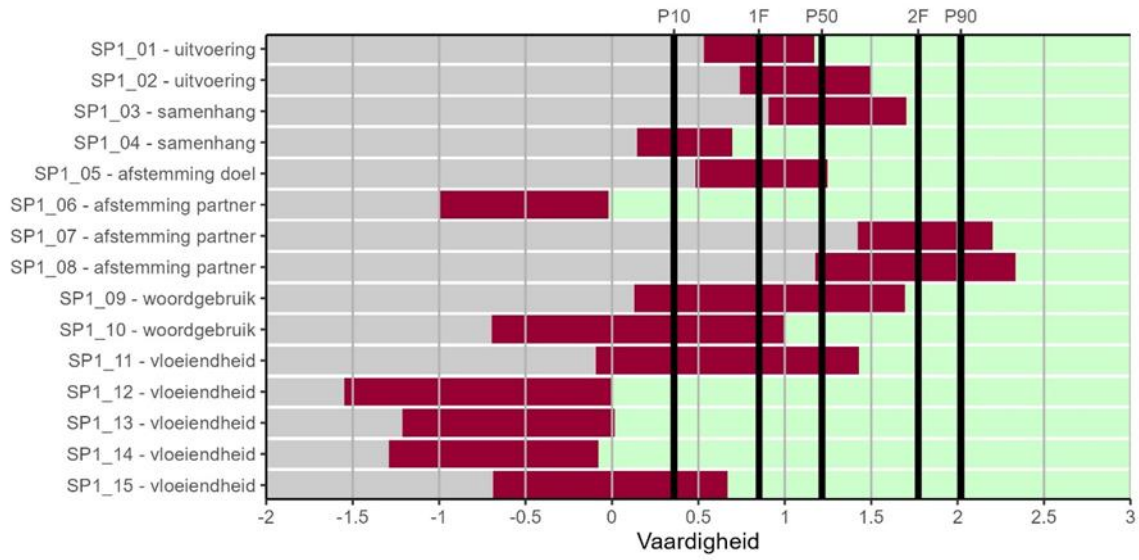
In Tabel 6.3.2 worden de bankscores – ten aanzien van spreekvaardigheid - weergegeven voor P10-, P50- en P90-leerlingen, evenals voor leerlingen die precies op de cesuren van de referentieniveaus zitten. De P10-leerling heeft een bankscore van 48,7%. De P90-leerling heeft een verwachte score van 90% van de maximaal te behalen score. De mediane leerling (P50) zit tussen de 1F- en 2F-cesuren in, en behaalt daarmee op basis van kans wel 1F, maar geen 2F.

Tabel 6.3.2: Bankscores spreekvaardigheid

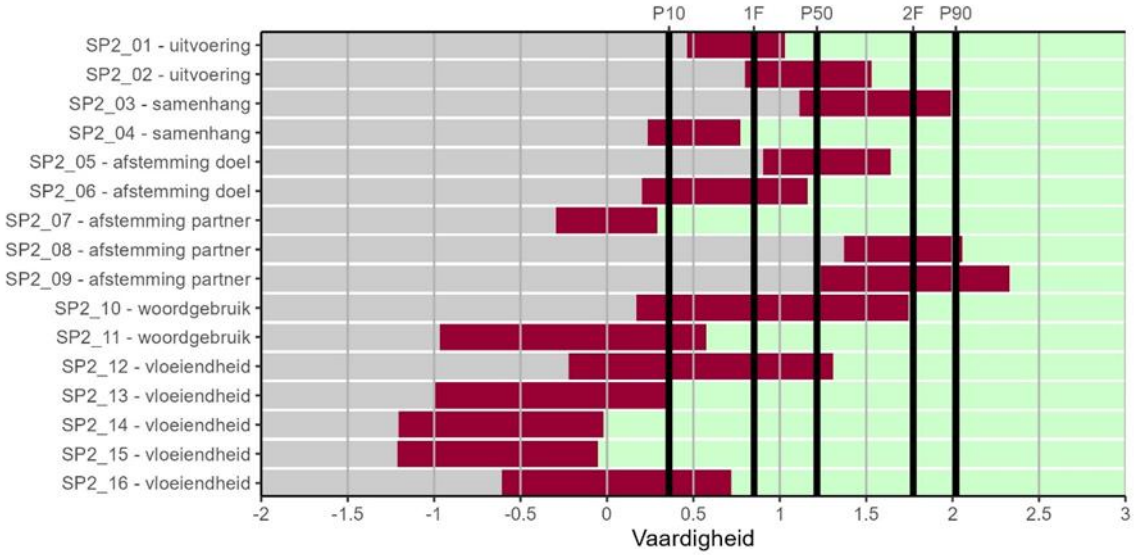
vaardigheidsniveau	bankscore%
P10	48,7
1F	64,6
P50	74,4
2F	86,3
P90	90,5

In Figuur 6.3.6 en 6.3.7 wordt de beheersing op itemniveau gepresenteerd, voor respectievelijk de naverteltaak en het betoog. De twee taken laten in beheersing een sterk vergelijkbaar patroon zien. Zo worden de items op vloeiendheid, verstaanbaarheid en grammaticale beheersing zeer goed beheerst, ook door P10-leerlingen. Daarentegen worden twee afstemming op gesprekspartner-items het minst goed beheerst in beide taken (SP1_07/SP2_08 (gesprekspartner boeien) en SP1_08/SP2_09 (non-verbale communicatie)).

Figuur 6.3.6: Beheersing van de vaardigheid op itemniveau - naverteltaak



Figuur 6.3.7: Beheersing van de vaardigheid op itemniveau - betoog



In Tabel 6.3.3 zijn de bankscores per spreektaak gepresenteerd. Hier kan vooral uit worden opgemaakt dat de twee taken in vergelijkbare mate werden beheerst.

Tabel 6.3.3: Bankscores spreektaken

	bankscore%	
vaardigheidsniveau	**naverteltaak**	**betoog**
P10	50,5	47
1F	65,7	63,7
P50	75,1	73,8
2F	86,6	86,1
P90	90,5	90,5

6.3.3 Gespreksvaardigheid

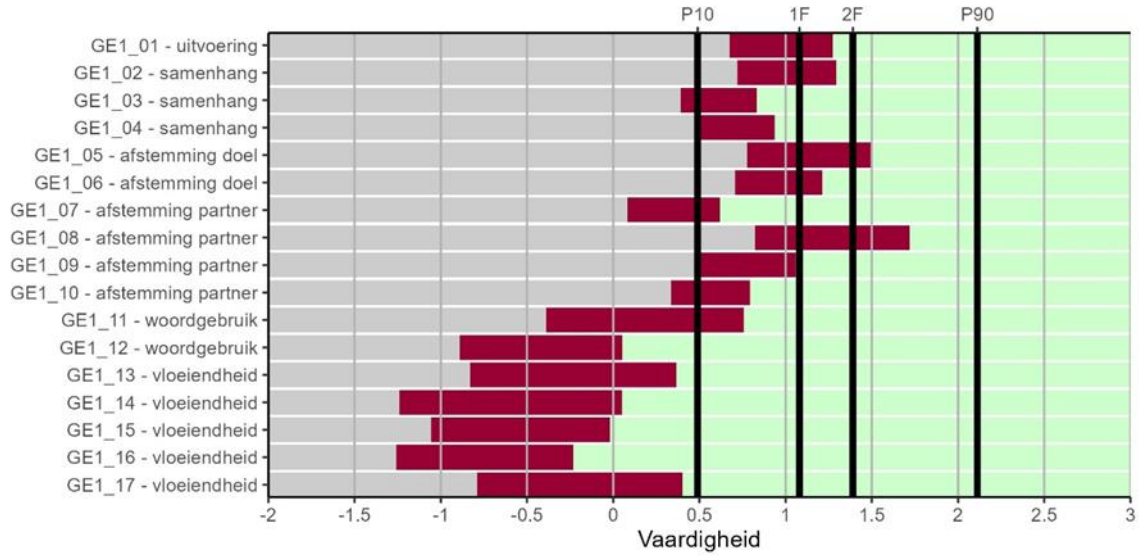
In Tabel 6.3.4 worden de bankscores – ten aanzien van gespreksvaardigheid - weergegeven voor P10-, P50- en P90-leerlingen, evenals voor leerlingen die precies op de cesuren van de referentieniveaus zitten. De P10-leerling heeft een bankscore van 55,5%. De P90-leerling heeft een bankscore van ongeveer 97%. De mediane leerling (P50) zit tussen de 1F- en 2F-cesuren in, en behaalt daarmee op basis van kans wel 1F, maar geen 2F.

Tabel 6.3.4: Bankscores gespreksvaardigheid

vaardigheidsniveau	bankscore%
P10	55,5
1F	79,7
P50	83,4
2F	88,0
P90	97,3

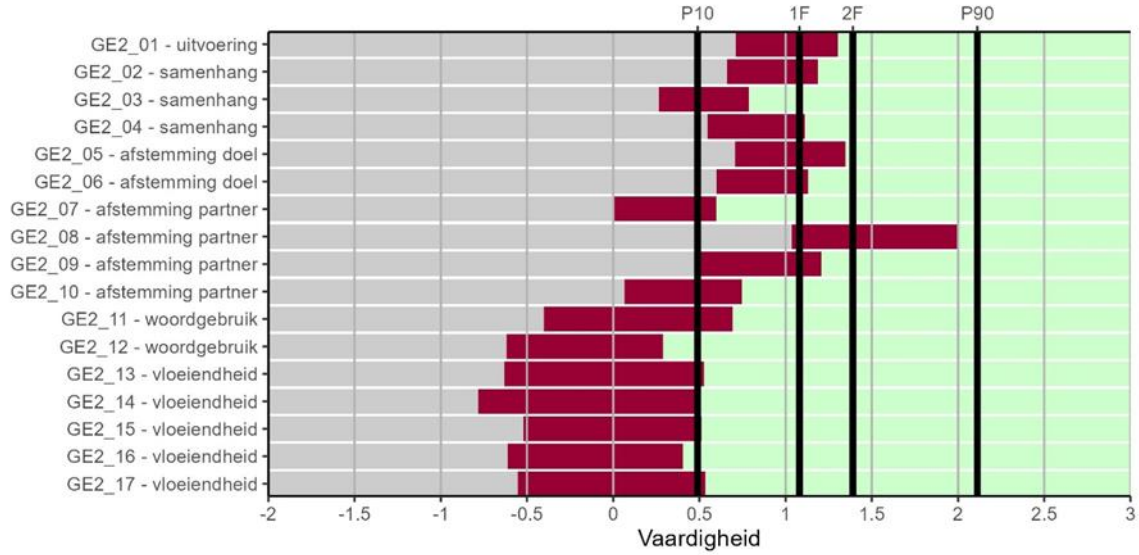
In Figuur 6.3.8 en 6.3.9 wordt de beheersing op itemniveau gepresenteerd, voor respectievelijk het samenwerkingsspel en het raadsel. De twee taken laten in beheersing een sterk vergelijkbaar patroon zien, zowel met elkaar als met de respectievelijke items van de spreektaken. Zo worden de items op vloeiendheid, verstaanbaarheid en grammaticale beheersing zeer goed beheerst, ook door P10-leerlingen.

Figuur 6.3.8: Beheersing van de vaardigheid op itemniveau - samenwerkingsspel



Noot. de P50-cesuur is niet weergegeven in bovenstaande afbeelding, in verband met de leesbaarheid

Figuur 6.3.9: Beheersing van de vaardigheid op itemniveau - raadsel



Noot. de P50-cesuur is niet weergegeven in bovenstaande afbeelding, in verband met de leesbaarheid

In Tabel 6.3.5 zijn de bankscores per gesprekstaak gepresenteerd. Hier kan uit worden opgemaakt dat het samenwerkingsspel en het raadsel in vergelijkbare mate werden beheerst.

Tabel 6.3.5: Bankscores gesprekstaken

	bankscore%	
Vaardigheidsniveau	**Samenwerkingsspel**	**Raadsel**
P10	55,3	55,8
1F	80,1	79,3
P50	83,8	83
2F	88,4	87,6
P90	97,6	96,9

7 Resultaten vragenlijsten

7.1 Inleiding

In dit hoofdstuk worden de resultaten op de leerling-, leerkracht- en schoolleidersvragenlijsten en de quickscan voor individuele leerlingen (ingevuld door leerkrachten) gepresenteerd. Daarbij worden achtergrond- en domeinspecifieke (onderwijs)kenmerken van leerlingen, leerkrachten en scholen vergeleken tussen bao, sbo en so.
De vragenlijsten waren identiek voor bao, sbo en so. Een deel van de domeinspecifieke kenmerken van leerlingen, leerkrachten en scholen is gemeten aan de hand van een construct bestaande uit een set van items, veelal in de vorm van stellingen die aan de respondent zijn voorgelegd. Er is in de vragenlijsten vooral gebruikgemaakt van bestaande constructen (uit Peil.Mondelinge taalvaardigheid (s)bo 2017 of ander peilingsonderzoeken) of een afgeleide daarvan. In paragraaf 2.5 is het instrumentarium reeds beschreven. Om na te gaan of deze constructen ook meetbaar waren bij deze peiling onder bao, sbo en so, zijn er componenten- en betrouwbaarheidsanalyses uitgevoerd. In hoofdstuk 5 worden de resultaten van deze analyses gepresenteerd.

7.2 Leerlingvragenlijst

Respons

In totaal hebben 4393 leerlingen deelgenomen aan het peilingsonderzoek. De leerlingvragenlijst is correct ingevuld door 4346 leerlingen (99%). De verdeling van deze leerlingen over de verschillende sectoren is weergegeven in tabel 7.2.1Een kleine groep leerlingen (n=6) was wel gestart met de vragenlijst, maar heeft meer dan de helft van alle vragen niet beantwoord. De resultaten van deze leerlingenzijn uit het vragenlijstbestand verwijderd. In deze paragraaf worden de resultaten op de vragen van 2895 bao-leerlingen, 1042 sbo-leerlingen en 409 so-leerlingen gepresenteerd.

Tabel 7.2.1: Aantallen respons op leerlingvragenlijst, naar schooltype

Type	N deelnemers	N respons	%
Regulier basisonderwijs (bao)	2926	2895	99%
Speciaal basisonderwijs (sbo)	1053	1042	99%
Speciaal onderwijs (so)	414	409	99%

Algemeen

In tabel 7.2.2 worden de achtergrondkenmerken van de leerlingen in percentages gepresenteerd.

Tabel 7.2.2: Achtergrondkenmerken leerlingen, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Geslacht*	**n=2781**	**n=1023**	**n=399**
jongen	50%	64%	82%
meisje	50%	37%	19%
Aantal boeken thuis	**n=2895**	**n=1042**	**n=409**
geen of weinig (0–10 boeken)	13%	31%	25%
een plank vol (11–25 boeken)	29%	35%	28%
een boekenkast vol (26–100 boeken)	33%	22%	27%
twee boekenkasten vol (101–200 boeken)	15%	6%	9%
drie of meer boekenkasten vol (meer dan 200 boeken)	11%	5%	11%
Frequentie thuis Nederlands gesproken	***n=2895***	***n=1042***	***n=409***
nooit	3%	4%	3%
soms	10%	14%	11%
meestal	21%	26%	33%
altijd	65%	55%	54%
Andere gesproken taal thuis (soms / meestal Nederlands) *meer antwoorden mogelijk*	***n=919***	***n=421***	***n=177***
Dialect of streektaal van het Nederlands	19%	18%	13%
Fries	5%	2%	5%
Een andere taal (bijvoorbeeld Turks of Pools)	79%	82%	85%
Andere gesproken taal thuis (nooit Nederlands) *meer antwoorden mogelijk*	**n=92**	**n=46**	**n=12**
Dialect of streektaal van het Nederlands	16%	15%	0%
Fries	29%	7%	33%
Een andere taal (bijvoorbeeld Turks of Pools)	54%	83%	67%
Eerst geleerde taal thuis	**n=2895**	**n=1042**	**n=409**
Nederlands	77%	67%	73%
Een andere taal	11%	17%	14%
Nederlands én een andere taal (tegelijk)	12%	16%	14%

*uit leerlinginformatieformulier

Domeinspecifiek

De domeinspecifieke kenmerken van de leerlingen worden in Tabel 7.2.3 t/m Tabel 7.2.8 gepresenteerd.

Tabel 7.2.3: Domeinspecifieke kenmerken – voorlezen, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Frequentie vroeger voorgelezen thuis*	**n=2895**	**n=1042**	**n=409**
(bijna) nooit	4%	11%	11%
1 of 2 keer per maand	3%	3%	4%
1 of 2 keer per week	20%	17%	12%
(bijna) elke dag	60%	43%	48%
Weet ik niet	13%	26%	24%
item gemiddelde#	3,55	3,24	3,28
Vroeger voorgelezen in Nederlands?	**n=2399**	**n=657**	**n=262**
nooit	3%	4%	3%
soms	3%	5%	6%
meestal	11%	9%	14%
altijd	82%	80%	75%
weet ik niet	2%	3%	3%
item gemiddelde##	3,74	3,69	3,66
Frequentie nu voorgelezen thuis	**n=2895**	**n=1042**	**n=409**
(bijna) nooit	88%	87%	87%
1 of 2 keer per maand	5%	4%	4%
1 of 2 keer per week	4%	4%	3%
(bijna) elke dag	3%	5%	6%
Nu voorgelezen in Nederlands?	**n=343**	**n=138**	**n=53**
nooit	3%	4%	4%
soms	8%	9%	8%
meestal	16%	15%	19%
altijd	74%	71%	70%

#Exclusief 'Weet niet', range 1 ((bijna) nooit) tot 4 ((bijna) elke dag)
##Exclusief 'Weet niet', range 1 (nooit) tot 4 (altijd).

Tabel 7.2.4: Domeinspecifieke kenmerken, activiteiten thuis, in percentages, naar schooltype

	% bao	*% sbo*	*% so*
Lezen van een boek (geen schoolboek), tijdschrift of krant	***n=2895***	***n=1042***	***n=409***
(bijna) nooit	28%	48%	46%
1 of 2 keer per maand	14%	8%	10%
1 of 2 keer per week	26%	19%	13%
(bijna) elke dag	28%	15%	23%
weet ik niet	4%	11%	7%
Luisteren naar een luisterboek of podcast	***n=2895***	***n=1042***	***n=409***
(bijna) nooit	75%	71%	79%
1 of 2 keer per maand	10%	5%	4%
1 of 2 keer per week	7%	7%	5%
(bijna) elke dag	4%	5%	7%
weet ik niet	4%	12%	6%
Kijken naar een televisieprogramma of film	***n=2895***	***n=1042***	***n=409***
(bijna) nooit	3%	7%	10%
1 of 2 keer per maand	8%	10%	12%
1 of 2 keer per week	31%	26%	22%
(bijna) elke dag	58%	54%	54%
weet ik niet	1%	3%	2%
Zelf een filmpje of vlog opnemen	***n=2895***	***n=1042***	***n=409***
(bijna) nooit	58%	54%	62%
1 of 2 keer per maand	14%	9%	10%
1 of 2 keer per week	13%	11%	12%
(bijna) elke dag	10%	16%	10%
weet ik niet	5%	10%	7%
Chatten of appen	***n=2895***	***n=1042***	***n=409***
(bijna) nooit	2%	7%	7%
1 of 2 keer per maand	1%	2%	3%
1 of 2 keer per week	6%	10%	11%
(bijna) elke dag	89%	77%	75%
weet ik niet	1%	3%	3%
Op internet informatie of nieuwsberichten opzoeken	***n=2895***	***n=1042***	***n=409***
(bijna) nooit	29%	39%	36%
1 of 2 keer per maand	19%	11%	14%

1 of 2 keer per week	30%	20%	22%
(bijna) elke dag	15%	18%	19%
weet ik niet	6%	12%	10%

*items zijn samen niet schaalbaar

Tabel 7.2.5: Domeinspecifieke kenmerken, thuis praten over.., in percentages, naar schooltype

	% bao	*% sbo*	*% so*
Praten over een boek dat ik heb gelezen	**n=2895**	**n=1042**	**n=409**
(bijna) nooit	58%	66%	67%
1 of 2 keer per maand	17%	8%	11%
1 of 2 keer per week	14%	10%	9%
(bijna) elke dag	4%	5%	5%
weet ik niet	7%	11%	8%
Praten over een televisieprogramma of film waarnaar we samen hebben gekeken	**n=2895**	**n=1042**	**n=409**
(bijna) nooit	28%	35%	39%
1 of 2 keer per maand	23%	15%	17%
1 of 2 keer per week	31%	22%	20%
(bijna) elke dag	11%	17%	16%
weet ik niet	6%	12%	9%
Praten over wat ik op of na school heb gedaan	**n=2895**	**n=1042**	**n=409**
(bijna) nooit	7%	17%	15%
1 of 2 keer per maand	4%	7%	5%
1 of 2 keer per week	20%	15%	17%
(bijna) elke dag	67%	57%	59%
weet ik niet	1%	4%	4%
Praten over het nieuws (bijvoorbeeld over het klimaat of de politiek)	**n=2895**	**n=1042**	**n=409**
(bijna) nooit	38%	51%	51%
1 of 2 keer per maand	21%	11%	12%
1 of 2 keer per week	25%	13%	17%
(bijna) elke dag	10%	11%	10%
weet ik niet	6%	13%	9%
Schaal Frequentie thuis praten over...*	**2,38**	**2,20**	**2,19**
Spreektaal Nederlands bij praten over bepaalde dingen thuis	**n=2763**	**n=917**	**n=368**

nooit	4%	5%	3%
soms	9%	14%	8%
meestal	19%	25%	29%
altijd	68%	57%	60%

*gemiddelde vier bovenstaande items, schaal 1 ((bijna) nooit) tot 4 ((bijna) elke dag).

Tabel 7.2.6: Domeinspecifieke kenmerken, taalrijkheid activiteiten ouders, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Lezen van een boek voor hun plezier	**n=2894**	**n=1038**	**n=408**
(bijna) nooit	31%	45%	42%
1 of 2 keer per maand	16%	10%	12%
1 of 2 keer per week	25%	16%	16%
(bijna) elke dag	20%	12%	16%
weet ik niet	9%	17%	14%
Brieven of mails schrijven of typen	***n=2894***	***n=1038***	***n=408***
(bijna) nooit	14%	26%	22%
1 of 2 keer per maand	13%	11%	11%
1 of 2 keer per week	25%	19%	21%
(bijna) elke dag	40%	29%	34%
weet ik niet	7%	16%	13%
De krant of een tijdschrift lezen (op papier of digitaal)	***n=2894***	***n=1038***	***n=408***
(bijna) nooit	28%	43%	42%
1 of 2 keer per maand	14%	11%	11%
1 of 2 keer per week	26%	15%	16%
(bijna) elke dag	23%	13%	15%
weet ik niet	9%	18%	16%
Kijken naar het journaal of een nieuwsprogramma	***n=2894***	***n=1038***	***n=408***
(bijna) nooit	13%	25%	27%
1 of 2 keer per maand	9%	7%	7%
1 of 2 keer per week	25%	16%	16%
(bijna) elke dag	46%	40%	40%
weet ik niet	7%	11%	9%
Praten over het nieuws (bijv. klimaat of politiek)	***n=2894***	***n=1038***	***n=408***
(bijna) nooit	27%	42%	41%
1 of 2 keer per maand	20%	11%	12%
1 of 2 keer per week	28%	14%	18%
(bijna) elke dag	14%	14%	14%
weet ik niet	11%	19%	14%
Schaal frequentie (taalrijke) activiteiten ouders*	**2,66**	**2,29**	**2,35**

*gemiddelde vijf items, schaal 1 ((bijna) nooit) tot 4 ((bijna) elke dag).

Tabel 7.2.7: Domeinspecifieke kenmerken, klasklimaat, naar schooltype

Kenmerk	% bao	% sbo	% so
Er is te veel lawaai om goed te kunnen horen wat de juf/meester zegt	**n=2894**	**n=1038**	**n=408**
(bijna) nooit	16%	26%	18%
1 of 2 keer per maand	12%	7%	9%
1 of 2 keer per week	38%	25%	28%
(bijna) elke dag	29%	33%	36%
weet ik niet	4%	10%	9%
De juf/meester moet lang wachten voordat de kinderen stil zijn	**n=2894**	**n=1038**	**n=408**
(bijna) nooit	8%	14%	13%
1 of 2 keer per maand	9%	6%	7%
1 of 2 keer per week	35%	25%	23%
(bijna) elke dag	45%	46%	55%
weet ik niet	3%	9%	3%
Kinderen luisteren niet naar wat de juf/meester zegt	**n=2894**	**n=1038**	**n=408**
(bijna) nooit	23%	22%	18%
1 of 2 keer per maand	17%	13%	11%
1 of 2 keer per week	34%	26%	26%
(bijna) elke dag	17%	24%	34%
weet ik niet	9%	15%	10%
Mijn juf/meester praat te zacht of te snel waardoor ik het niet kan volgen	**n=2894**	**n=1038**	**n=408**
(bijna) nooit	74%	67%	63%
1 of 2 keer per maand	10%	7%	6%
1 of 2 keer per week	9%	9%	13%
(bijna) elke dag	2%	7%	8%
weet ik niet	5%	10%	10%
De juf/meester gebruikt (moeilijke) woorden die ik niet begrijp	**n=2894**	**n=1038**	**n=408**
(bijna) nooit	67%	56%	58%
1 of 2 keer per maand	17%	11%	13%
1 of 2 keer per week	9%	13%	12%
(bijna) elke dag	2%	8%	8%
weet ik niet	5%	13%	9%
Schaal ervaren klasklimaat*	**2,27**	**2,33**	**2,47**

* gemiddelde van vijf items, schaal van 1 ((bijna) nooit) tot 4 ((bijna) elke dag).

Tabel 7.2.8: Domeinspecifieke kenmerken, spreekvrijheid in de klas, naar schooltype

Kenmerk	% bao	% sbo	% so
In mijn klas luisteren de kinderen goed naar elkaar	**n=2893**	**n=1037**	**n=408**
Nooit	2%	3%	6%
Soms	28%	43%	46%
Meestal	61%	39%	39%
Altijd	7%	11%	6%
Weet ik niet	2%	4%	3%
In mijn klas durven de kinderen vragen te stellen	**n=2893**	**n=1037**	**n=408**
Nooit	1%	2%	2%
Soms	7%	9%	8%
Meestal	41%	33%	36%
Altijd	47%	50%	49%
Weet ik niet	5%	6%	5%
In mijn klas word je geplaagd, gepest of uitgelachen als je een andere mening hebt dan de meesten	**n=2893**	**n=1037**	**n=408**
Nooit	45%	42%	42%
Soms	27%	29%	29%
Meestal	6%	8%	9%
Altijd	2%	5%	6%
Weet ik niet	20%	16%	14%
In mijn klas kan iedereen zeggen wat hij/zij van iets vindt	**n=2893**	**n=1037**	**n=408**
Nooit	2%	5%	9%
Soms	16%	20%	18%
Meestal	38%	27%	32%
Altijd	37%	38%	34%
Weet ik niet	6%	10%	7%
In mijn klas durven de kinderen over hun gevoelens te praten	**n=2893**	**n=1037**	**n=408**
Nooit	6%	8%	11%
Soms	31%	31%	26%
Meestal	30%	23%	28%
Altijd	9%	14%	13%
Weet ik niet	23%	25%	21%
Schaal *spreekvrijheid in de klas**	**3,07**	**2,99**	**2,93**

*gemiddelde vijf items, schaal van 1 (nooit) tot 4 (altijd), item 3 is omgedraaid voor de schaal

7.3 Quickscan

Respons

Aan de leerkrachten is gevraagd om een quickscan in te vullen voor de leerlingen die een luister-, spreek- of gesprekstaak hebben uitgevoerd. In hoofdstuk 2 is reeds toegelicht hoe deze quickscan is samengesteld. In totaal is voor 3150 leerlingen die één van de taken heeft uitgevoerd de quickscan ingevuld. Daarmee hebben we voor 82% van de leerlingen die een van de taken heeft uitgevoerd[10] een quickscan tot onze beschikking. In Tabel 7.3.1 wordt de respons naar schooltype getoond.

Tabel 7.3.1: Aantallen respons op quickscan, naar schooltype

Type	N	%
Regulier basisonderwijs (bao)	2198	70%
Speciaal basisonderwijs (sbo)	682	22%
Speciaal onderwijs (so)	270	9%
Totaal	3150	100%

In Tabel 7.3.2 tonen we de resultaten op de quickscan, waarbij onderscheid gemaakt wordt naar schooltype.

Tabel 7.3.2: Resultaten quickscan, naar schooltype

Kenmerk	bao	sbo	so
	n=2196	**n=680**	**n=270**
Externaliserend probleemgedrag *(gemiddelde score, van 1 geen tot 5 sterk internaliserend probleemgedrag))*	1,62	2,25	2,47
1 geen externaliserend gedrag	69%	48%	39%
2	12%	12%	16%
3	9%	16%	15%
4	7%	18%	19%
5 sterk externaliserend probleemgedrag	3%	7%	11%
	n=2195	**n=678**	**n=269**
Internaliserend probleemgedrag *(gemiddelde score, van 1 geen tot 5 sterk internaliserend probleemgedrag)*	1,90	2,85	3,20
1 geen internaliserend gedrag	52%	21%	12%
2	22%	21%	22%
3	14%	20%	20%
4	9%	28%	27%
5 sterk internaliserend probleemgedrag	3%	10%	19%

10 Met een geldige score op de taak.

Vervolg Tabel 7.3.2: Resultaten quickscan, naar schooltype

	bao	sbo	so
	n=2196	**n=680**	**n=270**
Introvert- extravert *(gemiddelde score)* 1= (sterk) introvert – 5 (sterk) extravert	3,81	3,50	3,60
1 introvert	4%	9%	5%
2	11%	16%	16%
3	22%	23%	22%
4	26%	19%	30%
5 extravert	37%	32%	28%
	n=2196	**n=680**	**n=270**
Spreekangst – geen spreekangst *(gemiddelde score)* 1= (sterk) spreekangstig– 5 niet spreekangstig	4,00	3,66	3,89
1 spreekangst	2%	7%	3%
2	9%	16%	11%
3	18%	17%	19%
4	26%	23%	26%
5 geen spreekangst	44%	37%	41%

Vervolg Tabel 7.3.2: Resultaten quickscan, naar schooltype

	bao	sbo	so
Communicatieproblemen 1 = niet aanwezig, 2 = in lichte mate, 3 = in behoorlijke mate, 4 = in ernstige mate			
	n=2169	**n=678**	**n=270**
Zwak taalbegrip (gemiddelde score)	1,33	2,02	1,66
1 niet aanwezig	76%	35%	55%
2 in lichte mate	16%	37%	28%
3 in behoorlijke mate	6%	18%	13%
4 in ernstige mate	2%	10%	4%
	n=2195	**n=682**	**n=270**
Zwakke uitdrukkingsvaardigheid (gemiddelde score)	1,33	2,01	1,84
1 niet aanwezig	75%	34%	43%
2 in lichte mate	18%	38%	35%
3 in behoorlijke mate	6%	19%	17%
4 in ernstige mate	1%	8%	5%
	n=2197	**n=679**	**n=270**
Zich niet kunnen verplaatsen in een ander (gemiddelde score)	1,28	1,85	2,20
1 niet aanwezig	79%	44%	27%
2 in lichte mate	15%	33%	34%
3 in behoorlijke mate	4%	19%	29%
4 in ernstige mate	1%	5%	10%
	n=2195	**n=675**	**n=269**
Zwak auditief geheugen / auditieve verwerking (gemiddelde score)	1,24	1,95	1,71
1 niet aanwezig	82%	34%	52%
2 in lichte mate	13%	42%	29%
3 in behoorlijke mate	4%	19%	15%
4 in ernstige mate	1%	5%	4%
Schaal communicatieproblemen *	**1,30**	**1,96**	**1,85**

*gebaseerd op gemiddelde vier items

Vervolg Tabel 7.3.2: Resultaten quickscan, naar schooltype

	bao	sbo	so
Belemmerende en stimulerende factoren 1=belemmerend 2=neutraal 3=stimulerend			
	n=2113	**n=616**	**n=253**
Inhibitie (gemiddelde score)	2,21	1,90	1,64
1 belemmerend	19%	36%	51%
2 neutraal	42%	38%	34%
3 stimulerend	39%	26%	15%
	n=2145	**n=680**	**n=253**
Emotieregulatie (gemiddelde score)	2,18	1,76	1,48
1 belemmerend	18%	43%	60%
2 neutraal	46%	38%	32%
3 stimulerend	36%	19%	8%
	n=2169	**n=680**	**n=270**
Planning en organisatie (gemiddelde score)	2,39	1,99	1,88
1 belemmerend	13%	29%	33%
2 neutraal	35%	43%	47%
3 stimulerend	52%	28%	20%
	n=2174	**n=675**	**n=270**
Zelfstandigheid (gemiddelde score)	2,51	2,17	2,11
1 belemmerend	9%	22%	24%
2 neutraal	31%	39%	42%
3 stimulerend	60%	39%	34%
	n=2172	**n=675**	**n=270**
Motivatie (gemiddelde score)	2,45	2,21	2,16
1 belemmerend	10%	20%	22%
2 neutraal	34%	38%	40%
3 stimulerend	56%	41%	38%
	n=2171	**n=672**	**n=269**
Concentratie en aandacht (*gemiddelde score)*	2,31	2,00	1,90
1 belemmerend	17%	31%	33%
2 neutraal	34%	39%	43%
3 stimulerend	49%	31%	24%
Schaal stimulerende factoren**	**2,34**	**2,01**	**1,87**

**gebaseerd op gemiddelde zes items

7.4 Leerkrachtvragenlijst

Respons

In totaal hebben 329 klassen deelgenomen aan het peilingsonderzoek, zie Tabel 7.4.1. De leerkrachtvragenlijst is correct ingevuld door 253 leerkrachten (77%). Een kleine groep leerkrachten (n=6) heeft minder dan de helft van de vragenlijst ingevuld en wordt daarmee buiten beschouwing gelaten. In deze paragraaf worden de resultaten op de vragen van 133 bao-leerkrachten, 85 sbo-leerkrachten en 38 so-leerkrachten gepresenteerd.

Tabel 7.4.1: Respons leerkrachtvragenlijst, naar schooltype

	N klassen	N respons	%
Regulier basisonderwijs (bao)	169	133	79%
Speciaal basisonderwijs (sbo)	108	82	76%
Speciaal onderwijs (so)	52	38	73%

Algemeen

In tabel 7.4.2 worden de achtergrondkenmerken van de leerkrachten in percentages gepresenteerd.

Tabel 7.4.2: Achtergrondkenmerken leerkrachten, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Geslacht	**n=133**	**n=82**	**n=38**
Man	26%	20%	24%
Vrouw	74%	79%	74%
Wil ik niet zeggen	0%	1%	3%
Leeftijd	**n=133**	**n=82**	**n=38**
Jonger dan 25	5%	2%	5%
25-29	15%	11%	11%
30-39	41%	18%	42%
40-49	26%	35%	29%
50-59	11%	23%	8%
60 jaar of ouder	3%	10%	5%
Lerarenopleiding(en) met diploma afgerond (meerkeuze)	**n=133**	**n=82**	**n=38**
Nog niet afgerond/stage	1%	2%	5%
PABO (4 jaar)	80%	68%	68%
Verkorte pabo (2-3) jaar	10%	16%	11%
Zij-instroom lerarenopleiding	5%	6%	8%
PA, kweekschool of kleuteropleiding	2%	5%	3%
Academische PABO	2%	0%	0%
Andere vooropleiding, namelijk..	3%	11%	11%

Aantal jaren werkzaam als leerkracht	n=133	n=82	n=38
Minder dan 5	17%	15%	13%
5-9	22%	12%	26%
10-19	38%	22%	34%
20-29	20%	40%	26%
30-39	3%	9%	0%
40 of meer	1%	2%	0%
Omvang dienstverband (fte)	**n=133**	**n=82**	**n=38**
0,4 – 0,59 fte	4%	1%	11%
0,6 – 0,79 fte	23%	16%	8%
0,8 –0,99 fte	29%	39%	29%
1 fte	44%	44%	53%
Gemiddeld aantal fte	**0,84**	**0,86**	**0,86**
Deelname professionaliseringactiviteit mondelinge taalvaardigheid in afgelopen twee jaar (vraag 17)	**n=133**	**n=82**	**n=38**
Ja	5%	11%	5%
Nee	96%	89%	95%

Domeinspecifiek

In Tabel 7.4.3 t/m 7.4.13 worden domeinspecifieke kenmerken uit de leerkrachtvragenlijst in beeld gebracht.

Tabel 7.4.3: Domeinspecifieke kenmerken, houding t.o.v. Mondelinge taalvaardigheid, gemiddelde score per item en percentages naar antwoordcategorie, naar schooltype

Kenmerk	bao (n=133)	sbo (n=82)	so (n=38)
Houding			
6.1 Mondelinge taalvaardigheid moet van groep 1 t/m eind groep 8 expliciet worden onderwezen *(gemiddelde score)*	3,4	3,7	3,4
1 zeer mee oneens	0%	0%	0%
2 enigszins mee oneens	6%	1%	11%
3 enigszins mee eens	48%	28%	39%
4 zeer mee eens	46%	71%	50%
6.2 Mondelinge taalvaardigheid is essentieel voor de lees- en schrijfontwikkeling van leerlingen *(gemiddelde score)*	3,7	3,7	3,7
1 zeer mee oneens	1%	0%	0%
2 enigszins mee oneens	3%	2%	0%
3 enigszins mee eens	23%	28%	34%
4 zeer mee eens	73%	70%	66%
6.3 Ik besteed voldoende aandacht aan onderwijs in mondelinge taalvaardigheid *(gemiddelde score)*	3,1	3,2	3,1
1 zeer mee oneens	0%	0%	0%
2 enigszins mee oneens	11%	10%	16%
3 enigszins mee eens	65%	59%	61%
4 zeer mee eens	24%	32%	24%
6.4 Een goede leesvaardigheid is belangrijker dan een goede mondelinge taalvaardigheid *(gemiddelde score)*	2,2	2,3	2,2
1 zeer mee oneens	14%	11%	13%
2 enigszins mee oneens	53%	52%	53%
3 enigszins mee eens	29%	33%	34%
4 zeer mee eens	4%	4%	0%
6.5 Een goede schrijfvaardigheid is belangrijker dan een goede mondelinge taalvaardigheid *(gemiddelde score)*	1,9	1,9	1,9
1 zeer mee oneens	18%	18%	16%
2 enigszins mee oneens	71%	71%	76%
3 enigszins mee eens	11%	11%	8%
4 zeer mee eens	1%	0%	0%
6.6 Onderwijs in mondelinge taalvaardigheid is voor alle leerlingen in mijn klas belangrijk *(gemiddelde score)*	3,6	3,8	3,6
1 zeer mee oneens	1%	0%	0%

2 enigszins mee oneens	5%	0%	0%
3 enigszins mee eens	26%	16%	42%
4 zeer mee eens	68%	84%	58%
Schaal relatieve belang mondelinge taalvaardigheid#	**2,1**	**2, 1**	**2,1**

schaal relatieve belang mtv gebaseerd op items 6.4 en 6.5

Tabel 7.4.4: Domeinspecifieke kenmerken, ervaren bekwaamheid t.o.v. Mondelinge taalvaardigheid, gemiddelde score per item en percentages naar antwoordcategorie, naar schooltype

	bao (n=133)	sbo (n=82)	so (n=38)
Ervaren bekwaamheid			
7.1 Ik weet in grote lijnen hoe de ontwikkeling van de mondelinge taalvaardigheid in groep 1 t/m 8 verloopt	2,9	3,0	3,0
1 zeer mee oneens	3%	1%	3%
2 enigszins mee oneens	17%	16%	8%
3 enigszins mee eens	65%	62%	79%
4 zeer mee eens	15%	21%	11%
7.2 Ik ben bekend met de beheersingsdoelen in het referentiekader voor mondelinge taalvaardigheid	2,6	2,6	2,7
1 zeer mee oneens	6%	6%	3%
2 enigszins mee oneens	37%	29%	29%
3 enigszins mee eens	47%	61%	63%
4 zeer mee eens	10%	4%	5%
7.3 Ik weet hoe ik de mondelinge taalvaardigheid van mijn leerlingen kan bevorderen	3,1	3,2	3,1
1 zeer mee oneens	1%	0%	0%
2 enigszins mee oneens	8%	10%	5%
3 enigszins mee eens	70%	57%	76%
4 zeer mee eens	22%	33%	18%
7.4 Ik vind het lastig om leerlingen te ondersteunen bij de ontwikkeling van hun mondelinge taalvaardigheid	2,2	2,2	2,3
1 zeer mee oneens	14%	16%	8%
2 enigszins mee oneens	53%	54%	61%
3 enigszins mee eens	32%	28%	26%
4 zeer mee eens	2%	2%	5%
7.5 Ik vind het moeilijk om de lessen in mondelinge taalvaardigheid goed vorm te geven	2,2	2,3	2,3
1 zeer mee oneens	21%	15%	11%
2 enigszins mee oneens	44%	49%	55%
3 enigszins mee eens	32%	32%	29%
4 zeer mee eens	2%	5%	5%
7.6 De wijze waarop ik gesprekken voer, luister en spreek, heeft invloed op hoe mijn leerlingen deze vaardigheden ontwikkelen	3,8	3,9	3,7
1 zeer mee oneens	0%	0%	0%

2 enigszins mee oneens	1%	0%	3%
3 enigszins mee eens	14%	13%	29%
4 zeer mee eens	86%	87%	68%
Schaal ervaren bekwaamheid mondelinge taalvaardigheid #	**2,9**	**2,9**	**2,8**

schaal ervaren bekwaamheid MTV gebaseerd op items 7.1 t/m 7.6 (items 7.4 en 7.5 omgedraaid voor schaling)

Tabel 7.4.5: Domeinspecifieke kenmerken, onderwijstijd en -aanbod t.o.v. Mondelinge taalvaardigheid, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Intentioneel aantal minuten besteed aan onderwijs in mondelinge taal	***n=133***	***n=82***	***n=38***
Ik besteed intentioneel geen tijd aan onderwijs in mondelinge taal	4%	6%	
Maximaal 15 minuten per week	5%	4%	5%
16 tot 30 minuten per week	23%	17%	18%
31 tot 60 minuten per week	25%	22%	29%
61 tot 90 minuten per week	25%	22%	21%
91 tot 120 minuten per week	8%	10%	18%
Meer dan 120 minuten per week	11%	20%	8%
Schaal onderwijstijd (1-7; geen tot meer dan 120 minuten per week)	**4,3**	**4,6**	**4,5**
Gemiddeld percentage tijd (intentioneel) besteed aan:	**n=128**	**n=77**	**n=38**
Gesprekken voeren	33 %	33%	34%
Luisteren	34%	36%	37%
Spreken	33%	31%	29%
Tijdsbesteding MTV in groep 8 t.o.v. leesvaardigheid	**n= 133**	**n=82**	**n=38**
Veel minder	19%	10%	3%
Minder	45%	39%	34%
Ongeveer even veel	26%	28%	47%
Meer	8%	23%	13%
Veel meer	2%	0%	3%
Tijdsbesteding MTV in groep 8 t.o.v. schrijfvaardigheid	**n= 133**	**n=82**	**n=38**
Veel minder	5%	0%	3%
Minder	26%	16%	8%
Ongeveer even veel	35%	21%	37%
Meer	30%	55%	39%
Veel meer	5%	9%	13%
Intentioneel aandacht besteed aan MTV bij taalonderwijs in andere subdomeinen en/of bij andere vakken (meer antwoorden mogelijk)	**n= 133**	**n=82**	**n=38**
Nee	8%	10%	8%
Ja, bij leesvaardigheid	59%	62%	58%
Ja, bij schrijfvaardigheid	34%	30%	24%
Ja, bij rekenvaardigheid	44%	50%	42%

Ja, bij kennis opbouwen bij de zaakvakken	82%	80%	74%
Anders, namelijk*	14%	6%	8%

* bij 'Anders, namelijk' worden de volgende domeinen en vakken genoemd: Engels, wereld oriënterende vakken, digitale vaardigheden, levensbeschouwing/filosofie en sociaal emotionele vaardigheden

Tabel 7.4.6: Domeinspecifieke kenmerken, gebruik taalmethode en vormgeving onderwijs, in percentages, naar schooltype

Kenmerk	%bao	%sbo	%so
Gebruik taalmethode voor vormgeving lessen in MTV	**n= 133**	**n= 82**	**n= 38**
Nee	18%	18%	3%
Soms	23%	33%	37%
Ja	59%	49%	61%
Manier van gebruik taalmethode voor MTV	**n= 109**	**n= 67**	**n= 37**
Ik bied de lesactiviteiten voor mondelinge taalvaardigheid uit de taalmethode aan zoals voorgeschreven in de methode	38%	21%	19%
Ik pas de lesactiviteiten voor mondelinge taalvaardigheid uit de taalmethode (meestal) aan	29%	48%	57%
Ik gebruik de lesactiviteiten uit de taalmethode als inspiratiebron voor de lessen	30%	28%	24%
Anders, namelijk*	3%	3%	0%
Slaat u lesactiviteiten voor MTV in taalmethode wel eens over? *Meer antwoorden mogelijk*	**n= 109**	**n= 67**	**n= 37**
Nee	30%	30%	35%
Ja, ik laat mondelinge taalvaardigheid in andere lesactiviteiten terugkomen	47%	45%	46%
Ja, ik ontwerp de lesactiviteiten voor mondelinge taalvaardigheid zelf	6%	15%	8%
Ja, ik vind andere taalonderdelen belangrijker	13%	13%	5%
Ja, deze vaardigheden worden niet getoetst	14%	3%	3%
Anders, namelijk…**	14%	12%	27%
Basis vormgeving onderwijs in MTV *meer antwoorden mogelijk*	***n=133***	***n=82***	***n=38***
Thema's/projecten	77%	74%	61%
De inhoud van de zaakvakken	77%	82%	82%
De kerndoelen	29%	30%	42%
De referentieniveaus	11%	17%	16%
Eigen materiaal	24%	44%	39%
Anders, namelijk…***	4%	12%	5%

*Andere manieren van gebruik taalmethode voor MTV: combinaties van of afwisseling tussen reeds genoemde manieren gebruikt.
**Om welke andere reden slaat u de lesactiviteiten voor mondelinge taalvaardigheid in de taalmethode wel eens over? O.a.: Het sluit niet aan bij (het niveau van) de doelgroep (sbo, so) of de lessen zijn niet aantrekkelijk voor de leerlingen / sluit niet aan bij de behoeften van de leerlingen.
***Andere basis vormgeving onderwijs MTV: klassengesprekken, gespreksvoering, sociale vaardigheidslessen.

Tabel 7.4.7: Domeinspecifieke kenmerken, Algemene didactiek MTV, gemiddelde score per item en percentages naar antwoordcategorie, naar schooltype

Kenmerk	bao	sbo	so
Algemene didactiek	***n=133***	***n=82***	***n=38***
Ik geef gerichte feedback aan leerlingen wanneer ze gesprekken voeren, luisteren of spreken, om zo hun prestaties hierin te verbeteren	3,1	3,2	3,1
1 helemaal niet van toepassing	1%	0%	0%
2 niet van toepassing	6%	1%	16%
3 van toepassing	73%	77%	58%
4 helemaal van toepassing	20%	22%	26%
Ik stem mijn taalgebruik af op het niveau van leerlingen, bijvoorbeeld door te letten op mijn spreektempo, intonatie en woordkeuze	3,5	3,7	3,5
1 helemaal niet van toepassing	0%	0%	0%
2 niet van toepassing	1%	1%	0%
3 van toepassing	46%	32%	47%
4 helemaal van toepassing	53%	67%	53%
Ik stel open, prikkelende vragen, zodat leerlingen worden uitgedaagd om actief mondelinge taal te gebruiken	3,2	3,3	3,3
1 helemaal niet van toepassing	0%	0%	0%
2 niet van toepassing	4%	2%	3%
3 van toepassing	68%	61%	68%
4 helemaal van toepassing	28%	37%	29%

Tabel 7.4.8: Domeinspecifieke kenmerken, Onderwijsaanbod gesprekken voeren: Frequentie gesprekstaken in de afgelopen 12 weken, in percentages, naar schooltype

Kenmerk	%bao	%sbo	%so
20.1 **Discussies** voeren (bijv. uitwisselen wat ze van een boek of film vinden of discussiëren over een omstreden of actueel onderwerp)	n=133	n=82	n=38
nooit	2%	5%	0%
1 of 2 keer	32%	32%	34%
3 tot 6 keer	36%	37%	45%
7 tot 12 keer	20%	17%	18%
Meer dan 12 keer	10%	10%	3%
20.2 **Debatteren** (bijv. een wedstrijddebat waarin een stelling wordt verdedigd of bestreden)	n=133	n=82	n=38
nooit	23%	44%	45%
1 of 2 keer	54%	40%	39%
3 tot 6 keer	21%	12%	13%
7 tot 12 keer	2%	4%	3%
Meer dan 12 keer	1%	0%	0%
20.3 **Overleggen** (bijv. over hoe ze samen een taak kunnen uitvoeren)	n=133	n=82	n=38
nooit	2%	0%	0%

1 of 2 keer	4%	7%	5%
3 tot 6 keer	23%	30%	29%
7 tot 12 keer	23%	20%	32%
Meer dan 12 keer	50%	43%	34%
20.4 **Brainstormen** (bijv. ideeën over de aanpak van een taak uitwisselen)	n=133	n=82	n=38
nooit	2%	4%	3%
1 of 2 keer	14%	22%	13%
3 tot 6 keer	38%	41%	50%
7 tot 12 keer	22%	17%	29%
Meer dan 12 keer	25%	16%	5%
20.5 **Interviews** afnemen en geven	n=133	n=82	n=38
nooit	42%	60%	47%
1 of 2 keer	50%	34%	37%
3 tot 6 keer	6%	4%	13%
7 tot 12 keer	2%	2%	3%
Meer dan 12 keer	0%	0%	0%
20.6 **Leergesprekken** (bijv. een gesprek met het doel kennis over een geschiedenisonderwerp te vergaren of vragen te formuleren voor een onderzoek)	n=133	n=82	n=38
nooit	11%	13%	5%
1 of 2 keer	32%	27%	34%
3 tot 6 keer	32%	23%	39%
7 tot 12 keer	15%	27%	16%
Meer dan 12 keer	9%	10%	5%
20.7 **Vrije gesprekken** (bijv. een kringgesprek om ervaringen en gevoelens uit te wisselen)	n=133	n=82	n=38
nooit	2%	0%	0%
1 of 2 keer	8%	2%	3%
3 tot 6 keer	13%	12%	16%
7 tot 12 keer	28%	15%	37%
Meer dan 12 keer	50%	71%	45%
Schaal frequentie gesprekstaken#	**3,1**	**3,0**	**2,9**

Schaal frequentie gesprekstaken: gebaseerd op gemiddelde items 20.1 t/m 20.7

Tabel 7.4.9: Domeinspecifieke kenmerken, Stellingen over gesprekken, gemiddelde score per item en percentages naar antwoordcategorie, naar schooltype

Kenmerk	bao	sbo	so
Stellingen gesprekken	**n=133**	**n=82**	**n=38**
21.1 Ik leer leerlingen expliciet aan hoe ze goed een gesprek kunnen voeren (bijv. beurt nemen, elkaars perspectief begrijpen, redeneren, argumenteren)	3,1	3,2	3,2
1 helemaal niet van toepassing	0%	0%	0%
2 niet van toepassing	9%	6%	3%
3 van toepassing	71%	66%	76%
4 helemaal van toepassing	20%	28%	21%
21.2 Ik laat leerlingen oefenen met verschillende gespreksdoelen, zoals uitleggen, discussiëren en argumenteren	2,9	2,9	3,0
1 helemaal niet van toepassing	1%	0%	0%
2 niet van toepassing	18%	23%	11%
3 van toepassing	67%	67%	82%
4 helemaal van toepassing	14%	10%	8%
21.3 Ik stimuleer een veilig gesprekssklimaat in de klas (bijvoorbeeld door leerlingen de ruimte te geven om hun mening te geven, aandachtig naar ze te luisteren, ze serieus te nemen en alle leerlingen aan te moedigen dit ook onderling te doen)	3,5	3,7	3,4
1 helemaal niet van toepassing	0%	0%	0%
2 niet van toepassing	0%	0%	5%
3 van toepassing	47%	30%	47%
4 helemaal van toepassing	53%	70%	47%
21.4 Ik laat leerlingen gesprekken voeren in kleine groepjes	3,4	3,1	3,0
1 helemaal niet van toepassing	0%	0%	0%
2 niet van toepassing	4%	12%	21%
3 van toepassing	54%	61%	55%
4 helemaal van toepassing	42%	27%	24%
21.5 Ik stimuleer alle leerlingen om tijdens gesprekken mee te denken en te praten	3,3	3,4	3,3
1 helemaal niet van toepassing	0%	0%	0%
2 niet van toepassing	2%	1%	0%
3 van toepassing	64%	62%	68%
4 helemaal van toepassing	35%	37%	32%
21.6 Ik sta toe dat leerlingen in een andere taal dan het Nederlands overleggen	1,6	1,5	1,8
1 helemaal niet van toepassing	51%	60%	37%
2 niet van toepassing	37%	29%	45%
3 van toepassing	8%	10%	16%
4 helemaal van toepassing	4%	1%	3%
Schaal Mate van effectieve instructie in gesprekken#	**3,3**	**3,3**	**3,2**

Gemiddelde score, 1= helemaal niet van toepassing, 4 = helemaal van toepassing
Schaal gebaseerd op items 21.1 t/m 21.5

Tabel 7.4.10: Onderwijsaanbod luisteren: Frequentie luistertaken in de afgelopen 12 weken, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
22.1 Luisteren naar een **verhaal** verteld door iemand die fysiek aanwezig is in het klaslokaal	n=133	n=82	n=38
nooit	3%	4%	8%
1 of 2 keer	14%	15%	16%
3 tot 6 keer	23%	26%	18%
7 tot 12 keer	19%	22%	29%
Meer dan 12 keer	41%	34%	29%
22.2 Luisteren naar een **gesprek** tussen mensen die fysiek aanwezig zijn in het klaslokaal	n=133	n=82	n=38
nooit	27%	26%	13%
1 of 2 keer	27%	24%	37%
3 tot 6 keer	23%	20%	16%
7 tot 12 keer	12%	11%	21%
Meer dan 12 keer	11%	20%	13%
22.3 Luisteren naar **gesproken tekst** op een digitaal apparaat die leerlingen kunnen terugspoelen (bijv. filmpjes of podcasts)	n=133	n=82	n=38
nooit	8%	15%	11%
1 of 2 keer	23%	15%	16%
3 tot 6 keer	23%	17%	21%
7 tot 12 keer	21%	29%	26%
Meer dan 12 keer	25%	24%	26%
Schaal frequentie luistertaken#	**3,2**	**3,3**	**3,3**

Schaal gebaseerd op items 22.1 t/m 22.3

Tabel 7.4.11: Stellingen over luisteren, gemiddelde score per item en percentages naar antwoordcategorie, naar schooltype

Kenmerk	bao	sbo	so
Stellingen Luisteren	n=133	n=82	n=38
23.1 Ik leer leerlingen expliciet luisterstrategieën aan (bijv. luisterdoel bepalen, onbekende woorden verduidelijken, voorspellingen doen, samenvatten)	2,9	2,7	2,9
1 helemaal niet van toepassing	2%	6%	0%
2 niet van toepassing	26%	28%	16%
3 van toepassing	56%	54%	79%
4 helemaal van toepassing	17%	12%	5%
23.2 Ik stimuleer een goed luisterklimaat in de klas (bijv. weinig afleiding en ruis)	3,5	3,5	3,3
1 helemaal niet van toepassing	0%	0%	0%
2 niet van toepassing	2%	1%	3%
3 van toepassing	50%	51%	63%
4 helemaal van toepassing	48%	48%	34%
23.3 Ik laat leerlingen aantekeningen maken tijdens het luisteren	2,9	2,5	2,6
1 helemaal niet van toepassing	1%	11%	5%
2 niet van toepassing	21%	37%	32%
3 van toepassing	61%	46%	61%
4 helemaal van toepassing	17%	6%	3%
23.4 Ik bied visuele ondersteuning aan bij het luisteren, bijvoorbeeld door middel van plaatjes, een woordweb, stroomdiagrammen, etc.	2,8	3,1	2,9
1 helemaal niet van toepassing	3%	0%	0%
2 niet van toepassing	23%	10%	24%
3 van toepassing	61%	71%	63%
4 helemaal van toepassing	13%	20%	13%
23.5 Ik bied luisterteksten herhaald aan bij het oefenen van begrijpend luisteren	2,4	2,4	2,7
1 helemaal niet van toepassing	13%	12%	5%
2 niet van toepassing	45%	41%	34%
3 van toepassing	34%	44%	45%
4 helemaal van toepassing	8%	2%	16%
23.6 Ik leg uit wat de moeilijke woorden die leerlingen horen betekenen	3,5	3,6	3,6
1 helemaal niet van toepassing	0%	1%	0%
2 niet van toepassing	4%	1%	0%
3 van toepassing	44%	35%	37%
4 helemaal van toepassing	52%	62%	63%
23.7 Ik deel luistertaken op in korte fragmenten voor leerlingen die luisteren lastig vinden	2,3	2,4	2,6
1 helemaal niet van toepassing	16%	18%	5%
2 niet van toepassing	44%	34%	45%

3 van toepassing	32%	37%	37%
4 helemaal van toepassing	8%	11%	13%
23.8 Ik geef leerlingen die extra ondersteuning nodig hebben bij luisteren instructie in kleine groepen of één-op-één	2,5	2,7	2,7
1 helemaal niet van toepassing	16%	15%	5%
2 niet van toepassing	34%	27%	42%
3 van toepassing	39%	34%	34%
4 helemaal van toepassing	11%	24%	18%
Schaal Mate van effectieve instructie in luisteren #	**2,8**	**2,8**	**2,9**

#Schaal gebaseerd op items 23.1 t/m 23.8

Tabel 7.4.12: Onderwijsaanbod spreken: Frequentie spreektaken in de afgelopen 12 weken, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
24.1 Een **beschrijving** geven (bijv. van een persoon, plaats, zaak of voorwerp)	n=133	n=82	n=38
nooit	11%	5%	8%
1 of 2 keer	37%	29%	32%
3 tot 6 keer	30%	43%	32%
7 tot 12 keer	14%	12%	21%
Meer dan 12 keer	9%	11%	8%
24.2 **Verslag** van een gebeurtenis uitbrengen	n=133	n=82	n=38
Nooit	4%	9%	11%
1 of 2 keer	32%	22%	21%
3 tot 6 keer	34%	34%	34%
7 tot 12 keer	18%	18%	21%
Meer dan 12 keer	13%	17%	13%
24.3 Een **verhaal** vertellen of navertellen	n=133	n=82	n=38
Nooit	8%	2%	5%
1 of 2 keer	32%	26%	37%
3 tot 6 keer	32%	32%	18%
7 tot 12 keer	20%	24%	32%
Meer dan 12 keer	9%	16%	8%
24.4 **Presenteren** (bijv. een spreekbeurt)	n=133	n=82	n=38
Nooit	2%	16%	16%
1 of 2 keer	42%	46%	55%
3 tot 6 keer	26%	21%	18%
7 tot 12 keer	19%	12%	8%
Meer dan 12 keer	12%	5%	3%
24.5 **Leerlingen uitleg** of **instructie laten geven** (bijv. de uitleg over een apparaat, instructie bij een spel of een routebeschrijving geven)	n=133	n=82	n=38
nooit	10%	13%	5%
1 of 2 keer	38%	38%	42%
3 tot 6 keer	24%	30%	29%
7 tot 12 keer	14%	10%	16%
Meer dan 12 keer	14%	9%	8%
Schaal frequentie spreektaken#	**2,9**	**3,0**	**2,9**

schaal gebaseerd op gemiddelde items 24.1 t/m 24.3 en 24.5

Tabel 7.4.13: Stellingen over spreken, gemiddelde score per item en percentages naar antwoordcategorie, naar schooltype

Kenmerk	bao	sbo	so
Stellingen spreken	n=133	n=82	n=38
25.1 Ik zorg bij spreektaken voor opbouw in taakeisen en complexiteit	2,5	2,5	2,6
1 helemaal niet van toepassing	8%	10%	0%
2 niet van toepassing	40%	37%	39%
3 van toepassing	51%	51%	58%
4 helemaal van toepassing	2%	2%	3%
25.2 Ik reflecteer na de lesactiviteit met leerlingen op de spreektaak en of de doelen zijn behaald	2,8	2,7	2,7
1 helemaal niet van toepassing	5%	7%	3%
2 niet van toepassing	18%	23%	34%
3 van toepassing	68%	60%	55%
4 helemaal van toepassing	10%	10%	8%
25.3 Ik laat leerlingen elkaar feedback geven op de uitvoering van een spreektaak	2,9	2,6	2,6
1 helemaal niet van toepassing	5%	7%	5%
2 niet van toepassing	21%	32%	37%
3 van toepassing	56%	54%	50%
4 helemaal van toepassing	19%	7%	8%
25.4 Ik modelleer (=hardop voordoen) hoe leerlingen spreektaken goed kunnen uitvoeren	3,0	3,1	3,1
1 helemaal niet van toepassing	3%	2%	0%
2 niet van toepassing	13%	10%	11%
3 van toepassing	61%	61%	71%
4 helemaal van toepassing	23%	27%	18%
25.5 Ik laat leerlingen spreektaken oefenen in kleine groepjes	2,8	2,6	2,7
1 helemaal niet van toepassing	5%	9%	3%
2 niet van toepassing	26%	27%	32%
3 van toepassing	57%	57%	55%
4 helemaal van toepassing	12%	7%	11%
Schaal Mate van effectieve instructie in spreken#	**2,8**	**2,7**	**2,7**

gebaseerd op gemiddelde items 25.1 t/m 25.5

7.5 Schoolleidersvragenlijst

Respons

In totaal hebben 214 scholen deelgenomen aan het peilingsonderzoek (zie tabel 7.5.1). De vragenlijst voor schoolleiders is correct ingevuld door 161 schoolleiders (75%). Een kleine groep schoolleiders (n=12) heeft minder dan de helft van de vragenlijst ingevuld en wordt daarmee buiten beschouwing gelaten. In deze paragraaf worden de resultaten op de vragen van 97 bao-schoolleiders, 42 sbo- schoolleiders en 21 so-schoolleiders gepresenteerd.

Tabel 7.5.1: Respons schoolleidersvragenlijst, naar schooltype

Schooltype	N scholen	N respons	%
Regulier basisonderwijs (bao)	131	97	74%
Speciaal basisonderwijs (sbo)	55	42	76%
Speciaal onderwijs (so)	28	21	75%

Algemeen

In tabel 7.5.2 worden de algemene kenmerken van de scholen uit de schoolleidersvragenlijst in beeld gebracht.

Tabel 7.5.2: Achtergrondkenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Thuistaal leerlingen van de school	***n=97***	***n=42***	***n=21***
Dominante thuistaal Nederlands	71%	66%	72%
Dominante thuistaal Nederlands dialect of streektaal	6%	6%	2%
Dominante thuistaal Fries	4%	1%	6%
Dominante thuistaal niet-Nederlands	19%	27%	21%
Taalrijkheid thuissituatie leerlingen	***n=97***	***n=42***	***n=21***
% leerlingen met uitgesproken taalrijke thuistaalsituatie	53%	32%	34%
% leerlingen met uitgesproken taalarme thuistaalsituatie	35%	62%	52%

Domeinspecifiek

In Tabel 7.5.3 t/m Tabel 7.5.12 worden domeinspecifieke kenmerken van de scholen gepresenteerd, voortgekomen uit de vragenlijst onder schoolleiders.

Tabel 7.5.3: Domeinspecifieke kenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Typering taalbeleidsplan school	**n=98**	**n=42**	**n=21**
In ons taalbeleidsplan of schoolplan wordt mondelinge taalvaardigheid niet expliciet benoemd	29%	38%	10%
In ons taalbeleidsplan of schoolplan is mondelinge taalvaardigheid globaal beschreven aan de hand van algemene (taal)doelen voor de hele school	62%	41%	76%
In ons taalbeleidsplan of schoolplan is mondelinge taalvaardigheid gedetailleerd beschreven aan de hand van concrete doelen per leerjaar of groepen van leerjaren	6%	14%	10%
Anders, namelijk …*	3%	7%	5%
Doelen taalbeleids-/schoolplan gebaseerd op (meerdere antwoorden mogelijk)	**n=66**	**n=23**	**n=18**
Kerndoelen Taal in het primair onderwijs (OCW)	46%	61%	61%
Referentieniveaus (1F/2F) voor Luisteren, Spreken en Gesprekken	32%	30%	44%
Leerlijnen en tussendoelen voor mondelinge communicatie	21%	39%	39%
De taalmethode	74%	48%	61%
Anders, namelijk….**	9%	17%	0%
Lesmethode/materialen voor MTV in groep 1 en 2	**n=97**	**n=42**	**n=21**
Vooral het kleuterprogramma en/of de taalmethode	35%	43%	33%
Vooral zelf ontwikkeld les- en leermateriaal	20%	12%	19%
Beide in ongeveer gelijke mate	36%	17%	43%
Anders, namelijk***	9%	29%	5%
Lesmethode/materialen voor MTV in groep 3 t/m 8	**n=97**	**n=42**	**n=21**
Vooral de taalmethode	79%	79%	86%
Vooral zelf ontwikkeld les- en leermateriaal	2%	5%	5%
Beide in ongeveer gelijke mate	9%	5%	10%
Anders, namelijk****	9%	12%	0%

*Typering taalbeleidsplan anders: Het taalbeleidsplan is in (door)ontwikkeling, logopedist in dienst
**Doelen taalbeleids-/schoolplan gebaseerd op anders: SLO-doelen, doelen opgesteld door logopedie, ZML leerlijnen,
***Lesmethode materialen voor MTV groep 1/2 anders: LOGO 3000, Leerlijnen Jonge Kind Parnassys, geen groep 1/2 aanwezig op de school
****Lesmethode/materialen voor MTV groep 3-8 anders: methoden voor wereld oriënterende vakken of sociaal-emotionele vaardigheden.

Tabel 7.5.4: Domeinspecifieke kenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Situatie leerlijn **luisteren**	n=97	n=42	n=21
De leerkrachten volgen de leerlijn uit de taalmethode	93%	86%	81%
Er is een eigen leerlijn uitgewerkt per leerjaar	2%	2%	10%
Er is een eigen leerlijn uitgewerkt per bouw	0%	0%	0%
Er is een eigen leerlijn uitgewerkt voor de hele school (doorgaande leerlijn)	1%	5%	10%
Anders, namelijk..*	4%	7%	0%
Situatie leerlijn **spreken**	n=97	n=42	n=21
De leerkrachten volgen de leerlijn uit de taalmethode	92%	83%	81%
Er is een eigen leerlijn uitgewerkt per leerjaar	2%	2%	10%
Er is een eigen leerlijn uitgewerkt per bouw	0%	2%	0%
Er is een eigen leerlijn uitgewerkt voor de hele school (doorgaande leerlijn)	0%	5%	10%
Anders, namelijk..*	6%	7%	0%
Situatie leerlijn **gesprekken**	n=97	n=42	n=21
De leerkrachten volgen de leerlijn uit de taalmethode	87%	83%	81%
Er is een eigen leerlijn uitgewerkt per leerjaar	2%	0%	10%
Er is een eigen leerlijn uitgewerkt per bouw	0%	5%	0%
Er is een eigen leerlijn uitgewerkt voor de hele school (doorgaande leerlijn)	1%	5%	10%
Anders, namelijk..*	10%	7%	0%
Afzonderlijke leerlijnen of geïntegreerde lijn	n=97	n=42	n=21
Afzonderlijke leerlijnen	2%	5%	5%
Een geïntegreerde leerlijn	70%	69%	71%
Deels afzonderlijk, deels geïntegreerd	20%	21%	24%
Anders, namelijk..**	8%	5%	0%

*Situatie leerlijn luisteren, spreken en gesprekken anders: Oriëntatie op nieuwe leerlijn, combinatie van verschillende leerlijnen in onder- en bovenbouw, geen vaste leerlijn

**Situatie afzonderlijke leerlijnen of geïntegreerde lijn anders: Opzoek naar een andere methode, niet duidelijk of gehanteerde methode afzonderlijke of geïntegreerde leerlijnen heeft, gebruik van kwaliteitskaart.

Tabel 7.5.5: Domeinspecifieke kenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao		% sbo		% so	
	n=97		n=42		n=21	
	Groep 1-2	groep 3-8	Groep 1-2	groep 3-8	groep 1-2	groep 3-8
Monitoring voortgang **luisteren**						
Nee	11%	27%	12%	24%	5%	14%
Ja, incidenteel	9%	31%	12%	36%	24%	43%
Ja, structureel	78%	39%	67%	33%	62%	43%
Anders, namelijk..*	1%	3%	10%	7%	10%	0%
Monitoring voortgang **spreken**						
Nee	9%	22%	12%	26%	5%	14%
Ja, incidenteel	11%	34%	10%	38%	24%	43%
Ja, structureel	78%	41%	69%	29%	62%	43%
Anders, namelijk..*	1%	3%	10%	7%	10%	0%
Monitoring voortgang **gesprekken**						
Nee	19%	33%	17%	29%	5%	14%
Ja, incidenteel	12%	29%	17%	40%	29%	48%
Ja, structureel	66%	34%	57%	24%	57%	38%
Anders, namelijk..*	3%	4%	10%	7%	10%	0%

* Bij de toelichting op 'Anders, namelijk..' voor groep 1 en 2 wordt voornamelijk genoemd dat er geen groep 1 en 2 op school aanwezig is. Bij de toelichting op 'Anders, namelijk… ' voor groep 3 t/m 8 wordt gemeld dat het alleen gebeurt wanneer het is opgenomen in de methodegebonden toetsen of dat het als geheel in het rapport wordt vermeld.

Tabel 7.5.6: Domeinspecifieke kenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao		% sbo		% so	
	groep 1-2	groep 3-8	groep 1-2	groep 3-8	groep 1-2	groep 3-8
Instrument monitoring voortgang **luisteren** *(meer antwoorden mogelijk)*	(n=85	n=68	n=33	n=29	n=18	n=18
Toetsen uit de (taal)methode	6%	75%	6%	55%	39%	67%
Methode-onafhankelijke toetsen (bijv. Cito)	7%	26%	18%	34%	17%	50%
Zelfontworpen toetsen	1%	3%	0%	3%	0%	0%
Observaties door groepsleerkracht	96%	71%	91%	79%	83%	83%
Observaties door taalcoördinator	2%	1%	12%	14%	0%	0%
Anders, namelijk..*	19%	3%	27%	10%	6%	6%
Instrument monitoring voortgang **spreken** *(meer antwoorden mogelijk)*	n=87	n=73	n=33	n=28	n=18	n=18
Toetsen uit de (taal)methode	6%	67%	6%	61%	39%	56%
Methode-onafhankelijke toetsen (bijv. Cito)	6%	23%	15%	21%	22%	39%
Zelfontworpen toetsen	1%	4%	0%	4%	0%	0%
Observaties door groepsleerkracht	97%	75%	91%	86%	83%	89%
Observaties door taalcoördinator	3%	3%	9%	14%	0%	0%
Anders, namelijk..*	18%	7%	30%	14%	6%	6%
Instrument monitoring voortgang **gesprekken** *(meer antwoorden mogelijk)*	n=76	n=61	n=31	n=27	n=18	n=18
Toetsen uit de (taal)methode	5%	66%	6%	63%	33%	56%
Methode-onafhankelijke toetsen (bijv. Cito)	7%	23%	13%	19%	22%	39%
Zelfontworpen toetsen	1%	3%	0%	4%	0%	0%
Observaties door groepsleerkracht	97%	79%	90%	89%	83%	89%
Observaties door taalcoördinator	1%	2%	13%	15%	0%	0%
Anders, namelijk..*	17%	5%	32%	11%	6%	6%

* Bij de 'Anders, namelijk..' toelichting voor groep 1 en groep 2 worden voornamelijk instrumenten (methode-onafhankelijke toetsen/volgsystemen) genoemd waarmee de voortgang gemonitord wordt; BOSOS, LOGO 3000, KIJK, IEP, leerlijnen in Parnassys. Ook een logopedist wordt genoemd. Ook bij de toelichting voor 'Anders, namelijk..' in groep 3 t/m 8 wordt de logopedist genoemd. Ook bestaande instrumenten, spreekbeurten/spreekopdrachten worden genoemd.

Tabel 7.5.7: Domeinspecifieke kenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao		% sbo		% so	
	n=97		n=42		n=21	
	groep 1-2	groep 3-8	groep 1-2	groep 3-8	groep 1-2	groep 3-8
Ondersteuning achterstand **luisteren**						
Nee	7%	15%	17%	10%	29%	19%
Ja, incidenteel	41%	58%	19%	45%	19%	38%
Ja, structureel	52%	27%	64%	45%	52%	43%
Ondersteuning achterstand **spreken**						
Nee	2%	9%	12%	7%	24%	14%
Ja, incidenteel	38%	52%	17%	40%	19%	38%
Ja, structureel	60%	39%	71%	52%	57%	48%
Ondersteuning achterstand **gesprekken**						
Nee	13%	19%	14%	19%	24%	14%
Ja, incidenteel	38%	54%	24%	43%	24%	43%

Tabel 7.5.8: Domeinspecifieke kenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Betrokken(en) bij de gerichte ondersteuning: (meer antwoorden mogelijk)	**n=59**	**n=31**	**n=12**
Groepsleerkracht(en)	92%	84%	100%
Taalcoördinator	27%	13%	8%
Interne begeleider	78%	55%	67%
Logopedist(e)	66%	90%	92%
Anders, namelijk..*	51%	16%	33%
Aard van de gerichte ondersteuning: (meer antwoorden mogelijk)	**n=59**	**n=31**	**n=12**
Verlengde instructie	69%	68%	100%
Extra oefening in groepjes	90%	94%	92%
Remedial Teaching	56%	55%	17%
Inzet extra oefenmateriaal	59%	65%	58%
Aangepaste taken/opdrachten	44%	71%	58%
Anders, namelijk..**	15%	10%	33%
Percentage onderwijstijd MTV van totale tijd taalonderwijs	**n=95**	**n=41**	**n=19**
Groep 1 en 2	50%	44%	51%
Groep 3 t/m 8	28%	31%	33%

* Genoemde andere betrokkenen zijn (ambulante begeleiding van) externe hulpverleners en specialisten, gedragswetenschappers, leerkrachten met speciale taak, onderwijsassistenten, *remedial teachers*.
** Bij andere gerichte ondersteuning wordt o.a. pre-teaching genoemd of andere vormen van de bovenstaande antwoordcategorieën.

7.5.9: Domeinspecifieke kenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Integratie onderwijsaanbod MTV met andere taaldomeinen	**n=95**	**n=40**	**n=19**
Op onze school is de integratie van het onderwijs in mondelinge taalvaardigheid met leerinhouden van andere taaldomeinen vastgelegd en geborgd.	25%	30%	16%
Onze school werkt aan een plan van aanpak om onderwijs in mondelinge taalvaardigheid (meer) te gaan integreren met onderwijs in andere taaldomeinen.	34%	43%	58%
Geen van beide	41%	28%	26%
Integratie onderwijsaanbod MTV met andere leergebieden dan taal	**n=95**	**n=40**	**n=19**
Op onze school is de integratie van het onderwijs in mondelinge taalvaardigheid met andere leergebieden dan taal vastgelegd en geborgd.	17%	23%	26%
Onze school werkt aan een plan van aanpak om onderwijs in mondelinge taalvaardigheid (meer) te integreren met onderwijs in andere leergebieden dan taal.	36%	45%	47%
Geen van beide	47%	33%	26%

Tabel 7.5.10: Domeinspecifieke kenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao			% sbo			% so		
groep	2	4	6	2	4	6	2	4	6
Tijd ingeroosterd voor onderwijs in MTV?									
Nee	55%	68%	66%	38%	57%	62%	65%	70%	75%
Ja, voor het leergebied mondelinge taalvaardigheid is specifiek tijd ingeroosterd	26%	12%	13%	43%	19%	14%	20%	15%	10%
Ja, voor één, twee of alle drie de domeinen is specifiek tijd ingeroosterd	13%	8%	9%	10%	12%	12%	10%	10%	10%
Anders, namelijk..*	6%	12%	12%	10%	12%	12%	5%	5%	5%
Minuten ingeroosterd voor onderwijs in MTV	186	128	121	296	220	153	180	103	130
Minuten ingeroosterd voor onderwijs in									
Luisteren	258	208	207	45	48	42	15	23	23
Spreken	253	236	219	23	28	52	15	25	25
Gesprekken	263	229	208	15	8	88	15	25	25

** Bij 'anders namelijk..' wordt voornamelijk genoemd dat er tijd is ingeroosterd voor de taalmethode, waar mondelinge taalvaardigheid deel van uitmaakt. Daarnaast wordt genoemd dat mondelinge taalvaardigheid wordt geïntegreerd binnen thema's en andere vakken.*

Tabel 7.5.11: Domeinspecifieke kenmerken scholen, in percentages, naar schooltype

Kenmerk	% bao	% sbo	% so
Bij hoeveel procent van de teamvergaderingen komen de onderstaande thema's aan de orde?			
(De veiligheid van) het spreekklimaat voor leerlingen in onze school	n=95	n=40	n=19
< 25%	85%	78%	79%
25-50%	9%	20%	5%
50-75%	4%	3%	16%
>75%	1%	0%	0%
De prestaties voor mondelinge taalvaardigheid op groepsniveau en/of schoolniveau	n=95	n=40	n=19
< 25%	79%	75%	79%
25-50%	15%	20%	16%
50-75%	6%	5%	5%
>75%	0%	0%	0%
De kwaliteit van het mondelinge taalvaardigheidsonderwijs (leerstof aanbod, leertijd, didactiek)	n=95	n=40	n=19
< 25%	75%	60%	79%
25-50%	20%	38%	21%
50-75%	4%	3%	0%
>75%	1%	0%	0%
De visie van onze school op het onderwijs in mondelinge taalvaardigheid	n=95	n=40	n=19
< 25%	80%	78%	74%
25-50%	14%	18%	21%
50-75%	5%	5%	5%
>75%	1%	0%	0%
Schaal aandacht voor mondelinge taalvaardigheid in teamvergaderingen*	**1,3**	**1,3**	**1,3**

**gemiddelde stelling 2 t/m 4, schaal 1 (<25%) – 4 (>75%).*

Tabel 7.5.12: Domeinspecifieke kenmerken scholen, gemiddelde score per item en percentages naar antwoordcategorie, naar schooltype

Kenmerk	bao (n=95)	sbo (n=40)	so (n=19)
Onze school heeft een duidelijke visie over de plaats van mondelinge taalvaardigheid in ons curriculum	2,1	2,5	2,4
1 zeer mee oneens	25%	15%	11%
2 enigszins mee oneens	37%	33%	42%
3 enigszins mee eens	37%	45%	42%
4 zeer mee eens	1%	8%	5%
Het belang van (meer) aandacht voor mondelinge taalvaardigheid in het onderwijs wordt door de leerkrachten op onze school breed gedragen	2,7	3,0	2,9
1 zeer mee oneens	11%	0%	0%
2 enigszins mee oneens	21%	20%	26%
3 enigszins mee eens	52%	60%	53%
4 zeer mee eens	17%	20%	21%
Het schoolbestuur heeft zicht op de kwaliteit van het mondelinge taalvaardigheidsonderwijs van onze school	2,1	2,2	2,5
1 zeer mee oneens	34%	33%	21%
2 enigszins mee oneens	27%	23%	32%
3 enigszins mee eens	37%	43%	26%
4 zeer mee eens	2%	3%	21%
Het schoolbestuur heeft in beeld welke leerlingen van onze school in het leergebied mondelinge taalvaardigheid een bepaald referentieniveau behalen	2,1	2,2	2,3
1 zeer mee oneens	39%	28%	32%
2 enigszins mee oneens	24%	33%	21%
3 enigszins mee eens	28%	30%	32%
4 zeer mee eens	8%	10%	16%
Schaal visie school*	**2,3**	**2,5**	**2,5**

Gemiddelde score item 1 t/m 4, 1=zeer mee oneens, 4= zeer mee eens

**Gemiddelde vier items*

8 Woordenschattoets

8.1 Inhoud en afname

De items uit de woordenschattoets zijn samengesteld door prof. dr. Marc Brysbaert van de Universiteit Gent op basis van de uitkomsten van Brysbaert et al. (2014), Bourdead'hui et al. (2020) en Monster et al. (2022). De toets bestond uit 43 items waarbij leerlingen een woord te zien kregen. Bij elk woord dienden zij aan te geven of het woord echt of nep was. In de peiling Digitale Geletterdheid po 2022[11] is deze woordenschattoets eerder ingezet en is de betrouwbaarheid van de woordenschattoets als hoog beoordeeld. Wel is daar gebleken dat twee woorden zich afwijkend gedroegen. Deze woorden zijn dan ook niet meegenomen in de woordenschattoets voor de peiling mondelinge taalvaardigheid. De woordenschattoets bestond daarmee uiteindelijk uit 41 woorden, zoals gepresenteerd in Tabel 8.2. In het onderzoek werd deze toets digitaal afgenomen. Items werden dichotoom gescoord. De totaalscore wordt berekend op basis van het aantal juiste antwoorden. Voor elk woord dat correct werd herkend als echt of nepwoord kreeg de leerling 1 punt.

Ontbrekende waarden

In totaal hebben we voor 99% van de deelnemers (N=4353) aan het onderzoek een geldige score op de woordenschattoets. Van de 4354 leerlingen die (een deel van) de woordenschattoets hebben gemaakt, hebben 4351 leerlingen de woordenschattoets volledig gemaakt. Eén leerling heeft meer dan 50% van de items van de woordenschattoets gemist. Vanwege het grote aantal ontbrekende waarden is besloten deze leerling bij verdere analyses buiten beschouwing te laten. Daarnaast blijven er nog 2 leerlingen over met een percentage ontbrekende waarden van 2% en 15 %. Dit gaat om een zeer kleine groep met een laag percentage missende waarden. Daarom is besloten deze leerlingen in de dataset te behouden. In totaal analyseren we de resultaten op de woordenschattoets van 4353 leerlingen. Op itemniveau komen deze ontbrekende waarden niet structureel voor bij specifieke items.

11 Inspectie van het Onderwijs. (2024). Peil.Digitale geletterdheid einde basisonderwijs 2021-2022. Inspectie van het Onderwijs.

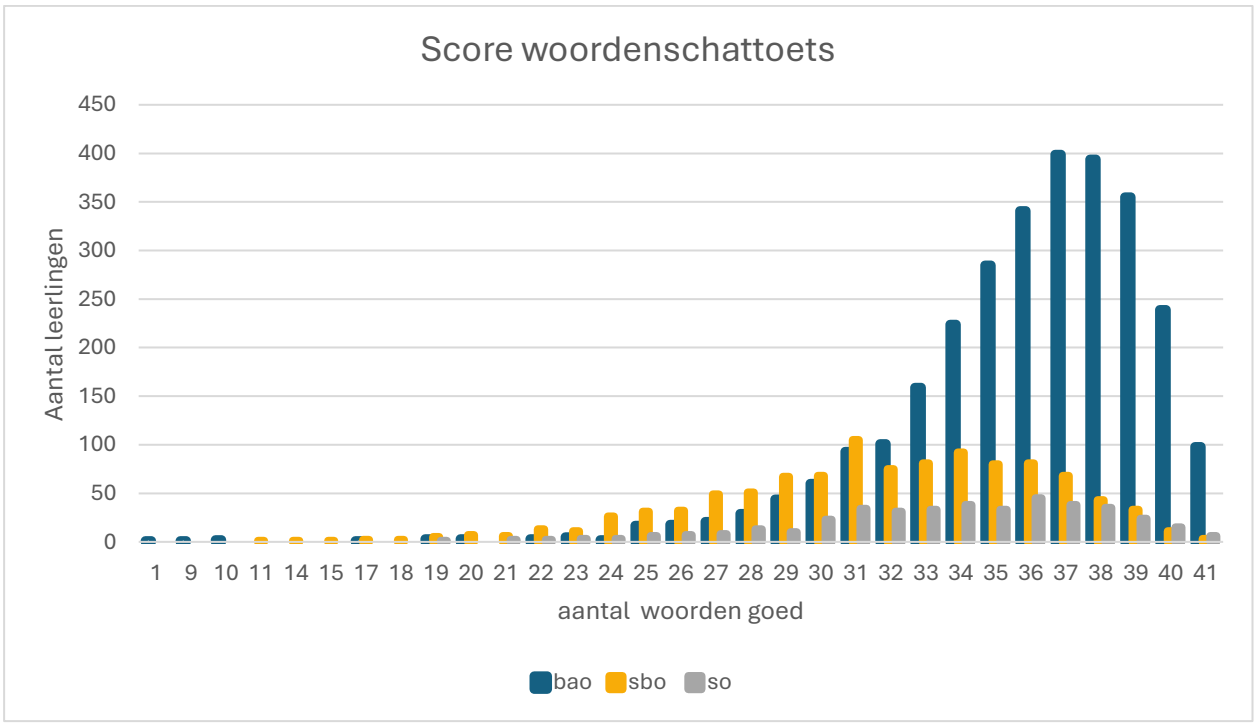
8.2 Resultaten

In Tabel 8.1 worden de resultaten van de woordenschattoets gepresenteerd. Gemiddeld hebben leerlingen 34,7 van de 41 items (85%) goed beoordeeld. Leerlingen in het regulier basisonderwijs hebben gemiddeld 88% van de woorden correct beoordeeld, so-leerlingen 82% en sbo-leerlingen 77%. 107 leerlingen (98 bao, 3 sbo en 6 so) hebben alle items van de toets correct beoordeeld. In Figuur 8.1 wordt de verdeling van de scores voor de verschillende schooltypen gepresenteerd. In Tabel 8.2 wordt per getoetst woord duidelijk welk deel van de leerlingen deze correct heeft beoordeeld. We maken daar geen onderscheid naar schooltype.

Tabel 8.1: Resultaten woordenschattoets naar schooltype

	bao	Sbo	so	Totaal
Gemiddeld aantal woorden correct (sd)	35,9 (3,6)	31,7 (4,6)	33,8 (4,1)	34,7 (4,3)
Percentage correct	88%	77%	82%	85%
N	2898	1045	410	4353

Figuur 8.1: Verdeling scores woordenschattoets, naar schooltype



Tabel 8.2: Beschrijvende statistieken woordenschattoets (totaal N= 4353)

		Incorrect		Correct	
	Bestaand = 1 Niet-bestaand=2	n	%	n	%
Vergallen	1	2829	65,0%	1524	35,0%
Donder	1	134	3,1%	4219	96,9%
Aanbol	2	611	14,0%	3742	86,0%
Afleggen	1	181	4,2%	4172	95,8%
Bijdraaien	1	492	11,3%	3861	88,7%
Meekmoedig	2	410	9,4%	3943	90,6%
Circuit	1	779	17,9%	3574	82,1%
Mandelaan	2	1325	30,4%	3028	69,6%
IJzeren	1	391	9,0%	3962	91,0%
Bip	2	583	13,4%	3770	86,6%
Aftershave	1	1594	36,6%	2759	63,4%
Intenhouten	2	1004	23,1%	3349	76,9%
Weldoener	1	1250	28,7%	3103	71,3%
Chimpansee	1	175	4,0%	4178	96,0%
Continent	1	266	6,1%	4087	93,9%
Substantie	1	692	15,9%	3661	84,1%
Beslag	1	92	2,1%	4261	97,9%
Loerscheer	2	292	6,7%	4061	93,3%
Fossiel	1	285	6,5%	4068	93,5%
Slan	2	300	6,9%	4053	93,1%
Gaas	1	898	20,6%	3455	79,4%
Rechtsperwiet	2	716	16,4%	3637	83,6%
Woerbel	2	569	13,1%	3784	86,9%
Suf	1	320	7,4%	4033	92,6%
Hieging	2	455	10,5%	3898	89,5%
Kinkeldruk	2	1162	26,7%	3191	73,3%
Bashouden	2	833	19,1%	3520	80,9%
Hindernis	1	197	4,5%	4156	95,5%
Gieten	1	133	3,1%	4220	96,9%
Afvragen	1	142	3,3%	4211	96,7%
Wild	1	89	2,0%	4264	98,0%
Ontgaan	1	708	16,3%	3644	83,7%
Duipen	2	885	20,3%	3468	79,7%
Huismus	1	774	17,8%	3579	82,2%
Kathedraal	1	668	15,3%	3685	84,7%
Ovaal	1	330	7,6%	4022	92,4%
Kluizenaar	1	1580	36,3%	2772	63,7%

Mondeling	1	398	9,1%	3954	90,9%
Stamsteger	2	605	13,9%	3747	86,1%
Geknoerd	2	821	18,9%	3531	81,1%
Ordenen	1	1365	31,4%	2987	68,6%

9 Samenhang prestaties en kenmerken van leerlingen, leerkrachten en onderwijsleerproces

9.1 Inleiding

Om na te gaan hoe verschillende kenmerken van leerlingen, leerkrachten, scholen en het onderwijsleerproces samenhangen met de prestaties van leerlingen, worden in dit hoofdstuk de resultaten van verschillende multilevel analyses gepresenteerd. Voordat de resultaten van deze analyses worden gepresenteerd (paragraaf 9.5.), wordt uiteengezet welke predictoren mee worden genomen in de modellen (paragraaf 9.2), wordt de imputatiestrategie toegelicht (paragraaf 9.3) en wordt besproken hoe de multilevel modellen worden opgebouwd (paragraaf 9.4).
In de multilevel analyse wordt de informatie verzameld met verschillende instrumenten bijeengebracht, om zo na te kunnen gaan welke leerling-, leerkracht- en schoolkenmerken van invloed zijn op de prestaties van leerlingen in de mondelinge taalvaardigheid. De leerlingkenmerken zijn afkomstig uit verschillende instrumenten, zoals de leerlingvragenlijst, de woordenschattoets, de quickscan en het leerlingformulier. Middels verschillende taken zijn de prestaties van leerlingen op het gebied van mondelinge taalvaardigheid in kaart gebracht. Op basis van deze taken konden er vaardigheidsscores voor de domeinen luisteren, spreken en gesprekken toegekend worden aan de leerlingen. Aangezien er gewerkt is met een incompleet toetsdesign, hebben leerlingen niet voor alle domeinen taken gemaakt. Wanneer een leerlingen geen taken voor een bepaald domein heeft gemaakt, kon er geen vaardigheidsscore toegekend worden. Een deel van de leerlingen heeft een score op twee van de drie vaardigheden, een zeer klein deel van de leerlingen heeft een score op alle drie de vaardigheden[12].
Ook voor de overige instrumenten geldt dat deze bij een deel van de leerlingen niet of slechts gedeeltelijk afgenomen zijn. In Tabel 9.1.1 wordt het aantal geldige responses per instrument getoond, uitgesplitst naar schooltype.
In paragraaf 9.2 wordt toegelicht welke variabelen uit de verschillende instrumenten zijn geselecteerd voor de multilevel analyse. In paragraaf 9.3 lichten we toe voor welke variabelen missende waarden worden geïmputeerd en brengen we het aantal waarnemingen waarvoor op de multilevel modellen geschat zullen worden in beeld.

12 560 leerlingen hebben een score op luisteren en spreken ; 999 leerlingen hebben een score op spreken en gesprekken; 559 leerlingen hebben een score op luisteren en gesprekken ; 35 leerlingen hebben een score op luisteren, spreken en gesprekken

Tabel 9.1.1: Aantal geldige responses per instrument

	bao	sbo	so	Totaal
Aantal leerlingen in de dataset	2926	1053	414	4393
Aantal leerlingen per instrument				
Leerlingvragenlijst	2895	1042	409	4346
Woordenschattoets	2898	1045	410	4353
Quickscan	2198	682	270	3150
Leerlingformulier13	2806	1023	399	4228
Vaardigheidsscores*	2619	872	349	3840
Luisteren	1258	373	120	1751
Spreken	1446	508	184	2138
Gesprekken	1463	446	230	2139
Leerkrachtvragenlijst*	2307	754	305	3366
Schoolleidersvragenlijst*	2199	777	312	3288

*In deze tabel tonen we het aantal leerlingen waarvoor de resultaten van de leerkracht- en schoolleidersvragenlijst gekoppeld konden worden.

9.2 Geselecteerde variabelen

Voorafgaan aan de imputatie en het uitvoeren van de multilevel analyses is een selectie gemaakt van variabelen die mee worden genomen als predictoren in de opbouw van de multilevel modellen (voor meer informatie over de opbouw zie paragraaf 9.3). De variabelen zijn gekozen op basis van inhoudelijke argumenten (bijv. wetenschappelijke literatuur) en zijn hieronder in Tabel 9.2.1 weergegeven.

13 Middels het leerlingformulier werd bij de school informatie over de leerling opgevraagd, zoals geslacht, leeftijd en LVS-scores.

Tabel 9.2.1: Geselecteerde predictoren

Algemene leerlingkenmerken	Domeinspecifieke leerlingkenmerken	Leerkracht- en schoolkenmerken	Kenmerken van de onderwijsleercontext
Geslacht	Thuis praten over … (schaal; (bijna) nooit tot (bijna) elke dag)	Aantal jaren werkzaam als leerkracht	Mate van effectieve instructie in luisteren/ spreken/ gesprekken (schaal; helemaal niet van toepassing tot helemaal van toepassing)[14]
Leeftijd in maanden	Ervaren klasklimaat (schaal; nooit problemen tot bijna elke dag)	Ervaren bekwaamheid in onderwijs in mondelinge taalvaardigheid (schaal; zeer mee oneens tot zeer mee eens)	Frequentie taken voor luisteren/ spreken/ gesprekken (schaal; nooit tot meer dan 12 keer)[14]
Woordenschat (aantal goed)	Ervaren spreekvrijheid in de klas (schaal)	Schoolweging (bo); percentage uitstroom pro/vso (sbo); percentage doorstroom naar vmbo g-t/havo/vwo (so)	Onderwijstijd besteed aan mondelinge taalvaardigheid in minuten
Frequentie waarin thuis Nederlands wordt gesproken	Communicatie-problemen (schaal; niet aanwezig tot in ernstige mate)		Monitoren van de voortgang in luisteren/ spreken/ gesprekken[14]
Aantal boeken thuis	Internaliserend probleemgedrag (schaal; geen tot (sterk) probleemgedrag)		Bieden van ondersteuning voor luisteren/ spreken/ gesprekken[14]
Taalrijkheid van de activiteiten van de ouders (schaal; (bijna) nooit tot (bijna) elke dag)	Externaliserend probleemgedrag (schaal; geen tot (sterk) probleemgedrag)		Mate van aandacht voor mondelinge taalvaardigheid in vergaderingen (schaal; minder dan 25% tot >75%)
	Spreekangst (sterke spreekangstig tot niet spreekangstig)		Aandacht voor mondelinge taalvaardigheid in visie school en bestuur (schaal; zeer mee oneens tot zeer mee eens)
	Stimulerende/ belemmerende factoren (schaal; belemmerend tot stimulerend)		

9.3 Imputatie

Voorafgaand aan de multilevel analyses zijn de ontbrekende waarden op de predictoren die mee worden genomen in de multilevel analyses geïmputeerd in R (pakket Mice; van Buuren & Groothuis-Oudshoorn, 2011), waarbij rekening is gehouden met de hiërarchische structuur van de data. Voorafgaand aan het

14 Voor de luistervaardigheidsmodellen worden de variabelen over luisteren meegenomen, voor de spreekvaardigheidsmodellen worden de variabelen over spreken meegenomen, voor de gespreksvaardigheidsmodellen worden de variabelen over gesprekken meegenomen.

imputatieproces zijn leerlingen met teveel missende waarden (50% of meer op de leerlingvragenlijst, leerkrachtvragenlijst en/ of schoolvragenlijst) verwijderd uit de dataset (1485 leerlingen). Ook leerlingen die geen enkele vaardigheidsscore hebben ontvangen voor één van de deelvaardigheden (Luisteren, Spreken en Gesprekken) zijn verwijderd uit de dataset (342 leerlingen[15]). Ook leerlingen voor wie het geslacht onbekend was zijn verwijderd uit de dataset (104 leerlingen), omdat het imputeren van geslacht niet wenselijk is.
Vanwege de relatief kleine aantallen leerlingen in het sbo en so is besloten om de imputatie voor de totale groep leerlingen (bao, sbo, so) als geheel uit te voeren met onderwijstype als een van de voorspellende variabelen. Ontbrekende waarden zijn geschat op basis van alle beschikbare predictoren (zie Tabel 9.2.1 in paragraaf 9.2). Alleen schoolweging (voor bao), percentage uitstroom naar vso/pro (sbo) en percentage uitstroom naar vmbo g-t/havo/vwo (so) zijn niet meegenomen als predictor, omdat deze variabele anders is voor de drie onderwijstypen en niet als één variabele opgenomen kan worden. Ook zijn de afhankelijke variabelen (WML-scores voor Luisteren, Spreken en Gesprekken) niet meegenomen als predictor. Het unieke klasnummer is gebruikt als clustervariabele.
De ontbrekende waarden van zowel de predictoren op het niveau van de leerling (niveau 1) als de predictoren op het niveau van de leerkracht/klas/school (niveau 2) zijn geïmputeerd volgens de Predictive Mean Matching-methode (PMM; pmm voor niveau 1 en 2lonly.pmm voor niveau 2), gebruikmakend van vijf imputatierondes. Mice gaat uit van de Fully Conditional Specification (FCS) methode, ook bekend als multivariate imputation by chained equations: elke variabele wordt om de beurt geïmputeerd middels een reeks van imputatiemodellen. PMM houdt rekening met het meetniveau van de variabelen die geïmputeerd dienen te worden, het is een generieke methode[16]. De random seed is vastgezet (26091983) waardoor het mogelijk is om de imputatie exact te repliceren en daarmee op dezelfde resultaten te komen. Er zijn vijf imputatierondes uitgevoerd die resulteren in vijf verschillende geïmputeerde datasets. Deze sets en de oorspronkelijke niet geïmputeerde dataset zijn bewaard als een mids-object zodat direct gepoolde multilevel analyses uitgevoerd kunnen worden in R.

9.4 Opzet multilevel analyses

Vanwege de hiërarchische structuur van de geïmputeerde datasets zijn multilevel modellen toegepast om de onderzoeksvragen te kunnen beantwoorden. De multilevel modellen betreffen steeds random intercept modellen (Snijders & Bosker, 2012; pp. 41). We onderscheiden twee niveaus: de leerlingen vormen niveau 1 en de leerkracht/klas is niveau 2. Er is gekozen om het schoolniveau niet mee te nemen, omdat er scholen zijn die slechts met één klas hebben deelgenomen aan het onderzoek en voor deze scholen het klasniveau gelijk is aan het schoolniveau. De gehanteerde twee niveaus impliceren dat de totale variantie in mondelinge taalvaardigheid (Luisteren, Spreken of Gesprekken) van de leerlingen wordt opgedeeld in variantie op niveau 1 (tussen leerlingen) en niveau 2 (tussen scholen).

15 In totaal hadden 553 leerlingen geen enkele vaardigheidsscore voor een van de deelvaardigheden, 211 leerlingen zijn in de vorige stap echter al verwijderd uit de dataset omdat zij teveel missende waarden op een van de vragenlijsten hadden.
16 Zie Tabel 7.2 in https://stefvanbuuren.name/fimd/sec-multioutcome.html

De afhankelijke variabele is in alle modellen een Weighted Maximum Likelihood (WML) schatter van de mondelinge taalvaardigheid (Luisteren, Spreken of Gesprekken) van individuele leerlingen, genaamd theta. Voor elk gepoold multilevel model is de proportie verklaarde totale variantie berekend volgens de volgende formule (Snijders & Bosker, 2012):

$$R^2 = 1 - \frac{\sigma^2 model + \tau^2 model}{\sigma^2 nulmodel + \tau^2 nulmodel}$$

waarbij $\sigma^2 model$ gelijk staat aan de variantie tussen leerlingen in het analysemodel, $\tau^2 model$ gelijk staat aan de variantie tussen klassen in het analysemodel, $\sigma^2 nulmodel$ gelijk staat aan de variantie tussen leerlingen in het nulmodel waarin alleen het type onderwijs is meegenomen als predictor en $\tau^2 nulmodel$ gelijk staat aan de variantie tussen klassen in het nulmodel waarin alleen een random intercept is meegenomen.

De vijf geïmputeerde datasets maken het toepassen van gepoolde multilevel analyses noodzakelijk voor het correct schatten van de regressiecoëfficiënten, varianties en hun standaardfouten (Snijders & Bosker, 2012, p 135). Er is gebruikgemaakt van het R-pakket *lmer* (*linear mixed effects* models) met de toevoeging van *with* om een serie van geneste statistische modellen te fitten. Er is vijf keer een random intercept model geschat waarna de schattingen van de parameters zijn samengenomen. Met *testEstimates(as.mitml.result)* zijn deze gepoolde resultaten van de modellen verkregen.

De analyses zijn per deelvaardigheid (Luisteren, Spreken en Gesprekken) en per onderwijstype (bao, sbo en so) uitgevoerd. Per combinatie van deelvaardigheid en onderwijstype is allereerst een nulmodel geschat (model 0; onderzoeksvraag 5), gevolgd door een model met algemene leerlingkenmerken (model 1; onderzoeksvraag 6), een model met domeinspecifieke leerlingkenmerken (model 2; onderzoeksvraag 7), een model met algemene en domeinspecifieke leerkracht- en schoolkenmerken (model 3; onderzoeksvraag 8, 9 en 10) en een model met kenmerken van de onderwijsleercontext (model 4; onderzoeksvraag 11). Vervolgens zijn, uit deze vier modellen, de predictoren met een p-waarde <.10 meegenomen in het samenstellen van een totaalmodel. Niet significante predictoren zijn vervolgens één voor één verwijderd uit het totaalmodel om zo tot een spaarzaam model met alleen significante predictoren te komen (onderzoeksvraag 12).

Voorafgaand aan de analyses zijn de predictoren gestandaardiseerd om zo vergelijkingen tussen predictoren mogelijk te maken. Alleen Geslacht, Monitoren van de voortgang en Bieden van ondersteuning zijn niet gestandaardiseerd omdat deze variabelen als factoren (met contrasten) zijn meegenomen in de analyses.

Voor ieder model is ook gekeken naar de Variance Inflation Factor (VIF) om de mate van multicollineariteit tussen de predictorvariabelen in de modellen te onderzoeken. Voor deze analyses zijn de modellen opnieuw geanalyseerd, maar met een enkele geïmputeerde dataset (de eerste). Waarden groter dan 5 zijn problematisch en kunnen een indicatie zijn voor problemen bij het schatten van de regressiecoëfficiënten. Er zijn geen VIF-waarden groter dan 5 gevonden in de analyses.

9.5 Resultaten multilevel analyses

Per model en per deelvaardigheid worden de resultaten van de predictoren (*fixed effecten*) gepresenteerd en geïnterpreteerd. Daarnaast worden de random effecten gepresenteerd (klas en leerling). De random effecten tonen de variatie op respectievelijk klas- en leerlingniveau, maar zijn moeilijk los te interpreteren omdat de waarde sterk afhankelijk is van het model en de variabelen in het model en de schaal van de afhankelijke variabele. De ICC laat zien hoeveel van de variantie wordt verklaard door verschillen op klasniveau en is daarmee proportioneel en beter te interpreteren. Een ICC < .10 suggereert dat de meeste variantie wordt verklaard door individuele verschillen; een ICC tussen .10 en .25 suggereert dat de klas enige invloed heeft, maar dat deze niet substantieel is (deze effecten zijn typisch voor onderwijskundige contexten); een ICC > .25 wijst op een sterke invloed van het groepsniveau (Hox, 2010). De R^2 geeft het percentage verklaarde variantie voor het totale model weer.

9.5.1 Regulier basisonderwijs (bao)

Hieronder worden achtereenvolgens de resultaten van het nulmodel en de modellen met algemene leerlingkenmerken (model 1), domeinspecifieke leerlingkenmerken (model 2), leerkracht- en schoolkenmerken (model 3) en kenmerken van het onderwijsleerproces (model 4) voor het regulier basisonderwijs gepresenteerd. Iedere tabel bevat de resultaten voor de drie verschillende deelvaardigheden (Luisteren, Spreken en Gesprekken). Als laatste is voor iedere deelvaardigheid het totaalmodel en het spaarzaam totaalmodel gepresenteerd.
Voor Luisteren geldt dat 809 leerlingen zijn meegenomen in de analyses; voor Spreken waren dit er 888; voor Gesprekken 902.

Nulmodel

In de tabel hieronder worden de resultaten van de nulmodellen gepresenteerd.

Tabel 9.5.1.1: Nulmodel bo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	*β*	SE	*p*	*β*	SE	*p*
Fixed effecten									
Intercept	1,295	0,074	<0,001	1,383	0,030	<0,001	1,376	0,028	<0,001
Random effecten									
Klas	0,262			0,052			0,045		
Leerling	0,762			0,343			0,302		
ICC	0,256			0,131			0,129		
R^2	-			-					

Het percentage variantie dat toe te schrijven is aan het klasniveau is groter voor Luisteren (25,6%) dan voor Spreken (13,1%) en Gesprekken (12,9%).

Model 1: algemene leerlingkenmerken

In de tabel hieronder worden de resultaten voor model 1 gepresenteerd.

Tabel 9.5.1.2: Model 1 (algemene leerlingkenmerken) bo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	ß	SE	*p*	*ß*	SE	*p*	*B*	SE	*p*
Fixed effecten									
Intercept	1,413	0,841	0,093	3,577	0,525	<0,000	2,578	0,510	<0,001
Geslacht-meisje*	0,075	0,059	0,203	0,061	0,040	0,121	0,028	0,038	0,467
Leeftijd in maanden	-0,002	0,006	0,759	-0,016	0,004	<0,000	-0,009	0,004	0,015
Woordenschat	0,273	0,038	<0,001	0,114	0,027	<0,000	0,066	0,026	0,012
Frequentie waarin thuis Nederlands wordt gesproken	-0,052	0,034	0,124	-0,005	0,022	0,805	-0,017	0,021	0,404
Aantal boeken thuis	0,104	0,034	0,002	0,018	0,022	0,416	0,002	0,021	0,938
Frequentie taalrijke activiteiten ouders	0,112	0,034	0,001	0,014	0,022	0,527	0,061	0,022	0,007
Random effecten									
Klas	0,178			0,039			0,032		
Leerling	0,667			0,326			0,298		
ICC	0,210			0,107			0,097		
R^2	0,175			0,076			0,049		
* Referentiecategorie geslacht: jongen									

Voor geen van de deelvaardigheden is Geslacht een significante voorspeller; meisjes en jongens behalen een gelijke score.

Leeftijd is geen significante voorspeller voor Luisteren, maar wel voor Spreken en Gesprekken. Het negatieve effect betekent dat naarmate leerlingen ouder zijn, ze een lagere vaardigheidsscore behalen voor Spreken en Gesprekken.

Woordenschat is voor alle drie de deelvaardigheden een significante voorspeller. Het positieve effect laat zien dat een grotere woordenschat samenhangt met hogere vaardigheidsscores voor Luisteren, Spreken en Gesprekken.

De frequentie waarmee thuis Nederlands wordt gesproken is voor geen van de deelvaardigheden een significante voorspeller.

Het aantal boeken thuis is alleen voor Luisteren een significante voorspeller. Het positieve effect laat zien dat de aanwezigheid van meer boeken in huis samengaat met een hogere vaardigheidsscore voor Luisteren.

De frequentie van taalrijke activiteiten die de ouders ondernemen is een significante voorspeller voor Luisteren en Gesprekken. Het positieve effect laat zien dat wanneer ouders vaker taalrijke activiteiten ondernemen, de vaardigheidsscores voor Luisteren en Gesprekken van leerlingen hoger zijn.

Opnieuw is de ICC voor Luisteren het grootst; 21,0% van de variantie is toe te schrijven aan het klasniveau. Dit betekent dat de klas enige invloed heeft op de luisterprestaties van de leerlingen. Voor Spreken en Gesprekken is de ICC lager; respectievelijk 10,7% en 9,7% van de variantie is toe te schrijven aan het

klasniveau. Ook voor Spreken betekent dit dat de klas enige invloed heeft op de prestaties. Voor Gesprekken is deze invloed klein en zijn verschillen vooral toe te wijzen aan individuele verschillen.

Model 2: domeinspecifieke leerlingkenmerken

In tabel 9.5.1.3 worden de resultaten voor model 2 gepresenteerd.

Tabel 9.5.1.3: Model 2 (domeinspecifieke leerlingkenmerken) bo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	*β*	SE	*p*	*β*	SE	*p*
Fixed effecten									
Intercept	1,214	0,065	<0.001	1,328	0,031	<0,000	1,332	0,029	<0,001
Thuis praten over	0,082	0,033	0,014	0,068	0,022	0,002	0,054	0,020	0,006
Ervaren klasklimaat	-0,037	0,041	0,367	-0,017	0,023	0,469	0,024	0,022	0,277
Spreekvrijheid in de klas	0,030	0,035	0,384	-0,026	0,022	0,240	0,003	0,021	0,865
Communicatie-problemen	-0,267	0,057	<0,001	-0,126	0,032	<0,001	-0,081	0,032	0,013
Internaliserend probleemgedrag	0,088	0,046	0,058	0,000	0,029	0,997	0,033	0,027	0,233
Externaliserend probleemgedrag	-0,032	0,061	0,613	-0,016	0,031	0,612	-0,050	0,028	0,069
Spreekangst	0,059	0,043	0,171	0,081	0,027	0,003	0,100	0,026	<0,001
Stimulerende/ belemmerende factoren	0,105	0,061	0,103	0,039	0,031	0,216	0,022	0,029	0,453
Random effecten									
Klas	0,182			0,051			0,043		
Leerling	0,701			0,315			0,284		
ICC	0,206			0,139			0,130		
R^2	0,138			0,073			0,058		

De schaal Thuis praten over is een significante voorspeller voor alle drie de deelvaardigheden. Het positieve effect laat zien wanneer er thuis vaker gepraat wordt over zaken, leerlingen hogere vaardigheidsscores behalen voor Luisteren, Spreken en Gesprekken.
Het ervaren klasklimaat, de spreekvrijheid in de klas en de stimulerende/ belemmerende factoren zijn voor geen van de deelvaardigheden een significante voorspeller.
Gerapporteerde communicatieproblemen is voor alle drie de deelvaardigheden een significante voorspeller. Het negatieve effect laat zien dat wanneer de leerkracht communicatieproblemen bij de leerling ervaart, dit samenhangt met lagere vaardigheidsscores voor Luisteren, Spreken en Gesprekken.
Internaliserend probleemgedrag is voor geen van de deelvaardigheden een significante voorspeller. Voor Luisteren is het effect echter wel marginaal significant en daarom wordt deze predictor voor Luisteren wel meegenomen in de opbouw van het totaalmodel.
Externaliserend probleemgedrag is voor geen van de deelvaardigheden een significante voorspeller. Voor Gesprekken is het effect echter wel marginaal significant en daarom wordt deze predictor voor Gesprekken wel meegenomen in de opbouw van het totaalmodel.

Spreekangst is voor Spreken en Gesprekken een significante voorspeller. Het positieve effect laat zien dat wanneer er volgens de leerkracht bij leerlingen minder sprake is van spreekangst, de vaardigheidsscores voor Spreken en Gesprekken hoger zijn.
Voor zowel Luisteren als Spreken en Gesprekken geldt dat de ICC laat zien dat de klas enige invloed heeft op de prestaties van de leerlingen.

Model 3: leerkracht- en schoolkenmerken

In de tabel 9.5.1.4 worden de resultaten voor model 3 gepresenteerd.

Tabel 9.5.1.4: Model 3 (leerkracht- en schoolkenmerken) bo-leerlingen, uitgesplitst naar deelvaardigheid

Model 3 (leerkracht- en schoolkenmerken) bo-leerlingen, uitgesplitst naar deelvaardigheid									
	Luisteren			Spreken			Gesprekken		
	B	SE	*p*	*B*	SE	*p*	*B*	SE	*p*
Fixed effecten									
Intercept	3,770	0,461	<0,001	1,936	0,229	<0,001	2,340	0,197	<0,000
Jaren werkzaam als leerkracht	0,038	0,075	0,615	0,008	0,031	0,787	-0,026	0,027	0,325
Ervaren bekwaamheid	-0,044	0,056	0,433	-0,043	0,028	0,118	-0,007	0,024	0,764
Schoolweging	-0,085	0,016	<0,001	-0,019	0,008	0,014	-0,033	0,007	<0,000
Random effecten									
Klas	0,161			0,047			0,030		
Leerling	0,761			0,343			0,302		
ICC	0,174			0,120			0,090		
R^2	0,100			0,013			0,043		

Het aantal jaren dat de leerkracht werkzaam is als leerkracht en de ervaren bekwaamheid van de leerkracht zijn voor geen van de deelvaardigheden een significante voorspeller.
De schoolweging is voor alle drie de deelvaardigheden een significante voorspeller. Het negatieve effect laat zien dat leerlingen op scholen met een hogere schoolweging (en dus hoe groter de kans op onderwijsachterstanden) lagere vaardigheidsscores behalen voor Luisteren, Spreken en Gesprekken is.
Voor zowel Luisteren als Spreken geldt dat de ICC laat zien dat de klas enige invloed heeft op de prestaties van de leerlingen. Voor Gesprekken is de ICC dermate klein dat de verschillen in prestaties vooral toe te schrijven zijn aan individuele verschillen.

Model 4: kenkerken van de onderwijsleercontext

In tabel 9.5.1.5 worden de resultaten voor model 4 gepresenteerd.

Tabel 9.5.1.5: Model 4 (kenmerken van de onderwijsleercontext) bo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	ß	SE	*p*	*ß*	SE	*p*	*ß*	SE	*p*
Fixed effecten									
Intercept	1,132	0,218	<0,001	1,383	0,110	<0,000	1,296	0,070	<0,000
Gepercipieerde toepassing van effectieve instructie in luisteren/ spreken/ gesprekken	-0,008	0,078	0,922	-0,041	0,033	0,207	0,008	0,030	0,777
Frequentie taken luisteren/ spreken/ gesprekken	0,129	0,077	0,095	-0,000	0,035	0,995	-0,002	0,029	0,950
Onderwijstijd mondelinge taalvaardigheid	0,128	0,080	0,109	-0,025	0,033	0,451	0,054	0,029	0,066
Monitoren van luisteren/ spreken/ gesprekken – ja, incidenteel*	0,102	0,196	0,603	-0,127	0,087	0,145	0,162	0,077	0,035
Monitoren van luisteren/ spreken/ gesprekken – ja, structureel*	0,119	0,217	0,584	-0,066	0,092	0,474	0,040	0,080	0,619
Bieden van ondersteuning bij luisteren/ spreken/ gesprekken – ja, incidenteel*	0,109	0,244	0,656	0,079	0,114	0,488	0,021	0,084	0,800
Bieden van ondersteuning bij luisteren/ spreken/ gesprekken – ja, structureel*	0,161	0,305	0,598	0,078	0,126	0,535	-0,010	0,097	0,915
Aandacht voor MTV in vergaderingen	-0,159	0,133	0,232	-0,058	0,040	0,146	0,085	0,034	0,013
Aandacht voor MTV in visie school en bestuur	0,028	0,106	0,792	0,047	0,039	0,231	-0,094	0,033	0,005
Random effecten									
Klas	0,271			0,057			0,037		
Leerling	0,761			0,341			0,302		
ICC	0,263			0,144			0,109		
R^2	-0,001			-0,008			0,023		
*Referentiecategorie: Nee									

De gepercipieerde toepassing van effectieve instructie, de frequentie waarin taken worden ingezet in het onderwijs, de onderwijstijd in mondelinge taalvaardigheid en het bieden van ondersteuning zijn geen

significante voorspellers voor de drie deelvaardigheden. Voor Luisteren is het effect van de frequentie van de taken marginaal significant en wordt dit effect dus wel meegenomen in de opbouw van het totaalmodel. Hetzelfde geldt voor de onderwijstijd voor de deelvaardigheid Gesprekken.
Het monitoren van de ontwikkeling is alleen een significante voorspeller voor de deelvaardigheid Gesprekken, waarbij alleen het contrast tussen 'nee' en 'incidenteel' significant is. Het positieve effect laat zien dat leerlingen op scholen waar incidenteel gemonitord wordt een hogere vaardigheidsscore voor Gesprekken behalen dan leerlingen op scholen waar niet gemonitord wordt. Er zijn geen verschillen tussen leerlingen op scholen waar niet en scholen waar structureel gemonitord wordt.
Aandacht voor mondelinge taalvaardigheid in teamvergadering is een significante voorspeller voor Gesprekken. Het positieve effect laat zien dat wanneer mondelinge taalvaardigheid vaker op de agenda van teamvergaderingen staat, de vaardigheidsscore voor Gesprekken hoger is.
Aandacht voor mondelinge taalvaardigheid in de visie van de school en het bestuur is een significante voorspeller voor Gesprekken. Het negatieve effect laat zien dat naarmate de schoolleider het sterker eens is met de stellingen binnen deze schaal, de vaardigheidsscore voor Gesprekken lager is.
Voor zowel Spreken als Gesprekken geldt dat de ICC laat zien dat de klas enige invloed heeft op de prestaties van de leerlingen. Voor Luisteren geldt dat ICC laat zien dat de klas een sterke invloed heeft op de prestaties van de leerlingen (ICC > .25).

Totaalmodel

Hieronder worden eerst de totaalmodellen voor Luisteren, Spreken en Gesprekken gepresenteerd. In de tabel daarna worden de resultaten van de spaarzame totaalmodellen gepresenteerd. De resultaten van de spaarzame modellen worden vervolgens besproken in de tekst.

Tabel 9.5.1.6: Totaalmodel voor bo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	ß	SE	*p*	*ß*	SE	*p*	*ß*	SE	*p*
Fixed effecten									
Intercept	2,565	0,466	<0,001	3,176	0,564	<0,001	0,285	0,539	<0,001
Leeftijd in maanden	-	-		-0,012	0,004	0,001	-0,006	0,004	0,070
Woordenschat	0,226	0,039	<0,001	0,077	0,027	0,005	0,038	0,026	0,139
Aantal boeken thuis	0,088	0,034	0,010	-	-	-	-	-	-
Frequentie taalrijke activiteiten ouders	0,094	0,038	0,014	-	-	-	0,022	0,024	0,355
Thuis praten over	0,015	0,036	0,680	0,059	0,021	0,005	0,038	0,021	0,068
Communicatie-problemen	-0,196	0,052	0,001	-0,096	0,030	0,001	-0,045	0,032	0,155
Internaliserend probleemgedrag	0,006	0,039	0,873	-	-	-	-	-	-
Externaliserend probleemgedrag	-	-	-	-	-	-	-0,050	0,023	0,030
Spreekangst	-	-	-	0,071	0,023	0,002	0,084	0,022	<0,001
Schoolweging	-0,048	0,015	0,002	-0,005	0,008	0,545	-0,021	0,007	0,001
Frequentie taken luisteren/ spreken/ gesprekken	0,030	0,058	0,605	-	-	-	-	-	-
Onderwijstijd mondelinge taalvaardigheid	-	-	-	-	-	-	0,035	0,023	0,141
Monitoren van luisteren/ spreken/ gesprekken – ja, incidenteel*	-	-	-	-	-	-	0,115	0,061	0,061
Monitoren van luisteren/ spreken/ gesprekken – ja, structureel*	-	-	-	-	-	-	-0,000	0,064	0,995
Aandacht voor MTV in vergaderingen	-	-	-	-	-	-	0,103	0,028	<0,001
Aandacht voor MTV in visie school en bestuur	-	-	-	-	-	-	-0,092	0,029	0,001
Random effecten									
Klas	0,129			0,043			0,021		
Leerling	0,654			0,311			0,285		
ICC	0,165			0,121			0,069		
R^2	0,235			0,104			0,188		
*Referentiecategorie: Nee									

Tabel 9.5.1.7: Spaarzaam totaalmodel voor bo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	ß	SE	*p*	*ß*	SE	*p*	*ß*	SE	*p*
Fixed effecten									
Intercept	2,634	0,423	<0,001	3,055	0,528	<0,001	3,155	0,519	<0,001
Leeftijd in maanden	-	-	-	-0,012	0,004	0,001	-0,008	0,003	0,027
Woordenschat	0,227	0,039	<0,001	0,079	0,027	0,004	0,056	0,024	0,021
Aantal boeken thuis	0,088	0,034	0,010	-	-	-	-	-	-
Frequentie taalrijke activiteiten ouders	0,102	0,033	0,002	-	-	-	-	-	-
Thuis praten over …	-	-	-	0,059	0,021	0,005	0,050	0,019	0,010
Communicatie-problemen	-0,195	0,049	<0,001	-0,098	0,030	0,001	-	-	-
Internaliserend probleemgedrag	-	-	-	-	-	-	-	-	-
Externaliserend probleemgedrag	-	-	-	-	-	-	-0,063	0,022	0,004
Spreekangst	-	-	-	0,071	0,023	0,002	0,093	0,021	<0,001
Schoolweging	-0,051	0,014	<0,001	-	-	-	-0,025	0,007	<0,001
Frequentie taken luisteren/ spreken/ gesprekken	-	-	-	-	-	-	-	-	-
Onderwijstijd mondelinge taalvaardigheid	-	-	-	-	-	-	-	-	-
Monitoren van luisteren/ spreken/ gesprekken – ja, incidenteel*	-	-	-	-	-	-	-	-	-
Monitoren van luisteren/ spreken/ gesprekken – ja, structureel*	-	-	-	-	-	-	-	-	-
Aandacht voor MTV in vergaderingen	-	-	-	-	-	-	0,092	0,028	0,001
Aandacht voor MTV in visie school en bestuur	-	-	-	-	-	-	-0,081	0,028	0,004
Random effecten									
Klas	0,127			0,042			0,025		
Leerling	0,652			0,311			0,284		
ICC	0,163			0,120			0,080		
R^2	0,239			0,106			0,110		
*Referentiecategorie: Nee									

Voor Luisteren geldt dat de woordenschat van een leerling, het aantal boeken dat een leerling thuis heeft, de frequentie van taalrijke activiteiten van ouders, de ervaren communicatieproblemen door de leerkracht en de schoolweging significante voorspellers zijn. Leerlingen met een grotere woordenschat, met meer boeken thuis en ouders die vaker taalrijke activiteiten ondernemen behalen hogere vaardigheidsscores voor Luisteren. Voor communicatieproblemen geldt dat leerlingen bij wie meer problemen worden ervaren, een lagere vaardigheidsscore voor Luisteren behalen. Ook voor de schoolweging geldt dat een hogere schoolweging samengaat met een lagere vaardigheidsscore voor Luisteren.

Voor Spreken geldt dat de leeftijd van de leerling, de woordenschat, de mate waarin thuis gepraat wordt over zaken, de ervaren communicatieproblemen en de spreekangst van de leerling significante voorspellers zijn. Leerlingen met een grotere woordenschat, bij wie thuis meer wordt gesproken en bij wie minder sprake is van spreekangst, behalen een hogere vaardigheidsscore voor Spreken. Oudere leerlingen en leerlingen bij wie meer communicatieproblemen worden ervaren, behalen een lagere vaardigheidsscore voor Spreken.
Voor Gesprekken geldt dat de leeftijd van de leerling, de woordenschat, de mate waarin thuis gepraat wordt over zaken, de mate waarin er sprake is van spreekangst, de mate waarin externaliserend probleemgedrag wordt ervaren door de leerkracht, de schoolweging, de mate waarin er aandacht is voor mondelinge taalvaardigheid in vergaderingen en de mate waarin er aandacht is voor mondelinge taalvaardigheid in de visie van de school en bestuur significante voorspellers zijn. Leerlingen met een grotere woordenschat en bij wie mindere sprake is van spreekangst behalen een hogere vaardigheidsscore voor Gesprekken. Ook leerlingen die op scholen zitten waar meer aandacht is voor mondelinge taalvaardigheid tijdens teamvergaderingen behalen een hogere vaardigheidsscore voor Gesprekken. Oudere leerlingen en leerlingen die meer externaliserend probleemgedrag laten zien behalen een lagere vaardigheidsscore. Leerlingen van scholen met een hoge schoolweging en leerlingen van scholen waar de directeur het sterker eens is met stellingen over aandacht voor mondelinge taalvaardigheid in de visie van de school en bestuur behalen een lagere vaardigheidsscore voor Gesprekken.
Voor zowel Luisteren als Spreken geldt dat de ICC laat zien dat de klas enige invloed heeft op de prestaties van de leerlingen. Voor Gesprekken is de ICC dermate klein dat de verschillen in prestaties vooral toe te schrijven zijn aan individuele verschillen.

9.5.2 Speciaal basisonderwijs (sbo)

Hieronder worden achtereenvolgens de resultaten van het nulmodel en de modellen met algemene leerlingkenmerken (model 1), domeinspecifieke leerlingkenmerken (model 2), leerkracht- en schoolkenmerken (model 3) en kenmerken van het onderwijsleerproces (model 4) voor het speciaal basisonderwijs gepresenteerd. Iedere tabel bevat de resultaten voor de drie verschillende deelvaardigheden (Luisteren, Spreken en Gesprekken). Als laatste is voor iedere deelvaardigheid het totaalmodel en het spaarzaam totaalmodel gepresenteerd.
Voor Luisteren geldt dat 211 leerlingen zijn meegenomen in de analyses; voor Spreken waren dit er 301; voor Gesprekken 264.

Nulmodel

In de tabel hieronder worden de resultaten van de nulmodellen gepresenteerd.

Tabel 9.5.2.1: Nulmodel sbo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	*β*	SE	*p*	*β*	SE	*p*
Fixed effecten									
Intercept	0,028	0,089	0,752	0,832	0,056	<0,000	0,953	0,053	<0,000
Random effecten									
Klas	0,085			0,078			0,056		
Leerling	0,951			0,498			0,424		
ICC	0,082			0,136			0,116		
R^2	-			-			-		

Kijkend naar de ICC kan gesteld worden dat voor Luisteren geldt dat de variantie in scores vooral toe te schrijven is aan individuele verschillen tussen leerlingen. Voor Spreken en Gesprekken geldt dat de klas van enige invloed is op de prestaties.

Model 1: algemene leerlingkenmerken

In Tabel 9.5.2.2 worden de resultaten van model 1 gepresenteerd.

Tabel 9.5.2.2: Model 1 (algemene leerlingkenmerken) sbo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	*β*	SE	*p*	*β*	SE	*p*
Fixed effecten									
Intercept	3,347	1,521	0,028	0,717	1,069	0,503	0,678	1,030	0,510
Geslacht-meisje*	-0,053	0,133	0,691	0,155	0,089	0,082	-0,023	0,088	0,789
Leeftijd in maanden	-0,021	0,010	0,046	0,001	0,007	0,856	0,002	0,007	0,726
Woordenschat	0,339	0,065	<0,001	0,184	0,042	<0,001	0,075	0,041	0,066
Frequentie waarin thuis Nederlands wordt gesproken	-0,006	0,061	0,927	0,019	0,047	0,697	0,000	0,044	0,996
Aantal boeken thuis	0,070	0,067	0,302	0,013	0,046	0,775	0,090	0,048	0,061

Frequentie taalrijke activiteiten ouders	0,139	0,064	0,030	-0,007	0,043	0,873	-0,003	0,046	0,946
Random effecten									
Klas	0,021			0,046			0,046		
Leerling	0,845			0,487			0,426		
ICC	0,024			0,087			0,098		
R^2	0,164			0,836			0,017		
* Referentiecategorie geslacht: jongen									

Voor geen van de deelvaardigheden is geslacht een significante voorspeller; meisjes en jongens behalen een gelijke score. Voor Spreken is het effect wel marginaal significant en daarom wordt geslacht wel meegenomen in de opbouw van het totaalmodel voor Spreken.
Leeftijd is alleen voor Luisteren een significante voorspeller. Het negatieve effect laat zien dat oudere leerlingen een lagere vaardigheidsscore voor Luisteren behalen. Voor Spreken en Gesprekken is leeftijd geen significante voorspeller.
Woordenschat is voor Luisteren en Spreken een significante voorspeller. Het positieve effect laat zien dat voor beide deelvaardigheden geldt dat leerlingen met een grotere woordenschat een hogere vaardigheidsscore behalen. Voor Gesprekken is woordenschat geen significante voorspeller.
De frequentie waarmee thuis Nederlands wordt gesproken is voor geen van de deelvaardigheden een significante voorspeller. Dit geldt ook voor het aantal boeken tuis. Voor Gesprekken is het aantal boeken dat een leerling thuis heeft wel marginaal significant. Een aantal boeken thuis wordt daarom voor deze deelvaardigheid meegenomen in de opbouw van het totaalmodel.
De frequentie van taalrijke activiteiten door de ouders is een significante voorspeller voor Luisteren. Het positieve effect laat zien dat leerlingen die ouder thuis vaker taalrijke activiteiten zien of horen ondernemen een hogere vaardigheidsscore behalen. De frequentie van de taalrijke activiteiten door de ouders is geen significante voorspeller voor Spreken en Gesprekken.
Kijkend naar de ICC kan opnieuw gesteld worden dat voor Luisteren geldt dat de variantie in scores vooral toe te schrijven is aan individuele verschillen tussen leerlingen. Voor Spreken en Gesprekken geldt dat de klas van enige invloed is op de prestaties.

Model 2: domeinspecifieke leerlingkenmerken

In Tabel 9.5.2.3 hieronder worden de resultaten van model 2 gepresenteerd.

Tabel 9.5.2.3: Model 2 (domeinspecifieke leerlingkenmerken) sbo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	ß	SE	*p*	*ß*	SE	*p*	*ß*	SE	*p*
Fixed effecten									
Intercept	0,176	0,106	0,101	0,985	0,063	<0,001	1,095	0,067	<0,001
Thuis praten over	0,102	0,072	0,153	0,032	0,042	0,440	-0,006	0,043	0,883
Ervaren klasklimaat	-0,098	0,067	0,140	-0,088	0,040	0,030	-0,053	0,044	0,229
Spreekvrijheid in de klas	0,021	0,076	0,783	0,023	0,042	0,574	-0,048	0,042	0,256
Communicatie-problemen	-0,340	0,104	0,004	-0,100	0,048	0,040	-0,111	0,047	0,020
Internaliserend probleemgedrag	0,052	0,113	0,652	-0,079	0,049	0,107	-0,074	0,051	0,146
Externaliserend probleemgedrag	-0,056	0,114	0,629	-0,003	0,049	0,951	-0,032	0,045	0,478
Spreekangst	0,036	0,127	0,785	0,133	0,048	0,006	0,092	0,048	0,058
Stimulerende/ belemmerende factoren	-0,086	0,121	0,484	0,032	0,061	0,601	-0,010	0,056	0,853
Random effecten									
Klas	0,070			0,063			0,077		
Leerling	0,865			0,445			0,380		
ICC	0,074			0,125			0,168		
R^2	0,097			0,118			0,048		

De mate waarin leerlingen thuis praten over zaken, de spreekvrijheid in de klas, de mate van internaliserend probleemgedrag, de mate van externaliserend probleemgedrag en de mate van aanwezigheid van stimulerende/ belemmerende factoren zijn geen significante voorspellers voor Luisteren, Spreken en Gesprekken.
Het ervaren klasklimaat is een significante voorspeller voor Spreken. Het negatieve effect laat zien dat hoe vaker er problemen worden ervaren, des te lager de vaardigheidsscore voor Spreken is. Het ervaren klasklimaat is geen significante voorspeller van Luisteren en Gesprekken.
Gerapporteerde communicatieproblemen is voor alle drie de deelvaardigheden een significante voorspeller. Het negatieve effect laat zien dat het ervaren van communicatieproblemen door de leerkracht samengaat met lagere vaardigheidsscores voor Luisteren, Spreken en Gesprekken.
Spreekangst is voor Spreken een significante voorspeller. Het positieve effect laat zien dat leerlingen met meer spreekangst een lagere vaardigheidsscore voor Spreken behalen. Spreekangst is geen significante voorspeller voor Luisteren en Gesprekken, maar wordt in de opbouw van het totaalmodel voor Gesprekken wel meegenomen, omdat het effect marginaal significant is.
Kijkend naar de ICC kan opnieuw gesteld worden dat voor Luisteren geldt dat de variantie in scores vooral toe te schrijven is aan individuele verschillen tussen leerlingen. Voor Spreken en Gesprekken geldt dat de klas van enige invloed is op de prestaties.

Model 3: leerkracht- en schoolkenmerken

In Tabel 9.5.2.4 worden de resultaten van model 3 gepresenteerd.

Tabel 9.5.2.4: Model 3 (leerkracht- en schoolkenmerken) sbo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	ß	SE	*p*	*ß*	SE	*p*	*ß*	SE	*p*
Fixed effecten									
Intercept	-0,268	0,309	0,385	1,010	0,198	<0,001	0,970	0,187	<0,001
Jaren werkzaam als leerkracht	0,032	0,091	0,721	-0,000	0,049	0,997	-0,014	0,047	0,767
Ervaren bekwaamheid	0,066	0,101	0,512	-0,059	0,064	0,355	-0,025	0,070	0,720
Percentage uitstroom vso	0,006	0,006	0,343	-0,003	0,004	0,377	-0,000	0,003	0,987
Random effecten									
Klas	0,094			0,079			0,065		
Leerling	0,952			0,499			0,424		
ICC	0,090			0,136			0,133		
R^2	-0,010			-0,003			-0,019		

Geen van de predictoren is een significante voorspeller voor een van de deelvaardigheden.
Kijkend naar de ICC kan opnieuw gesteld worden dat voor Luisteren geldt dat de variantie in scores vooral toe te schrijven is aan individuele verschillen tussen leerlingen. Voor Spreken en Gesprekken geldt dat de klas van enige invloed is op de prestaties.

Model 4: kenmerken van de onderwijsleercontext

In Tabel 9.5.2.5 hieronder worden de resultaten van model 4 gepresenteerd.

Tabel 9.5.2.5: Model 4 (kenmerken van de onderwijsleercontext) sbo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	ß	SE	*p*	*ß*	SE	*p*	*ß*	SE	*p*
Fixed effecten									
Intercept	0,603	0,814	0,458	0,875	0,132	<0,001	1,294	0,122	<0,000
Mate van effectieve instructie in luisteren/ spreken/ gesprekken	-0,169	0,128	0,188	0,002	0,082	0,984	-0,026	0,070	0,705
Frequentie taken luisteren/ spreken/ gesprekken	0,037	0,128	0,775	0,047	0,074	0,523	0,021	0,064	0,743
Onderwijstijd mondelinge taalvaardigheid	0,144	0,127	0,257	-0,041	0,074	0,582	-0,069	0,056	0,219
Monitoren van luisteren/ spreken/ gesprekken – ja, incidenteel*	-0,245	0,458	0,593	-0,100	0,194	0,607	-0,067	0,165	0,687
Monitoren van luisteren/ spreken/ gesprekken – ja, structureel*	-0,212	0,430	0,622	-0,035	0,202	0,861	0,010	0,169	0,953
Bieden van ondersteuning bij luisteren/ spreken/ gesprekken – ja, incidenteel*	-0,566	0,743	0,446	-0,114	0,229	0,618	-0,328	0,150	0,028

Bieden van ondersteuning bij luisteren/ spreken/ gesprekken – ja, structureel*	-0,400	0,706	0,571	0,073	0,223	0,745	-0,446	0,158	0,005
Aandacht voor MTV in vergaderingen	0,119	0,128	0,352	0,075	0,063	0,229	0,021	0,058	0,709
Aandacht voor MTV in visie school en bestuur	-0,025	0,157	0,872	-0,107	0,077	0,164	-0,044	0,065	0,497
Random effecten									
Klas	0,144			0,093			0,038		
Leerling	0,945			0,498			0,423		
ICC	0,132			0,157			0,083		
R^2	-0,051			0,752			0,040		
*Referentiecategorie: Nee									

Voor Luisteren en Spreken geldt dat geen van de predictoren een significante voorspeller is voor de prestaties van leerlingen. Voor Gesprekken is alleen het bieden van ondersteuning een significante voorspeller. De groep leerlingen van wie de schoolleider aan heeft gegeven geen ondersteuning te bieden is afgezet tegen de groep leerlingen van die schoolleider aan heeft gegeven dat leerlingen incidenteel ondersteuning krijgen wanneer nodig en afgezet tegen de groep leerlingen van wie de schoolleider heeft aangegeven structureel ondersteuning te bieden wanneer nodig. De negatieve effecten laten zien dat de vaardigheidsscore voor Gesprekken van leerlingen op scholen waar incidenteel of structureel ondersteuning wordt geboden lager is dan van leerlingen op scholen waar geen ondersteuning wordt geboden.
De resultaten met betrekking tot de variabele bieden van ondersteuning moeten met voorzichtigheid geïnterpreteerd worden. De groep leerlingen van wie de schoolleider aangeeft geen ondersteuning te bieden is klein (21 leerlingen voor Luisteren, 35 leerlingen voor Spreken en 92 leerlingen voor Gesprekken), zeker wanneer deze uitgesplitst wordt naar groepen leerlingen met een vaardigheidsscore voor Luisteren, Spreken en Gesprekken (3 leerlingen vallen in Nee categorie voor ondersteuning en hebben een luisterscore; 28 leerlingen vallen in Nee categorie voor ondersteuning en hebben een spreekscore; 59 leerlingen vallen in Nee categorie voor ondersteuning en hebben een gesprekscore). Er is besloten om dit model nog een keer te analyseren, maar dan zonder de variabelen Bieden van ondersteuning. Resultaten zijn in Tabel 9.5.2.6 terug te vinden.

Tabel 9.5.2.6: Model 4 (kenmerken van de onderwijsleercontext, zonder Bieden ondersteuning) sbo leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	*β*	SE	*p*	*β*	SE	*p*
Fixed effecten									
Intercept	0,044	0,322	0,892	0,875	0,132	<0,001	1,121	0,108	<0,000
Mate van effectieve instructie in luisteren/ spreken/ gesprekken	-0,178	0,121	0,142	-0.000	0,080	>0,999	0,019	0,073	0,799

Frequentie taken luisteren/ spreken/ gesprekken	0,046	0,119	0,388	0,046	0,073	0,527	0,013	0,069	0,848
Onderwijstijd mondelinge taalvaardigheid	0,152	0,120	0,205	-0,036	0,072	0,619	-0,064	0,060	0,283
Monitoren van luisteren/ spreken/ gesprekken – ja, incidenteel*	-0,082	0,380	0,829	-0,083	0,181	0,646	-0,256	0,161	0,112
Monitoren van luisteren/ spreken/ gesprekken – ja, structureel*	-0,103	0,386	0,791	-0,026	0,181	0,884	-0.182	0,163	0,265
Aandacht voor MTV in vergaderingen	0,117	0,124	0,346	0,071	0,061	0,242	-0,007	0,062	0,904
Aandacht voor MTV in visie school en bestuur	-0,050	0,141	0,721	-0,108	0,075	0,153	0,006	0,067	0,929
Random effecten									
Klas	0,119			0,087			0,058		
Leerling	0,948			0,498			0,423		
ICC	0,111			0,148			0,120		
R^2	-0,002			-0,016			-0,002		
*Referentiecategorie: Nee									

De resultaten veranderen niet door het verwijderen van bieden van ondersteuning uit het model. Geen van de overgebleven kenmerken van de onderwijsleercontext is een significante voorspeller van Luisteren, Spreken en Gesprekken.

Totaalmodel

Hierna worden eerst de totaalmodellen voor Luisteren, Spreken en Gesprekken gepresenteerd. In de tabel daarna worden de resultaten van de spaarzame totaalmodellen gepresenteerd. De resultaten van de spaarzame modellen worden vervolgens besproken in de tekst.

Tabel 9.5.2.7: Totaalmodel voor sbo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	ß	SE	*p*	*ß*	SE	*p*	*ß*	SE	*p*
Fixed effecten									
Intercept	3,112	1,504	0,039	0,945	0,067	<0,001	1,094	0,065	<0,001
Geslacht*	-	-	-	0,145	0,085	0,082	-	-	-
Leeftijd in maanden	-0,019	0,010	0,064	-	-	-	-	-	-
Woordenschat	0,273	0,082	0,002	0,128	0,044	0,004	0,017	0,042	0,681
Aantal boeken thuis	-	-	-	-	-	-	0,074	0,045	0,104
Frequentie taalrijke activiteiten thuis	0,143	0,060	0,018	-	-	-	-	-	-
Ervaren klasklimaat	-	-	-	-0,072	0,038	0,057	-	-	-
Communicatie-problemen	-0,156	0,100	0,179	-0,046	0,046	0,320	-0,107	0,047	0,024
Spreekangst	-	-	-	0,169	0,041	<0,001	0,103	0,040	0,011
Random effecten									
Klas	0,033			0,038			0,073		
Leerling	0,810			0,447			0,381		
ICC	0,040			0,079			0,161		
R^2	0,187			0,158			0,054		
*Referentiecategorie geslacht: jongen **Referentiecategorie: Nee									

Tabel 9.5.2.8: Spaarzaam totaalmodel voor sbo-leerlingen, uitgesplitst naar deelvaardigheid

	Luisteren			Spreken			Gesprekken		
	ß	SE	*p*	*ß*	SE	*p*	*ß*	SE	*p*
Fixed effecten									
Intercept	3,332	1,488	0,025	0,998	0,056	<0,001	1,073	0,065	<0,001
Geslacht*	-	-	-	-	-	-	-	-	-
Leeftijd in maanden	-0,021	0,010	0,041	-	-	-	-	-	-
Woordenschat	0,273	0,082	0,002	0,167	0,039	<0,001	-	-	-
Aantal boeken thuis	-	-	-	-	-	-	-	-	-
Frequentie taalrijke activiteiten thuis	0,151	0,061	0,014	-	-	-	-	-	-
Ervaren klasklimaat	-	-	-	-	-	-	-	-	-
Communicatie-problemen	-	-	-	-	-	-	-0,122	0,043	0,004
Spreekangst	-	-	-	0,176	0,040	<0,001	0,106	0,041	0,010
Random effecten									
Klas	0,023			0,042			0,079		
Leerling	0,835			0,452			0,379		
ICC	0,027			0,085			0,173		
R^2	0,172			0,142			0,046		
*Referentiecategorie geslacht: jongen									

Voor Luisteren geldt dat de leeftijd van een leerling, de woordenschat van een leerlingen en de frequentie van taalrijke activiteiten van ouders significante voorspellers zijn. De positieve effecten van woordenschat en de frequentie van taalrijke activiteiten van de ouders laten zien dat leerlingen met een grotere woordenschat en leerlingen die ouders vakere taalrijke activiteiten zien en horen ondernemen een hogere vaardigheidsscore behalen. Het negatieve effect van leeftijd laat zien dat oudere leerlingen een lagere vaardigheidsscore voor Luisteren behalen.
Voor Spreken geldt dat de woordenschat van een leerling en de mate waarin er sprake is van spreekangst significante voorspellers zijn. Leerlingen met een grotere woordenschat en leerlingen bij wie minder sprake is van spreekangst behalen een hogere vaardigheidsscore voor Spreken.
Voor Gesprekken geldt dat de ervaren communicatieproblemen en de mate waarin sprake is van spreekangst significante voorspellers zijn. Het negatieve effect van communicatieproblemen, laat zien dat leerlingen bij wie meer communicatieproblemen worden ervaren een lagere vaardigheidsscore behalen. Het positieve effect van spreekangst laat zien leerlingen die minder spreekangst ervaren een hogere vaardigheidsscore behalen.

9.5.3 Speciaal onderwijs (so)

Voor Luisteren geldt dat voor 91 leerlingen data beschikbaar zijn; voor Spreken zijn dit er 112; voor Gesprekken 116. Het aantal leerlingen voor wie data beschikbaar zijn, is klein. Dit kan een aantal gevolgen hebben voor (het interpreteren van) de resultaten. De kleine steekproefgrootte vermindert de statistische kracht en betrouwbaarheid van regressiemodellen, wat kan leiden tot onnauwkeurige, niet-repliceerbare en moeilijk te interpreteren resultaten. Voor de volledigheid zijn de resultaten van de analyses wel opgenomen in Bijlage 2, maar wordt er geen interpretatie van de resultaten gegeven. Ook zijn er geen totaalmodellen samengesteld voor deze groep leerlingen.

10 Samenvatting

Het doel van het peilingsonderzoek Mondelinge taalvaardigheid is om vast te stellen wat de gespreks-, spreek- en luistervaardigheid van leerlingen aan het eind van het (speciaal) basisonderwijs en speciaal onderwijs is. Daarnaast is vastgesteld hoe het onderwijsleerproces voor mondelinge taalvaardigheid er op de deelnemende scholen uitziet en geeft het een beeld van de kenmerken van leerlingen, leerkrachten en scholen die gerelateerd zijn aan de mondelinge taalvaardigheid van leerlingen. De gespreks-, spreek- en luistervaardigheid van leerlingen is gemeten met verschillende taken. Daarnaast zijn vragenlijsten afgenomen bij de leerlingen, hun leerkrachten en schoolleiders om informatie te verzamelen over achtergrondkenmerken en het onderwijsleerproces. Hierna vatten we de antwoorden op alle onderzoeksvragen samen.

10.1 Leerlingprestaties

1. Hoeveel procent van de leerlingen einde basisonderwijs en van de schoolverlaters speciaal basisonderwijs en speciaal onderwijs beheerst de referentieniveaus (1F en 2F) voor gesprekken voeren, spreken en luisteren en wat is de vaardigheidsverdeling op de onderliggende schaal?

Aan de hand van verschillende luistertoetsversies en verschillende spreek- en gesprekstaken is de luister-, spreek- en gespreksvaardigheid in kaart gebracht aan de hand van drie vaardigheidsschalen. Per domein zijn de vaardigheidsverdelingen tussen de schooltypen met elkaar vergeleken. Daarnaast zijn, waar mogelijk, de 1F- en 2F-referentiecesuren overgezet van de peilingen mondelinge taalvaardigheid bao (2017) en mondelinge taalvaardigheid so/sbo (2018). De bao-peiling uit 2017 en de so/sbo-peiling uit 2018 werden gebruikt voor het overzetten van cesuren ten aanzien van het domein luistervaardigheid. De so/sbo-peiling werd gebruikt voor het overzetten van de cesuren ten aanzien van de domeinen spreekvaardigheid en gespreksvaardigheid.

Bij luistervaardigheid was er een significant verschil gevonden tussen de drie schooltypen. Het grootste verschil zat tussen bao enerzijds en sbo en so anderzijds (effectgrootte: groot). Leerlingen uit het so scoorden net wat hoger dan leerlingen uit het sbo (effectgrootte: klein). In totaal behaalde voor bao 8% van de leerlingen <1F, 47,7% 1F en 44,3% 2F. Van de sbo-leerlingen behaalde 58,7% <1F, 30% 1F, en 11,3% 2F. Tot slot behaalde 40% van de so-leerlingen <1F, 50% 1F en 10% 2F.

Bij spreekvaardigheid was er een significant verschil gevonden tussen de drie schooltypen. Tussen bao enerzijds en sbo en so anderzijds werd het grootste effect gevonden (effectgrootte: groot). Het kleine verschil in prestatie tussen de sbo- en so-leerlingen was niet significant. In totaal behaalde voor bao 18,4% van de leerlingen <1F, 55,9% 1F en 25,7% 2F. Van de sbo-leerlingen behaalde 46,5% <1F, 45,1% 1F, en 8,5% 2F. Tot slot behaalde 42,4% van de so-leerlingen <1F, 43,5% 1F en 14,1% 2F.

Tot slot, bij gespreksvaardigheid was er een significant verschil gevonden tussen de drie schooltypen. Het grootste verschil zat tussen bao enerzijds en sbo en so anderzijds (effectgrootte: groot). Het kleine verschil in prestatie tussen de sbo- en so-leerlingen was niet significant. In totaal behaalde voor bao 31,3% van de leerlingen <1F, 24,5% 1F en 44,2% 2F. Van de sbo-leerlingen behaalde 62,1% <1F, 17,3% 1F, en 20,6% 2F. Tot slot behaalde 65,2% van de so-leerlingen <1F, 12,6% 1F en 22,2% 2F.

Concluderend, voor alle drie domeinen scoorden de leerlingen uit het bao overtuigend hoger dan de leerlingen uit het so/sbo. De so-leerlingen presteerden iets beter op luistervaardigheid dan de sbo-leerlingen, en bij spreek- en gespreksvaardigheid waren de prestaties tussen de sbo- en so-leerlingen weinig verschillend.

2. Hoe kan de prestatie van leerlingen op de referentieniveaus voor luisteren, spreken en gesprekken voeren worden getypeerd naar kenmerken van de taakuitvoering?

De leerlingprestaties zijn geduid in termen van wat leerlingen beheersen op het cesuurpunt van <1F/1F (de zogenaamde 1F-leerling) en op het cesuurpunt 1F/2F (de zogenaamde 2F-leerling). Voor ieder domein zijn voor de 1F- en 2F-leerling de bankscores berekend, i.e., de verwachte somscore indien de 1F- en 2F-leerling alle items hadden gemaakt.

Voor luistervaardigheid geldt dat de 1F-leerling 49,9% van de maximaal te behalen punten zou hebben behaald. Voor de 2F-leerling geldt een percentage van 76,7%. De 1F-leerling beheerst de meeste items niet goed, dan wel matig. De 2F-leerling daarentegen beheerst de meeste items matig tot goed. Het kenmerk van de taakuitvoering *interpreteren* wordt net iets beter beheerst dan het kenmerk van de taakuitvoering *begrijpen*. Ten aanzien van de aard van de activiteit worden de kijk/luister-items (i.e., *kijken & luisteren*-items en de *kritisch kijken en luisteren*) beter beheerst dan de luister-items (*luisteren*-items en de *kritisch luisteren*). *Dialoog*- en *monoloog*-items worden nagenoeg vergelijkbaar beheerst. Tot slot worden 1F-items beter beheerst dan de 2F-items. De 1F/2F-items en de 2F-items worden in vergelijkbare mate beheerst..

Voor spreekvaardigheid geldt dat de 1F-leerling 64,6% van de maximaal te behalen punten zou hebben behaald. Voor de 2F-leerling geldt een percentage van 86,3%. Uitgesplitst per taak haalt een 1F-leerling een percentage van 65,7% op de naverteltaak en van 63,7% op het betoog. De 2F-leerling haalt een percentage van 86,6% op de naverteltaak en van 86,1% op het betoog. De twee taken worden dus in vergelijkbare mate beheerst. Op beide taken werden de items op *vloeiendheid, verstaanbaarheid en grammaticale beheersing* zeer goed beheerst. De 1F-leerling beheerste deze items matig tot goed, en de 2F-leerling beheerste deze items allemaal goed. De items uit de domeinen *uitvoering, samenhang, afstemming op doel, afstemming op publiek* en *woordgebruik* werden – relatief gezien - minder goed beheerst. Desalniettemin waren er voor de 1F-leerling slechts drie items die niet werden beheerst bij de naverteltaak en vier items die niet werden beheerst bij het betoog. De 2F-leerling beheerste – op twee items na – alle items van de naverteltaak goed. Op drie items na werden alle items van het betoog goed beheerst.

Voor gespreksvaardigheid geldt dat de 1F-leerling 79,7% van de maximaal te behalen punten zou hebben behaald. Voor de 2F-leerling geldt een percentage van 88,0%. Uitgesplitst per taak haalt een 1F-leerling 80,1% op het samenwerkingsspel en 79,3% op het raadsel. De 2F-leerling haalt 88,4% op het samenwerkingsspel en 87,6% op het raadsel. Het samenwerkingsspel en het raadsel worden dus in vergelijkbare mate beheerst. Net als bij spreekvaardigheid werden de items op *vloeiendheid, verstaanbaarheid en grammaticale beheersing* zeer goed beheerst. Zowel de 1F- als de 2F-leerling beheerste deze items allemaal goed. De items uit de domeinen *uitvoering, samenhang, afstemming op doel, afstemming op publiek* en *woordgebruik* werden – relatief gezien - minder goed beheerst.

Desalniettemin beheerste de 1F-leerling de items matig tot goed, en beheerste de 2F-leerling - op drie items na - alle items goed.

3. Wat is de onderlinge samenhang in prestaties op de drie subdomeinen?

Het afnamedesign was zo opgezet dat leerlingen verschillende taken/items maakten uit verschillende domeinen, daardoor was het mogelijk om de samenhang tussen de verschillende domeinen te evalueren. Hiertoe werden correlaties berekend tussen de verschillende vaardigheidsscores van de leerlingen. De taken en toetsen meten de vaardigheid echter niet zonder meetfout, ze zijn immers niet 100% betrouwbaar. De vaardigheidsscores bevatten zodoende meetfout, en dus zijn correlaties tussen de vaardigheidsscores geen correlaties tussen zuiver gemeten constructen. Om de samenhang tussen de domeinen luistervaardigheid, spreekvaardigheid en gespreksvaardigheid te evalueren is daarom gebruik gemaakt van latente correlaties, en niet van ruwe correlaties, welke gecorrigeerd zijn voor meetfout. De latente correlatie tussen luistervaardigheid en spreekvaardigheid was 0,381. Tussen luistervaardigheid en gespreksvaardigheid was de latente correlatie 0,420 en de latente correlatie tussen spreekvaardigheid en gespreksvaardigheid was 0,460.

13. Welke verschillen zijn er tussen de leerlingprestaties einde bo, sbo en so anno 2024 en de leerlingprestaties in Peil.Mondelinge taalvaardigheid einde bao 2016-2017 en Peil.Mondelinge taalvaardigheid einde s(b)o 2017-2018, ook gecorrigeerd voor eventuele verschillen in achtergrondkenmerken van de leerlingpopulaties?

Aangezien bij ieder domein taken/items werden afgenomen die ook in de peiling mondelinge taalvaardigheid bao (2017) of mondelinge taalvaardigheid so/sbo (2018) werden afgenomen, kon de vaardigheid van de leerlingen uit deze peilingen deels worden vergeleken met de vaardigheid van de leerlingen uit de huidige peiling.
Bij luistervaardigheid waren de items afkomstig uit beide eerdere peilingen. Daarom werden de prestaties van de bao-leerlingen van de huidige peiling eerst vergeleken met de prestaties van de leerlingen uit de bao-peiling van 2017. Hieruit bleek dat de bao-leerlingen uit de huidige peiling hoger scoorden dan de leerlingen uit de vorige peiling (effectgrootte: middelgroot). Daarna zijn de prestaties van de sbo- en so-leerlingen uit de huidige peiling vergeleken met de prestaties van de leerlingen uit de sbo/so-peiling van 2018. Hieruit volgde dat de prestaties van de sbo-leerlingen zo goed als onveranderd waren gebleven, en dat de so-leerlingen ietwat minder goed scoorden in de huidige peiling dan in de vorige (effectgrootte: klein).
Ten aanzien van de spreek- en gespreksvaardigheid is enkel geankerd naar de so/sbo-peiling van 2018. Geen spreek- en gesprekstaken die zijn gebruikt in de bao-peiling van 2017 zijn eveneens gebruikt in de huidige peiling. Daardoor was het alleen mogelijk om een trendanalyses uit te voeren tussen de sbo- en so-populaties van beide jaargangen ten aanzien van de spreek- en gespreksvaardigheid. Bij spreekvaardigheid presteerden zowel de sbo- als de so-leerlingen in de huidige peiling lager dan de leerlingen uit de so/sbo-peiling van 2018 (effectgroottes: klein/middelgroot). Voor gespreksvaardigheid gold eveneens dat de so-leerlingen in de huidige peiling minder goed hadden gepresteerd dan de so-leerlingen uit de so-sbo-peiling

van 2018 (effectgrootte: klein/middelgroot). De gespreksvaardigheid van de sbo-leerlingen is daarentegen onveranderd gebleven door de tijd heen.
Samengevat, de luistervaardigheid van de bao-leerlingen is verbeterd, de luister- en gespreksvaardigheid van de sbo-leerlingen zijn zo goed als onveranderd gebleven. Tot slot zijn de spreekvaardigheid van de sbo-leerlingen en de luister-, spreek- en gespreksvaardigheid van de so-leerlingen achteruitgegaan.

10.2 Onderwijsleerproces op scholen

Hoe kan het onderwijs in mondelinge taalvaardigheid op de deelnemende scholen worden getypeerd (in elk geval het luister- en spreekklimaat in de klas, de aanwezigheid van geconcretiseerde leerdoelen en expliciete instructie voor gesprekken voeren, luisteren en spreken, het volgen van de vaardigheid van de leerlingen op deze subdomeinen, de aan deze subdomeinen intentioneel bestede onderwijstijd en de integratie van het onderwijs in deze subdomeinen met zowel andere taaldomeinen als andere leergebieden)?

Met de vragenlijsten voor de leerlingen, de leerkrachten en de schoolleiders is een groot aantal kenmerken van het onderwijs in mondelinge taalvaardigheid, zowel in de klas als op de hele school, in kaart gebracht. In hoofdstuk 7 zijn alle vragenlijstresultaten beschreven. Hierna vatten we puntsgewijs de bevindingen samen over de specifiek benoemde thema's in de onderzoeksvraag.
Het luister- en spreekklimaat in de klas. Het niet kunnen volgen (horen, begrijpen) van de juf of meester en het ervaren van onrust en lawaai in de klas, komt volgens de meeste leerlingen een paar keer per maand tot een paar keer per week voor. De spreekvrijheid in de klas (leerlingen luisteren naar elkaar en durven te zeggen wat ze vinden en voelen) beoordelen de leerlingen in bao, sbo en so als tamelijk positief, met een gemiddelde van ongeveer 3,0 op een schaal van 1 (nooit) tot 4 (altijd). Vrijwel alle ondervraagde leerkrachten in groep 8 c.q. de schoolverlatersgroepen onderschrijven de stelling dat ze een veilig luister- en gesprekssklimaat in de klas stimuleren, bijvoorbeeld door te zorgen voor weinig ruis en afleiding en door leerlingen de ruimte te geven hun mening te geven. Volgens ongeveer vier vijfde van de schoolleiders komt (de veiligheid van) het spreekklimaat voor leerlingen hooguit bij een kwart van de teamvergaderingen aan de orde.
De aanwezigheid van geconcretiseerde leerdoelen. Gevraagd naar de plaats van het leergebied mondelinge taalvaardigheid in het taalbeleids- of schoolplan geven schoolleiders in bao, sbo en so het vaakst als antwoord dat mondelinge taalvaardigheid globaal is beschreven aan de hand van algemene (taal)doelen voor de hele school. In sbo komt het ook vrij vaak voor dat mondelinge taalvaardigheid niet expliciet benoemd is in het schoolplan. Op een kleine minderheid van de scholen (6-14%) is mondelinge taalvaardigheid gedetailleerd beschreven in het taalbeleids- of schoolplan, aan de hand van concrete doelen per leerjaar of groepen van leerjaren. Als er sprake is van doelen voor mondelinge taalvaardigheid in het schoolplan, zijn deze meestal op meerdere bronnen gebaseerd. Het vaakst zijn de doelen gebaseerd op de taalmethode van de school en/of op de Kerndoelen Taal in het primair onderwijs. Ook de referentieniveaus voor Gesprekken voeren, Luisteren en Spreken worden vrij veel genoemd (30-44%). In sbo en so worden ook de leerlijnen en tussendoelen voor mondelinge communicatie geregeld genoemd als basis voor de schooldoelen voor mondelinge taalvaardigheid.
Expliciete instructie. Bijna alle leerkrachten zijn het eens met de stelling dat mondelinge taalvaardigheid in groep 1 tot en met groep 8 expliciet moet worden onderwezen en verreweg de meesten vinden ook dat ze

voldoende aandacht besteden aan onderwijs in mondelinge taalvaardigheid. Het merendeel geeft aan gerichte feedback te geven aan leerlingen om hun prestaties in gesprekken voeren, luisteren en spreken te verbeteren en leerlingen uit te dagen om actief mondelinge taal te gebruiken. Ongeacht het schooltype beoordelen de leerkrachten hun instructie voor elk van de drie subdomeinen als behoorlijk effectief (gemiddelde van rond de 3,0 op een schaal van 1 tot 4). Als gesprekstaak kiezen leerkrachten in bao, sbo en so relatief vaak voor het voeren van een vrij gesprek of overleg. Als luister- en spreektaak wordt het luisteren naar respectievelijk vertellen van een verhaal in/aan de klas door de meeste leerkrachten vaak ingezet.

Het volgen van de vaardigheid van de leerlingen. De voortgang van leerlingen in mondelinge taalvaardigheid wordt in groep 1 en 2 structureel gemonitord op een meerderheid van de bao-, sbo- en so-scholen[17]. In groep 3 tot en met 8 is het beeld minder eenduidig. Een kwart tot een derde van de bao-scholen monitort helemaal niet, van de overige bao-scholen monitort er ongeveer dertig tot veertig procent. Ruim een kwart van de sbo-scholen en 14 procent van de so-scholen monitort de mondelinge taalvaardigheid in groep 3 tot en met 8 niet. Als instrument voor de monitoring noemt een groot deel van de scholen de observaties door de groepsleerkracht. In groep 3 tot en met 8 worden ook toetsen uit de (taal)methode veel genoemd, in frequentie gevolgd door methode-onafhankelijke toetsen, bijvoorbeeld van Cito.

De intentioneel bestede onderwijstijd. Gemiddeld besteden de leerkrachten van groep 8 c.q. de schoolverlatersgroep zo'n 60 tot 90 minuten per week intentioneel (met vooraf opgestelde doelen) aan onderwijs in mondelinge taalvaardigheid. Een derde deel (in bao) tot een kwart (in sbo en so) besteedt er echter minder dan 30 minuten per week aan, terwijl er ook 10% (in bao en so) respectievelijk 20% (in sbo) meer dan 120 minuten per week aan besteedt. De bestede tijd wordt nagenoeg evenredig verdeeld over de drie subdomeinen gesprekken voeren, luisteren en spreken.

Twee derde deel van de bao-leerkrachten besteedt (veel) minder onderwijstijd aan mondelinge taalvaardigheid dan aan leesvaardigheid, terwijl ruim de helft van de leerkrachten in sbo en so evenveel of zelfs (veel) meer tijd aan onderwijs in mondelinge taalvaardigheid dan in leesvaardigheid besteedt. Ongeacht het schooltype wordt door de meerderheid van de leerkrachten evenveel of meer tijd besteed aan onderwijs in mondelinge taalvaardigheid dan in schrijfvaardigheid. Ruim driekwart van de leerkrachten besteedt ook intentioneel aandacht aan mondelinge taalvaardigheid tijdens het onderwijs in de zaakvakken, zo'n 60% doet dat ook tijdens het onderwijs in leesvaardigheid. Een (ruime) minderheid besteedt ook bij het onderwijs in rekenvaardigheid en schrijfvaardigheid intentioneel aandacht aan mondelinge taalvaardigheid.

Overigens blijkt uit de resultaten van de schoolvragenlijst dat op ruim twee derde deel van de scholen voor bao, sbo en so in het weekrooster voor groep 4 en 6 geen specifieke tijd is ingeroosterd voor het leergebied mondelinge taalvaardigheid of voor één of meer van de subdomeinen afzonderlijk. In groep 2 staat mondelinge taalvaardigheid bij 45% van de bao-scholen, 62% van de sbo-scholen en 35% van de so-scholen op het weekrooster.

De integratie met andere taaldomeinen en leergebieden. Hiervoor is al benoemd dat een aanzienlijk deel van de leerkrachten van groep 8 c.q. de schoolverlatersgroep intentioneel onderwijstijd besteedt aan

17 Mits s(b)o-school groepen 1 en 2 heeft.

mondelinge taalvaardigheid bij andere taaldomeinen (lezen, schrijven) of andere vakken (rekenen, zaakvakken). Van de schoolleiders in bao meldt een kwart dat op hun school de integratie van het onderwijs in mondelinge taalvaardigheid met leerinhouden van andere taaldomeinen is vastgelegd en geborgd; in sbo en so gaat het om respectievelijk 30% en 16%. Een derde deel van de bao-scholen, 43% van de sbo-scholen en 58% van de so-scholen werken aan een plan van aanpak voor deze integratie.
De integratie van het onderwijs in mondelinge taalvaardigheid met andere leergebieden dan taal is op 17% van de basisscholen vastgelegd en geborgd; 36% werkt aan een plan van aanpak daarvoor. Op ongeveer een kwart van de sbo- en so-scholen is de integratie van mondelinge taalvaardigheid met andere leergebieden dan taal vastgelegd en geborgd; bij circa 46% wordt gewerkt aan een plan van aanpak hiervoor.

10.3 Samenhang kenmerken en vaardigheid

In hoofdstuk 9 is aan de hand van multilevel analyses de samenhang tussen de leerlingprestaties en kenmerken van leerlingen, leerkrachten, scholen en het onderwijsleerproces onderzocht. Leerlingen in het speciaal onderwijs zijn hierbij buiten beschouwing gelaten, omdat het aantal so-leerlingen daarvoor te klein was.
Allereerst is nagegaan hoeveel variabiliteit toe te schrijven is aan klassen en hoeveel aan leerlingen (*onderzoeksvraag 5*). Voor leerlingen in het bao geldt dat het percentage variantie dat toe te schrijven is aan het klasniveau groter is voor Luisteren (25,6%) dan voor Spreken (13,1%) en Gesprekken (12,9%). Voor leerlingen in het sbo is het percentage variantie dat toe te schrijven is aan het klasniveau minder groot: Luisteren (8,2%), Spreken (13,6%) en Gesprekken (11,2%). Ook geldt voor de sbo-leerlingen dat de variabiliteit vooral toe te schrijven is aan het leerlingniveau.
Vervolgens is, door het opstellen van verschillende multilevel modellen, nagegaan welke verschillen in leerlingprestaties er zijn naar leerlingkenmerken (*onderzoeksvragen 6 en 7*), leerkracht- en schoolkenmerken (onderzoeksvragen 8, 9 en 10) en het onderwijsleerproces (*onderzoeksvraag 11*). Ten slotte zijn de kenmerken in samenhang meegenomen (*onderzoeksvraag 12*). Hieronder wordt een samenvatting gegeven van de gevonden resultaten. Hierbij worden eerst de resultaten van de bo-leerlingen voor de drie deelvaardigheden samengevat en vervolgens de resultaten voor de sbo-leerlingen.

10.3.1 Basisonderwijs

De **leeftijd van de leerling** is een significante voorspeller voor Spreken en Gesprekken, maar niet voor Luisteren. Voor Spreken en Gesprekken geldt dat oudere leerlingen, gemiddeld een lagere vaardigheidsscore behalen. De breedte van de **woordenschat** (het aantal woorden dat een leerling kent) is een significante voorspeller voor zowel Luisteren, Spreken als Gesprekken. Leerlingen die meer woorden kennen behalen hogere vaardigheidsscores voor alle drie de deelvaardigheden. De frequentie waarmee **ouders** thuis **taalrijke activiteiten** ondernemen (bijvoorbeeld hoe vaak leerlingen ouders zien lezen of hoe vaak ouders praten over het nieuws) is een significante voorspeller voor Luisteren en Gesprekken; leerlingen die ouders dit vaker zien of horen doen behalen een hogere score voor Luisteren en Gesprekken. Voor Spreken is deze variabele geen significante voorspeller. Het **aantal boeken thuis** is een significante

voorspeller voor Luisteren; als het aantal boeken dat een leerling thuis heeft groter is, is ook de vaardigheid in Luisteren hoger.
De mate waarin leerlingen **thuis praten over** bijvoorbeeld datgene wat ze lezen, het nieuws of wat ze gedaan hebben op school (schaalscore) is een significante voorspeller voor zowel Luisteren, Spreken als Gesprekken. Leerlingen die thuis meer praten over deze zaken, hebben hogere vaardigheidsscores voor de drie deelvaardigheden. **Communicatieproblemen**, zoals gerapporteerd door de leerkracht, is ook een significante voorspeller voor alle drie de deelvaardigheden; sterker ervaren problemen hangen samen met lagere vaardigheidsscores. **Spreekangst**, zoals ervaren voor de leerkracht, is een significante voorspeller voor Spreken en Gesprekken; meer spreekangst hangt samen met lagere vaardigheidsscores voor Spreken en Gesprekken. Voor Gesprekken geldt dat de mate van **externaliserend probleemgedrag** zoals gerapporteerd door de leerkracht een marginaal significante voorspeller is. Hoewel deze dus niet significant is, wordt de mate van externaliserend probleemgedrag wel meegenomen in het totaalmodel voor Gesprekken.
Op schoolniveau blijkt alleen de **schoolweging** van de school een significante voorspeller. Voor zowel Luisteren, Spreken als Gesprekken geldt dat een hogere schoolweging (en daarmee de kans op onderwijsachterstanden) samenhangt met lagere vaardigheidsscores van de leerlingen voor de drie deelvaardigheden.
Kijkend naar de kenmerken van het onderwijsleerproces, zijn er alleen significante samenhangen gevonden voor de deelvaardigheid Gesprekken. Voor Gesprekken geldt dat leerlingen op scholen waar de **ontwikkeling** in gesprekken incidenteel **gemonitord** wordt een hogere vaardigheidsscore behalen dan leerlingen op scholen waar de ontwikkeling in gesprekken niet gemonitord wordt. Er is geen verschil in vaardigheid tussen leerlingen van scholen waar de ontwikkeling in gesprekken structureel gemonitord wordt en leerlingen van scholen waar de ontwikkeling in gesprekken niet gemonitord wordt. De resultaten laten ook zien dat op scholen met meer **aandacht** is **voor mondelinge taalvaardigheid tijdens teamvergaderingen**, leerlingen een hogere vaardigheidsscore behalen voor Gesprekken. **Aandacht voor mondelinge taalvaardigheid in de visie van de school en bestuur**, ten slotte, blijkt ook een significante voorspeller voor de vaardigheidsscore in Gesprekken. Leerlingen op scholen waar de schoolleider het eens is met stellingen als ‘onze school heeft een duidelijke visie over de plaats van mondelinge taalvaardigheid in ons curriculum’ en ‘het schoolbestuur heeft zicht op de kwaliteit van het mondelinge taalvaardigheidsonderwijs van onze school’ behalen een lagere vaardigheidsscore voor Gesprekken.
Het totaalmodel, waarin alle (marginaal) significante voorspellers zijn meegenomen laat zien dat, in samenhang, de volgende kenmerken significante voorspellers zijn voor **Luisteren**: de grootte van de woordenschat, het aantal boeken thuis, de frequentie waarmee ouders volgens de leerlingen taalrijke activiteiten ondernemen, de mate van ervaren communicatieproblemen bij de leerling door de leerkracht en de schoolweging van de school. Voor **Spreken** geldt dat de volgende kenmerken significant zijn: de leeftijd van de leerling, de grootte van de woordenschat, de mate waarin leerlingen thuis praten over verschillende zaken, de mate van ervaren communicatieproblemen bij de leerling door de leerkracht en de mate waarin er sprake is van spreekangst bij de leerling volgens de leerkracht. Voor **Gesprekken**, ten slotte, geldt dat de volgende kenmerken significante voorspellers zijn: de leeftijd van de leerling, de grootte van de woordenschat, de mate waarin leerlingen thuis praten over verschillende zaken, de mate waarin er volgens de leerkracht sprake is van externaliserend probleemgedrag, de mate waarin er sprake is van

spreekangst bij de leerling volgens de leerkracht, de schoolweging, de mate waarin er aandacht is voor MTV in vergaderingen en de mate waarin er aandacht is voor MTV in de visie van de school en het bestuur.

10.3.2 Speciaal basisonderwijs

De **leeftijd van de leerling** is een significante voorspeller voor Luisteren, maar niet voor Spreken en Gesprekken. Voor Luisteren geldt dat oudere leerlingen gemiddeld lagere vaardigheidsscores behalen. De grootte van de **woordenschat** is een significante voorspeller van zowel Luisteren als Spreken. Leerlingen die meer woorden kennen behalen hogere vaardigheidsscores voor Luisteren en Spreken. Ook de frequentie waarmee **ouders** thuis **taalrijke activiteiten** ondernemen (bijvoorbeeld hoe vaak leerlingen ouders zien lezen of hoe vaak ouders praten over het nieuws) is een significante voorspeller van Luisteren; leerlingen die ouders dit vaker zien of horen doen behalen een hogere score voor Luisteren.
Communicatieproblemen, zoals gerapporteerd door de leerkracht, is ook een significante voorspeller voor alle drie de deelvaardigheden. Leerlingen voor wie sterker communicatieproblemen worden ervaren, behalen lagere vaardigheidsscores voor Luisteren, Spreken en Gesprekken. **Spreekangst**, zoals ervaren voor de leerkracht, is een significante voorspeller van Spreken; leerlingen bij wie minder sprake is van spreekangst bij de leerlingen, behalen een hogere vaardigheidsscore voor Spreken. Voor Gesprekken is spreekangst een marginaal significante voorspeller. Hoewel deze dus niet significant is, wordt spreekangst wel meegenomen in het totaalmodel voor Gesprekken. Tenslotte is het **ervaren klasklimaat** door de leerling een significante voorspeller voor Spreken. Leerlingen die vaker problemen ervaren in de klas, behalen een lagere vaardigheidsscore voor Spreken.
Geen van de leerling- en schoolkenmerken en kenmerken van het onderwijsleerproces waren een significante voorspeller voor Luisteren, Spreken en Gesprekken.
Het totaalmodel, waarin alle (marginaal) significante voorspellers zijn meegenomen laat zien dat, in samenhang, de volgende kenmerken significante voorspellers zijn voor **Luisteren**: de leeftijd van de leerling, de grootte van de woordenschat en de frequentie waarmee ouders volgens de leerlingen taalrijke activiteiten ondernemen. Voor **Spreken** geldt dat de volgende kenmerken significant zijn: de grootte van de woordenschat en de mate waarin er sprake is van spreekangst bij de leerling volgens de leerkracht. Voor **Gesprekken**, ten slotte, geldt dat de volgende kenmerken significante voorspellers zijn: de mate van ervaren communicatieproblemen bij de leerling door de leerkracht en de mate waarin er sprake is van spreekangst bij de leerling volgens de leerkracht.

11 Literatuur

Bechger, T. M., Maris, G., Verstralen, H. H. F. M., Verstralen, & Béguin, A. A. (2003). Using Classical Test Theory in Combination with Item Response Theory. *Applied Psychological Measurement, 27*(5), 319-334. https://doi.org/10.1177/0146621603257518.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Routledge. https://doi.org/10.4324/9780203771587.

Evers, A., Lucassen, W., Meijer, R. R., & Sijtsma, K. (2009). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Nederlands Instituut van Psychologen/Commissie Testaangelegenheden Nederland.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory (Vol. 2). Sage.

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd edition). Routledge.

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

Keuning, J., Schouwstra, S., Scheltinga, F. & Van der Lubbe, M. (2019). Game-based Spoken Interaction Assessment in Special Need Children. In: B.P. Veldkamp and C. Sluijter (Eds.), *Theoretical and Practical Advances in Computer-based Educational Measurement. Methodology of Educational Measurement and Assessment* (pp. 361-380). Springer. https://doi.org/10.1007/978-3-030-18480-3.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd edition). Springer Science + Business Media. https://doi.org/10.1007/978-1-4939-0317-7.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310.

Langen, A.. van, Druten-Frietman, L. van, Wolbers, M., Teunissen, C., Strating, H., Dood, C., Geelen, A., Binsbergen, M. (2017). *Peilingsonderzoek Mondelinge Taalvaardigheid in het basisonderwijs*. Nijmegen: Expertisecentrum Nijmegen/KBA Nijmegen/Bureau ICE/Radboud Universiteit Nijmegen/ResearchNed.

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*(4), 764-766. https://doi.org/10.1016/j.jesp.2013.03.013.

Maris, G., Bechger, T., Koops, J., Partchev, I. (2024). *dexter: Data Management and Analysis of Tests*. R package version 1.4.0. https://CRAN.R-project.org/package=dexter.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149-174. https://doi.org/10.1007/BF02296272.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied psychological measurement*, *16*(2), 159-176. https://doi.org/10.1177/014662169201600206.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

Ritzema, E. S., Pander Maat, H., Stukker, N, Hemker, B. T., Naayer, H. M., Lampe, T., Van Gelderen, A., De Boer, H. (2019). Peiling Schrijfvaardigheid einde (speciaal) Basisonderwijs 2019. Technische rapportage. GION Onderwijs/Onderzoek.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage Publishing

Van der Veen, C., van Kruistum, C., & Michaels, S. (2015). Productive classroom dialogue as an activity of shared thinking and communicating: A commentary on Marsal. *Mind, Culture, and Activity: An International Journal, 22*(4), 320-325. https://doi.org/10.1080/10749039.2015.1071398.

Verhelst, N. D., & Glas, C. A. (1995). The one parameter logistic model. In *Rasch models: Foundations, recent developments, and applications* (pp. 215-237). Springer New York. https://doi.org/10.1007/978-1-4612-4230-7_12.

Verhelst, N. D., Glas, C. A. W., & Verstralen, H. H. F. M. (1993). *OPLM: One parameter logistic model. Computer program and manual.* Cito.

Warm, T. A. (1989). Weighted likelihood estimation of ability in the item response theory. *Psychometrika, 54*, 427-450. https://doi.org/10.1007/BF02294627.

12 Bijlagen

12.1 Bijlage 1. Aantallen en p-waarden per item

Fragment	item_id	Aantal	p-waarde
Internet (2F)	12304	360	0,611
Internet (2F)	12305	360	0,342
Internet (2F)	12306	356	0,680
Internet (2F)	12309	348	0,385
Zebrastrepen (1F)	128VR02	701	0,746
Zebrastrepen (1F)	128VR03	698	0,705
Zebrastrepen (1F)	128VR04	683	0,912
Bacterien (2F)	13156	932	0,555
Bacterien (2F)	13158	910	0,518
Bacterien (2F)	13160	897	0,678
Bacterien (2F)	13164	888	0,757
Bacterien (2F)	13165	879	0,833
Internet (2F)	13473	353	0,918
Ridders (1F)	138VR01	415	0,523
Ridders (1F)	138VR04	415	0,720
Ridders (1F)	138VR05	416	0,608
Ridders (1F)	138VR06	415	0,713
Ridders (1F)	138VR07	413	0,438
Melkmeisje (2F)	143VR01	780	0,727
Melkmeisje (2F)	143VR04	778	0,717
Melkmeisje (2F)	143VR05	781	0,817
Melkmeisje (2F)	143VR06	767	0,593
Tekenares (1F)	688	861	0,849
Tekenares (1F)	689	857	0,869
Tekenares (1F)	691	852	0,703
Pindakaas (1F/2F)	755	464	0,894
Pindakaas (1F/2F)	756	460	0,713
Pindakaas (1F/2F)	758	453	0,523
Pindakaas (1F/2F)	760	449	0,695
Pindakaas (1F/2F)	762	442	0,710
Tekenares (1F)	898	860	0,741
julia (1F)	PO.NeLu.1F.v1.1i	961	0,594
julia (1F)	PO.NeLu.1F.v1.2b	955	0,396
julia (1F)	PO.NeLu.1F.v1.5i	964	0,776
goud (1F)	PO.NeLu.1F.v5.1i	455	0,796

goud (1F)	PO.NeLu.1F.v5.2i	441	0,943
goud (1F)	PO.NeLu.1F.v5.3b	429	0,517
contant geld (2F)	PO.NeLu.2F.a2.1b	466	0,841
contant geld (2F)	PO.NeLu.2F.a2.2b	460	0,343
bank (2F)	PO.NeLu.2F.v4.1i	750	0,859
bank (2F)	PO.NeLu.2F.v4.2b	740	0,764
Gooi (2F)	PO.NeLu.2F.v5.3b	787	0,896
Gooi (2F)	PO.NeLu.2F.v5.4i	793	0,741

12.2 Bijlage 2: Resultaten multilevel modellen speciaal onderwijs

Hieronder worden achtereenvolgens de resultaten van het nulmodel en de modellen met algemene leerlingkenmerken (model 1), domeinspecifieke leerlingkenmerken (model 2), leerkracht- en schoolkenmerken (model 3) en kenmerken van het onderwijsleerproces (model 4) voor het speciaal onderwijs gepresenteerd. Iedere tabel bevat de resultaten voor de drie verschillende deelvaardigheden (Luisteren, Spreken en Gesprekken). Hieronder volgend de resultaten van de verschillende modellen voor het speciaal onderwijs. Vanwege het relatief kleine aantal leerlingen per subvaardigheid (91 voor Luisteren; 112 voor Spreken; 116 voor Gesprekken) zijn deze tabellen niet opgenomen in het rapport (zie ook paragraaf 9.5.3). De kleine steekproefgrootte vermindert de statistische kracht en betrouwbaarheid van regressiemodellen, wat kan leiden tot onnauwkeurige, niet-repliceerbare en moeilijk te interpreteren resultaten. Voor de volledigheid worden hieronder de tabellen gepresenteerd, maar wordt er geen interpretatie van de resultaten gegeven. Ook zijn er geen totaalmodellen samengesteld voor deze groep leerlingen.

Nulmodel

In de tabel hieronder worden de resultaten van de nulmoddelen gepresenteerd.

Tabel 2.1 Nulmodel so-leerlingen, uitgesplitst naar subvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	β	SE	*p*	β	SE	*p*
Fixed effecten									
Intercept	0,134	0,116	0,248	0,927	0,074	<0,001	0,893	0,072	<0,001
Random effecten									
Klas	0,043			0,021			0,084		
Leerling	0,884			0,494			0,328		
ICC	0,047			0,041			0,204		
R^2	-			-			-		

Model 1: algemene leerlingkenmerken

In de tabel hieronder worden de resultaten van model 1 gepresenteerd.

Tabel 2.2 Model 1 (algemene leerlingkenmerken) so-leerlingen, uitgesplitst naar subvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	β	SE	*p*	β	SE	*p*
Fixed effecten									
Intercept	-0,813	2,224	0,715	2,012	1,549	0,194	1,304	1,067	0,222
Geslacht-meisje*	-0,013	0,251	0,960	0,092	0,191	0,629	-0,031	0,123	0,802
Leeftijd in maanden	0,008	0,015	0,610	-0,008	0,010	0,474	-0,003	0,007	0,710
Woordenschat	0,422	0,103	<0,001	0,119	0,078	0,126	0,002	0,063	0,973

Frequentie waarin thuis Nederlands wordt gesproken	-0,119	0,105	0,261	0,046	0,084	0,581	-0,045	0,052	0,390
Aantal boeken thuis	0,072	0,095	0,453	0,156	0,069	0,023	0,041	0,053	0,431
Frequentie taalrijke activiteiten thuis	-0,035	0,093	0,706	-0,062	0,071	0,382	0,005	0,053	0,929
Random effecten									
Klas	0,000				0,008		0,076		
Leerling	0,778				0,484		0,341		
ICC	0,00				0,016		0,183		
R^2	0,161				0,045		-0,012		
* Referentiecategorie geslacht: jongen									

Model 2: domeinspecifieke leerlingkenmerken

In de tabel hieronder worden de resultaten van model 2 gepresenteerd.

Tabel 2.3 Model 2 (domeinspecifieke leerlingkenmerken) so-leerlingen, uitgesplitst naar subvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	β	SE	*p*	β	SE	*p*
Fixed effecten									
Intercept	0,375	0,176	0,039	0,949	0,112	<0,001	1,022	0,094	<0,001
Thuis praten over	-0,185	0,090	0,040	0,055	0,067	0,411	0,063	0,049	0,193
Ervaren klasklimaat	-0,124	0,097	0,211	0,067	0,075	0,369	0,061	0,051	0,235
Spreekvrijheid in de klas	0,063	0,095	0,508	-0,063	0,068	0,358	0,018	0,051	0,728
Communicatie-problemen	-0,023	0,108	0,041	-0,030	0,078	0,701	-0,071	0,057	0,219
Internaliserend probleemgedrag	0,088	0,137	0,526	0,061	0,080	0,449	0,008	0,062	0,891
Externaliserend probleemgedrag	0,186	0,104	0,079	-0,032	0,073	0,658	0,052	0,054	0,333
Spreekangst	0,060	0,116	0,607	0,256	0,076	0,001	0,035	0,059	0,548
Stimulerende/ belemmerende factoren	0,316	0,175	0,090	0,051	0,113	0,651	0,186	0,083	0,027
Random effecten									
Klas	0,021			0,003			0,060		
Leerling	0,655			0,474			0,318		
ICC	0,030			0,005			0,158		
R^2	0,271			0,074			0,083		

Model 3: leerkracht- en schoolkenmerken

In de tabel hieronder worden de resultaten van model 3 gepresenteerd.

Tabel 2.4 Model 3 (leerkracht- en schoolkenmerken) so-leerlingen, uitgesplitst naar subvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	β	SE	*p*	β	SE	*p*
Fixed effecten									
Intercept	-0,240	0,227	0,290	0,737	0,232	0,001	0,756	0,184	<0,001
Jaren werkzaam als leerkracht	0,111	0,134	0,407	-0,013	0,087	0,879	0,103	0,089	0,249
Ervaren bekwaamheid	0,156	0,162	0,337	-0,078	0,092	0,397	0,046	0,088	0,603
Percentage doorstroom vmbo g-t/havo/vwo	0,010	0,007	0,139	0,004	0,006	0,442	0,004	0,005	0,400
Random effecten									
Klas	0,000			0,031			0,089		
Leerling	0,874			0,494			0,326		
ICC	0,000			0,060			0,214		
R^2	0,057			-0,019			-0,007		

Model 4: kenmerken van de onderwijsleercontext

In de tabel hieronder worden de resultaten van model 4 gepresenteerd.

Tabel 2.5 Model 4 (Kenmerken van de onderwijsleercontext) so-leerlingen, uitgesplitst naar subvaardigheid

	Luisteren			Spreken			Gesprekken		
	β	SE	*p*	β	SE	*p*	β	SE	*p*
Fixed effecten									
Intercept	0,260	0,232	0,197	0,264	0,546	0,154	1.087	0,373	0,004
Mate van effectieve instructie in luisteren/ spreken/ gesprekken	0,163	0,226	0,470	-0,033	0,116	0,778	0,056	0,092	0,541
Frequentie taken luisteren/ spreken/ gesprekken	-0,015	0,155	0,920	-0,03	0,104	0,775	0,147	0,13	0,259
Onderwijstijd mondelinge taalvaardigheid	-0,052	0,130	0,687	0,001	0,094	0,993	-0,098	0,1	0,329
Monitoren van luisteren/ spreken/ gesprekken – ja, incidenteel*	†	†	†	-0,009	0,289	0,975	-0,136	0,267	0,609
Monitoren van luisteren/ spreken/ gesprekken – ja, structureel*	†	†	†	-0,272	0,35	0,436	-0,092	0,295	0,756
Bieden van ondersteuning bij luisteren/ spreken/ gesprekken – ja, incidenteel*	-0,190	0,347	0,584	0,315	0,482	0,513	-0,081	0,316	0,797
Bieden van ondersteuning bij luisteren/ spreken/ gesprekken – ja, structureel*	-0,128	0,272	0,639	0,524	0,499	0,294	-0,06	0,389	0,878
Aandacht voor MTV in vergaderingen	0,189	0,088	0,031	0,021	0,072	0,770	0,037	0,086	0,668
Aandacht voor MTV in visie school en bestuur	-0,194	0,138	0,159	-0,214	0,124	0,084	-0,03	0,136	0,826
Random effecten									
Klas	0,003			0,002			0,115		
Leerling	0,908			0,496			0,324		
ICC	0,004			0,005			0,262		
R^2	0,017			0,033			-0,065		

*Referentiecategorie: Nee

† Voor Luisteren geldt dat geen enkele schoolleider de categorie 'Nee' heeft gekozen. Dit betekent dat er geen leerlingen tot de referentiecategorie behoren. Daarop is besloten om voor Luisteren deze variabele niet op te nemen in de analyses.