



Inspectie van het Onderwijs
*Ministerie van Onderwijs, Cultuur en
Wetenschap*

**TECHNISCH RAPPORT PEIL.MONDELINGE
TAALVAARDIGHEID EINDE SPECIAAL
(BASIS)ONDERWIJS**

2017-2018

november 2019

INHOUD

Inleiding 3

1 Het onderwijsleerproces mondelinge taalvaardigheid 4

- 1.1 Databronnen 4
- 1.2 Werkwijze 4
 - 1.2.1 Samengestelde variabelen 5
 - 1.2.2 Nieuwe variabelen 5
 - 1.2.3 Significantie toetsing 7

2 Prestaties mondelinge taalvaardigheid 9

- 2.1 Databronnen 9
 - 2.1.1 Luisteren 9
 - 2.1.2 Spreken 10
- 2.2 Werkwijze 12
 - 2.2.1 Weging SO 12
 - 2.2.2 Indeling in zorgprofielen 12
 - 2.2.3 Significantie toetsing 13
 - 2.2.4 Bepalen van grensitems 17

3 Verschillen tussen scholen en leerlingen 20

- 3.1 Databronnen 20
- 3.2 Werkwijze 22

Inleiding

Dit is het technisch rapport bij het peilingsonderzoek Mondelinge taalvaardigheid in het speciaal basisonderwijs (sbo) en speciaal onderwijs (so; voormalig cluster 4). Dit peilingsonderzoek is in opdracht van het Nationaal Regieorgaan Onderwijsonderzoek en de Inspectie van het Onderwijs uitgevoerd door een consortium bestaande uit Kennisinstituut voor de Taalontwikkeling Amsterdam (ITTA) en Cito BV. Voor de werving van de scholen is samengewerkt met de CED groep. In de technische rapportage van het uitvoerend consortium (Keuning, Rietdijk, Schouwstra, & Scheltinga, 2019) worden de wijze van dataverzameling, het instrumentarium en de eerste gegevensverwerking ter beantwoording van de onderzoeksvragen uitgebreid verantwoord.

In dit technisch rapport staan de keuzes die gemaakt zijn door de schrijvers van de publieksrapportage centraal. De verdere gegevensverwerking en – analyse die ten grondslag liggen aan het publieksrapport worden hier verantwoord.

Het doel van het peilingsonderzoek Mondelinge taalvaardigheid einde speciaal (basis)onderwijs is om een beeld te schetsen van de luister-, spreek- en gespreksvaardigheid van schoolverlaters in het s(b)o. Waar mogelijk worden leerlingprestaties vergeleken met prestaties gemeten in het regulier basisonderwijs (Inspectie van het Onderwijs, 2019) en in een vorige peiling naar mondelinge taalvaardigheid in het sbo (Krom, van Berkel, van der Schoot, Sijtsma, Hemker, & Marsman, 2011). Daarnaast beschrijft het rapport hoe het onderwijsleerproces voor mondelinge taalvaardigheid er op de deelnemende scholen uitziet. Tot slot wordt gekeken naar de prestatieverschillen tussen leerlingen en scholen en naar hoe de prestaties samenhangen met de manier waarop scholen het onderwijsleerproces op het gebied van mondelinge taalvaardigheid vormgeven.

De onderzoeksvragen die het uitgangspunt vormen voor het publieksrapport zijn:

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

De opbouw van dit technisch rapport volgt de hoofstukindeling van het publieksrapport. We onderbouwen per hoofdstuk uit deel B (de Resultaten), de gegevens die we hebben gebruikt en onze werkwijze.

1 Het onderwijsleerproces mondelinge taalvaardigheid

In dit hoofdstuk van de publieksrapportage wordt het onderwijsleerproces van de sbo- en so-scholen in dit onderzoek in kaart gebracht. Deze beschrijving is gebaseerd op interviews die gehouden zijn met 62 van de 68 scholen uit de steekproef; 50 in het sbo (van in totaal 289 sbo-vestigingen) en 12 in het so (van de 130 so-scholen die voorheen als cluster 4 of cluster 3/4 scholen bekend stonden).

1.1 Databronnen

De interviews op de scholen namen in de regel een half uur tot een uur in beslag en zijn waar mogelijk opgenomen met een recorder.

De volgende aspecten van het onderwijsleerproces kwamen in de interviews aan de orde:

- onderwijsaanbod en methodegebruik (6 items)
- doelgerichtheid en planmatigheid (7 items)
- onderwijstijd (1 item)
- monitoring en planning (8 items)
- differentiatie (9 items)
- teamontwikkeling (4 items)

De items zijn afzonderlijk voor luisteren, spreken en gesprekken nagelopen (met uitzondering van het onderdeel teamontwikkeling), omdat het onderwijs bij de ene vaardigheid mogelijk anders is ingericht dan bij de andere. De gegevens zijn na afloop gekwantificeerd om scholen op itemniveau te kunnen vergelijken. Ook is er een beknopte kwalitatieve analyse gedaan van de notities die de interviewer bij ieder gesprek heeft gemaakt.

Het thema onderwijstijd bleek na verwerking van de interviews niet geschikt voor verder gebruik in de analyses. De overige items bleken allemaal wel bruikbaar voor de beschrijving van het onderwijsleerproces.

1.2 Werkwijze

Als de resultaten van de subdomeinen luisteren, spreken en gesprekken voeren niet significant van elkaar verschilden, zijn ze gezamenlijk beschreven. Dit is in bijna alle gevallen gebeurd. De betreffende variabelen zijn in die gevallen samengenomen door het gemiddelde over de subdomeinen te berekenen, waarbij ten minste één van de waarden een geldige score moest zijn.

De beschrijvingen in het hoofdstuk betreffen overwegend frequentieverdelingen, zowel in percentages als in aantallen. In een aantal gevallen zijn er nieuwe variabelen samengesteld. Ook zijn er enkele uitsplitsingen getoetst op significante verschillen.

1.2.1

Samengestelde variabelen

De 12 scholen voor speciaal onderwijs hebben in hun beantwoording geen onderscheid laten zien tussen de verschillende subdomeinen. Een antwoord voor het subdomein luisteren, gold in alle gevallen ook voor de subdomeinen spreken en gesprekken. Vragen over teamontwikkeling zijn niet voor de afzonderlijke subdomeinen gesteld. De vraag over onderwijstijd is in dit onderzoek niet meegenomen (zie hierboven). Scholen voor speciaal basisonderwijs differentieerden hun antwoorden in enkele gevallen naar subdomein. Een antwoord voor subdomein luisteren was niet in alle gevallen ook het antwoord voor het subdomein spreken of gesprekken. Deze verschillen waren in geen enkel geval significant, waardoor we de antwoorden op de subdomeinen samen konden nemen. In onderstaande tabel staan *general least squares* analyses om verschillen in antwoorden tussen subdomeinen te toetsen. In de tabel zijn enkel die vragen opgenomen die verschil in antwoorden lieten zien tussen de subdomeinen.

Tabel 1.2.1 GLS analyse sbo; antwoorden op de verschillende subdomeinen, p-waarde bij Sphericity Assumed

Variabele code	Vraag	p-waarde
Onderwijsaanbod en methodegebruik (in totaal V1 – V6)		
INT_AANB_V5	Wordt het onderwijs in mtv ingebed in de taalles?	0,37
INT_AANB_V6	Wordt het onderwijs in mtv geïntegreerd in andere vakken?	0,61
Doelgerichtheid en planmatigheid (in totaal V1 – V7)		
INT_DOEL_V1	Heeft de school (expliciet) doelen voor mtv?	0,61
INT_DOEL_V2	Wordt gebruikgemaakt van een leerlijn?	0,23
INT_DOEL_V3	Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)	0,37
INT_DOEL_V4	Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)	0,37
INT_DOEL_V5	Wordt gebruikgemaakt van herhaald aanbod met feedback?	0,37
Monitoring en planning (in totaal V1 – V8)		
INT_VOLG_V1	Worden leervorderingen in de lessen gemonitord en geregistreerd?	0,37
INT_VOLG_V2	Worden leervorderingen in het schooljaar gemonitord en geregistreerd?	0,14
INT_VOLG_V3	Wordt gebruikgemaakt van een leerling-administratiepakket?	0,72
INT_VOLG_V4	Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?	0,37
Differentiatie (in totaal V1 – V9)		
INT_DIFF_V4	Is er differentiatie naar instructie?	0,37

1.2.2

Nieuwe variabelen

In het hoofdstuk is gerapporteerd over een aantal nieuw samengestelde variabelen. De samenstelling daarvan lichten we hieronder toe.

Nieuwe variabele: sterk doelgericht

Dit is een combinatie van antwoorden op vier afzonderlijke vragen:

- 1 Heeft de school (expliciet) doelen voor MTV?

- 2 Zijn er einddoelen voor schoolverlaters geformuleerd?
 - 3 Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)
 - 4 Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)
- Als vraag 1 en 2 is ja en vraag 3 of vraag 4 is ja, dan is de nieuwe variabele ja. In alle andere gevallen is de nieuwe variabele nee.

Nieuwe variabele: schoolbrede doelen benoemd in taalplan

Dit is een combinatie van antwoorden op twee afzonderlijke vragen:

- 1 Heeft de school (expliciet) schoolbrede doelen voor MTV?
 - 2 Is mondelinge taalvaardigheid benoemd in het taalbeleids- of schoolplan?
- Als vraag 1 en 2 is ja, dan is de nieuwe variabele ja; als vraag 1 en 2 is nee, dan is de nieuwe variabele nee. Als vraag 1 is ja en vraag 2 is nee, of andersom, dan geldt geen ja of nee.

Nieuwe variabele: registratie leervorderingen

Dit is een combinatie van antwoorden op twee afzonderlijke vragen:

- 1 Worden leervorderingen in de lessen gemonitord en geregistreerd?
 - 2 Worden leervorderingen in het schooljaar gemonitord en geregistreerd?
- Als vraag 1 is ja of vraag 2 is ja, dan is de nieuwe variabele ja. In alle andere gevallen is de nieuwe variabele nee.

Nieuwe variabele: instrumentgebruik voor toetsing

Dit is een combinatie van antwoorden op drie afzonderlijke vragen:

- 1 Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?
 - 2 Wordt gebruikgemaakt van methodetoetsen?
 - 3 Wordt gebruikgemaakt van zelfgemaakte instrumenten?
- Als vraag 1 is ja of vraag 2 is ja of vraag 3 is ja, dan is de nieuwe variabele ja. In alle andere gevallen is de nieuwe variabele nee.

Nieuwe variabele: systematisch volgen

Dit is een combinatie van scores op twee nieuwe variabelen:

- 1 registratie leervorderingen
 - 2 instrumentgebruik voor toetsing
- Als 1 is ja en 2 is ja, dan is de nieuwe variabele ja. In alle andere gevallen is de nieuwe variabele nee.

Nieuwe variabele: registratie leervorderingen en resultaten gebruiken voor planning

Dit is een combinatie van scores en antwoorden op een nieuwe variabele en een bestaande vraag:

- 1 registratie leervorderingen
 - 2 Wordt in de planning gebruikgemaakt van leerresultaten van leerlingen?
- Als 1 is ja en vraag 2 is ja, dan is de nieuwe variabele ja. In alle andere gevallen is de nieuwe variabele nee.

Nieuwe variabele: differentiëren

Dit is een combinatie van antwoorden op drie afzonderlijke vragen:

- 1 Is er differentiatie naar instructie?

2 Is er differentiatie naar leerstof?

3 Is er differentiatie naar (streef)doelen?

Als vraag 1 is ja of vraag 2 is ja of vraag 3 is ja, dan is de nieuwe variabele ja. In alle andere gevallen is de nieuwe variabele nee.

Nieuwe variabele: extra ondersteuning

Dit is een combinatie van antwoorden op twee afzonderlijke vragen:

1 Is er extra ondersteuning door een specialist binnen de school?

2 Is er extra ondersteuning door een specialist buiten de school?

Als vraag 1 is ja of vraag 2 is ja, dan is de nieuwe variabele ja. In alle andere gevallen is de nieuwe variabele nee.

1.2.3 *Significantie toetsing*

Het werken met niveaugroepen (ja/nee) en het systematisch volgen (ja/nee) is voor de sbo-scholen getoetst op drie vormen van differentiëren. In onderstaande tabel staan de resultaten van de chi-kwadraat toetsing.

Tabel 1.2.3a Uitkomsten van de chi-kwadraattoetsen systematisch volgen naar differentiëren

<i>Differentiëren naar instructie</i>			
	<i>n</i>	<i>Chi kwadraat</i>	<i>p-waarde</i>
<i>Systematisch volgen</i>			
Nee	33	3,37	0,07*
Ja	17		
<i>Differentiëren naar leerstof</i>			
	<i>n</i>	<i>Chi kwadraat</i>	<i>p-waarde</i>
<i>Systematisch volgen</i>			
Nee	32	4,13	0,04**
Ja	17		
<i>Differentiëren naar (streef)doelen</i>			
	<i>n</i>	<i>Chi kwadraat</i>	<i>p-waarde</i>
<i>Systematisch volgen</i>			
Nee	32	1,52	0,22
Ja	16		

** *p significant <0,05; * p significant <0,10*

SBO-scholen die hun leerlingen systematisch volgen differentiëren significant vaker naar instructie en leerstof.

Tabel 1.2.3a Uitkomsten van de chi-kwadraattoetsen werken met niveaugroepen naar differentiëren

<i>Differentiëren naar instructie</i>			
	<i>n</i>	<i>Chi kwadraat</i>	<i>p-waarde</i>
<i>Niveaugroepen</i>			
Nee	31	7,86	0,01**
Ja	19		
<i>Differentiëren naar leerstof</i>			
	<i>n</i>	<i>Chi kwadraat</i>	<i>p-waarde</i>
<i>Niveaugroepen</i>			
Nee	30	10,40	0,00**
Ja	19		
<i>Differentiëren naar (streef)doelen</i>			
	<i>n</i>	<i>Chi kwadraat</i>	<i>p-waarde</i>
<i>Niveaugroepen</i>			
Nee	31	2,43	0,12
Ja	17		

** *p* significant <0,05; * *p* significant <0,10

SBO-scholen die werken met niveaugroepen differentiëren significant vaker naar instructie en leerstof.

2 Prestaties mondelinge taalvaardigheid

In dit hoofdstuk in het publieksrapport beschrijven we de prestaties van s(b)o-leerlingen binnen de verschillende subdomeinen van mondelinge taalvaardigheid. Ook besteden we aandacht aan verschillen met de prestaties voor luisteren uit eerder peilingsonderzoek (2007) en de prestaties van leerlingen aan het einde van het regulier basisonderwijs op het gebied van mondelinge taalvaardigheid (Inspectie van het Onderwijs, 2019).

2.1 Databronnen

2.1.1 *Luisteren*

De luistervaardigheid van leerlingen is in dit onderzoek op twee verschillende manieren gemeten. Ten eerste via een klassikale pen-en-papier taak. Hierin beantwoordden leerlingen vragen over kijk- en luisterfragmenten die via het digibord werden gepresenteerd. Om rekening te houden met niveauverschillen tussen leerlingen zijn er drie verschillende toetsboekjes gemaakt (met gedeeltelijk overlappende opgaven), passend bij de drie verschillende leerroutes.

In de tweede taak speelden leerlingen een bestaand samenwerkingsspel uit Duitsland (het zogeheten Fischerspiel, voor dit peilingsonderzoek enigszins aangepast). Doel van dit spel is om met een groepje van vier leerlingen alle vissen naar de haven te brengen voordat de storm losbarst. Dit spel kent geen individuele winnaars of verliezers: de leerlingen winnen of verliezen samen van het spel. Deze taak is primair ontwikkeld om de gespreksvaardigheid van leerlingen in kaart te brengen, maar in dit onderzoek werd tegelijkertijd ook de luistervaardigheid van leerlingen gemeten. Tijdens het spel kregen de leerlingen namelijk een mondelinge instructie met diverse opdrachten. Door de opdrachten goed uit te voeren, konden ze laten zien of ze de instructie hadden begrepen. Een toetsleider scoorde op een observatieformulier of dit het geval was. De twee opdrachten samen testen twee van de drie taken die in het referentiekader bij het subdomein luisteren onderscheiden worden, namelijk luisteren naar instructies (samenwerkingsspel) en luisteren naar radio en televisie en gesproken tekst op internet (klassikale taak). De taak luisteren als lid van een live publiek is in dit onderzoek niet als een afzonderlijke taak meegenomen (zie voor een toelichting Prenger en Damhuis, 2016).

Prestaties pen-en-papiertaak

De 3 toetsboekjes (1 per leerroute) uit de pen-en-papiertaak bevatten een verschillend aantal opgaven, namelijk 28, 24 en 20. Om de scores van leerlingen op de toetsboekjes met elkaar te kunnen vergelijken, zijn deze vertaald naar vaardigheidsscores. De vaardigheidsscores zijn vervolgens omgezet naar een bankscore (Keuning et al., 2019). Deze bankscores kunnen geïnterpreteerd worden als een beheersingsniveau en liggen tussen de 0 en 100 procent.

In het publieksrapport is uitsluitend gerekend met de bankscores, omdat deze makkelijk interpreteerbaar zijn en geen overschatting geven van de populatievariantie (zoals bij vaardigheidsscores het geval is; zie Keuning et al., 2019). Gemiddelde bankscores zijn uitgerekend per kenmerk van de taakuitvoering

(bijv. begrijpen), tekstkenmerk (bijv. 1F) en fragmentkenmerk (bijv. audio) om onderlinge vergelijkingen mogelijk te maken.

Prestaties in de gesprekstaak

Tijdens het spelen van het Fischerspiel werd de luistertaak 'luisteren naar instructie' gemeten. In totaal zijn vier beoordelingsaspecten op een driepuntsschaal gescoord (0 =lage vaardigheid tot 2 = hoge vaardigheid):

1. Instructie toepassen (taakkenmerk: begrijpen)
2. Herhaling van de instructie (taakkenmerk: begrijpen)
3. Mondeling samenvatten (taakkenmerk: samenvatten)
4. Mening herkennen (taakkenmerk: interpreteren)

In totaal konden leerlingen dus maximaal 8 punten halen op deze luistertaak. In het publieksrapport werd de ruwe totaalscore gebruikt om prestaties te analyseren.

2.1.2

Spreken

In dit peilingsonderzoek is de spreekvaardigheid gemeten aan de hand van twee taken. In de eerste taak kregen leerlingen een nieuwsfragment te zien uit het Jeugdjournaal waarin wordt verteld over een babyotter die gevonden is in Alaska. Vervolgens werd de leerlingen gevraagd om aan de toetsleider, die niet meekeek naar het filmpje, zo precies mogelijk na te vertellen wat er in het filmpje te zien is. Om leerlingen met een zwakker werkgeheugen niet te benadelen, kregen alle leerlingen bij het navertellen een geheugensteuntje, bestaande uit een aantal kernwoorden uit het filmfragment.

Bij de tweede taak is gebruikgemaakt van het bestaande educatieve spel Story Cubes. Leerlingen kregen negen dobbelstenen met daarop verschillende afbeeldingen. Deze dobbelstenen gooiden ze, waarna ze vijf afbeeldingen uitkozen om een verhaal over te vertellen. Het verhaal moest beginnen met "Er was eens...". De leerling speelde het spel meerdere keren. De eerste keer was om te oefenen, de tweede en derde keer werden beoordeeld.

Alle leerlingen voerden beide spreektaken uit. De afname vond plaats in een aparte, zoveel mogelijk prikkelarme, ruimte. De toetsleider begeleidde de afname en beoordeelde de leerlingen op de spreektaak door middel van een gestandaardiseerd beoordelingsmodel. De toetsleider vulde het beoordelingsmodel voor elke leerling zoveel mogelijk tijdens het spel in¹. De spreekvaardigheid werd beoordeeld aan de hand van vijf kenmerken van de taakuitvoering die in het referentiekader worden onderscheiden:

- samenhang (SH)
- afstemming op doel (AD)
- afstemming op publiek (AP)
- woordgebruik en woordenschat (WW)
- vloeiendheid, verstaanbaarheid en grammaticale beheersing (VVG)

Tot slot is uitvoering (UV) van de opdracht in het algemeen als randvoorwaarde aan het beoordelingsmodel toegevoegd. In dit beoordelingsmodel zijn - onderliggend aan deze kenmerken van de taakuitvoering - vijftien beoordelingsaspecten

¹ De beoordelaarsovereenstemming werd geëvalueerd in een aanvullend onderzoek, waarbij expertbeoordelaars een deel van de leerlingen achteraf op basis van de beschikbare filmopnames, opnieuw beoordeelden. Dit onderzoek gaf geen aanleiding om oordelen van bepaalde toetsleiders uit te sluiten van analyse (zie Keuning et al., 2019, p. 49)

gespecificeerd. Deze werden door de toetsleider gescoord op een driepuntsschaal: onvoldoende (0 punten), voldoende (1 punt) of goed (2 punten).

In het construeren van een totaalscore is aan de zeven technische beoordelingsaspecten minder gewicht (0,43) toegekend dan aan de acht inhoudelijke beoordelingsaspecten (1,50), waardoor deze items voor 20% meetelden in de totaalscore (zie Keuning et al., 2019 voor een toelichting). In totaal konden leerlingen zestig punten halen (30 punten per taak). Deze totaalscore werd in het rapport gebruikt om leerlingprestaties te analyseren. Aanvullend werden de totaalscores per taak gebruikt om een analyse te doen naar eventuele interactie-effecten tussen referentieniveaugroepen en taken of zorgprofielen en taken. Voor de analyse van leerlingprestaties per kenmerk van de taakuitvoering zijn gemiddelde (ongewogen) proportiescores per taakkenmerk berekend, door de gemiddelde scores per taakkenmerk te delen door de maximaal haalbare score per aspect (2).

Gesprekken

Om de gespreksvaardigheid te meten, is gebruik gemaakt van een bestand, coöperatief spel uit Duitsland: het Fischerspiel. Leerlingen moeten in dit spel samen een aantal vissen vangen en naar de haven brengen, voordat de storm losbarst. Het spel kent geen individuele winnaars of verliezers, je wint of verliest als groep. Het spel is voor dit onderzoek iets vereenvoudigd, zodat leerlingen van uiteenlopende niveaus de taak goed begrijpen en hun gespreksvaardigheid tot uiting komt.

Het spel werd zoveel mogelijk gespeeld in groepjes van vier. De toetsleider stelde zich tijdens het spel op als coach. Hij of zij zorgde ervoor dat alle leerlingen evenveel aan bod kwamen en zich veilig en betrokken voelden. De toetsleider kon hiervoor vooraf bepaalde interventies doen (Keuning et al., 2019; bijlage 6). In totaal duurde het spel zo'n 30 tot 45 minuten.

De toetsleider trad ook op als beoordelaar: hij/zij scoorde de leerlingen op een driepuntsschaal (0 - onvoldoende, 1 - voldoende, 2 - goed) op zeventien verschillende beoordelingsaspecten die deel uitmaakten van een gestandaardiseerd beoordelingsmodel. Dit beoordelingsmodel was – net als bij het onderdeel spreken – gebaseerd op de vijf kenmerken van de taakuitvoering die in het referentiekader voor gesprekken worden onderscheiden, aangevuld met de uitvoering van het gesprek (UV) in het algemeen. Leerlingen konden in totaal maximaal 34 punten halen. Net als bij het onderdeel spreken, vulde de toetsleider het beoordelingsmodel voor elke leerling zoveel mogelijk tijdens het spel in².

Ook bij het construeren van een totaalscore voor het onderdeel gesprekken is aan de technische aspecten van gespreksvaardigheid minder gewicht toegekend (7 aspecten met gewicht 0,49 = 20% van de score) dan aan de inhoudelijke beoordelingsaspecten (10 items met gewicht 1,36 = 80% van de score). Voor de analyse van leerlingprestaties per kenmerk van de taakuitvoering zijn gemiddelde (ongewogen) proportiescores per taakkenmerk berekend, door de gemiddelde scores per taakkenmerk te delen door de maximaal haalbare score per aspect (2).

² Ook bij het onderdeel gesprekken voeren, is de kwaliteit van de beoordelingen onderzocht in een aanvullend onderzoek. Hiervoor werd een deel van de leerlingen achteraf, op basis van de beschikbare filmopnames, door een expertbeoordelaar beoordeeld. Het aanvullend onderzoek gaf geen aanleiding om oordelen van bepaalde toetsleiders uit te sluiten van analyse.

2.2 Werkwijze

De werkwijze om te komen tot beantwoording van de onderzoeksvragen wordt voor een groot deel beschreven in het technisch rapport van het uitvoerend consortium (Keuning et al., 2019). In de hiernavolgende paragrafen beschrijven we de aanvullende analyses die zijn gedaan en geven we aanvullingen op of beschrijven we hoe we afwijken van de reeds beschreven werkwijze.

2.2.1 *Weging SO*

In de steekproef van so-scholen waren leerlingen met leerroute 2 oververtegenwoordigd ten opzichte van de gehele populatie van so-leerlingen. Omdat dit achtergrondkenmerk (theoretisch gezien) verband kan houden met leerlingprestaties, is er voor gekozen om in de populatieschattingen voor so een weging naar leerroute toe te passen. Tabel 2.2.1a geeft de wegingsfactor.

Tabel 2.2.1a Wegingsfactor

	wegingsfactor
Leerroute 1	1,221
Leerroute 2	0,763
Leerroute 3	1,496

Analyses waarbij weging is toegepast zijn:

- Algemene tellingen referentieniveau (figuur 2.2a)
- Scoreverdelingen per subdomein (histogrammen, bijv. 2.3.1a)

In de analyses waar binnen of tussen referentieniveaugroepen en zorgprofielen wordt vergeleken, is er niet gewogen naar leerroute.

So-leerlingen van wie de leerroute niet bekend was (aan de scholen is gevraagd om voorafgaand aan het onderzoek aan te geven in welke leerroute leerlingen geplaatst waren), zijn niet meegenomen in de gewogen analyses. Dit geldt voor 41 leerlingen bij het onderdeel luisteren, 36 leerlingen bij spreken en 38 leerlingen bij het onderdeel gesprekken voeren. In de analyses waarbij we vergelijkingen maken tussen of binnen referentiegroepen <1F, 1F en 1S/2F (bij onderdeel luisteren) en zorgprofielen zijn deze leerlingen wél meegenomen.

2.2.2 *Indeling in zorgprofielen*

Leerlingkenmerken zijn in het peilingsonderzoek verzameld via een Quickscan die door de leerkracht per leerling werd ingevuld. Het zorgprofiel van de leerling werd geïnventariseerd door te vragen naar de aanwezigheid van een aantal beperkingen. Leerkrachten konden voor de volgende beperkingen aangeven of deze zijn gediagnosticeerd of werden vermoed bij de leerling.

- attention deficit (hyperactivity) disorder (ADHD/ADD)
- oppositionele-opstandige stoornis (Oppositional Defiant Disorder - ODD) of een normoverschrijdendgedragsstoornis (Conduct Disorder - CD)
- autisme spectrum stoornis (ASS)
- angst- en/of stemmingsstoornis
- hechtingsstoornis
- obsessieve-compulsieve stoornis (Obsessive Compulsive Disorder-OCD)/dwanghandelingen
- Gilles de la Tourette
- een andere psychiatrische stoornis

De gediagnosticeerde stoornissen zijn vervolgens gegroepeerd naar bestaande zorgprofielen uit Passende perspectieven taal: profielschetsen (Langberg, Jansma, Fortgens & Van der Schie, 2012), namelijk: ADHD, ASS en andere beperkingen/een combinatie van beperkingen.

In de analyses die aan het publieksrapport ten grondslag liggen, zijn we dus uitgegaan van gediagnosticeerde enkelvoudige beperkingen. Leerlingen met meerdere gediagnosticeerde beperkingen zijn ingedeeld in het zorgprofiel 'combinatie van beperkingen'. Naast de zorgprofielen onderscheidt Passende perspectieven taal nog een aantal andere profielen met mogelijke gevolgen voor de taalontwikkeling, waaronder een profiel 'lagere cognitie'. Ook dit profiel nemen wij mee in het peilingsonderzoek, gebaseerd op een vraag uit de Quickscan of er bij de leerling een IQ lager dan 70 is vastgesteld.

2.2.3 *Significantie toetsing*

In het rapport is er getoetst op verschillen in prestaties tussen a) schooltype (sbo vs. so), b) referentieniveaugroepen, c) taakkenmerken, en d) zorgprofielen.

In hiernavolgende paragrafen wordt beschreven welke analysemethoden zijn gebruikt en worden de uitkomsten van de analyses samengevat.

Prestaties luisteren, spreken en gesprekken voeren samen

Om verschillen in prestaties van sbo- versus so-leerlingen te toetsen, is een chi-kwadraattoets gebruikt.

Tabel 2.2.3a Aantal leerlingen in het sbo en so dat een bepaald referentieniveau haalt en uitkomsten van een chi-kwadraattoets op onderlinge verschillen

	<i>sbo</i>	<i>so (gewogen)</i>	<i>Chi kwadraat</i>	<i>p-waarde</i>
<i>Luisteren</i>				
<1F	672	71		
1F	469	140	45,84	<0,05
<i>Spreken</i>				
<1F	366	49		
1F	787	179	9,52	<0,05
<i>Gesprekken</i>				
<1F	657	101		
1F	465	96	3,64	0,06

Op de onderdelen luisteren en spreken presteren so-leerlingen beter dan sbo-leerlingen; een significant groter deel van hen behaalt niveau 1F.

Prestaties luistervaardigheid

Om verschillen in beheersing van verschillende taakkenmerken en tekst/fragmentkenmerken te toetsen, is gebruik gemaakt van een reeks ANOVA's voor herhaalde metingen met als 'within-subject variabele' het betreffende taak- of tekstkenmerk. Post hoc tests met Bonferroni correctie zijn gebruikt om te achterhalen op welke taak- of tekstkenmerken precies de significante verschillen zijn te vinden.

Tabel 2.2.3b Hoofdeffecten taak/tekstkenmerk

	F	Df	p-waarde	Post-hoc
Taakkenmerk: begrijpen/interpreteren/samenvatten	197773	(1, 1465)	p < 0,001	Alle contrasten p < 0,001
Fragmentkenmerk: audio vs. video x	5328,42	(1, 1390)	p < 0,001	-
Fragmentkenmerk: dialoog vs. monoloog	315,83	(1, 1390)	p < 0,001	-
Tekstkenmerk: <1F, 1F, 2F	56794,82	(1, 1679)	p < 0,001	Alle contrasten p < 0,001

NB: met Greenhouse Geisser correctie, vanwege violatie van de assumptie op sphericiteit. Vrijheidsgraden zijn wel afgerond op gehelen.

Interacties met referentieniveaugroep zijn getoetst door aan de ANOVA's als 'between-subject' variabele de referentieniveaugroep toe te voegen en dus zogeheten mixed ANOVA's uit te voeren.

Tabel 2.2.3c Interactie-effecten taak/tekstkenmerk x referentieniveaugroep

	F	Df	p-waarde
Taakkenmerk x referentieniveaugroep	288,13	(2, 1442)	p < 0,001
Fragmentkenmerk: audio vs. video x referentieniveaugroep	1816,30	(2, 1390)	p < 0,001
Fragmentkenmerk: dialoog vs. monoloog x referentieniveaugroep	1023,69	(2,1390)	p < 0,001
Tekstkenmerk: <1F, 1F, 2F x referentieniveaugroep	2406,27	(2, 1679)	p < 0,001

NB: met Greenhouse Geisser correctie, vanwege violatie van de assumptie op sphericiteit. Vrijheidsgraden zijn wel afgerond op gehelen.

Het is niet mogelijk om via post hoc toetsen te achterhalen op welk taakkenmerk beheersingsniveaus significant van elkaar verschillen. Leerlingen in referentieniveaugroep 1F presteren namelijk beter dan leerlingen die het niveau 1F niet behalen en dat geldt voor alle kenmerken van de taakuitvoering. Daarom wordt de beheersing naar taakkenmerken relatief getoetst: waarop scoren leerlingen in beide referentieniveaugroepen het beste en op welk taakkenmerk het slechtst?

Voor dit doel werd met een repeated measures ANOVA per referentieniveaugroep, met als 'within subjects variabele' taak/tekstkenmerk, het beheersingsniveau per taakkenmerk binnen elke referentieniveaugroep geanalyseerd. In het rapport wordt gespecificeerd op welke taak- of tekstkenmerken relatieve beheersingsniveaus tussen groepen verschillen.

Interacties met zorgprofielen werden getoetst met een repeated measures ANCOVA met als 'within subjects variabele' taak/tekstkenmerk en als 'between subjects variabele' zorgprofiel. Daarnaast is geslacht meegenomen als covariaat. Jongens en meisjes zijn namelijk ongelijk verdeeld over de zorgprofielen (zie Tabel 2.2.3d) wat een vertekend beeld geeft van de prestaties naar zorgprofiel.

Tabel 2.2.3d Verdeling van geslacht naar zorgprofiel

		GESLACHT	
		Jongen	Meisje
Geen	N	316	215
	%	38,6%	52,4%
ADHD	N	122	43
	%	14,9%	10,5%
ASS	N	112	24
	%	13,7%	5,9%
Lagere cognitie	N	72	76
	%	8,8%	18,5%
Anders/combinatie van beperkingen	N	196	52
	%	24,0%	12,7%

Tabel 2.2.3e geeft een overzicht van de interactie-effecten.

Tabel 2.2.3e Interactie-effecten taak/tekstkenmerk x zorgprofiel

	F	Df	p-waarde
Taakkenmerk x zorgprofiel	19,46	(4, 1290)	p < 0,001
Fragmentkenmerk: audio vs. video x zorgprofiel	46,50	(4, 1222)	p < 0,001
Fragmentkenmerk: dialoog vs. monoloog x zorgprofiel	47,68	(4, 1222)	p < 0,001
Tekstkenmerk: <1F, 1F, 2F x zorgprofiel	18,34	(4, 1305)	P < 0,001

NB: met Greenhouse Geisser correctie, vanwege violatie van de assumptie op sphericiteit (de variantie van de verschil-scores tussen de kenmerken (bijvoorbeeld van begrijpen – interpreteren en interpreteren – samenvatten) moet gelijk zijn). Vrijheidsgraden zijn wel afgerond op gehelen.

Ook hier geldt dat we repeated measures ANOVA's per zorgprofiel hebben uitgevoerd om inzicht te krijgen in relatieve beheersingsniveaus: wat kunnen leerlingen met een bepaald zorgprofiel goed/minder goed, kijkend naar het beheersingsniveau van de andere taakkenmerken? En verschilt dit van leerlingen met een ander zorgprofiel? In het rapport wordt aangegeven hoe relatieve beheersingsniveaus tussen zorgprofielen van elkaar verschillen.

Prestaties spreekvaardigheid

In het rapport wordt de spreekvaardigheid per kenmerk van de taakuitvoering weergegeven. Kenmerken van de taakuitvoering voor het onderdeel spreken zijn: Uitvoering, Samenhang, Afstemming op doel, Afstemming op publiek en Woordgebruik en woordenschat en Vloeiendheid, verstaanbaarheid en grammaticaliteit.

Om verschillen in vaardigheid tussen deze kenmerken van de taakuitvoering te toetsen, is een repeated measures ANOVA uitgevoerd. Kenmerk van de taakuitvoering ('within subjects factor') had een significant effect op de spreekvaardigheid, $F(5, 6639) = 465,88, p < 0,001$. Post hoc tests met Bonferroni correctie toonden aan dat de vaardigheid op alle kenmerken van de taakuitvoering onderling significant verschilde, alle p 's < 0,02.

Tabel 2.2.3f Beschrijvende statistieken bij de taakkenmerken van het onderdeel spreken (vaardigheidsniveau in proportie t.o.v. het maximum)

	Gemiddelde	Std. Deviatie	N
Uitvoering	0,73	0,22	1417
Samenhang	0,74	0,24	1417
Afstemming op doel	0,85	0,19	1417
Afstemming op het publiek	0,63	0,17	1417
Woordgebruik en woordenschat	0,80	0,24	1417
Vloeiendheid, verstaanbaarheid en grammaticaliteit	0,84	0,17	1417

Vervolgens zijn beurtelings 'referentieniveaugroepen' en 'zorgprofielen' als 'between subjects factor' aan de analyse toegevoegd, om de interactie tussen deze groepen en kenmerken van de taakuitvoering te toetsen. Beide interacties zijn significant, met $F(5, 6603) = 64,21, p < 0,001$ voor het interactie-effect tussen referentieniveaugroepen en taakkenmerken en $F(19, 5859) = 3,99, p < 0,001$ voor het interactie-effect tussen zorgprofielen en taakkenmerken (waarbij in de analyse 'geslacht' opgenomen is als covariaat).

Met repeated measures ANOVA's (met als within-subjects factor 'taakkenmerk') per referentieniveau en zorgprofiel is vervolgens inzichtelijk gemaakt wat het relatieve vaardigheidsniveau is per kenmerk van de taakuitvoering binnen een bepaalde groep leerlingen. In het rapport beschrijven we waar relatieve vaardigheidsniveaus tussen groepen van elkaar afwijken.

Prestaties gesprekvaardigheid

Ook de gespreksvaardigheid werd gemeten aan de hand van de kenmerken van de taakuitvoering: Uitvoering, Samenhang, Afstemming op doel, Afstemming op de gesprekspartner(s), Woordgebruik en woordenschat en Vloeiendheid, verstaanbaarheid en grammaticaliteit.

Tabel 2.2.3g geeft het vaardigheidsniveau van leerlingen op de verschillende kenmerken van de taakuitvoering weer.

Tabel 2.2.3g Beschrijvende statistieken bij de taakkenmerken van het onderdeel gesprekken (vaardigheidsniveau in proportie t.o.v. het maximum)

	Gemiddelde	Std. Deviatie	N
Uitvoering	0,62	0,34	1357
Samenhang	0,75	0,22	1357
Afstemming op doel	0,67	0,31	1357
Afstemming op de gesprekspartner(s)	0,74	0,23	1357
Woordgebruik en woordenschat	0,88	0,22	1357
Vloeiendheid, verstaanbaarheid en grammaticaliteit	0,89	0,16	1357

Om verschillen in vaardigheid tussen deze kenmerken van de taakuitvoering te toetsen, is een repeated measures ANOVA uitgevoerd. Kenmerk van de taakuitvoering ('within subjects factor') had een significant effect op de gespreksvaardigheid, $F(5, 5018) = 472,69, p < 0,001$. Post hoc tests met Bonferroni correctie toonden aan dat de vaardigheid op alle kenmerken van de taakuitvoering

onderling significant verschilde (alle p 's $< 0,04$), m.u.v. de kenmerken Samenhang en Afstemming op gesprekspartners, $p = 0,09$.

Ook hier zijn beurtelings 'referentieniveaugroepen' en 'zorgprofielen' als 'between subjects factor' aan de analyse toegevoegd, om de interactie tussen deze groepen en kenmerken van de taakuitvoering te toetsen. Beide interacties zijn significant, met $F(4, 5262) = 143,92$, $p < 0,001$ voor het interactie-effect tussen referentieniveaugroepen en taakkenmerken en $F(15, 4557) = 3,82$, $p < 0,001$ voor het interactie-effect tussen zorgprofielen en taakkenmerken (waarbij in de analyse 'geslacht' opgenomen is als covariaat).

Met repeated measures ANOVA's (met als 'within-subjects factor' 'taakkenmerk') per referentieniveau en zorgprofiel is gekeken naar relatieve vaardigheidsniveaus per kenmerk van de taakuitvoering binnen een bepaalde groep leerlingen; waar relatieve vaardigheidsniveaus tussen groepen van elkaar afwijken, worden deze gerapporteerd.

2.2.4 *Bepalen van grensitems Luisteren*

In het rapport kijken we naar de exacte opgaven uit de pen-en-papiertaak die sommige groepen leerlingen nog nét wel en net niet meer beheersen. Dit noemen we de 'grensitems'. Op die manier geven de opgaven een beeld van wat leerlingen die de referentieniveaus beheersen, kennen en kunnen.

Grensitems zijn bepaald door te kijken naar de beheersing op itemniveau, dat bepaald is met behulp van het item respons model (zie pagina 86 in het Technisch rapport van Keuning et al., 2019). Voor elk item staat in het Technisch rapport beschreven hoeveel vaardigheid nodig is om 50 of 80 procent kans te hebben op het correct maken van dat item en vervolgens is gekeken hoe dat zich verhoudt tot de hoeveelheid vaardigheid die nodig is om referentieniveau 1F of 2F te beheersen. Voor het makkelijkste item uit de pen-en-papiertaak ligt het vaardigheidsniveau dat benodigd is voor beheersing dus het laagst en voor het moeilijkste het hoogst. Voor items op de grens van 1F, ligt het punt op de vaardigheidsverdeling waarop voor het item geen beheersing (kans op een correct antwoord kleiner dan 50%) overgaat in matige beheersing (minimaal 50% kans op een goed antwoord) net boven het punt waar ook het cesuurpunt voor 1F ligt op de (lineair getransformeerde) latente vaardigheidsschaal. Voor items op de grens van 1F en 2F, ligt het beheersingspunt net boven het cesuurpunt voor 2F op de vaardigheidsschaal (zie o.a. figuur 8.5 in het Technisch rapport van Keuning et al., 2019).

Spreken en gesprekken

In de spreek- en gesprekstaak is ieder kenmerk van de taakuitvoering beoordeeld aan de hand van meerdere aspecten. Zo komt de score voor het kenmerk *samenhang* bijvoorbeeld tot stand op basis van scores op de beoordelingsaspecten *samenhang op macro- en mesoniveau: opbouw* en *samenhang op microniveau: signaalwoorden*. Als we hierop inzoomen, kunnen we achterhalen welke specifieke beoordelingsaspecten leerlingen die het fundamentele niveau 1F niet beheersen met name onderscheiden van leerlingen die dit wel doen.

Gemiddelde aspectscores zijn vergeleken tussen referentieniveaugroepen middels onafhankelijke t-toetsen (independent sample t-test). Omdat verreweg de meeste aspectscores significant verschilden tussen referentieniveaugroepen, is de grootte van het effect (Cohen's d) gebruikt om de aspecten die de meest onderscheidende

scores tussen <1F-leerlingen en 1F-leerlingen opleverde, te selecteren. Tabel 2.2.4a. geeft voor het onderdeel spreken een overzicht van de gemiddelde scores per beoordelingsaspect, het verschil in score tussen <1F en 1F en de grootte van het effect in Cohen's *d*. Tabel 2.2.4b geeft hetzelfde weer voor het onderdeel gesprekken.

Tabel 2.2.4a Gemiddelde score per beoordelingsaspect (maximale score 2) voor het onderdeel spreken binnen elk referentieniveau, verschillen in scores tussen referentiegroepen en effectgrootte (Cohen's *d*) van deze verschillen. De aspecten met het grootste onderscheidend vermogen (gemeten naar Cohen's *d* van het verschil) zijn grijs gemarkeerd.

Kenmerk taakuitvoering	Beoordelingsaspect	Gemiddelde score		<1F vs 1F	
		<1F	1F	verschil in score	Cohen's <i>d</i>
Taak: Dobbelspel					
UV	Volledigheid	1,55	1,91	-0,35	0,68
	Inhoudelijke kwaliteit	0,81	1,49	-0,68	1,21
SH	Samenhang op macro- en mesoniveau: opbouw	1,00	1,62	-0,62	1,17
	Samenhang op microniveau: signaalwoorden	1,13	1,78	-0,65	1,02
AD	Spreekdoel tot uiting brengen	1,81	1,98	-0,17	0,45
AP	Taalgebruik afstemmen	1,65	1,94	-0,29	0,61
	Publiek boeien	0,48	1,13	-0,65	1,08
	Non-verbale communicatie	0,85	1,20	-0,35	0,84
WW	Woordgebruik	1,05	1,77	-0,71	1,20
	Woordenschat	1,29	1,86	-0,57	1,02
VVG	Vloeiendheid	1,24	1,73	-0,49	0,78
	Uitspraak en intonatie	1,45	1,86	-0,41	0,70
	Grammatica: vervoeging	1,38	1,84	-0,46	0,80
	Grammatica: verbuiging	1,54	1,85	-0,30	0,57
	Grammatica: zinsbouw	1,17	1,73	-0,56	1,07
Taak: Jeugdjournaalfragment					
UV	Volledigheid	1,19	1,72	-0,53	0,79
	Inhoudelijke kwaliteit	0,64	1,43	-0,79	1,38
SH	Samenhang op macro- en mesoniveau: opbouw	0,77	1,60	-0,83	1,42
	Samenhang op microniveau: signaalwoorden	1,00	1,83	-0,83	1,30
AD	Spreekdoel tot uiting brengen	0,94	1,70	-0,76	1,29
AP	Taalgebruik afstemmen	1,65	1,97	-0,32	0,67

	Publiek boeien	0,32	0,93	-0,62	1,14
	Non-verbale communicatie	0,79	1,12	-0,32	0,76
WW	Woordgebruik	0,90	1,73	-0,82	1,34
	Woordenschat	1,18	1,82	-0,63	1,09
VVG	Vloeiendheid	1,11	1,68	-0,57	0,85
	Uitspraak en intonatie	1,48	1,88	-0,40	0,73
	Grammatica: vervoeging	1,47	1,87	-0,40	0,78
	Grammatica: verbuiging	1,60	1,92	-0,33	0,64
	Grammatica: zinsbouw	1,13	1,73	-0,60	1,16

Tabel 2.2.4b Gemiddelde score per beoordelingsaspect (maximaalscore 2) voor het onderdeel gesprekken binnen elk referentieniveau, verschillen in scores tussen referentiegroepen en effectgrootte (Cohen's *d*) van deze verschillen. De aspecten met het grootste onderscheidend vermogen (gemeten naar Cohen's *d* van het verschil) zijn grijs gemarkeerd.

Kenmerk taakuitvoering	Beoordelingsaspect	Gemiddelde score		<1F vs 1F	
		<1F	1F	verschil in score	Cohen's <i>d</i>
IK	Kwaliteit van de inhoud	0,90	1,69	-0,79	1,50
BS	Gesprek onderhouden	1,02	1,75	-0,73	1,33
	Passendheid van beurten nemen	1,34	1,89	-0,55	1,05
	Inbreng gesprekspartners integreren	1,39	1,91	-0,52	1,05
AD	Doelgerichtheid	0,90	1,77	-0,87	1,67
	Begrijpelijk maken	1,05	1,85	-0,80	1,57
AG	Taalgebruik afstemmen	1,49	1,94	-0,46	1,05
	Non-verbale communicatie	1,05	1,56	-0,51	1,01
	Actief luistergedrag	1,08	1,85	-0,78	1,60
	Reageren op gesprekspartner	1,30	1,91	-0,61	1,19
WW	Woordgebruik	1,53	1,96	-0,43	1,00
	Woordenschat	1,66	1,96	-0,30	0,78
VVG	Vloeiendheid	1,60	1,89	-0,29	0,64
	Uitspraak en intonatie	1,64	1,93	-0,28	0,67
	Grammatica: vervoeging	1,75	1,95	-0,19	0,56
	Grammatica: verbuiging	1,79	1,94	-0,16	0,48
	Grammatica: zinsbouw	1,57	1,90	-0,33	0,79

3 Verschillen tussen scholen en leerlingen

In dit hoofdstuk van het publieksrapport gaan we in op verschillen in prestaties tussen leerlingen en scholen. We maken een koppeling tussen prestaties en kenmerken van het onderwijsleerproces en (algemene) achtergrondkenmerken van de leerlingen en scholen. Daarmee proberen we een antwoord te vinden op de vragen:

- Hoe groot zijn verschillen tussen scholen en leerlingen in de prestaties op de subdomeinen luisteren, spreken en gesprekken?
- Hangen deze prestatieverschillen samen met kenmerken van het onderwijsleerproces op het gebied van de mondelinge taalvaardigheid?
- Of zijn het vooral domeinspecifieke (zoals de thuistaal) of algemene achtergrondkenmerken (zoals het geslacht) van leerlingen en scholen (zoals schoolgrootte) die samenhang vertonen met prestatieverschillen?

3.1 Databronnen

Voor de analyses die ten grondslag liggen aan dit hoofdstuk zijn de leerlingprestaties in bankscores (luisteren) en in ruwe scores (spreken en gesprekken) gekoppeld aan de variabelen uit het interview (Hoofdstuk 1), de Quickscan (Deel C, Hoofdstuk 1) en aan algemene achtergrondkenmerken (zoals de regio waarin de school zich bevindt) die zijn opgehaald.

Vervolgens zijn ontbrekende waarden in de variabelen (m.u.v. de uitkomstmaten voor luisteren, spreken en gesprekken) geïmputeerd volgens de methode van joint modeling multiple imputation (zie Keuning et al., 2019). Voordat waarden werden geïmputeerd werden eerst scholen zonder interview ($n=6$ scholen, 92 leerlingen), leerlingen zonder quickscan ($n=6$ scholen, 18 leerlingen) en leerlingen met zeer incomplete quickscans ($n= 2$ scholen, 132 leerlingen) uit de dataset verwijderd. Er bleven 1131 leerlingen over (911 sbo; 220 so) en 54 scholen (43 sbo; 11 so). Daarna is er in het imputatiemodel een selectie gemaakt uit alle beschikbare variabelen. Deze selectie werd gebaseerd op inhoudelijke en statistische gronden. Zo zijn variabelen die zeer kleine groepen opleveren (zoals bijvoorbeeld de vraag of een hechtingsstoornis gediagnosticeerd is) niet gekozen, tenzij ze speciaal belangrijk worden geacht (zoals herkomst). Op schoolniveau is per aspect van het onderwijsleerproces een keuze gemaakt voor variabelen die inhoudelijk het meest relevant zijn en een eenduidige interpretatie hebben, en ook hier is erop gelet dat de variabelen geen al te kleine groepen scholen opleveren. Deze selectie is belangrijk, omdat deze de bovengrens vormt voor de selectie van predictoren voor de meerniveau-analyses. In tabel 3.1a (afkomstig uit het Technisch rapport van Keuning et al., 2019) wordt een overzicht gegeven van de variabelen die meegenomen zijn bij het imputeren van ontbrekende gegevens.

Ten behoeve van de meerniveau-analyses zijn aan deze set variabelen nog de volgende schoolkenmerken toegevoegd:

Schoolkenmerken:

- Achtergrondkenmerken: stedelijkheid, schoolgrootte, regio
- Aantal activiteiten die aan bod komen tijdens spreken/luisteren/gesprekken²
- Aantal MTV onderwerpen die aan bod komen tijdens het teamoverleg³
- Het percentage leerlingen met leerroute 3 per school

2 Deze variabelen zijn (i.v.m. missings) alleen toegevoegd als ze een duidelijke modelbijdrage leverden.

Tabel 3.1a Overzicht van variabelen die zijn meegenomen bij het imputeren van missende waarden

Omschrijving variabele	Type	Specificatie in jomo	% missing
Schoolnummer		cluster-ID	---
Bankscore Luisteren	continu	imputeren, niv. 1	9.37
Ruwe score Spreken	continu	imputeren, niv. 1	6.90
Ruwe score Gesprekken	continu	imputeren, niv. 1	12.73
Leerweg, als 1, 2, 3 of 4 = "anders"	polytoom	imputeren, niv. 1	2.03
Leerkracht-inschatting uitstroomprofiel	polytoom	imputeren, niv. 1	2.39
Thuis taal	dichotoom	imputeren, niv. 1	2.92
Herkomst	dichotoom	imputeren, niv. 1	4.07
Leerachterstand in taal/lezen	polytoom	imputeren, niv. 1	5.48
Leerachterstand informatieverwerking	polytoom	imputeren, niv. 1	9.55
Mate waarin er sprake is van stimulerende factoren	continu	imputeren, niv. 1	3.09
Onderwijsbehoefte: extra structuur	dichotoom	imputeren, niv. 1	1.24
Onderwijsbehoefte: stimulering zelfvertrouwen	dichotoom	imputeren, niv. 1	1.24
Onderwijsbehoefte: meer verantwoordelijkheid	dichotoom	imputeren, niv. 1	1.24
Laag IQ (versus hoog/gemiddeld)	dichotoom	imputeren, niv. 1	1.15
ADHD/ADD gediagnosticeerd	dichotoom	imputeren, niv. 1	0.97
Autisme of een verwante stoornis gediagnosticeerd	dichotoom	imputeren, niv. 1	0.97
Andere stoornis gediagnosticeerd	dichotoom	imputeren, niv. 1	0.97
Wordt de methode volledig gevolgd?	dichotoom	imputeren, niv. 2	2.12
Wordt het onderwijs in X ingebed in de taalles?	dichotoom	imputeren, niv. 2	5.13
Wordt het onderwijs in X geïntegreerd in andere vakken?	dichotoom	imputeren, niv. 2	5.13
Zijn aan elk doel taalactiviteiten gekoppeld? (doelgerichte aanpak)	dichotoom	imputeren, niv. 2	3.63
Zijn aan elke taalactiviteit doelen gekoppeld? (activiteitgerichte aanpak)	dichotoom	imputeren, niv. 2	4.86
Zijn er einddoelen voor schoolverlaters geformuleerd?	dichotoom	imputeren, niv. 2	3.01
Wordt gebruikgemaakt van genormeerde toetsen of observatielijsten?	dichotoom	imputeren, niv. 2	11.85
Worden leervorderingen in de lessen en/of in de loop van het schooljaar gemonitord en geregistreerd?	dichotoom	imputeren, niv. 2	0.71
Is er differentiatie naar leerstof?	dichotoom	imputeren, niv. 2	0.35
Is er differentiatie naar (streef)doelen?	dichotoom	imputeren, niv. 2	5.31
Is mondelinge taalvaardigheid benoemd in het school- of taalbeleidsplan?	dichotoom	imputeren, niv. 2	10.43
Zijn er intervisiemomenten of wordt collegiale consultatie toegepast?	dichotoom	imputeren, niv. 2	1.24
Is er interne ondersteuning? (Bijvoorbeeld een taalcoördinator)	dichotoom	imputeren, niv. 2	3.98
Is er externe ondersteuning? (Bijvoorbeeld een onderwijsadviseur)	dichotoom	imputeren, niv. 2	5.39
Geslacht	dichotoom	covariaat, niv. 1	---
Schooltype (sbo of so-4)	dichotoom	covariaat, niv. 2	---
Vindt er convergente differentiatie plaats?	dichotoom	covariaat, niv. 2	---
Vindt er divergente differentie plaats?	dichotoom	covariaat, niv. 2	---
Wordt er gewerkt met niveaugroepen?	dichotoom	covariaat, niv. 2	---
Is er differentiatie naar instructie?	dichotoom	covariaat, niv. 2	---

3.2 Werkwijze

Intraclasscorrelatie

Om in kaart te brengen welk deel van de verschillen in prestaties (gemeten naar de totaalscore voor de drie subdomeinen) verklaard wordt door schoolverschillen en welk deel door leerlingverschillen, is er een leeg meerniveaumodel (een model zonder predictoren) uitgevoerd met enkel een random intercept voor scholen. Uit dit '0-model' is de intraclasscorrelatie te berekenen. Deze geeft aan welke proportie van de gevonden verschillen verklaard kan worden door schoolkenmerken.

$$ICC = \frac{\text{Schoolvariantie}}{\text{Schoolvariantie} + \text{Leerlingvariantie}}$$

Tabel 3.2a Intraclasscorrelaties luisteren, spreken en gesprekken

Onderdeel		Variantie	ICC
Luisteren	School	41,01	0,15
	Leerling	231,47	
Spreken	School	16,64	0,20
	Leerling	67,29	
Gesprekken	School	6,17	0,15
	Leerling	34,87	

Uit de intraclasscorrelatie kunnen we afleiden dat 15% van de verschillen in luistervaardigheid, 20% van de verschillen in spreekvaardigheid en 15% van de verschillen in gespreksvaardigheid verklaard wordt door verschillen op het niveau van de school; het resterende deel wordt verklaard door verschillen op het niveau van de leerling.

Meerniveaumodellen

Om na te gaan welke schoolkenmerken de verschillen in leerlingprestaties verklaren, zijn meerdere meerniveaumodellen uitgevoerd (Snijders & Bosker, 2012). In het eerste model zijn enkel de achtergrondkenmerken (antecedenten) opgenomen. Met dit model controleren we voor verschillen tussen scholen die ontstaan door verschillen in leerlingpopulaties en andere achtergrondkenmerken. We beschrijven in het rapport dan ook eerst welke achtergrondkenmerken van scholen en leerlingen samenhangen met de prestaties.

Vervolgens hebben we het schooltype (sbo vs. so) toegevoegd aan het model. Na correctie voor leerlingkenmerken bleek het schooltype geen significant effect te hebben op leerlingprestaties in het meerniveau-model. De variabele schooltype is daarom weer verwijderd uit het model. Aan het model met achtergrondkenmerken zijn tot slot de onderwijsleerprocesvariabelen toegevoegd. Tabel 4.2.2 toont de drie eindmodellen voor luistervaardigheid, spreekvaardigheid en gespreksvaardigheid.

Tabel 3.2b Meerniveau-analyses luistervaardigheid, spreekvaardigheid en gespreksvaardigheid

Subdomein Mondelinge taalvaardigheid						
Kenmerken	Luisteren		Spreken		Gesprekken	
	<i>b</i>	SE (<i>b</i>)	<i>b</i>	SE (<i>b</i>)	<i>b</i>	SE (<i>b</i>)
Fixed effecten						
Intercept	62,08	4,96	44,55	2,95	27,55	2,2
Onderwijsleerproces						
De methode wordt volledig gevolgd	-0,13	1,95	-0,96	1,40	-0,68	1,17
Er zijn einddoelen voor schoolverlaters geformuleerd	2,9	2,66	3,28	1,41	1,67	1,4
Leerlingvorderingen worden in de lessen en/of in de loop van het schooljaar gemonitord en geregistreerd	0,54	2,35	0,56	1,39	-0,37	1,17
Er wordt gewerkt met niveaugroepen	-2,37	2,73	-1,1	1,58	0,08	1,27
Er is differentiatie naar leerstof	-1,64	4,4	0,07	1,94	-1,22	1,63
Er is differentiatie naar (streef)doelen	-1,32	3,4	-2,71	1,83	-1,17	1,41
Mondelinge taalvaardigheid is benoemd in het school- of taalbeleidsplan	0,54	2,75	0,55	1,22	-1,05	1,27
Er zijn intervisiemomenten of collegiale consultatie wordt toegepast	-1,09	2,05	-3,45	1,35	-0,7	1,09
Leerlingachtergrond						
Meisjes	-2,08	0,97	-0,38	0,56	-0,11	0,42
Herkomst Nederland (versus anders)	-0,31	2,24	-1,04	1,27	0,08	0,94
Laag IQ (vastgesteld <70)	-14,75	1,24	-5,52	0,73	-1,7	0,54
Stimulerende factoren	5,55	1,1	1,82	0,65	1,76	0,65
ADHD/ADD	1,24	1,03	0,84	0,59	-0,49	0,44
ASS	4,0	1,25	0	0,68	-0,86	0,49
Overige beperkingen/combinatie van beperkingen	1,07	1,49	0,63	0,81	0,18	0,6
Thuis taal Nederlands (versus niet-Nederlands)	3,54	1,74	2,67	1,05	1,28	0,77
Schoolachtergrond						
Percentage leerlingen met leerroute 3 (vso arbeid/pro)	-14,24	6,34	-1,27	2,97	-1,33	2,66
Regio: West vs. rest van Nederland*	0,09	3,67	-2,93	1,59	-3,77	1,64
Urbanisatiegraad stedelijk (versus niet-stedelijk)	-2,42	2,16	-0,16	1,56	-2,15	1,73
Schoolgrootte	0,02	0,01	0,02	0,01	0,01	0,01
Random effecten						
Schoolniveau variantie	16,97		10,69		6,71	
Leerlingniveau variantie	187,96		61,78		33,29	
Totaal variantie (% verklaard)	204,94 (25)		72,47 (14)		40 (3)	
ICC (0 model)	0,15		0,20		0,15	
Aantal scholen	54		54		54	
Aantal leerlingen	1131		1131		1131	
NB: Sign. codes: donker grijs < 0.001 grijs < 0.01 lichtgrijs < 0.05 onderstreept < 0.10						

*Voor regio zijn alle contrasten gecheckt. Alleen het significante contrast is hier opgenomen.

4 Referenties

- Inspectie van het Onderwijs (2019). *Peil. Mondelinge taalvaardigheid einde basisonderwijs*. Utrecht: Inspectie van het Onderwijs.
- Krom, R., Berkel, S. van, Schoot, F. van der, Sijtsma, J., Hemker, B., & Marsman, M. (2011). *Balans van het luisteronderwijs in het basis- en speciaal basisonderwijs 4. Uitkomsten van de vierde peiling in 2007*. Arnhem: Cito.
- Keuning J., Rietdijk S., Schouwstra S., & Scheltinga F. (2019). *Mondelinge Taalvaardigheid in het s(b)o. Technisch rapport peiling 2018*. Amsterdam: ITTA/Cito.
- Langberg, M., Jansma, S., Fortgens, C., & van der Schie, V. (2012). *Passende perspectieven taal*. Profielschetsen. Enschede: SLO.
- Prenger, J., & Damhuis, R. (2016). *Mondelinge taalvaardigheid in het basisonderwijs. Domeinbeschrijving ten behoeve van peilingsonderzoek*. Enschede: SLO.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles/London/New Delhi/Singapore/Washington D.C.: SAGE Publications Inc.

Colofon

Inspectie van het Onderwijs
Postbus 2730 | 3500 GS Utrecht
www.onderwijsinspectie.nl

Een exemplaar van deze publicatie is te downloaden vanaf de website van de Inspectie van het Onderwijs: www.onderwijsinspectie.nl.

© Inspectie van het Onderwijs | November 2019