



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

TECHNISCH RAPPORT PEIL.MONDELINGE TAALVAARDIGHEID

EINDE BASISONDERWIJS, SPECIAAL BASISONDERWIJS EN
SPECIAAL ONDERWIJS 2023-2024

December 2025

Inhoudsopgave

Inhoudsopgave	2
1 Inleiding	3
2 Werkwijze	4
2.1 Databronnen.....	4
2.2 Statistische toetsing	4
3 Het onderwijsleerproces	6
3.1 Schoolbeleid	6
3.2 Onderwijsaanbod.....	6
3.3 Lespraktijk	7
3.4 Zicht op ontwikkeling en differentiatie.....	9
4 Houding en gedrag ten opzichte van taal	11
4.1 Leerlingkenmerken	11
4.2 Leerkrachtkenmerk: ervaren bekwaamheid	13
5 Prestaties Mondelinge Taalvaardigheid	14
5.1 Databronnen.....	14
5.2 Prestaties luisteren, spreken en gesprekken voeren	16
5.3 Prestaties luistervaardigheid	16
5.4 Spreekvaardigheid.....	21
5.5 Gespreksvaardigheid.....	23
6 Verschillen tussen scholen en leerlingen	25
6.1 Ontbrekende waarden.....	25
6.2 Intraklassecorrelatie	25
6.3 Meerniveaumodellen	26
7 Deel C	27
Literatuurlijst	28

1 Inleiding

Dit is het technisch rapport (verder: TR) bij het peilingsonderzoek Mondelinge taalvaardigheid einde basisonderwijs (bo), speciaal basisonderwijs (sbo) en speciaal onderwijs (so). De peiling betreft het schooljaar 2023-2024. Dit 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 Stichting Cito.

Bij het peilingsonderzoek horen 2 technische rapportages. In het TR van het uitvoerend consortium is een uitgebreide verantwoording opgenomen van de wijze van dataverzameling, het instrumentarium en de eerste gegevensverwerking en -analyse ter beantwoording van de onderzoeksvragen (Gijsel et al., 2025). Het onderhavige TR, dat gemaakt is door de onderzoekers van de inspectie, dient ter verantwoording van de aanvullende gegevensverwerking die is uitgevoerd ten behoeve van het publieksrapport (Inspectie van het Onderwijs, 2025). Zo rapporteren we over een aantal aangepaste variabelen en over de statistische toetsing van verschillen tussen onderwijssoorten en samenhang tussen variabelen.

Het doel van het peilingsonderzoek Mondelinge taalvaardigheid einde bo, sbo en so is het in kaart brengen van de mondelinge taalvaardigheid van leerlingen aan het einde van het bo, sbo en so. Wat kennen en kunnen deze leerlingen op het gebied van mondelinge taalvaardigheid? Daarnaast beschrijft het rapport hoe het onderwijsleerproces voor mondelinge taalvaardigheid er op de deelnemende scholen uitziet en geeft het inzicht in een aantal aan mondelinge taalvaardigheid gerelateerde kenmerken van leerlingen en leerkrachten. Tot slot wordt er gekeken naar de prestatieverschillen tussen leerlingen en klassen en naar hoe de prestaties samenhangen met kenmerken van leerlingen, leerkrachten en scholen.

De onderzoeksvragen die in het publieksrapport opeenvolgend worden beantwoord zijn:

- 1) Wat zijn kenmerken van het onderwijsleerproces op het gebied van mondelinge taalvaardigheid?
- 2) Wat zijn kenmerken van de houding en het gedrag van leerlingen en leerkrachten ten aanzien van mondelinge taalvaardigheid?
- 3) Wat kunnen de leerlingen op het gebied van mondelinge taalvaardigheid?
- 4) Hoe hangen de prestaties van leerlingen op het gebied van mondelinge taalvaardigheid samen met kenmerken van de leerlingen, leerkrachten en het onderwijsleerproces?

De opbouw van dit TR is als volgt: we bespreken eerst kort de werkwijze (databronnen en statistische aanpak). Vervolgens worden voor elk hoofdstuk uit deel B en deel C van het publieksrapport (Inspectie van het Onderwijs, 2025) de gemiddelden, percentages en statistische toetsing gerapporteerd die niet in het TR van het consortium staan. Alle analyses zijn uitgevoerd met behulp van R (R Core Team, 2025).

2 Werkwijze

2.1 Databronnen

De gebruikte databronnen zijn uitgebreid beschreven in het TR van het uitvoerend consortium (Gijssel et al., 2025). De tabellen in het onderhavige TR zijn gebaseerd op dezelfde datasets met daarin de antwoorden op de leerling-, leerkracht- en schoolleidersvragenlijsten en de door het consortium bewerkte toetsdata. Voor verdere informatie over de databronnen verwijzen we dan ook naar Gijssel et al. (2025).

2.2 Statistische toetsing

Om verschillen tussen de onderwijssoorten (bo, sbo en so) te onderzoeken zijn statistische toetsen uitgevoerd. Om het aantal toetsen en daarmee het risico op vals positieven te beperken, zijn de verschillen tussen onderwijssoorten alleen getoetst op schaalscores en niet op de individuele stellingen onderliggend aan deze schaalscores. Stellingen die op zichzelf staan en geen onderdeel uitmaken van een schaal (zoals onderwijstijd) zijn ook vergeleken tussen onderwijssoorten. Bij alle toetsen hanteerden we $p < 0,05$ als grenswaarde voor het verwerpen van de nulhypothese.

De onderwijssoorten in het bo bestonden uit 3 groepen: bo, sbo en so. Wanneer de so groep te klein was voor toetsing, zijn alleen bo en sbo vergeleken. Om de schaalscores tussen de onderwijssoorten te vergelijken zijn steeds 3 stappen gevolgd. In de eerste stap is door middel van de Shapiro-Wilk test bepaald of de schaalscore normaal verdeeld was, en door middel van Levene's test of de varianties van de groepen gelijk verdeeld waren. In het geval van normaal verdeelde data en homogene varianties zijn als tweede stap de verschillen tussen de 3 groepen vergeleken met een ANOVA test (onafhankelijke t-toets bij 2 groepen).

Wanneer er sprake was van niet-normaal verdeelde data, met of zonder homogene varianties, zijn de verschillen tussen groepen getoetst met de Kruskal-Wallis test (Mann-Whitney U toets bij 2 groepen). Er waren geen gevallen waarin de data wel normaal verdeeld waren, maar de varianties niet homogeen waren. Bij een significant effect van groep, bekeken we in de laatste stap tussen welke groepen de schaalscores van elkaar verschilden met post-hoc toetsen. Als post-hoc toets gebruikten we dezelfde toets als voor het vergelijken van de 3 groepen. We rapporteren in dit TR de uitkomsten van deze tweede en eventueel derde stap.

In het geval van individuele stellingen die geen onderdeel uitmaken van een schaal zijn de percentages tussen onderwijssoorten in het bo vergeleken met een Kruskal-Wallis test als het gaat om ordinale data en een Chikwadraat-toets als het gaat om specifieke antwoordcategorieën (zoals percentage "ja").

Daarnaast zijn er toetsen uitgevoerd om de samenhang tussen variabelen te onderzoeken. In alle gevallen betrof het niet-normaal verdeelde data. Hiervoor is de Spearman's correlatiecoëfficiënt (ρ) berekend.

Tot slot is er waar mogelijk een vergelijking gemaakt met de resultaten van het peilingsonderzoek Mondelinge taalvaardigheid einde basisonderwijs 2016-2017 (Inspectie van het Onderwijs, 2019a). Omdat de vraagstelling en de wijze van

dataverzameling tussen het vorige peilingsonderzoek en het voorliggende peilingsonderzoek verschilt, is dit slechts beperkt mogelijk. Wanneer identieke schaalscores beschikbaar waren, zijn deze statistisch getoetst en is in dit TR een tabel opgenomen met de relevante toetsstatistieken; in andere gevallen zijn vergelijkingen gemaakt met eerder gerapporteerde uitkomsten zónder verdere toetsing. Deze vergelijkingen worden in het publieksrapport in globale termen besproken.

Er wordt geen vergelijking gemaakt met de resultaten van het peilingsonderzoek Mondelinge taalvaardigheid einde speciaal (basis)onderwijs 2018 (Inspectie van het Onderwijs, 2019b). Om zicht te krijgen op het onderwijsleerproces, zijn in 2018 interviews afgenomen met leerkrachten en/of intern begeleiders, waarvan de uitkomsten vervolgens zijn gekwantificeerd. Deze uitkomsten zijn niet (direct) te vergelijken met de resultaten van de schoolleiders- of leerkrachtvragenlijsten die voor de huidige peiling zijn afgenomen.

3 Het onderwijsleerproces

3.1 Schoolbeleid

3.1.1 Visie op het onderwijs in mondelinge taalvaardigheid

Tabel 3.1.1a: Score voor mate van visie op het onderwijs in mondelinge taalvaardigheid (bo n=95; sbo n=40)

Onderwijssoort	Gemiddelde	P10	P90
bo	2,26	1,25	3,25
sbo	2,46	1,50	3,02

Tabel 3.1.1b: Output Mann-Whitney test schaal visie onderwijs in mondelinge taalvaardigheid - effect van onderwijssoort (2 groepen)

Schaal	Groepen	W	p	Significantie
Visie	bo vs. sbo	1.680	0,221	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

3.1.2 Professionalisering

Tabel 3.1.2a: Score voor mate waarin thema's met betrekking tot mondelinge taalvaardigheid aan de orde komen in teamvergaderingen (bo n=95; sbo n=40)

Onderwijssoort	Gemiddelde	P10	P90
bo	1,29	1,00	2,00
sbo	1,33	1,00	2,00

Tabel 3.1.2b: Output Mann-Whitney test schaal thema's mondelinge taalvaardigheid in teamvergaderingen - effect van onderwijssoort (2 groepen)

Schaal	Groepen	W	p	Significantie
Aandacht MTV in teamvergaderingen	bo vs. sbo	1.680	0,221	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

3.2 Onderwijsaanbod

3.2.1 Onderwijstijd

Tabel 3.2.1a: Percentage onderwijstijd voor mondelinge taalvaardigheid van totale tijd taalonderwijs (bo n=95; sbo n=41)

Schalen	Onderwijssoort	Gemiddelde	P10	P90
Groep 1 en 2	bo	49,61	20,00	80,00
	sbo	43,90	0,00	80,00
Groep 3 t/m 8	bo	27,79	10,00	50,00
	sbo	30,61	10,00	65,00

Tabel 3.2.1b: Output Mann-Whitney test schaal onderwijstijd groep 1 en 2 - effect van onderwijssoort (2 groepen)

Schaal	Groepen	W	p	Significantie
Percentage onderwijstijd MTV	bo vs. sbo	2.138	0,367	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 3.2.1c: Output Mann-Whitney test schaal onderwijstijd groep 3 t/m 8 - effect van onderwijssoort (2 groepen)

Schaal	Groepen	W	p	Significantie
Percentage onderwijstijd MTV	bo vs. sbo	1.757	0,363	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 3.2.1d: Output Mann-Whitney test schaal onderwijstijd groep 1 en 2 - trendvergelijking (2016-2017 - 2023-2024)

Schaal	Groepen	W	p	Significantie
Percentage onderwijstijd MTV	16-17 vs. 23-24	5.690	0,270	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 3.2.1e: Output Mann-Whitney test schaal onderwijstijd groep 3 t/m 8 - trendvergelijking (2016-2017 - 2023-2024)

Schaal	Groepen	W	p	Significantie
Percentage onderwijstijd MTV	16-17 vs. 23-24	5.680	0,280	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

3.2.2

Methode en materiaal

Tabel 3.2.2a: Gebruikte lesmethode of materialen voor mondelinge taalvaardigheid (bo n=97; sbo n=36-42)

Groep	Antwoordoptie	Onderwijssoort	Percentage
Groep 3 t/m 8	vooral de taalmethode (en/of het kleuterprogramma)	bo	79,4
		sbo	78,6
	beide in ongeveer gelijke mate	bo	9,3
		sbo	4,8
	vooral zelf ontwikkeld les- en leermateriaal	bo	2,1
		sbo	4,8
anders	bo	9,3	
	sbo	11,9	
Groep 1 en 2	vooral de taalmethode (en/of het kleuterprogramma)	bo	35,1
		sbo	50,0
	beide in ongeveer gelijke mate	bo	36,1
		sbo	19,4
	vooral zelf ontwikkeld les- en leermateriaal	bo	19,6
		sbo	13,9
	anders	bo	9,3
		sbo	16,7

Noot: 6 sbo-schoolleiders gaven aan geen groep 1 en 2 te hebben op hun school, deze antwoorden zijn daarom niet meegenomen.

3.3 Lespraktijk

3.3.1 Effectieve instructie

Tabel 3.3.1a: Score voor mate van toepassing van effectieve instructie in de 3 subdomeinen (bo n=133; sbo n=82; so n=38)

Schalen	Onderwijssoort	Gemiddelde	P10	P90
Luisteren	bo	2,84	2,38	3,38
	sbo	2,85	2,38	3,49
	so	2,91	2,50	3,41
Spreken	bo	2,80	2,20	3,40
	sbo	2,71	2,00	3,20
	so	2,75	2,20	3,20
Gesprekken	bo	3,26	2,80	3,80
	sbo	3,26	2,80	3,60
	so	3,18	2,80	3,66

Tabel 3.3.1b: Output kruskal wallis test schaal effectieve instructie luisteren - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Stellingen luisteren	bo vs. sbo vs. so	1,620	2	0,445	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 3.3.1c: Output kruskal wallis test schaal effectieve instructie spreken - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Stellingen spreken	bo vs. sbo vs. so	1,780	2	0,411	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 3.3.1d: Output kruskal wallis test schaal effectieve instructie gesprekken - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Stellingen gesprekken	bo vs. sbo vs. so	1,440	2	0,487	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 3.3.1e: Spearman correlatie schalen effectieve instructie luisteren, spreken, gesprekken voor 3 onderwijssoorten

Schaal 1	Schaal 2	Onderwijs-type	Spearman's rho	p	n	Significantie
Spreken	Luisteren	bo	0,535	0,000	133	***
Gesprekken	Luisteren		0,420	0,000	133	***
	Spreken		0,414	0,000	133	***
Spreken	Luisteren	sbo	0,464	0,000	82	***
Gesprekken	Luisteren		0,332	0,002	82	**
	Spreken		0,460	0,000	82	***
Spreken	Luisteren	so	0,474	0,003	38	**
Gesprekken	Luisteren		0,528	0,001	38	***
	Spreken		0,356	0,028	38	*

3.3.2 Frequentie taken

Tabel 3.3.2a: Score voor frequentie oefenen taken in de afgelopen 12 weken (bo n=133; sbo n=82; so n=38)

Schalen	Onderwijssoort	Gemiddelde	P10	P90
Luisteren	bo	3,22	2,00	4,33
	sbo	3,26	2,00	4,67
	so	3,27	2,00	4,43
Spreken	bo	2,89	1,75	4,00
	sbo	2,99	2,00	4,22
	so	2,93	1,92	4,25
Gesprekken	bo	3,05	2,29	3,71
	sbo	2,98	2,29	3,57
	so	2,94	2,39	3,61

Tabel 3.3.2b: Output kruskal wallis test schaal frequentie taken luisteren - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Freq. luistertaken	bo vs. sbo vs. so	0,110	2	0,945	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 3.3.2c: Output kruskal wallis test schaal frequentie taken spreken - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Freq. spreektaken	bo vs. sbo vs. so	0,790	2	0,675	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 3.3.2d: Output kruskal wallis test schaal frequentie taken gesprekken - effect van onderwijssoort (3 groepen)

Schaal	Groepen	F	df	p	Significantie
Freq. gesprekstaken	bo vs. sbo vs. so	0,890	2	0,413	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 3.3.2e: Spearman correlatie schalen frequentie taken luisteren, spreken, gesprekken voor 3 onderwijssoorten

Schaal 1	Schaal 2	Onderwijs-type	Spearman's rho	p	n	Significantie
Spreken	Luisteren	bo	0,477	0,000	133	***
Gesprekken	Luisteren		0,412	0,000	133	***
	Spreken	sbo	0,583	0,000	133	***
Spreken	Luisteren		0,568	0,000	82	***
Gesprekken	Luisteren	so	0,419	0,000	82	***
	Spreken		0,566	0,000	82	***
Spreken	Luisteren	so	0,480	0,002	38	**
Gesprekken	Luisteren		0,443	0,005	38	**
	Spreken		0,550	0,000	38	***

3.4 Zicht op ontwikkeling en differentiatie

3.4.1 Bijhouden van leervorderingen

Tabel 3.4.1a: Volgen van vorderingen luisteren, spreken en gesprekken van leerlingen in groep 1 en 2 (bo n=97; sbo n=36)

Groep	Antwoordoptie	Onderwijssoort	Percentage
Spreken	structureel	bo	78,4
		sbo	80,6
	incidenteel	bo	11,3
		sbo	11,1
	geen	bo	9,3
		sbo	5,6
anders	bo	1,0	
	sbo	2,8	
Luisteren	structureel	bo	78,4
		sbo	77,8
	incidenteel	bo	9,3
		sbo	13,9
	geen	bo	11,3
		sbo	5,6
anders	bo	1,0	
	sbo	2,8	
Gesprekken	structureel	bo	66,0
		sbo	66,7
	incidenteel	bo	12,4
		sbo	19,4
	geen	bo	18,6
		sbo	11,1
anders	bo	3,1	
	sbo	2,8	

Noot: 6 sbo-scholleiders gaven aan geen groep 1 en 2 te hebben op hun school, deze antwoorden zijn daarom niet meegenomen.

3.4.2

Gerichte ondersteuning van leerlingen bij achterblijvende prestaties

Tabel 3.4.2a: Gerichte ondersteuning van leerlingen bij achterblijvende prestaties in groep 1 en 2 (bo n=97; sbo n=36)

Groep	Antwoordoptie	Onderwijssoort	Percentage
Spreken	structureel	bo	59,8
		sbo	80,6
	incidenteel	bo	38,1
		sbo	19,4
	geen	bo	2,1
sbo		0,0	
Luisteren	structureel	bo	51,5
		sbo	72,2
	incidenteel	bo	41,2
		sbo	22,2
	geen	bo	7,2
sbo		5,6	
Gesprekken	structureel	bo	48,5
		sbo	69,4
	incidenteel	bo	38,1
		sbo	27,8
	geen	bo	13,4
sbo		2,8	

Noot: 6 sbo-scholleiders gaven aan geen groep 1 en 2 te hebben op hun school, deze antwoorden zijn daarom niet meegenomen.

4 Houding en gedrag ten opzichte van taal

4.1 Leerlingkenmerken

4.1.1 *Thuis taal*

Tabel 4.1.1a: Andere talen die leerlingen thuis spreken wanneer ze thuis nooit, soms of meestal Nederlands spreken (meerdere antwoorden mogelijk) (bo n=1.011; sbo n=467; so n=189)

Antwoordoptie	Onderwijssoort	Percentage
Een andere taal (bijvoorbeeld Turks of Pools)	bo	77,2
	sbo	81,8
	so	83,6
Een dialect of streektaal van het Nederlands	bo	18,6
	sbo	17,8
	so	12,2
Fries	bo	6,9
	sbo	2,8
	so	6,3

4.1.2 *Taalactiviteiten thuis*

Tabel 4.1.2a: Score voor de mate waarin leerlingen hun ouders diverse taalactiviteiten zien ondernemen (bo n=2.783; sbo n=928; so n=378)

Schaal	Onderwijssoort	Gemiddelde	P10	P90
Taalactiviteiten ouders	bo	2,66	1,60	3,60
	sbo	2,29	1,00	3,40
	so	2,35	1,40	3,40

Tabel 4.1.2b: Output kruskal wallis test en posthoc schaal taalactiviteiten ouders - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Taalactiviteiten ouders	bo vs. sbo vs. so	176,440	2	0,000	***
	bo vs. sbo	148,030	1	0,000	***
	bo vs. so	53,900	1	0,000	***
	sbo vs. so	1,420	1	0,234	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 4.1.2c: Score voor de mate waarin leerlingen thuis praten over diverse onderwerpen (bo n=2.867; sbo n=1.012; so n=402)

Schaal	Onderwijssoort	Gemiddelde	P10	P90
Thuis praten over	bo	2,38	1,50	3,25
	sbo	2,20	1,00	3,25
	so	2,19	1,25	3,25

Tabel 4.1.2d: Output kruskal wallis test en posthoc schaal thuis praten over - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Thuis praten over	bo vs. sbo vs. so	65,880	2	0,000	***
	bo vs. sbo	49,150	1	0,000	***
	bo vs. so	27,920	1	0,000	***
	sbo vs. so	0,090	1	0,764	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

4.1.3 *Luisteren en spreken in de klas*

De schaal luisterklimaat is omgedraaid ten opzichte van het TR van het consortium (Gijsel et al., 2025). Een hoge score betekent een goed luisterklimaat.

Tabel 4.1.3a: Scores voor door leerlingen ervaren luisterklimaat en spreekvrijheid in de klas (bo n=2.779-2.847; sbo n=981-986; so n=390-395)

Schalen	Onderwijssoort	Gemiddelde	P10	P90
Luisterklimaat	bo	2,73	2,00	3,60
	sbo	2,67	1,75	3,80
	so	2,53	1,60	3,60
Spreekvrijheid	bo	3,07	2,40	3,60
	sbo	2,99	2,33	3,60
	so	2,93	2,25	3,60

Tabel 4.1.3b: Output kruskal wallis test en posthoc schaal luisterklimaat - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Luisterklimaat	bo vs. sbo vs. so	29,910	2	0,000	***
	bo vs. sbo	6,180	1	0,013	*
	bo vs. so	27,930	1	0,000	***
	sbo vs. so	8,920	1	0,003	**

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 4.1.3c: Output kruskal wallis test en posthoc schaal spreekvrijheid - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Spreekvrijheid	bo vs. sbo vs. so	38,500	2	0,000	***
	bo vs. sbo	20,350	1	0,000	***
	bo vs. so	25,070	1	0,000	***
	sbo vs. so	2,790	1	0,095	

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 4.1.3d: Spearman correlatie schalen luisterklimaat en spreekvrijheid

Onderwijssoort	Spearman's rho	p	n	Significantie
bo	0,288	0,000	2.748	***
sbo	0,297	0,000	948	***
so	0,345	0,000	380	***

4.2 Leerkrachtkenmerk: ervaren bekwaamheid

Tabel 4.2a: Score voor de door leerkrachten ervaren bekwaamheid op het gebied van onderwijs in mondelinge taalvaardigheid (bo n=133; sbo n=82; so n=38)

Schaal	Onderwijssoort	Gemiddelde	P10	P90
Ervaren bekwaamheid	bo	2,86	2,20	3,40
	sbo	2,89	2,22	3,58
	so	2,85	2,60	3,20

Tabel 4.2b: Output kruskal wallis test schaal ervaren bekwaamheid - effect van onderwijssoort (3 groepen)

Schaal	Groepen	Chi-kwadraat	df	p	Significantie
Ervaren bekwaamheid	bo vs. sbo vs. so	0,290	2	0,865	

5 Prestaties Mondelinge Taalvaardigheid

In dit hoofdstuk presenteren we de tabellen met gegevens die ten grondslag liggen aan hoofdstuk 3 van het publieksrapport. In dat hoofdstuk beschrijven we het vaardigheidsniveau van leerlingen op het gebied van mondelinge taalvaardigheid.

5.1 Databronnen

Het domein mondelinge taalvaardigheid wordt in het Referentiekader Taal onderverdeeld in 3 subdomeinen: luisteren, spreken en gesprekken voeren. In deze peiling zijn opgaven ontwikkeld om de vaardigheid van leerlingen op deze 3 subdomeinen in kaart te brengen.

5.1.1 *Luistervaardigheid*

De luistervaardigheid is in de eerste plaats gemeten met een luistertoets. Leerlingen kregen audio- en videofragmenten voorgelegd. Bij deze fragmenten werden enkele vragen gesteld, die de leerlingen hebben beantwoord. Er waren 4 verschillende toetsversies. Elke toetsversie had een unieke samenstelling van fragmenten. Fragmenten kwamen in meerdere toetsversies voor.

Bij de luistertoets maakte elke leerling 1 van de 4 toetsversies. Om de prestaties van leerlingen te vergelijken, zijn de scores van leerlingen door middel van itemresponsetheorie op 1 vaardigheidsschaal gezet. De vaardigheidsscores zijn vervolgens vertaald naar een bankscore (Gijssel et al., 2025). Deze bankscore kan geïnterpreteerd worden als een score die uitdrukt welke totaal score een leerling zou hebben behaald als deze leerling *alle* luisteritems (en niet alleen de items in de gemaakte toetsversie) zou hebben gemaakt. De bankscores zijn vervolgens gebruikt in de rapportage in het publieksrapport, uitgedrukt in percentage van de maximaal te behalen score.

De toetsitems in de luistertoets verschilden op verschillende punten. De fragmenten bestonden in de eerste plaats uit audio- en videofragmenten. Ten tweede verschilden de fragmenten in de gespreksvorm die in het fragment gehanteerd werd (een monoloog of een dialoog). Ten derde was de moeilijkheid van de tekst die in het fragment gesproken werd (uitgedrukt in referentieniveaus: <1F, 1F of 2F) verschillend. Daarnaast waren er verschillen tussen de toetsitems in het onderdeel van de luistervaardigheid wat zij beoogden te meten: begrijpen, interpreteren of samenvatten.

Naast de luistertoets is de luistervaardigheid ook gemeten met een beoordeling in een praktijksituatie. Bij aanvang van de gesprekstaak *Samenwerkingsspel* kregen leerlingen een instructie van de toetsleider. De toetsleider stelde op verschillende momenten vragen over de instructie aan de leerlingen om onder meer te beoordelen of de leerlingen goed naar de instructie hadden geluisterd. De toetsleider beoordeelde de leerlingen op de onderstaande 4 beoordelingsaspecten, waar een leerling telkens 0 (onvoldoende), 1 (voldoende) of 2 (goed) punten kon scoren.

- Begrijpen: toepassen van de instructie.
- Begrijpen: herhalen van de instructie.
- Samenvatten: mondeling samenvatten.
- Interpreteren: meningen herkennen.

5.1.2 *Spreekvaardigheid*

In deze peiling is de spreekvaardigheid van leerlingen door middel van 2 taken in kaart gebracht: de *Naverteltaak* en het *Betoog*. Een deel van de leerlingen nam deel aan slechts 1 spreektaak. Sommige leerlingen maakten beide taken en andere namen deel aan geen van beide spreektaken. Om de prestaties van leerlingen op het gebied van spreekvaardigheid toch te kunnen vergelijken, zijn de scores van leerlingen die aan ten minste 1 spreektaak hebben deelgenomen op een vaardigheidsschaal voor spreekvaardigheid geplaatst. Net als bij de luistervaardigheid zijn de resulterende vaardigheidsscores vervolgens omgezet in een bankscore (Gijsel et al., 2025), als percentage. Dit percentage drukt uit hoeveel procent van het totaal aantal te behalen punten een leerling zou hebben behaald als deze leerling aan beide spreektaken had deelgenomen en op alle beoordelingsaspecten beoordeeld zou zijn.

Bij de *Naverteltaak* kregen leerlingen een fragment uit het Jeugdjournaal te zien. Vervolgens werd aan de leerlingen gevraagd om na te vertellen wat zij zojuist in het videofragment gezien hadden. Een toetsleider beoordeelde de leerlingen daarbij op 15 beoordelingsaspecten. Elk beoordelingsaspect was gekoppeld aan 1 kenmerk van de taakuitvoering:

- Uitvoering
- Samenhang
- Afstemming op het doel
- Afstemming op de gesprekspartner
- Woordgebruik en woordenschat
- Vloeiendheid, verstaanbaarheid en grammaticale beheersing

In de spreektaak *Betoog* kregen leerlingen de vraag om een plan te maken voor een fictief schoolreisje. Vervolgens werd aan de leerlingen gevraagd om een (virtuele) leerkracht te overtuigen van hun plan voor het schoolreisje. De toetsleider beoordeelde de leerling op deze spreektaak op dezelfde beoordelingsaspecten die bij de naverteltaak gebruikt zijn. Daarnaast is er een aanvullend beoordelingsaspect gebruikt dat de kwaliteit van de argumentatie beoordeelde, als onderdeel van het kenmerk van de taakuitvoering "afstemming op doel".

In de vorige peiling naar mondelinge taalvaardigheid in het so en sbo hadden de verschillende beoordelingsaspecten een andere weging. Om de beoordelingsaspecten in deze peiling op dezelfde wijze mee te nemen in de vaardigheidsscore, is er besloten om enkele beoordelingsaspecten zwaarder te laten wegen (Gijsel et al., 2025).

5.1.3 *Gespreksvaardigheid*

Om de gespreksvaardigheid van leerlingen te meten, zijn er in deze peiling 2 gesprekstaken ingezet: het *Samenwerkingsspel* en het *Raadsel*. Veel leerlingen namen slechts aan 1 van beide gesprekstaken deel. Daarnaast was er een groep leerlingen die beide gesprekstaken uitvoerde. Sommige leerlingen deden geen van beide gesprekstaken. Om de gespreksvaardigheid van leerlingen te analyseren, zijn de scores van leerlingen op 1 vaardigheidsschaal voor gespreksvaardigheid geplaatst. Op een vergelijkbare wijze als bij de subdomeinen luisteren en spreken, zijn de vaardigheidsscores voor gesprekken voeren ook vertaald in een bankscore (Gijsel et al., 2025). Deze bankscore drukt uit hoeveel procent van het totaal aantal te behalen punten een leerling zou hebben behaald als deze leerling aan beide gesprekstaken zou hebben deelgenomen.

Bij het *Samenwerkingsspel* moesten leerlingen samen een bordspel spelen. Hierbij konden leerlingen individueel niet winnen of verliezen. Bij de uitvoering van het spel moesten de leerlingen overleggen over de strategie om het doel van het spel te behalen. Hierbij beoordeelde de toetsleider de leerlingen op 17 beoordelingsaspecten. Elk beoordelingsaspect had betrekking op 1 kenmerk van de taakuitvoering:

- Uitvoering
- Woord nemen en bijdragen aan samenhang
- Afstemming op doel
- Afstemming op gesprekspartners
- Woordgebruik en woordenschat
- Vloeiendheid, verstaanbaarheid, en grammaticale beheersing

In de gesprekstaak *Raadsel* kregen de leerlingen een situatie voorgelegd van een jongen die in een bepaalde situatie was beland. Als groep moesten de leerlingen een logische verklaring vinden voor het scenario. Elke leerling kreeg hierbij een tekst en een fysiek object als hint. De toetsleider beoordeelde de gespreksvaardigheid van de leerlingen door dezelfde 17 beoordelingsaspecten te beoordelen die bij het *Samenwerkingsspel* zijn gebruikt.

Ook bij de gesprekstakingen is een afwijkend gewicht aan de beoordelingsaspecten toegekend, zodat de weging van de beoordelingsaspecten vergelijkbaar is met de weging zoals die in de vorige peiling voor mondelinge taalvaardigheid in het sbo en so is gehanteerd. Dit komt er op neer dat de technische beoordelingsaspecten minder zwaar meewegen dan de inhoudelijke beoordelingsaspecten (Gijssels et al., 2025).

5.2 Prestaties luisteren, spreken en gesprekken voeren

In het publieksrapport wordt per subdomein het percentage leerlingen dat een gegeven referentieniveau haalt gevisualiseerd. Hierbij wordt een uitsplitsing gemaakt naar onderwijssoort. In Tabel 5.2a worden de percentages die aan dit figuur ten grondslag liggen in tabelvorm weergegeven. Vanwege afrondingen kan het voorkomen dat de percentages niet volledig optellen tot 100 procent.

Tabel 5.2a: Percentage leerlingen wat een bepaald referentieniveau haalt per onderwijssoort en subdomein

onderwijssoort	subdomein	<1F	1F	2F	totaal
bo	Luisteren	8,0	47,7	44,3	1.258
	Spreken	18,4	55,9	25,7	1.446
	Gesprekken	31,3	24,5	44,2	1.463
sbo	Luisteren	58,7	30,0	11,3	373
	Spreken	46,5	45,1	8,5	508
	Gesprekken	62,1	17,3	20,6	446
so	Luisteren	40,0	50,0	10,0	120
	Spreken	42,4	43,5	14,1	184
	Gesprekken	65,2	12,6	22,2	230

5.3 Prestaties luistervaardigheid

5.3.1 Luistertoets

In Tabel 5.3.1a worden beschrijvende statistieken bij de verdeling van de vaardigheidsscores op het subdomein luisteren gerapporteerd. In deze tabel wordt vermeld bij hoeveel leerlingen er een vaardigheidsscore berekend is. Ook is uit deze tabel op te maken wat de laagste, hoogste en gemiddelde bankscore

van de leerlingen was. De spreiding van de bankscores is tot uitdrukking gebracht in de standaarddeviatie.

Tabel 5.3.1a: Verdeling van de scores (bankscore) op de luistertoets

N	Minimum	Maximum	Gemiddelde	Standaarddeviatie
1751	8,66	98,61	66,24	18,92

In het subdomein luisteren wordt er onderscheid gemaakt tussen 3 kenmerken van de taakuitvoering: begrijpen, interpreteren en samenvatten. In Tabel 5.3.1b wordt per kenmerk van de taakuitvoering gerapporteerd wat de gemiddelde proportie beheersing van de leerlingen is op de items die bij het betreffende kenmerk hoorden. Deze gemiddelde proporties zijn uitgesplitst naar de verschillende referentieniveaugroepen.

Tabel 5.3.1b: Gemiddelde proportie beheersing per kenmerk (begrijpen, interpreteren en samenvatten) per referentieniveaugroep

kenmerk	<1F	1F	2F
Begrijpen	0,34	0,61	0,83
Interpreteren	0,39	0,68	0,87
Samenvatten	0,50	0,80	0,94

Naast dat de items van de luistertoets betrekking hadden op verschillende kenmerken van de taakuitvoering, verschilden de luisteritems ook op andere aspecten. Zo verschilden de items in type (kritisch kijken en luisteren, kijken en luisteren, kritisch luisteren, en luisteren), de in het fragment gehanteerde gespreksvorm (dialoog of monoloog) en het niveau van de gesproken tekst (uitgedrukt in referentieniveaus). Ook voor deze verschillen tussen items is de beheersing van leerlingen in kaart gebracht door de gemiddelde bankscore per kenmerk per referentieniveaugroep uit te rekenen. Tabel 5.3.1c toont de gemiddelde proportie beheersing per type fragment. Tabel 5.3.1d toont de gemiddelde proportie beheersing, uitgesplitst naar de verschillende gespreksvormen. In Tabel 5.3.1e wordt de proportie beheersing van leerlingen weergegeven voor verschillende niveaus van de tekst die in het fragment wordt gesproken.

Tabel 5.3.1c: Gemiddelde proportie beheersing per type fragment ((kritisch) kijken en/of luisteren) per referentieniveaugroep

kenmerk	<1F	1F	2F
kijken & luisteren	0,42	0,70	0,89
kritisch kijken & luisteren	0,45	0,76	0,92
kritisch luisteren	0,18	0,43	0,70
luisteren	0,27	0,53	0,77

Tabel 5.3.1d: Gemiddelde proportie beheersing per gespreksvorm van het fragment (dialogo of monoloog) per referentieniveaugroep

kenmerk	<1F	1F	2F
dialogo	0,39	0,67	0,86
monoloog	0,36	0,64	0,85

Tabel 5.3.1e: Gemiddelde proportie beheersing per tekstniveau van het fragment (uitgedrukt in referentieniveau) per referentieniveaugroep

kenmerk	<1F	1F	2F
1F	0,40	0,69	0,88
1F/2F	0,30	0,60	0,84
2F	0,36	0,62	0,83

Voor elk van de kenmerken van de luisteritems is het verschil in beheersing getoetst. Hierbij is gebruik gemaakt van een reeks ANOVA's voor herhaalde meting met als *within-subject* variabele het betreffende kenmerk. Interacties met de referentieniveaugroep zijn getoetst door aan de ANOVA's als *between-subject* variabele de referentieniveaugroep toe te voegen. Het resultaat is een mixed ANOVA-model per kenmerk.

Onderstaande tabellen tonen per kenmerk de resultaten van het mixed ANOVA-model. Tabel 5.3.1f toont het model voor de kenmerken van de taakuitvoering, Tabel 5.3.1g het model voor de gespreksvorm die in het betreffende fragment gebruikt werd (dialogo of monoloog), Tabel 5.3.1h het model voor niveau van de gesproken tekst uitgedrukt in referentieniveaus, en Tabel 5.3.1i het model voor de verschillen tussen audio- en videofragmenten.

Tabel 5.3.1f: ANOVA-resultaten voor taakkenmerken (begrijpen, interpreteren, samenvatten) en bankscore

Effect	DF _n	DF _d	SS _n	SS _d	F	p	Sig
(Intercept)	1	1.748	2.096,355	28,249	129.718,858	0,000	***
referentieniveau	2	1.748	150,463	28,249	4.655,193	0,000	***
kenmerk	2	3.496	19,161	0,751	44.609,576	0,000	***
referentieniveau:kenmerk	4	3.496	1,054	0,751	1.227,069	0,000	***

Significantie: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 5.3.1g: ANOVA-resultaten voor fragmentkenmerk (dialogo of monoloog) en bankscore

Effect	DF _n	DF _d	SS _n	SS _d	F	p	Sig
(Intercept)	1	1.748	1.259,009	19,092	115.269,939	0,000	***
referentieniveau	2	1.748	105,600	19,092	4.834,164	0,000	***
fragmentkenmerk_dialogo_monoloog	1	1.748	0,449	0,017	45.487,083	0,000	***
referentieniveau:fragmentkenmerk_dialogo_monoloog	2	1.748	0,025	0,017	1.244,173	0,000	***

Significantie: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 5.3.1h: ANOVA-resultaten voor referentieniveau van de tekst (<1F, 1F en 2F) en bankscore

Effect	DF _n	DF _d	SS _n	SS _d	F	p	Sig
(Intercept)	1	1.748	1.799,353	29,762	105.682,465	0,000	***
referentieniveau	2	1.748	168,680	29,762	4.953,579	0,000	***
tekstkenmerk_referentieniveau	2	3.496	5,515	0,257	37.519,028	0,000	***
referentieniveau:tekstkenmerk_refnivea	4	3.496	0,788	0,257	2.680,631	0,000	***

Significantie: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabel 5.3.1i: ANOVA-resultaten voor fragmentkenmerk ((kritisch) kijken/luisteren) en bankscore

Effect	DF _n	DF _d	SS _n	SS _d	F	p	Sig
(Intercept)	1	1.748	2.176,317	39,793	95.599,203	0,000	***
referentieniveau	2	1.748	223,367	39,793	4.905,927	0,000	***
fragmentkenmerk_kijk_luist	3	5.244	75,199	2,421	54.299,798	0,000	***
referentieniveau:fragmentkenmerk_kijk_luist	6	5.244	2,806	2,421	1.012,951	0,000	***

Significantie: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Per item in de luistertoets wordt in Tabel 5.3.1j een indicatie van de moeilijkheid van het item gegeven. De p-waarde beschrijft de proportie leerlingen die de vraag goed heeft beantwoord. Dus hoe hoger de p-waarde, hoe eenvoudiger het item. Deze waarde is berekend voor de verschillende referentieniveaugroepen. Hiermee is inzichtelijk gemaakt welke items relatief beter beheerst worden naargelang het vaardigheidsniveau van de leerling toeneemt. Het verschil in p-waarde, alsmede de effectgrootte (Cohen's d) tussen referentieniveaugroepen wordt hier ook weergegeven.

Tabel 5.3.1j: P-waarde per item in de luistertoets

kenmerk	item	p-waarde			<1F - 1F		1F - 2F	
		<1F	1F	2F	verschil	Cohen's d	verschil	Cohen's d
Begrijpen	12309	0,08	0,29	0,53	0,20	0,47	0,24	0,51
	128VR03	0,41	0,65	0,80	0,24	0,50	0,16	0,36
	128VR03_so	0,46	0,85	1,00	0,39	0,92	0,15	0,46
	13158	0,18	0,37	0,73	0,19	0,40	0,36	0,78
	13164	0,27	0,68	0,92	0,41	0,89	0,23	0,60
	13165	0,52	0,78	0,94	0,27	0,62	0,16	0,47
	138VR04	0,55	0,88	0,97	0,33	0,74	0,09	0,31
	138VR05	0,47	0,69	0,91	0,21	0,44	0,22	0,54
	138VR07	0,25	0,55	0,86	0,31	0,67	0,31	0,69
	143VR05	0,67	0,86	0,98	0,19	0,47	0,12	0,43
	755	0,72	0,85	0,98	0,13	0,35	0,12	0,44
	756	0,36	0,60	0,90	0,24	0,48	0,30	0,72
	758	0,06	0,44	0,69	0,38	0,80	0,24	0,51
	760	0,17	0,61	0,87	0,45	0,94	0,26	0,60
	PO.NeLu.1F.v1.2b	0,11	0,27	0,58	0,17	0,39	0,31	0,66
	PO.NeLu.1F.v5.3b	0,18	0,36	0,74	0,18	0,38	0,38	0,81

	PO.NeLu. 2F.a2.1b	0,38	0,81	0,96	0,44	1,08	0,14	0,44	
	PO.NeLu. 2F.a2.2b	0,19	0,27	0,45	0,08	0,19	0,18	0,39	
	PO.NeLu. 2F.v4.2b	0,50	0,87	1,00	0,37	0,87	0,13	0,50	
	PO.NeLu. 2F.v5.3b	0,62	0,89	0,98	0,27	0,77	0,09	0,35	
Interpreteren	12304	0,21	0,48	0,79	0,27	0,56	0,31	0,69	
	12305	0,12	0,17	0,54	0,05	0,13	0,36	0,82	
	12306	0,46	0,51	0,88	0,05	0,10	0,37	0,88	
	128VR02	0,36	0,70	0,92	0,33	0,72	0,22	0,57	
	128VR04	0,55	0,93	0,99	0,37	1,17	0,07	0,33	
	13156	0,17	0,40	0,79	0,22	0,47	0,39	0,87	
	13160	0,17	0,57	0,88	0,40	0,83	0,31	0,74	
	13473	0,75	0,91	0,95	0,16	0,54	0,03	0,12	
	138VR01	0,32	0,66	0,95	0,34	0,72	0,30	0,74	
	138VR06	0,53	0,91	0,95	0,37	0,86	0,05	0,18	
	143VR04	0,45	0,82	0,97	0,36	0,82	0,15	0,46	
	143VR06	0,30	0,66	0,93	0,36	0,77	0,27	0,66	
	688	0,46	0,84	0,98	0,38	0,96	0,13	0,46	
	689	0,53	0,87	0,97	0,34	0,91	0,10	0,36	
	691	0,39	0,66	0,85	0,27	0,57	0,19	0,45	
	762	0,27	0,60	0,91	0,33	0,68	0,32	0,78	
	898	0,37	0,72	0,88	0,35	0,77	0,15	0,38	
		PO.NeLu. 1F.v1.1i	0,29	0,45	0,81	0,16	0,32	0,36	0,80
		PO.NeLu. 1F.v1.5i	0,40	0,76	0,87	0,35	0,80	0,11	0,28
		PO.NeLu. 1F.v5.1i	0,41	0,74	0,92	0,33	0,75	0,17	0,47
	PO.NeLu. 1F.v5.2i	0,80	0,93	0,98	0,13	0,48	0,05	0,23	
	PO.NeLu. 2F.v4.1i	0,70	0,92	1,00	0,23	0,61	0,08	0,37	
	PO.NeLu. 2F.v5.4i	0,40	0,65	0,94	0,26	0,54	0,28	0,73	
Samenvatten	143VR01	0,51	0,79	0,97	0,28	0,60	0,18	0,53	

In het publieksrapport wordt in hoofdstuk 3 de luistervaardigheidsschaal nader toegelicht door de positie van enkele opgaven uit de luistertoets op de vaardigheidsschaal te projecteren.

Door middel van itemresponstheorie is voor leerlingen die de luistertoets gemaakt hebben een vaardigheidsscore berekend. Omdat in deze peiling ook items gebruikt worden die ook in de vorige peiling naar mondelinge taalvaardigheid zijn afgenomen, was het mogelijk om de prestatiestandaarden, die in de vorige peiling bepaald zijn, over te zetten naar de huidige peiling (Gijssels et al., 2025). Daarom is ook uit te drukken welke vaardigheidsscore een leerling had moeten behalen om het referentieniveau 1F of 2F te beheersen.

Voor de selectie van items voor figuur 3.3d in het publieksrapport is gekeken naar de toetsitems die de grenswaarden voor 1F en 2F het dichtst benaderden. Om de volle breedte van de vaardigheidsschaal in kaart te brengen, is ook een moeilijk en makkelijk item geselecteerd. Hierbij is erop gelet dat de in het fragment gesproken tekst ook duidelijk te transcriberen moest zijn voor het rapport.

5.3.2 *Luisteren in de gesprekstaak*

Naast de luistertoets is de luistervaardigheid van de leerlingen ook in kaart gebracht door middel van het beoordelen van het luisteren van de leerlingen in een praktijksituatie. De verdeling van de ruwe totaalscores die de leerlingen op dit onderdeel behaalden wordt weergegeven in Tabel 5.3.2a. Hierbij wordt een uitsplitsing gemaakt naar onderwijssoort.

Tabel 5.3.2a: Verdeling behaalde scores voor het onderdeel luisteren in praktijksituatie, uitgesplitst naar onderwijssoort

Onderwijssoort	0	1	2	3	4	5	6	7	8
bo	8	2	11	21	39	57	139	185	372
sbo	5	6	13	21	23	35	37	53	58
so	0	2	7	9	19	14	28	18	25

5.4 **Spreekvaardigheid**

In Tabel 5.4a wordt de verdeling van de bankscore voor de spreekvaardigheid tot uitdrukking gebracht middels het minimum, maximum en gemiddelde percentage van de bankscore van de leerlingen. De spreiding is weergegeven door middel van de standaarddeviatie.

Tabel 5.4a: Verdeling van de scores (bankscore) op het onderdeel spreken

N	Minimum	Maximum	Gemiddelde	Standaarddeviatie
2138	0,6	99,1	71,9	16,8

De spreekvaardigheid is gemeten door beoordelingsaspecten die betrekking hadden op meerdere kenmerken van de taakuitvoering. In Tabel 5.4b wordt de gemiddelde score gerapporteerd die leerlingen zouden halen als zij op alle beoordelingsaspecten behorend bij het betreffende kenmerk van de taakuitvoering zouden zijn beoordeeld. Dit is uitgesplitst naar het referentieniveau dat de leerlingen behaalden.

Tabel 5.4b: Gemiddelde bankscore per referentieniveaugroep voor het onderdeel spreken

kenmerk	<1F	1F	2F
Afstemming op doel	0,44	0,76	0,96
Afstemming op gesprekspartner	0,43	0,65	0,84
Samenhang	0,45	0,77	0,93
Uitvoering	0,40	0,78	0,96
Vloeiendheid, verstaanbaarheid en grammaticale beheersing	0,78	0,89	0,95
Woordgebruik en woordenschat	0,64	0,79	0,89

Ook op het niveau van het individuele beoordelingsaspect zijn uitspraken te doen over de scores van leerlingen. Tabel 5.4c rapporteert per beoordelingsaspect voor het onderdeel spreken de gemiddelde score die leerlingen op verschillende referentieniveaus behaalden. In deze tabel is ook opgenomen hoe groot het verschil is tussen aaneengrenzende referentieniveaugroepen. Hierbij moet rekening gehouden worden met het feit dat sommige items op een schaal van 1-2-3 zijn gescoord, daar waar andere een score op de schaal 0-3-6 hebben. Ook wordt de grootte van het effect tussen 2 referentieniveaugroepen uitgedrukt in Cohen's *d*.

Tabel 5.4c: Gemiddelde scores per beoordelingsaspect, met verschillen in beheersing tussen referentieniveaugroepen uitgelicht.

kenmerk	beoordelingsaspect	gemiddelde score			<1F - 1F		1F - 2F	
		<1F	1F	2F	verschil	Cohen's d	verschil	Cohen's d
Afstemming op doel	SP1_05	2,61	4,94	5,91	2,33	1,57	0,97	0,76
	SP2_05	1,75	3,88	5,74	2,13	1,39	1,87	1,40
	SP2_06	3,58	4,89	5,80	1,31	0,87	0,91	0,71
Afstemming op gesprekspartner	SP1_06	5,45	5,92	5,99	0,48	0,60	0,06	0,15
	SP1_07	1,30	2,53	4,40	1,23	0,75	1,87	1,13
	SP1_08	2,34	2,96	4,09	0,62	0,50	1,13	0,91
	SP2_07	5,13	5,85	5,93	0,72	0,69	0,08	0,13
	SP2_08	1,05	2,58	4,90	1,53	0,93	2,32	1,38
	SP2_09	2,06	2,92	4,34	0,86	0,70	1,42	1,12
Samenhang	SP1_03	1,93	3,87	5,42	1,95	1,22	1,54	1,02
	SP1_04	3,84	5,71	5,97	1,87	1,33	0,27	0,34
	SP2_03	1,68	3,31	4,94	1,63	1,14	1,63	1,19
	SP2_04	3,78	5,62	5,90	1,84	1,25	0,28	0,31
Uitvoering	SP1_01	2,73	5,00	5,86	2,27	1,37	0,86	0,66
	SP1_02	2,10	4,31	5,87	2,21	1,41	1,57	1,16
	SP2_01	3,02	5,26	5,87	2,24	1,38	0,60	0,51
	SP2_02	1,97	4,17	5,81	2,20	1,42	1,64	1,22
Vloeiendheid, verstaanbaarheid en grammaticale beheersing	SP1_11	1,12	1,60	1,83	0,48	0,81	0,23	0,48
	SP1_12	1,69	1,86	1,97	0,16	0,40	0,11	0,35
	SP1_13	1,59	1,93	1,98	0,34	0,86	0,05	0,21
	SP1_14	1,66	1,93	1,97	0,27	0,70	0,05	0,20
	SP1_15	1,27	1,86	1,97	0,60	1,29	0,11	0,35
	SP2_12	1,25	1,64	1,77	0,39	0,67	0,14	0,28
	SP2_13	1,61	1,82	1,96	0,21	0,45	0,14	0,39
	SP2_14	1,65	1,93	1,96	0,27	0,70	0,03	0,12
	SP2_15	1,70	1,92	1,96	0,21	0,56	0,05	0,18
Woordgebruik en woordenschat	SP1_09	0,87	1,57	1,85	0,71	1,27	0,28	0,58
	SP1_10	1,26	1,73	1,94	0,46	0,91	0,21	0,53
	SP2_10	0,84	1,54	1,91	0,70	1,28	0,37	0,81
	SP2_11	1,41	1,84	1,93	0,42	0,92	0,09	0,27

Om verschillen in vaardigheid tussen de kenmerken van de taakuitvoering te toetsen, is een repeated-measures ANOVA uitgevoerd. Het kenmerk van de taakuitvoering is daarin als within-subjects factor en de referentieniveaugroep als between-subjects factor meegenomen. De resultaten van deze toetsing zijn opgenomen in Tabel 5.4d.

Tabel 5.4d: ANOVA-resultaten voor kenmerk van de taakuitvoering en bankscore

Effect	DF _n	DF _d	F	p	Sig
referentieniveau	2	2.135	3.564	0,000	***
kenmerk	5	10.675	11.511	0,000	***
referentieniveau:kenmerk	10	10.675	3.444	0,000	***

Significantie: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Om te illustreren wat de spreekvaardigheden van leerlingen op verschillende referentieniveaus zijn, zijn in hoofdstuk 3 enkele citaten uit de spreektaken opgenomen. Hierbij is gebruik gemaakt van fragmenten die voor de dubbele beoordeling zijn opgenomen. Voor deze citaten zijn leerlingen geselecteerd die binnen de uitgelichte referentieniveaugroep dicht bij de gemiddelde score op het betreffende kenmerk van de taakuitvoering presteerden.

5.5 Gespreksvaardigheid

Voor de leerlingen die aan ten minste 1 gesprekstaak hadden deelgenomen is aan de hand van hun vaardigheidsscore een bankscore berekend. De minimum, maximum en gemiddelde bankscore voor de gespreksvaardigheid, uitgedrukt in een percentage van de maximale bankscore (percentage beheersing), zijn in Tabel 5.5a gepresenteerd, samen met de standaarddeviatie en het aantal leerlingen voor wie een bankscore berekend is.

Tabel 5.5a: Verdeling van de scores (percentage beheersing) op het onderdeel gesprekken voeren

N	Minimum	Maximum	Gemiddelde	Standaarddeviatie
2139	0,9	99,2	79,2	17,5

Evenals bij de gesprekstaken, is ook voor de gespreksvaardigheid een uitsplitsing te maken naar kenmerk van de taakuitvoering. Tabel 5.5b toont, uitgesplitst naar referentieniveaugroep, de gemiddeld proportie beheersing op elk kenmerk van de taakuitvoering.

Tabel 5.5b: Gemiddelde proportie beheersing per referentieniveaugroep voor het onderdeel gesprekken voeren

kenmerk	<1F	1F	2F
Afstemming op doel	0,50	0,78	0,93
Afstemming op gesprekspartners	0,63	0,83	0,93
Kwaliteit van de uitvoering	0,50	0,78	0,93
Vloeiendheid, verstaanbaarheid en grammaticale beheersing	0,84	0,92	0,95
Woord nemen en bijdragen aan samenhang	0,60	0,87	0,96
Woordgebruik en woordenschat	0,82	0,91	0,95

De kenmerken van de taakuitvoering zijn gemeten door middel van meerdere beoordelingsaspecten. De gemiddelde score per beoordelingsaspect zijn weergegeven in Tabel 5.5c. Hier wordt per beoordelingsaspect en per referentieniveaugroep de gemiddelde bankscore weergegeven. Tussen verschillende referentieniveaugroepen wordt ook het verschil in gemiddelde score gerapporteerd, samen met de effectgrootte (Cohen's *d*). Hierbij moet rekening gehouden worden met het feit dat sommige beoordelingsaspecten op een schaal van 1-2-3 zijn gescoord, terwijl andere een score op de schaal 0-3-6 hebben.

Tabel 5.5c: Gemiddelde scores per beoordelingsaspect, met verschillen in beheersing tussen referentieniveaugroepen uitgelicht.

kenmerk	Beoordelings-aspect	gemiddelde score			<1F - 1F		1F - 2F	
		<1F	1F	2F	verschil	Cohen's d	verschil	Cohen's d
Afstemming op doel	GE1_05	2,72	4,25	5,47	1,53	0,98	1,23	0,96
	GE1_06	2,99	5,07	5,82	2,08	1,15	0,75	0,72
	GE2_05	2,79	4,60	5,68	1,81	1,03	1,08	0,89
	GE2_06	3,49	5,26	5,84	1,77	1,01	0,58	0,59
Afstemming op gesprekspartners	GE1_07	4,95	5,78	5,94	0,83	0,60	0,16	0,28
	GE1_08	2,90	3,74	4,91	0,84	0,58	1,17	0,79
	GE1_09	3,62	5,34	5,87	1,72	1,00	0,53	0,60
	GE1_10	4,53	5,58	5,88	1,05	0,61	0,30	0,38
	GE2_07	5,11	5,71	5,92	0,61	0,48	0,21	0,31
	GE2_08	2,63	3,19	4,17	0,57	0,41	0,97	0,66
	GE2_09	3,55	4,82	5,62	1,27	0,72	0,80	0,64
	GE2_10	4,65	5,47	5,88	0,82	0,51	0,41	0,46
Kwaliteit van de uitvoering	GE1_01	2,99	4,95	5,82	1,96	1,21	0,87	0,84
	GE2_01	3,10	4,75	5,75	1,65	0,96	1,00	0,88
Vloeiendheid, verstaanbaarheid en grammaticale beheersing	GE1_13	1,75	1,87	1,96	0,13	0,29	0,08	0,33
	GE1_14	1,81	1,89	1,96	0,08	0,22	0,06	0,25
	GE1_15	1,83	1,97	1,98	0,14	0,41	0,01	0,06
	GE1_16	1,86	1,96	1,99	0,10	0,32	0,02	0,16
	GE1_17	1,68	1,94	1,99	0,26	0,61	0,05	0,28
	GE2_13	1,71	1,87	1,92	0,16	0,36	0,05	0,17
	GE2_14	1,73	1,85	1,90	0,12	0,28	0,05	0,16
	GE2_15	1,71	1,88	1,95	0,17	0,37	0,06	0,23
	GE2_16	1,77	1,90	1,95	0,14	0,33	0,05	0,19
Woord nemen en bijdragen aan samenhang	GE1_02	2,91	4,71	5,72	1,80	0,99	1,01	0,87
	GE1_03	4,69	5,45	5,79	0,76	0,45	0,33	0,35
	GE1_04	4,15	5,58	5,94	1,43	0,84	0,36	0,50
	GE2_02	3,13	4,93	5,78	1,80	0,92	0,85	0,73
	GE2_03	4,92	5,40	5,78	0,49	0,32	0,38	0,39
	GE2_04	3,93	5,23	5,77	1,30	0,81	0,54	0,52
Woordgebruik en woordenschat	GE1_11	1,60	1,87	1,98	0,27	0,55	0,10	0,43
	GE1_12	1,84	1,95	1,99	0,11	0,32	0,04	0,26
	GE2_11	1,64	1,88	1,96	0,23	0,50	0,08	0,33
	GE2_12	1,82	1,95	1,97	0,12	0,34	0,02	0,11

De verschillen in vaardigheid tussen de kenmerken van de taakuitvoering zijn statistisch getoetst met een repeated-measures ANOVA. Het kenmerk van de taakuitvoering is daarin als within-subjects factor en de referentieniveaugroep als between-subjects factor meegenomen. De resultaten van deze toetsing zijn gepresenteerd in Tabel 5.5d.

Tabel 5.5d: ANOVA-resultaten voor kenmerk van de taakuitvoering en bankscore

Effect	DF _n	DF _d	F	p	Sig
referentieniveau	2	2.136	1.947	0,000	***
kenmerk	5	10.680	7.586	0,000	***
referentieniveau:kenmerk	10	10.680	3.132	0,000	***

Significantie: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Ook voor de gesprekstaken zijn de vaardigheden van leerlingen geïllustreerd met enkele citaten uit de gesprekstaken. Hiervoor is dezelfde methode als bij de spreekstaken gebruikt (zie bespreking in paragraaf 5.4).

6 Verschillen tussen scholen en leerlingen

In hoofdstuk 4 van het publieksrapport gaan we in op de verschillen in prestaties tussen leerlingen en scholen. We maken een koppeling tussen de prestaties van leerlingen enerzijds en kenmerken van het onderwijsleerproces en algemene en domeinspecifieke achtergrondkenmerken van leerlingen en scholen anderzijds. Deze koppeling tussen prestaties en variabelen is afzonderlijk gemaakt voor elk van de 3 subdomeinen van mondelinge taalvaardigheid.

De verdeling van het aantal klassen wat per school deelnam verschilde per subdomein. In Tabel 6.1a worden deze aantallen per subdomein weergegeven. Voor alle subdomeinen geldt dat het merendeel van de scholen met slechts 1 klas aan de peiling deelnam.

Tabel 6.1a: Verdeling van het aantal deelnemende klassen per school, uitgesplitst naar subdomein

Aantal klassen per school	Luistervaardigheid	Spreekvaardigheid	Gespreksvaardigheid
1	58	137	136
2	33	60	64
3	5	13	11
4	4	2	1
Totaal	100	212	212

6.1 Ontbrekende waarden

De data die nodig zijn om dit inzichtelijk te maken, zijn verzameld door middel van vragenlijsten. Daarnaast zijn er ook kenmerken opgehaald, zoals de schoolweging. Bij sommige leerlingen ontbreken enkele waarden voor de verzamelde variabelen. Deze ontbrekende waarden zijn geïmputeerd, waarbij de geneste structuur van de data is meegenomen. Voor een beschrijving van de wijze waarop de data geïmputeerd zijn, zie Gijssel et al. (2025).

6.2 Intraklassecorrelatie

Om de vraag te beantwoorden welk deel van de variantie verklaard wordt door verschillen tussen klassen en welk deel door verschillen tussen leerlingen, is eerst door middel van een leeg meerniveaumodel (dat wil zeggen, een model zonder predictoren) de intraklassecorrelatie (icc) berekend. De icc geeft aan welk deel van de gevonden verschillen verklaard kan worden door verschillen tussen klassen. In Tabel 6.2a wordt per subdomein en per onderwijssoort aangegeven wat de variantie op klas- en leerlingniveau was. Voor elk van deze combinaties wordt vervolgens ook benoemd wat de interklassecorrelatie was.

Tabel 6.2a: Intraklassecorrelaties voor de onderdelen luisteren, spreken en gesprekken voeren, uitgesplitst naar onderwijstype

subdomein	onderwijstype	niveau	variantie	icc
luisteren	bo	klas	0,22	0,22
	bo	leerling	0,79	
	sbo	klas	0,19	0,17
	sbo	leerling	0,95	
	so	klas	0,22	0,18
	so	leerling	1,01	
spreeken	bo	klas	0,05	0,12

	bo	leerling	0,35	
	sbo	klas	0,07	0,13
	sbo	leerling	0,45	
	so	klas	0,12	0,19
	so	leerling	0,50	
gesprekken	bo	klas	0,05	0,14
	bo	leerling	0,28	
	sbo	klas	0,07	0,18
	sbo	leerling	0,35	
	so	klas	0,13	0,29
	so	leerling	0,32	

6.3 Meerniveaumodellen

Om na te gaan welke schoolkenmerken de verschillen in leerlingprestaties verklaren, zijn door het consortium meerdere meerniveaumodellen uitgevoerd. Deze zijn per onderwijssoort en per subdomein uitgevoerd. Aangezien er in het so voor te weinig leerlingen data beschikbaar was voor dit type analyses, zijn de analyses alleen voor bo en sbo gedaan. Dit heeft uiteindelijk geleid tot 6 verschillende modellen: voor elke onderwijssoort (bo en sbo) 1 model per subdomein. In elk model was de vaardigheidsscore voor de leerling op het betreffende subdomein de afhankelijke variabele. Elk multilevelmodel is 5 keer geschat; 1 keer voor elk van de 5 geïmputeerde datasets. De modeluitkomsten voor deze 5 modellen (1 per geïmputeerde dataset) zijn gemiddeld.

De modellen zijn tot stand gekomen door variabelen in groepen als predictor aan het model toe te voegen. Dit heeft tot 4 verschillende modellen geleid voor elk van de 3 subdomeinen. De uitkomsten van deze modellen worden gerapporteerd in Gijsel et al. (2025).

- Model 1: algemene leerlingkenmerken.
- Model 2: domeinspecifieke leerlingkenmerken.
- Model 3: leerkracht- en schoolkenmerken.
- Model 4: kenmerken van het onderwijsleerproces.

Na deze afzonderlijke modellen is een totaalmodel geconstrueerd. Predictoren met een p-waarde lager dan 0.1 in de modellen 1 tot en met 4 zijn hierin meegenomen. Vervolgens zijn, om tot een spaarzaam model te komen, niet-significante predictoren stapsgewijs verwijderd. De resultaten hiervan staan in hoofdstuk 9 van Gijsel et al. (2025).

7 Deel C

7.1.1 Algemene achtergrondkenmerken van leerlingen

Probleemgedrag

Tabel 7.1.1a: Scores voor de mate van internaliserend en externaliserend probleemgedrag van leerlingen volgens leerkrachten (bo n=2195-2196; sbo n=678-680; so n=269-270)

Schalen	Onderwijssoort	Gemiddelde	P10	P90
Internaliserend probleemgedrag	bo	1,90	1,00	4,00
	sbo	2,85	1,00	5,00
	so	3,20	1,00	5,00
Externaliserend probleemgedrag	bo	1,62	1,00	3,00
	sbo	2,25	1,00	4,00
	so	2,47	1,00	5,00

Spreekbereidheid

Tabel 7.1.1b: Scores voor de mate van spreekangst en extraversie van leerlingen volgens leerkrachten (bo n=2194-2196; sbo n=679-680; so n=270)

Schalen	Onderwijssoort	Gemiddelde	P10	P90
Spreekangst	bo	2,00	1,00	4,00
	sbo	2,34	1,00	4,00
	so	2,11	1,00	4,00
Extraversie	bo	3,81	2,00	5,00
	sbo	3,50	2,00	5,00
	so	3,60	2,00	5,00

Communicatieproblemen

Tabel 7.1.1c: Scores voor de door leerkrachten gerapporteerde communicatieproblemen van leerlingen (bo n=2198; sbo n=682; so n=270)

Schaal	Onderwijssoort	Gemiddelde	P10	P90
Communicatieproblemen	bo	1,30	1,00	2,00
	sbo	1,96	1,00	3,00
	so	1,85	1,00	2,75

Belemmerende en stimulerende factoren

Tabel 7.1.1d: Scores voor de door leerkrachten geobserveerde stimulerende factoren voor de mondelinge taalvaardigheid van leerlingen (bo n=2177; sbo n=681; so n=270)

Schaal	Onderwijssoort	Gemiddelde	P10	P90
Stimulerende factoren	bo	2,34	1,50	3,00
	sbo	2,01	1,33	2,83
	so	1,87	1,17	2,50

Literatuurlijst

Gijssel, M., Swart, N., Stuij, E., Bruggink, M., Jenniskens, T., Van Langen, A., Koman, F. & Scheer, E., van der. (2025). *Technisch rapport Peil.Mondelinge taalvaardigheid 2023-2024*. KBA.

Inspectie van het Onderwijs. (2019a). *Peil.Mondelinge taalvaardigheid. Einde basisonderwijs 2016-2017*. Inspectie van het Onderwijs.

Inspectie van het Onderwijs. (2019b). *Peil.Mondelinge taalvaardigheid. Einde speciaal (basis)onderwijs 2017-2018*. Inspectie van het Onderwijs.

R Core Team (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.