

## University of Groningen

### Peiling Engels einde basisonderwijs 2018

Ritzema, Lieneke; Hemker, Bas; Naaijer, Harm; Lowie, Wander; Corda, Alessandra; van Aken, Margreet; Rekers-Mombarg, Lyset

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2019

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Ritzema, L., Hemker, B., Naaijer, H., Lowie, W., Corda, A., van Aken, M., & Rekers-Mombarg, L. (2019). *Peiling Engels einde basisonderwijs 2018: Technische rapportage*. GION onderwijs/onderzoek.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# PEILING ENGELS EINDE BASISONDERWIJS 2018 TECHNISCHE RAPPORTAGE

---

Lieneke Ritzema | Bas Hemker | Harm Naayer | Wander Lowie

Alessandra Corda | Magreet van Aken | Lyset Rekers-Mombarg



## Colofon

Peil Engels is uitgevoerd door een consortium bestaande uit GION Onderwijs/Onderzoek, Cito, Hogeschool van Amsterdam, Applied Linguistics (RuG) en de lerarenopleiding van de RuG. Het project is uitgevoerd in opdracht van de Inspectie van het Onderwijs. Bij vermelding van dit rapport in een publicatie dient gerefereerd te worden aan:

Ritzema, E.S., Hemker, B.T., Naayer, H.M., Lowie, W., Aken, van M., Corda, A.C., Rekers-Mombarg, L. T. M., (2019). *Peiling Engels einde Basisonderwijs 2018. Technische rapportage*. Groningen: GION Onderwijs/Onderzoek.

<b>Projectmanagement</b>	Lieneke Ritzema, Roel Bosker
<b>Werving</b>	Harm Naayer, Denise Leijdelmeijer
<b>Logistiek afnames</b>	Harm Naayer, Denise Leijdelmeijer
<b>Inhoudelijke input</b>	Alessandra Corda, Magreet van Aken, Wander Lowie, Louise Taylor
<b>Kwaliteitswaarborging afnames</b>	Ton Koet, Margot Mol, Joost Ides, Bregtje Seton, Vera Hukker, Rolf Plomp, Bianca Seinen, Peta Eisberg, Hester Louter, Marjolein van de Plas, Mara van der Ploeg.
<b>Ontwerp toetsen</b>	Magreet van Aken, Bas Hemker
<b>Ontwerp gesprekstaken</b>	Magreet van Aken, Alessandra Corda, Wander Lowie, Louise Taylor
<b>Beoordeling gesprekstaken</b>	Alessandra Corda, Wander Lowie, Louise Taylor, Lieneke Ritzema
<b>Ontwerp vragenlijsten</b>	Cito, Lieneke Ritzema, Harm Naayer
<b>Psychometrie en methodologie</b>	Bas Hemker
<b>Analyse onderzoeksvragen</b>	Bas Hemker, Lyset Rekers-Mombarg
<b>Standaardsetting</b>	Margreet van Aken, Bas Hemker
<b>Redactie</b>	Lieneke Ritzema, Harm Naayer, Bas Hemker
<b>Eindredactie databestand</b>	Lieneke Ritzema, Bas Hemker, Djurre Zijssling

**ISBN** 90-367-xxxxx-3 (E-book)

© Oktober 2019. GION Onderwijs/Onderzoek

Niets van deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van de Directeur van het Instituut.





## Inhoudsopgave

Inhoudsopgave	5
Voorwoord	9
Leeswijzer	10
Hoofdstuk 1. Theoretisch kader en instrumentarium	13
1.1 Engels in het primair onderwijs: kerndoelen en ERK	13
1.2 Context en prioritering van vaardigheden	19
1.3 Consequenties voor de huidige peiling	20
1.4 ERK-beschrijvingen van de getoetste vaardigheden	20
1.5 Gespreksvaardigheid: beschrijving en theoretische achtergrond	23
1.6 Woordenschat: beschrijving en theoretische achtergrond	32
1.7 Luistervaardigheid: beschrijving en theoretische achtergronden	39
1.8 Leesvaardigheid: beschrijving en theoretische achtergrond	43
1.9 Spreekvaardigheid	47
1.10 De huidige peiling in relatie tot de vorige peiling	49
Hoofdstuk 2. Steekproef, werving en onderzoeksgroep	51
2.1 Steekproefopzet	51
2.2 Bronnen en definitief steekproefkader	54
2.3 Opzet en uitvoering van de werving	59
2.4 Gerealiseerde steekproef: scholen	64
2.5 Gerealiseerde steekproef: leerlingen	69
Hoofdstuk 3. Design	75
3.1 Uitgangspunten	75
3.2 Algemene opzet	76
3.3 Toetsdesign per design-onderdeel	76
3.4 Toets-design over vaardigheden heen	78
3.5 Aantallen Leerlingen	79
Hoofdstuk 4. Voorbereiding en uitvoering	83
4.1 Werving en selectie van (hoofd)testleiders	83
4.2 Training van (hoofd)testleiders	84
4.3 Pilotonderzoek	85
4.4 Voorbereiding van de scholen	87

4.5	Organisatie van de afnames	89
4.6	Verloop van een gemiddelde afnamedag	90
Hoofdstuk 5. Kwaliteit van de meetinstrumenten, vragenlijsten en schalen		93
5.1	Toetsen: Lezen, Luisteren en Woordenschat	93
5.2	Taken: Spreekvaardigheid en Gespreksvaardigheid	98
5.3	Vragenlijst leerlingen, scholen en leerkrachten	107
5.4	Constructie van de vaardigheidsschalen	115
Hoofdstuk 6. Kenmerken deelnemers		123
6.1	Kenmerken van de deelnemende scholen	123
6.2	Kenmerken van de deelnemende leerkrachten	128
6.3	Aanbod van Engels in de les	133
6.4	Resultaten vragenlijst leerlingen	150
Hoofdstuk 7. Standaardbepaling		161
7.1	Het expertpanel	162
7.2	Methode	163
7.3	Procedure	167
7.4	Resultaten	169
7.5	Woordenschat: geen standaarden, wel een referentiescore	178
Hoofdstuk 8. Resultaten - beantwoording van de onderzoeksvragen		181
8.1	Weging en gebruikte leerlingresultaten	182
8.2	Leerlingresultaten	184
8.3	ERK-prestaties gemeten vaardigheden en relatieve woordenschatomvang	186
8.4	Resultaten interactie (paarscores)	188
8.5	Samenhang tussen de vaardigheden	191
8.6	Koppeling prestaties leerlingen aan school- en individuele kenmerken	191
8.7	Meerniveau analyses	197
8.8	Resultaten meerniveau analyses	201
Hoofdstuk 9. Trendvergelijking leerlingprestaties Engels		215
9.1	Trendvergelijking lezen, luisteren en woordenschat	215
9.2	Trendvergelijking spreekvaardigheid	220
Hoofdstuk 10. Verdiepend onderzoek		224
10.1	Randvoorwaarden voor het onderzoek	224
10.2	De selectie van scholen	225
10.3	Werving van scholen	226

10.4	De onderzoeksgroep	226
10.5	Instrumentarium en training inspecteurs	228
10.6	Dataverwerking	233
10.7	Resultaten verdiepend onderzoek: interviews en observaties	235
10.8	De invulling van de lessen Engels	239
10.9	Opleiding en professionele ontwikkeling	260
10.10	Inbedding van en focus op Engels in de school	263
10.11	Conclusie en discussie	264
Hoofdstuk 11. Overzicht van databestanden en variabelen		273
11.1	Opbouw van databestanden	273
11.2	Overzicht van variabelen in het hoofdbestand	274
Literatuurlijst		279
Bijlage 1. Scoreformulier gesprekstaken		285
Bijlage 2. Beoordelingsformulieren voor parenbeoordeling		292
Bijlage 3. Beoordelingsformulier spreekvaardigheid		295
Bijlage 4. Brief naar besturen van scholen in de steekproef		297
Bijlage 5. Wervingsmail richting scholen		299
Bijlage 6. Intakeformulier scholen		300
Bijlage 7. Leerkrachtvragenlijst		312
Bijlage 8. Vragenlijst voor de leerlingen		328
Bijlage 9. Resultaten voor de opgaven in de vragenlijstschalen		336
Bijlage 10. Toelichting op selectie van voorbeelden ten behoeve van de standaardbepaling bij gesprekken		338
Bijlage 11. Operationalisatie van de predictoren in de meerniveau-analyses		340
Bijlage 12. Lesobservatieformulier Peil Engels (verdiepend onderzoek)		344
Bijlage 13. Interview invulformulier (verdiepend onderzoek)		353
Bijlage 14. Observatie instrument (lange versie)		358
Bijlage 15. Overzicht van variabelen in het leerlingvragenlijstbestand		361
Bijlage 16. Overzicht van variabelen in het intakebestand		364
Bijlage 17. Overzicht van variabelen in het leerkrachtbestand		366
Bijlage 18: Overzicht van variabelen m.b.t. gesprekken op paarniveau		373



## Voorwoord

Voor u ligt het technisch rapport van de vijfde peiling Engels einde basisonderwijs, Peil.Engels. Dit peilingsonderzoek heeft plaatsgevonden in de periode januari – maart 2018 en is uitgevoerd door een consortium bestaande uit GION Onderwijs/Onderzoek, Cito, Applied Linguistics (RuG), Hogeschool van Amsterdam en de lerarenopleiding RuG. De peiling is uitgezet door NRO en is uitgevoerd onder regie van de Inspectie van het Onderwijs. Doel van het peilingsonderzoek is om in brede zin de vaardigheden Engels van de leerlingen in kaart te brengen, de ontwikkelingen in die vaardigheden over de tijd vast te stellen en om een beeld te krijgen van het onderwijsaanbod, in relatie tot de resultaten van de leerlingen. In de peiling zijn vier vaardigheden en competenties gemeten: leesvaardigheid, luistervaardigheid, mondelinge vaardigheid (spreek- en gespreksvaardigheid) en woordenschat. Uitgangspunt bij de peiling was de koppeling aan de descriptorren van het Europees Referentiekader (ERK), zowel in de constructie van het instrumentarium als ook in de interpretatie van de resultaten. Dit is gedaan door standaarden te bepalen op de ERK-niveaus A1, A2 en B1 voor leesvaardigheid, luistervaardigheid en gespreksvaardigheid<sup>1</sup>. Door het ERK als uitgangspunt te nemen kenmerkt de peiling zich door een focus op de communicatieve en sociale functie van taal, met andere woorden op het taalgebruik van leerlingen. Hierbij is gestreefd naar zo authentiek mogelijke situaties. Vanwege deze communicatieve focus wordt in deze peiling gespreksvaardigheid bij leerlingen gemeten met behulp van parentaken: gesprekjes tussen twee leerlingen. Deze methodiek van toetsing van mondelinge vaardigheden is niet alleen innovatief ten opzichte van eerdere peilingen, maar ook interessant vanuit het bredere perspectief van taaltoetsenonderzoek. Naast vaardigheden Engels zijn ook de attitude ten opzichte van Engels, de durf van leerlingen om Engels te gebruiken, het buitenschoolse contact met Engels en de eigen inschatting van de vaardigheden Engels van leerlingen gepeild.

Naast het in kaart brengen van de huidige stand van zaken omtrent de vaardigheden Engels van de leerlingen in groep 8 is een tweede doelstelling van deze peiling een trendvergelijking met de vorige peiling, die in 2012 heeft plaatsgevonden. De vaardigheden lezen, luisteren en woordenschat zijn daarom deels op dezelfde manier gemeten als in 2012. Ditzelfde geldt voor de mondelinge vaardigheden: hoewel de focus in deze peiling lag op het meten van gespreksvaardigheid (interactievaardigheid van individuele leerlingen in termen van ERK én interactievaardigheid van *paren* leerlingen), is spreekvaardigheid (monologisch spreken) gemeten aan de hand van een zogenaamde ankertaak, die ook in 2012 is afgenomen. Hoewel met behulp van de resultaten van de huidige peiling een vergelijking met de resultaten van de vorige peiling kan worden gemaakt, moet er worden opgemerkt dat het voorbij de scope van de huidige peiling en dit rapport gaat om verschillen tussen 2012 en 2018 te verklaren. Tenslotte zijn analyses uitgevoerd ter verklaring van mogelijke verschillen in vaardigheid tussen leerlingen. Hiertoe is onderzocht of bepaalde kenmerken van de leerlingen, hun leerkrachten en de scholen en het onderwijsaanbod in Engels voorspellend zijn voor de vier onderscheiden vaardigheden lezen, luisteren, woordenschat en gespreksvaardigheid.

---

<sup>1</sup> Voor de competentie woordenschat zijn geen ERK-beschrijvingen vastgesteld, waardoor voor woordenschat geen standaarden bepaald konden worden en deze resultaten niet in termen van ERK-niveaus worden weergegeven.

## Leeswijzer

In het huidige rapport wordt een verantwoording gegeven van de theoretische achtergrond, opzet en het design van de peiling Engels. Daarnaast wordt de gegevensverzameling beschreven, evenals de kwaliteit van het instrumentarium, de analyse van de gegevens tot schalen, de analyses voor het vaststellen van verbanden tussen resultaten en voorspellers en de databestanden. Dit rapport is bedoeld als achtergrond bij de publicatie van de Inspectie van het Onderwijs over de peiling Engels in 2018. De rapportage is tevens bedoeld voor de gebruikers van de data die in deze peiling zijn verzameld en die voor secundaire analyse via DANS aan onderzoekers ter beschikking worden gesteld. Zij zouden in deze rapportage voldoende achtergrond moeten kunnen vinden om met de data te kunnen werken.

In het eerste hoofdstuk wordt een korte schets gegeven van de theoretische achtergronden die van invloed waren op de opzet van deze peiling, meer specifiek op de constructie van het instrumentarium. In dit hoofdstuk worden de kerndoelen Engels beschreven en de rol die het ERK kan spelen om de beheersingsniveaus van de moderne vreemde taal, in dit geval het Engels, te kunnen vaststellen. Daarnaast wordt aandacht besteed aan de context van Engels in het po en hoe deze context, de door experts vastgestelde prioritering van de te peilen vaardigheden en de communicatieve insteek, de opzet van deze peiling hebben beïnvloed. Daarna wordt in hoofdstuk 1 per vaardigheid weergegeven hoe deze uitgangspunten zijn uitgewerkt in het instrumentarium. Hoofdstuk 1 wordt afgesloten met een bespreking van de verschillen tussen de vorige en huidige peiling. In het tweede hoofdstuk wordt het steekproefkader, de wijze van steekproeftrekking en de werving van scholen toelicht. Daarbij wordt weergegeven op welke manier is geprobeerd om een representatieve steekproef te realiseren en wordt op zowel school- als leerling-niveau beschreven in hoeverre dit is gelukt. Het derde hoofdstuk bevat een toelichting op het gekozen onderzoeksdesign en de uitgangspunten die daaraan ten grondslag lagen. Gekozen is voor een incompleet design als gevolg van de beperkte tijd die beschikbaar was voor de afname van de peiling. Een peiling vergt een ruime voorbereiding. Deze voorbereiding en de uitvoering van de peiling staan centraal in hoofdstuk 4. Beschreven wordt hoe de testleiders zijn voorbereid op de afnames, hoe de instrumenten tijdens pilots zijn getest (en welke aanpassingen als gevolg van deze pilots hebben plaatsgevonden) en hoe de scholen zijn voorbereid op de daadwerkelijke peilingsafname. Het vijfde hoofdstuk is gewijd aan de kwaliteit van het instrumentarium. In dit enigszins technische hoofdstuk wordt de invloed van de pilotresultaten op het definitieve instrumentarium besproken. Ook wordt de psychometrische kwaliteit van dit definitieve instrumentarium beschreven. Hoofdstuk 5 bevat daarnaast informatie over de analyses en schaling die zijn uitgevoerd ten behoeve van de vragenlijsten. Tenslotte richt hoofdstuk 5 zich op de constructie van vaardigheidsschalen voor de toetsen en praktische taken en op de manier waarop vaardigheidsscores zijn gecombineerd ten behoeve van vervolgonderzoek. Kenmerken van de school, leerkracht en leerlingen alsmede van het onderwijsleerproces staan centraal in hoofdstuk 6. Hoofdstuk 7 gaat in op de manier waarop er, na afname en analyse van de lees-, luister- en gesprekstaken, door experts standaarden zijn bepaald op de ERK-niveaus A1, A2 en B1. Een alternatief voor de interpretatie van de woordenschatresultaten wordt ook beschreven. De resultaten van de daadwerkelijke afnames worden besproken in hoofdstuk 8. Voor de vier vaardigheden worden de resultaten van de leerlingen besproken, zowel in termen van vaardigheidsscores als van ERK-niveaus. In dit hoofdstuk worden tevens verschillende statistische modellen besproken om zicht te krijgen op prestatieverschillen. Hierbij is onderzocht of kenmerken van leerlingen, hun leerkrachten, structurele kenmerken van de scholen en het onderwijsleerproces voorspellend zijn voor de prestaties van leerlingen. In hoofdstuk 9 staat de trendvergelijking centraal.

De leerlingresultaten in 2018 voor woordenschat, lees-, luister- en spreekvaardigheid worden in dit hoofdstuk vergeleken met de leerlingresultaten in 2012, het jaar van de vorige peiling. Hoofdstuk tien, gaat in op de opzet en resultaten van het verdiepende onderzoek dat na afloop van het hoofdonderzoek heeft plaatsgevonden. In september - oktober 2018 zijn bovengemiddeld presterende scholen bezocht. Deze scholen zijn geselecteerd uit vier combinatieclusters, die zijn gebaseerd op het startmoment Engels en het percentage gewichtenleerlingen op een school<sup>2</sup>. Het laatste hoofdstuk, hoofdstuk 11, bevat een beschrijving van de opgeleverde bestanden en een overzicht van variabelen. Deze informatie is met name voor gebruikers van de data voor secundaire analyse van toepassing. Op basis van deze informatie kunnen gebruikers relevante variabelen voor hun eigen analyse. Na een referentielijst treft u de bijlagen aan.

---

<sup>2</sup> De gewichtenregeling is een regeling in het Nederlandse basisonderwijs en richt zich op het verminderen van onderwijsachterstanden van risicoleerlingen. Het 'gewicht' van een leerling bepaalt de subsidie die een school voor de leerling krijgt, en is gebaseerd op factoren in de thuissituatie die ervoor zorgen dat de leerling een vergroot risico loopt op een onderwijsachterstand. Tegenwoordig wordt het gewicht van een leerling berekend op basis van het opleidingsniveau van de ouders/verzorgers. Het basisgewicht van een leerling is 0,0, leerling met verhoogde risico's op onderwijsachterstanden hebben een gewicht van 0.3 en leerlingen met een (zeer) hoog risico een gewicht van 1,2.





## Hoofdstuk 1. Theoretisch kader en instrumentarium

Het belang van Engels wordt in onze moderne maatschappij steeds groter. Dit geldt ook voor leerlingen in het primair onderwijs, voor wie een goede voorbereiding op de kernvakken in het voortgezet onderwijs essentieel is voor hun verdere schoolloopbaan en hun toekomst. Met behulp van de peiling Engels einde basisonderwijs is de stand van zaken wat betreft de competenties Engels van Nederlandse leerlingen in groep 8 vastgesteld, mede in vergelijking tot de vorige peiling. Daarnaast is het onderwijs Engels op de basisscholen in kaart gebracht.

In dit hoofdstuk worden de theoretische achtergronden beschreven en de wijze waarop die van invloed waren op het ontwikkelde instrumentarium. Allereerst worden verschillende invullingen van Engels in het basisonderwijs beschreven. Daarna wordt aandacht besteed aan de kerndoelen voor Engels en niveau-beschrijvingen binnen het Europees Referentiekader, ook wel ERK genoemd (paragraaf 1.1). Paragraaf 1.2 richt zich op de context waarin Engels in het po wordt gegeven, het uitgevoerde curriculum en de prioritering van vaardigheden door experts in het werkveld, waarna in paragraaf 1.3 de consequenties hiervan voor de huidige peiling worden besproken. Beschreven wordt dat de communicatieve insteek waarin taal in zo authentiek mogelijke situaties wordt gebruikt - zoals verondersteld binnen het ERK - leidend was bij deze peiling. Deze koppeling met het ERK gold zowel voor de constructie van het instrumentarium als voor de interpretatie van de resultaten. De focus op dit communicatieve aspect komt tevens naar voren in de keuze om gesprekstaken te construeren, waarbij leerlingen met elkaar in gesprek gaan. Na een globale beschrijving van de te meten ERK-niveaus per vaardigheid – binnen deze peiling werden de niveaus A1, A2 en B1 gemeten – worden in paragraaf 1.5 tot 1.9 gespreksvaardigheid, woordenschat, luistervaardigheid, leesvaardigheid en spreekvaardigheid beschreven. Spreekvaardigheid is enkel met behulp van een ankertaak gepeild. Per vaardigheid bevat deze beschrijving een theoretische achtergrond, een weergave van het daarvan afgeleide instrumentarium, de specifieke ERK-beschrijvingen van de in de peiling gemeten sub-vaardigheden, informatie over afname en procedure, alsmede over de beoordeling. In paragraaf 1.10 wordt tenslotte weergegeven hoe de huidige peiling zich verhoudt tot de vorige peiling, die in 2012 is uitgevoerd.

### 1.1 Engels in het primair onderwijs: kerndoelen en ERK

Engels speelt een belangrijke rol in de Nederlandse samenleving en is sinds 1986 ook een verplicht vak in het basisonderwijs in Nederland. In Rose (2016) wordt het algemene doel van onderwijs Engels in het basisonderwijs als volgt omschreven “het leggen van een eerste basis om te kunnen communiceren met moedertaalsprekers of anderen die buiten de school Engels spreken”.

Beleidsmatig en chronologisch zijn deze de belangrijkste veranderingen geweest rondom Engels op de basisschool:

- 1986 – verplicht Engels in de groepen 7 en 8, met een aangeraden totale hoeveelheid lessen tussen de 80 en 100. Het aantal lessen Engels is niet vastgelegd in de wet. Scholen zijn vrij in de wijze waarop ze de taal aanbieden en dit zorgt voor grote onderlinge verschillen.

- 1993 – voor het eerst worden kerndoelen Engels in het basisonderwijs geformuleerd.
- 2006 – nieuwe kerndoelen Engels
- 2014 – in 2014-2015 is een pilot tweetalig primair onderwijs (TPO) gestart (t/m 2019) op een beperkt aantal scholen. Leerlingen krijgen binnen het TPO vanaf groep 1 30% tot 50% van hun lessen in het Engels.
- 2016 - vanaf 1 januari 2016 mogen scholen tot maximaal 15% van de onderwijstijd in het Engels aanbieden.

Sinds begin jaren '90 zijn steeds meer basisscholen Engels eerder dan in groep 7 en 8 gaan aanbieden. Naast de reguliere EIBO-variant in groep 7 en 8 (EIBO= Engels in het basisonderwijs) is ook een variant VVTO (=vroeg vreemdetalenonderwijs) ontstaan. Door ondersteuning van het Europees Platform, later NUFFIC, konden scholen die overgingen op VVTO subsidie ontvangen voor de startfase. Het aantal scholen met vormen van VVTO wordt nu geschat tussen 15 en 20% van het totaal. Een derde categorie scholen zijn de scholen die Engels vanaf groep 5 aanbieden, de zogenaamde vervroegd EIBO-scholen. Geschat wordt dat dit ongeveer 17% van de scholen is (Thijs, Tuin & Trimbos, 2011).<sup>3</sup> Door de grote verschillen tussen en binnen de varianten kan het instroomniveau in het voortgezet onderwijs enorm variëren. De verschillen belemmeren de invoering van doorlopende leerlijnen Engels po-vo. Een interessante lopende ontwikkeling is curriculum.nu, een groot project gericht op de vernieuwing van het curriculum van po en vo waarbij Engels/moderne vreemde talen één van de leergebieden is waarover wordt nagedacht. Voor het eerst werken in dit project vo- en po-leerkrachten en deskundigen Engels samen.

De beheersing van het Engels door de Nederlandse bevolking is goed in vergelijking met andere Europese landen. Er zijn veel aanwijzingen dat het hoge niveau grotendeels bereikt wordt dankzij het contact met het buitenschoolse Engels – jonge kinderen zien en horen de taal op televisie, op internet, via digitale spellen en natuurlijk via muziek. Het verband tussen contact met Engels buiten school en prestaties in het Engels is onderzocht in het Early Language Learning in Europe (ELLiE) project<sup>4</sup>. Dit onderzoek laat zien dat buitenschools contact<sup>5</sup> met Engels een goede voorspeller is (samen met linguïstische afstand tussen de talen) van de luister- en leesvaardigheden van de jonge taalleerder (van 10 – 11 jaar). De mate en intensiteit van dit contact zal niet voor elke leerling hetzelfde zijn. Het onderzoek van Muñoz, Cadierno en Casas (2018) laat zien dat informeel contact met Engels buiten school een grote invloed heeft op receptieve grammaticale kennis bij 7-9 jarigen. De effecten van buitenschools Engels op Nederlandse kinderen worden op dit moment onderzocht in het ORWELL project (ORal and WRitten English Language Learning).<sup>6</sup> Naast het aantal lessen op school, een vroege of late start en het contact met het Engels buiten school is natuurlijk ook de kwaliteit van de leerkracht van invloed op de prestaties van leerlingen. Zo is het beheersingsniveau van het Engels van leerlingen met een leerkracht die een native speaker is

<sup>3</sup> Hoewel de vervroegd EIBO-variant in dit hoofdstuk wel wordt beschreven is deze variant in de huidige peiling niet als aparte variant meegenomen. In deze peiling wordt enkel onderscheid gemaakt tussen EIBO en VVTO scholen. Reden hiervoor was dat er geen betrouwbare bronnen beschikbaar zijn voor een accurate indicatie van het aantal vervroegd EIBO-scholen, laat staan welke specifieke scholen dit zijn.

<sup>4</sup> [www.ellieresearch.eu](http://www.ellieresearch.eu)

<sup>5</sup> Evenals Muñoz, Cadierno en Casas (2018) is gekozen om de term “contact” te gebruiken in plaats van “blootstelling” omdat deze laatste niet per se de productieve vaardigheden omvat die bij activiteiten zoals online spelletjes wel ingezet worden.

<sup>6</sup> Zie [www.lbbo.nl/userdata/Brochure-ORWELL.pdf](http://www.lbbo.nl/userdata/Brochure-ORWELL.pdf)

of een non-native speaker met een goede beheersing van het Engels (een *proficient user*, oftewel iemand met C-niveau binnen het ERK) hoger dan die van leerlingen met een leerkracht met een B-niveau binnen het ERK (de zogenaamde *independent user*), ook wanneer gecontroleerd wordt voor het aantal minuten per week waarin op school aandacht aan Engels is besteed (Unsworth et al, 2014). Een mogelijke verklaring hiervoor zou kunnen zijn dat leerkrachten die het Engels beter beheersen meer lexicale variatie en grammaticale complexiteit tentoonspreiden. Veel onderzoeken naar impact van de kwaliteit van de leerkracht zijn overigens gebaseerd op eigen inschatting van de leerkracht zelf en lesobservaties, niet op resultaten van een taalvaardigheidstest. Ook de gegevens over de hoeveelheid schoolse en buitenschoolse input komen uit vragenlijsten. In onze peiling wordt eveneens door middel van zelfinschatting, vragenlijsten, observaties en interviews onderzocht wat de impact van deze factoren is. Daarbij wordt een koppeling met het ERK gemaakt. Het ERK is een internationaal raamwerk dat niveaus van competenties beschrijft voor vreemde talen, ontwikkeld vanuit de Raad van Europa in Straatsburg.

Leerkrachten is verder aangeboden op vrijwillige basis en op eigen gelegenheid gebruik te maken van de digitale test *Aptis for teachers* (APTIS), een beroepsspecifieke test voor de Engelse taalvaardigheid ontwikkeld door de British Council. Deze toets is geschikt voor het meten van de A t/m C niveaus van het ERK en is dus toereikend voor het meten van het niveau van de leerkrachten. Binnen de kennisbasis pabo van voorjaar 2018 is namelijk voorgeschreven dat dit niveau minimaal B2 niveau zou moeten behelzen.<sup>7</sup> In de peiling konden leerkrachten zichzelf op een deel van deze APTIS toetsen: (a) grammatica en vocabulaire (het kerngedeelte, afkomstig uit General Aptis) en (b) spreekvaardigheid (O'Sullivan & Dunlea, 2015). De keuze voor de APTIS is in lijn met beleid van alle hogescholen om de Engelse vaardigheden bij hun studenten voortaan verplicht met APTIS te meten. De respons op het APTIS-aanbod was echter laag. Van de 28 leerkrachten die zich hadden aangemeld, hebben 15 leerkrachten aan de test deelgenomen. Van deze 15 leerkrachten hebben tien de test daadwerkelijk afgerond. Dit aantal is te laag om nadere analyses uit te voeren en er wordt in dit rapport niet over gerapporteerd.

### 1.1.1 Kerndoelen

De kerndoelen in Nederland, die het wettelijk kader vormen voor het domein Engels in het primair onderwijs, zijn gericht op communicatie. In 2006 zijn de volgende vier beknopte, globale kerndoelen geformuleerd:

- Kerndoel 13: De leerlingen leren informatie te verwerven uit eenvoudige gesproken en geschreven Engelse teksten.
- Kerndoel 14: De leerlingen leren in het Engels informatie te vragen of geven over eenvoudige onderwerpen en zij ontwikkelen een attitude waarbij ze zich durven uit te drukken in die taal.
- Kerndoel 15: De leerlingen leren de schrijfwijze van enkele eenvoudige woorden over alledaagse onderwerpen.
- Kerndoel 16: De leerlingen leren om woordbetekenissen en schrijfwijzen van Engelse woorden op te zoeken met behulp van het woordenboek.

---

<sup>7</sup> Zie [www.10voordeleraar.nl/documents/kennisbases\\_pabo/kennisbasispabo.pdf](http://www.10voordeleraar.nl/documents/kennisbases_pabo/kennisbasispabo.pdf)

Aangezien de kerndoelen geen beschrijvingen voor beheersingsniveaus weergeven, voldoen basisscholen hier altijd aan, ongeacht het startmoment, de onderwijstijd of de invulling van het vak. Nadere handreikingen voor het werkveld worden gegeven in de curriculaire uitwerking door SLO.<sup>8</sup> Deze suggereert dat de focus op gesproken taal zou moeten liggen. Schrijf- en leesvaardigheid in het Engels krijgen in het primair onderwijs minder aandacht. Het aanbod Engels in het primair onderwijs is daarbij vooral gericht op de belevingswereld van jonge kinderen, met thema's als hobby's, de woonomgeving en dieren. Nuffic heeft, in samenwerking met een aantal Pabo's, de kerndoelen en de uitwerking van deze kerndoelen in tussendoelen en leerlijnen (TULE), verder doorontwikkeld voor VVTO scholen met streefniveaus voor wat betreft de beheersing van de taal.<sup>9</sup> In deze nadere uitwerking wordt de link gelegd met het Europees Referentiekader voor de Talen (ERK), die niet genoemd wordt in de kerndoelen en in de TULE-uitwerking daarvan.

### 1.1.2. Het Europees Referentiekader

Er bestaat geen directe koppeling tussen de kerndoelen en het ERK, omdat de kerndoelen geen beschrijving van het beoogde niveau weergeven. Toch kunnen de kerndoelen Engels voor het Nederlandse basisonderwijs in relatie worden gebracht met de ERK niveaus. Met het ERK is een instrument gecreëerd waarmee degene die de taal leert (zoals de leerling) zelf inzicht krijgt in zijn of haar taalvaardigheid en de progressie hierin kan volgen. Het ERK geeft aan wat degene die de taal leert, kan. Het ERK biedt bovendien een inzichtelijk kader om internationale samenwerking te bevorderen tussen onderwijsinstanties in verschillende landen, geeft een basis voor herkenning van taalkwalificaties en helpt onderwijsinstanties in benchmarking. Het ERK heeft door de niveauomschrijvingen de kwaliteit van taalonderwijs inzichtelijker gemaakt en helpt om taaltoetsen met elkaar te vergelijken. Het ERK is in 2001 gepubliceerd en wordt nu ook buiten Europa toegepast als hulpmiddel om niveaus te definiëren, ten behoeve van lesgeven en toetsing. Er worden zes niveaus van beheersing omschreven binnen het ERK, waarbij A1 het laagste niveau is en C2 het hoogste.

A1 / A2	basic user
B1 / B2	independent user
C1 / C2	proficient user

In het ERK wordt per taalvaardigheid een beheersingsniveau aangeduid. Ten opzichte van de traditionele verdeling in vier taalvaardigheden (spreken, luisteren, lezen, schrijven) maakt het ERK onderscheid tussen monologisch en dialogisch spreken: gesproken productie en gesproken interactie.

Het ERK is gebaseerd op *real-life* communicatieve taken. Per vaardigheid en per niveau zijn er gedetailleerde "*can do*" descriptors geformuleerd. Deze "*can do*" descriptors gaan uit van wat iemand wel kan in plaats van vanuit welke fouten iemand die de taal leert, maakt (=deficiëntie perspectief). Hierbij is het doel duidelijk gericht op de taal als communicatiemiddel. De taal heeft hierbij een sociale functie om de interactie tussen individuen te bevorderen waarbij degenen die de taal leren eigenlijk als taalgebruikers moeten worden gezien. Ze leren de taal gebruiken in plaats van

<sup>8</sup> <http://tule.slo.nl/Engels/F-KDEngels.html>.

<sup>9</sup> <https://www.nuffic.nl/documents/109/standaard-vroeg-vreemdetalenonderwijs-engels.pdf>

dat ze de taal leren. Het primaire doel van deze descriptoren is transparantie en congruentie, in de eerste plaats dus voor degene die de taal leert zelf, door middel van zelfreflectie. Hieronder enkele voorbeelden van de globale descriptoren (Taalprofielen 2015):

Gesprekken voeren A1 niveau: *Ik kan deelnemen aan een eenvoudig gesprek, wanneer de gesprekspartner bereid is om zaken in een langzamer spreektempo te herhalen of opnieuw te formuleren en mij helpt bij het formuleren van wat ik probeer te zeggen. Ik kan eenvoudige vragen stellen en beantwoorden die een directe behoefte of zeer vertrouwde onderwerpen betreffen.*

Gesprekken voeren B1 niveau: *Ik kan de meeste situaties aan die zich kunnen voordoen tijdens een reis in het gebied waar de betreffende taal wordt gesproken. Ik kan onvoorbereid deelnemen aan een gesprek over onderwerpen die vertrouwd zijn, of mijn persoonlijke belangstelling hebben, of die betrekking hebben op het dagelijks leven (bijvoorbeeld familie, vrijetijdsbesteding, werk, reizen en actuele gebeurtenissen).*

De ERK-descriptoren omvatten niet alleen niveauomschrijvingen van taal maar ook van communicatiestrategieën. Het verwerven van een taal wordt gezien als circulair in plaats van lineair: door taken uit te voeren leert degene die de taal leert bepaalde competenties en verwerft hij ook bepaalde strategieën zoals plannen, uitvoeren, monitoren en verbeteren.

### 1.1.3 Het ERK in het Nederlandse schoolstelsel

Het ERK wordt ingezet in het voortgezet onderwijs in Nederland als raamwerk voor het expliciteren van leeropbrengsten. In het primair onderwijs is, zoals boven genoemd, nog geen koppeling gemaakt tussen leeropbrengsten en ERK-niveaus. In principe is het ERK sectoronafhankelijk en zou het voor zowel jonge als oudere leerders kunnen worden gebruikt. Maar de beschrijvingen zijn voornamelijk ontwikkeld voor (jong)volwassenen, die al ver zijn in hun cognitieve ontwikkeling. Samen met de verruiming van het aanbod van vreemde talen, voornamelijk Engels, in het primair onderwijs in Europese landen groeide de behoefte aan nadere uitwerkingen en een aanscherping van de ERK beschrijvingen voor jonge taalleerders.

De ERK descriptoren zijn in Nederland verder door SLO uitgewerkt in 2015 in de publicatie *Taalprofielen* (SLO, 2015), op basis van de Nederlandse vertaling van het ERK. Deze publicatie bevat veel voorbeelden die taalleerders en docenten meer grip geven op de ERK-niveaus.

Het eerder aangehaalde voorbeeldniveau A1;

*Ik kan deelnemen aan een eenvoudig gesprek, wanneer de gesprekspartner bereid is om zaken in een langzamer spreektempo te herhalen of opnieuw te formuleren en mij helpt bij het formuleren van wat ik probeer te zeggen. Ik kan eenvoudige vragen stellen en beantwoorden die een directe behoefte of zeer vertrouwde onderwerpen betreffen.*

is dan verder uitgewerkt met voorbeelden van receptieve en productieve tekstkenmerken:

*A1 niveau Vloeiendheid: Kan zeer korte, geïsoleerde, voornamelijk kant-en-klare uitingen hanteren, met veel pauzes om naar uitdrukkingen te zoeken, minder vertrouwde woorden uit te spreken en de communicatie te herstellen.*

*A1 niveau bereik van de woordenschat: Heeft een zeer elementaire woordenschat die bestaat uit geïsoleerde woorden en eenvoudige uitdrukkingen met betrekking op persoonlijke gegevens en bepaalde concrete situaties.*

De kerndoelen beschrijven geen beheersingsniveaus. Het ERK doet dat wel en de beschrijvingen van het ERK kunnen nuttig worden ingezet voor lesgeven en toetsing. Daarom heeft SLO in 2015 op verzoek van het ministerie van OCW een concept-leerplankader ontwikkeld om de kerndoelen Engels verder te concretiseren in ERK-uitstroomniveaus (Rose, 2016). In het leerplankader worden de volgende eindniveaus verwacht voor de meer intensieve vormen van Engels op de basisschool, maar wordt geen inschatting gegeven voor de reguliere EIBO (Tabel 1.1).

Tabel 1.1: Inschatting voor ERK-niveaus leerkrachten in VVTO en vervroegd EIBO

Type onderwijs	Domein	Niveau
VVTO - vanaf groep:		
1-2	luisteren en begrijpen	B1
1-2	spreken en gesprekken voeren	A2-B1
4	lezen en begrijpen	A2-B1
5	Schrijven	A1-A2
vervroegd EIBO - vanaf groep:		
5-6	luisteren en spreken	A2
5-6	spreken en gesprekken voeren	A1-A2
5-6	lezen en begrijpen	A1-A2
5-6	Schrijven	A1

Tabel 1.1 laat twee aspecten zien: het uitstroomprofiel per vaardigheid per variant en een eventuele opbouw in het tijdstip waarop de vaardigheden worden aangeboden. Dit laatste is vooral voor het VVTO van belang. In het VVTO zou in jaargroep 1-2 met name aandacht moeten worden besteed aan luisteren en begrijpen, alsmede spreken en gesprekken voeren. Vanaf groep 4 zou het aanbod in VVTO-scholen zich (ook) op lezen en begrijpen moeten richten, waarna vanaf groep 5 aandacht zou kunnen worden geschonken aan schrijfvaardigheid. De vervroegd EIBO-variant kent deze opbouw niet: in deze variant wordt verondersteld dat het aanbod Engels in groep 5 en 6 alle vaardigheden bevat.

De eerdergenoemde wens tot aanscherping van de ERK beschrijvingen voor jonge leerlingen heeft overigens in februari 2018 geleid tot de publicatie *CEFR Companion Volume with New Descriptors* (Council of Europe, Language Policy Programme, 2018). In deze publicatie is tevens een pre-A1 niveau beschreven en zijn descriptoren specifiek uitgewerkt voor twee groepen jonge leerlingen (tussen 7-10 en 11-15 jaar oud). Deze descriptoren kunnen een verdere impuls geven aan de toepassing van het ERK in het basisonderwijs (zie ook de uitgebreide verzameling in *Collated Representative Samples of Descriptors of Language Competences Developed for Young Learners, 2018* (Council of Europe, 2018). Enkele voorbeelden van het pre-A1 niveau:

- can make simple purchases where pointing or other gesture can support the verbal reference;
- can ask and tell day, time of day and date;

- can use some basic greetings;
- can say yes, no, excuse me, please, thank you, sorry;
- can fill in uncomplicated forms with personal details, name, address, nationality, marital status;
- can write a short, simple postcard.

De publicatie *CEFR Companion Volume with New Descriptors* kwam uit tijdens de toetsafnames voor deze peiling. Het pre-A1 niveau is derhalve niet in de constructie van het instrumentarium en de standaardbepaling meegenomen. Voor de toekomst verdient het echter aanbeveling om te onderzoeken of het wenselijk is om ook een pre-A1 niveau te meten.

## 1.2 Context en prioritering van vaardigheden

Op basis van de vorige peiling blijkt dat het uitgevoerde curriculum in het po zich in 2012 kenmerkte door een focus op luistervaardigheid, woordenschat en spreekvaardigheid (met name naspreken van nieuwe woorden, het voeren van standaardgesprekjes tussen leerlingen (oftewel het (min of meer) voorlezen van voorgestructureerde dialogen) en het voeren van een gesprekje in het Engels met de leerkracht (Geurts & Hemker, 2013; Rose, 2016). De vorige peiling liet ook zien dat de lestijd verschilde van zeven uur per jaar tot 90 uur per jaar en dat er in VVTO-scholen veel meer tijd aan Engels werd besteed (Geurts & Hemker, 2013). Er werd een diversiteit aan methodes gebruikt, waarvan het merendeel zich focust op communicatie. De methodes geven weinig handreikingen voor differentiatie naar niveau. Ook eerder onderzoek laat zien dat differentiatie in de vorm van remediëring slechts beperkt voorkomt: 17% van de leerkrachten gaf aan dergelijke activiteiten uit te voeren (Thijs, Trimbos & Tuin, 2011). Om de beheersingsniveaus vast te stellen toetsten de meeste scholen (70%) in 2012 weliswaar met toetsen, maar deze geven geen indicatie voor een doorgaande lijn richting vo. Gestandaardiseerde toetsen Engels (IEP, Cito) werden weinig gebruikt en op 18% van de scholen werd het Engels niet getoetst. Ook was opvallend dat leerkrachten aangaven relatief veel tijd aan spreekvaardigheid te besteden, maar dit weinig te toetsen (Geurts & Hemker, 2013).

De resultaten van het merendeel van de gebruikte toetsen (de methodetoetsen) zijn weinig informatief voor het voortgezet onderwijs (Rose, 2016). Uit een verkenning (Thijs, Tuin, & Trimbos, 2011) en de domeinbeschrijving rond Engels in het basisonderwijs (Rose, 2016) die door de SLO zijn uitgevoerd, bleek dat leerlingen het voortgezet onderwijs binnenstromen met zeer uiteenlopende beheersingsniveaus Engels. Met de toevoeging van het tweetalig primair onderwijs (TPO) op sommige scholen zullen deze verschillen vermoedelijk alleen maar toenemen. De lessen in de onderbouw van het voortgezet onderwijs sluiten door al deze verschillen vaak niet aan bij de kennis en vaardigheden die veel van hun leerlingen tijdens het primair onderwijs hebben opgedaan (Geurts & Hemker, 2013; Bodde & Schokkenbroek, 2013). Dit doet geen recht aan de leerlingen en werkt ook demotiverend (SLO, 2018).

Voorafgaand aan de huidige peiling is een expertmeeting georganiseerd waarin zeven te peilen vaardigheden en taalcompetenties Engels zijn geprioriteerd. Dit is gedaan door per vaardigheid/competentie te indiceren hoe relevant en hoe realistisch dit was voor Engels in het primair onderwijs. De unanieme uitkomst was dat (1) gespreksvaardigheid, (2) woordenschat, (3) luistervaardigheid het hoogst werden geprioriteerd. Met name gespreksvaardigheid werd belangrijk geacht, aangezien deze vaardigheid een 'realistisch beeld geeft van de taalvaardigheid van een



leerling' en dat het als gevolg van de communicatieve aanpak van methoden realistisch zou zijn dat hier veel aandacht aan besteed zou worden. Leesvaardigheid (4) werd met name relevant geacht in relatie tot de overgang naar het voortgezet onderwijs. Daarnaast hebben de experts de wens uitgesproken om de peilingsresultaten te koppelen aan ERK-beschrijvingen.

### 1.3 Consequenties voor de huidige peiling

De prioritering van vaardigheden door de experts en de wens om de peiling te koppelen aan ERK-niveaus en –beschrijvingen zijn leidend geweest voor de opzet van deze peiling. Door zowel de constructie van het instrumentarium als de interpretatie van de resultaten te koppelen aan het ERK kenmerkt deze peiling zich door een focus op de communicatieve en sociale functie van taal, met andere woorden op het taalgebruik van leerlingen en dan met name in zo authentiek mogelijke situaties. Vanwege de communicatieve focus wordt in deze peiling gespreksvaardigheid bij leerlingen gemeten. Geprobeerd is om hiervoor zo authentiek mogelijke situaties te creëren. Dit heeft voor het toetsen van mondelinge vaardigheid geleid tot parentaken, gesprekjes tussen twee leerlingen. Het op deze manier meten van gespreksvaardigheid was om twee redenen vernieuwend ten opzichte van vorige peilingen: voor het eerst werd in een landelijke peiling onder jonge leerlingen gespreksvaardigheid Engels gemeten, en voor het eerst werden de toetsinstrumenten aan ERK-niveaus gekoppeld. De methodiek van toetsing van mondelinge vaardigheden in deze peiling is niet alleen innovatief ten opzichte van eerdere peilingen, maar ook interessant vanuit het bredere perspectief van taaltoetsenonderzoek.

In deze peiling Engels is er dus, in tegenstelling tot eerdere peilingen, in de interpretatie van de resultaten een koppeling gemaakt tussen de gemeten vaardigheden (lezen, luisteren, gespreksvaardigheid) en de ERK-niveaus A1, A2 en B1. Dit is gedaan aan de hand van een standaardbepaling. De afgelopen jaren zijn er landelijke en internationale peilingen geweest om het ERK-niveau in verschillende taalvaardigheden te meten aan het eind van het voortgezet onderwijs. Het koppelen van de resultaten van de peiling aan het eind van het basisonderwijs aan de ERK-niveaus vult het reeds beschikbare beeld van de Engels competentie van Nederlandse leerlingen aan. Naast taalvaardigheden is vocabulairekennis (woordenschat) gemeten. Aangezien vocabulairekennis niet rechtstreeks aan ERK-niveaus kan worden gekoppeld, heeft voor woordenschat geen standaardbepaling plaatsgevonden.

### 1.4 ERK-beschrijvingen van de getoetste vaardigheden

Als gevolg van de koppeling met het ERK zijn voor de toetsconstructie voor luister-, lees- en gespreksvaardigheid de richtlijnen gevolgd van internationale handreikingen op het gebied van toetsen en beoordelen met het ERK (zoals *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) - A Manual*, Council of Europe 2009<sup>10</sup> en *Manual for Language Test Development and Examining for use with the CEFR*, Council of Europe 2011.<sup>11</sup> Ook de nationale handreikingen (*Toetsen en beoordelen met het ERK*, SLO 2011)<sup>12</sup> zijn gebruikt. Deze handleidingen behandelen de verschillende stappen die in het proces van

---

<sup>10</sup> <https://rm.coe.int/1680667a2d>

<sup>11</sup> <https://rm.coe.int/manual-for-language-test-development-and-examining-for-use-with-the-ce/1680667a2b>

<sup>12</sup> <http://www.erk.nl/docent/toetsing/toetsen-en-beoordelen-met-het-ERK.pdf>

de toetsconstructie genomen moeten worden vanuit het perspectief van het ERK. Voor de selectie van concrete, te meten (sub)vaardigheden is gebruik gemaakt van Taalprofielen (SLO, 2015). Bij de constructie van de toetsmatrijzen is de Nederlandse uitwerking van het Europees Referentiekader alsmede Taalprofielen (2015) geraadpleegd. Bij woordenschat en spreekvaardigheid is een andere procedure gevolgd. Woordenschat kent, zoals genoemd, geen ERK-beschrijvingen. Bij woordenschat zijn, voor zover mogelijk, de principes van het zogenaamde CATTS-model leidend geweest, die worden besproken in paragraaf 1.6. Daarnaast is het *English Vocabulary Profile* geraadpleegd. Het EVP wordt beschikbaar gesteld door Cambridge University Press op basis van het Cambridge Learner Corpus. Het EVP geeft informatie over welke woorden en woordbetekenissen bekend zijn bij en gebruikt worden door leerlingen in elk van de ERK niveaus.

Spreekvaardigheid is enkel gemeten met een ankertaak om een trendvergelijking met de vorige peiling mogelijk te maken. Deze ankertaak was niet aan de hand van ERK-beschrijvingen ontwikkeld.

Er is voor gekozen om toetsen te construeren die de niveaus A1, A2 en B1 van het ERK meten. Er wordt uitgegaan van deze voor het po relatief ruime range aan niveaus, omdat uit eerdere peilingen en onderzoeken bleek dat er grote verschillen zijn tussen leerlingen in Engelse taalvaardigheid. Ook de eerder besproken uitstroomprofielen die door het SLO zijn opgesteld rechtvaardigden deze keuzes. Om logistieke redenen was de toetsafname beperkt tot twee uur in totaal. In deze beschikbare tijd was het niet mogelijk om uitgebreide taaltoetsen af te nemen per vaardigheid, met opdrachten die over alle descriptoren per niveau gingen. We kunnen dus geen uitspraak doen over alle aspecten van de beheersing van een bepaald niveau, maar alleen over de descriptoren die voor de toetsopdrachten rechtstreeks relevant zijn. In deze paragraaf wordt per vaardigheid een globale ERK-beschrijving gegeven van de beoogde beheersingsniveaus. In de paragrafen 1.5.2, 1.7.3 en 1.8.2 wordt specifiek weergegeven welke subvaardigheden zijn getoetst.

#### ***Globale ERK beschrijving gespreksvaardigheid***

De toets gespreksvaardigheid is geconstrueerd aan de hand van taalactiviteiten en spreesituaties zoals beschreven in het ERK voor de niveaus A1, A2, B1. Het ERK maakt onderscheid tussen spreekvaardigheid (monologisch spreken) en gespreksvaardigheid (dialogisch spreken). In Taalprofielen (2015), worden voor beide vaardigheden op alle ERK-niveaus *can-do* descriptoren onderscheiden op de volgende schalen: woordenschat, grammaticale correctheid, vloeiendheid, uitspraak, samenhang, afstemming op doel en publiek. Bij gespreksvaardigheid zijn er ook *can-do* descriptoren voor interactie en interactiestrategieën. In Tabel 1.2 staat per gemeten ERK-niveau weergegeven wat taalgebruikers op het gebied van interactie en productie beheersen.

Tabel 1.2: ERK – beschrijving gesproken taal

	A1	A2	B1
Interactie	Ik kan deelnemen aan een eenvoudig gesprek, wanneer de gesprekspartner bereid is om zaken in een langzamer spreektempo te herhalen of opnieuw te formuleren en mij helpt bij het formuleren van wat ik probeer te zeggen. Ik kan eenvoudige vragen stellen en beantwoorden die een directe behoefte of zeer vertrouwde onderwerpen betreffen.	Ik kan communiceren over eenvoudige en alledaagse taken die een eenvoudige en directe uitwisseling van informatie over vertrouwde onderwerpen en activiteiten betreffen. Ik kan zeer korte sociale gesprekken aan, alhoewel ik gewoonlijk niet voldoende begrijp om het gesprek zelfstandig gaande te houden	Ik kan de meeste situaties aan die zich kunnen voordoen tijdens een reis in een gebied waar de betreffende taal wordt gesproken. Ik kan onvoorbereid deelnemen aan een gesprek over onderwerpen die vertrouwd zijn, of mijn persoonlijke belangstelling hebben of die betrekking hebben op het dagelijks leven (bijvoorbeeld familie, hobby's, werk, reizen en actuele gebeurtenissen).
Productie	Ik kan eenvoudige uitdrukkingen en zinnen gebruiken om mijn woonomgeving en de mensen die ik ken, te beschrijven	Ik kan een reeks uitdrukkingen en zinnen gebruiken om in eenvoudige bewoordingen mijn familie en andere mensen, leefomstandigheden, mijn opleiding en mijn huidige of meest recente baan te beschrijven.	Ik kan uitingen op een simpele manier aan elkaar verbinden, zodat ik ervaringen en gebeurtenissen, mijn dromen, verwachtingen en ambities kan beschrijven. Ik kan in het kort redenen en verklaringen geven voor mijn meningen en plannen. Ik kan een verhaal vertellen, of de plot van een boek of film weergeven en mijn reacties beschrijven.

### *Globale ERK beschrijving Luistervaardigheid*

In Taalprofielen (SLO, 2015) worden de volgende schalen uitgewerkt voor luistervaardigheid: woordenschat en zinsbouw, tempo en articulatie, tekstlengte, en luisterstrategieën. De eerste drie schalen betreffen tekstkenmerken, de laatste gaat om de middelen die de luisteraar inzet om zoveel mogelijk van een tekst te kunnen begrijpen, ook als er onbekende woorden worden gebruikt. Bekendheid met het onderwerp en de pregnantie van de context zijn daarbij belangrijke factoren. In Tabel 1.3 wordt de globale ERK-beschrijving voor luisteren weergegeven voor het A1-, A2- en B1-niveau.

Tabel 1.3: ERK beschrijving luistervaardigheid

	A1	A2	B1
Luisteren	Ik kan vertrouwde woorden en basiszinnen begrijpen die mezelf, mijn familie en directe concrete omgeving betreffen, wanneer de mensen langzaam en duidelijk spreken.	Ik kan zinnen en de meest frequente woorden begrijpen die betrekking hebben op gebieden die van direct persoonlijk belang zijn (bijvoorbeeld basisinformatie over mezelf en mijn familie, winkelen, plaatselijke omgeving, werk). Ik kan de belangrijkste punten in korte, duidelijke eenvoudige boodschappen en aankondigingen volgen.	Ik kan de hoofdpunten begrijpen wanneer in duidelijk uitgesproken standaardtaal wordt gesproken over vertrouwde zaken die ik regelmatig tegenkom op mijn werk, school, vrije tijd enz. Ik kan de hoofdpunten van veel radio- of tv-programma's over actuele zaken of over onderwerpen van persoonlijk of beroepsmatig belang begrijpen, wanneer er betrekkelijk langzaam en duidelijk gesproken wordt.

### *Globale ERK beschrijving leesvaardigheid*

In het ERK worden de beheersingsniveaus voor leesvaardigheid gedefinieerd aan de hand van tekstkenmerken als woordfrequentie, visuele presentatie, tekstlengte en leesstrategieën. In Tabel 1.4 wordt de ERK-beschrijving voor lezen van de 3 getoetste niveaus weergegeven.

Tabel 1.4: ERK Beschrijving leesvaardigheid

	A1	A2	B1
Lezen	Ik kan vertrouwde namen, woorden en zeer eenvoudige zinnen begrijpen, bijvoorbeeld in mededelingen, op posters en in catalogi.	Ik kan zeer korte eenvoudige teksten lezen. Ik kan specifieke voorspelbare informatie vinden in eenvoudige, alledaagse teksten zoals advertenties, folders, menu's en dienstregelingen en ik kan korte, eenvoudige, persoonlijke brieven begrijpen.	Ik kan teksten begrijpen die hoofdzakelijk bestaan uit hoogfrequente, alledaagse of aan mijn werk gerelateerde taal. Ik kan de beschrijving van gebeurtenissen, gevoelens en wensen in persoonlijke brieven begrijpen.

## 1.5 Gespreksvaardigheid: beschrijving en theoretische achtergrond

Mondelinge taalvaardigheid is het resultaat van verschillende componenten: uitspraak, vloeiendheid (inclusief tempo), lexicale en grammaticale kennis (De Jong, Steinell, Florijn, Schoonen, & Hulstijn, 2012). Kormos (2014) onderscheidt drie belangrijke verschillen tussen taalproductie in de eerste taal en in de vreemde taal: onvolledige kennis van de vreemde taal, invloed van de eerste taal op de productie in de vreemde taal en een lager spreektempo in de vreemde taal. Het is daarom essentieel dat vreemdetaallearners communicatieve strategieën leren inzetten om problemen op te lossen zoals gebrek aan taalmiddelen, tijdsdruk, ervaren gebreken in de eigen taalproductie en in het decoderen van de taalproductie van de gesprekspartner (Dörnyei & Scott 1997). Op deze manier

ervaren ze dat zich verstaanbaar maken in de vreemde taal belangrijk is voor succesvolle interactie in de vreemde taal, belangrijker dan bijvoorbeeld accentloos spreken (Saito, Webb, Trofimovich, & Isaacs, 2016). Communicatieve strategieën zijn de middelen die vreemdetaallearners inzetten om taaluitingen aan te passen om ze zo begrijpelijker te maken voor de gesprekspartner (Gass & Mackey, 2006; Long, 1996).

Dit alles heeft implicaties voor het lesgeven en het toetsen van spreekvaardigheid. In de vreemde-taalles zou met spreektaken moeten worden geoefend die zoveel mogelijk de authentieke communicatieve situaties in de vreemde taal nabootsen. Spreekvaardigheid zou vervolgens moeten worden getoetst aan de hand van vergelijkbare spreektaken, en de beoordeling daarvan zou gebaseerd moeten worden op de prestaties in de verschillende taalcomponenten. Zoals reeds meermaals genoemd bieden de beschrijvingen van het ERK, gebaseerd op deze inzichten uit de tweedetaalverwervingstheorie, een belangrijk aanknopingspunt voor het structureren van lessen en ontwikkelen van toetsen.

Sinds 1980 zijn verschillende theoretische modellen voor communicatieve taaltoetsen ontwikkeld, gebaseerd op modellen van communicatieve taalcompetenties. Fulcher (2010) laat zien hoe de eerste modellen (Canale & Swain 1980, Bachman 1990) in de loop der jaren verder zijn ontwikkeld en verfijnd. Er werd niet alleen gestreefd naar een meer nauwkeurige beschrijving van de linguïstische (de gebruikte grammaticale structuren en het lexicon) en sociolinguïstische competentie (bijvoorbeeld het gebruik van een formeel/informeel register), maar ook naar de integratie met de interactionele competentie (bijvoorbeeld het vermogen om adequaat te reageren en om een gesprek gaande te houden). Tegelijkertijd toonden onderzoeken over tweede-taalverwerving een overtuigend verband tussen gesproken interactie en taalverwerving, waarbij interactie een leidend principe bleek in de verwerving van woordenschat en taalproductie (Long, 1996; Mackey, 1999). Interactie kreeg dus een prominentere rol zowel bij de modellen voor communicatieve taalcompetenties (zie o.a. Celce-Murcia, 2007) als bij de modellen voor communicatieve taaltoetsen. Ondanks deze prominentere rol maken veel onderzoeken naar taaltoetsen waarin interactie tussen de sprekers wordt gemeten gebruik van criteria voor correctheid, fluency, communicatieve vaardigheid, coherentie en complexiteit, maar niet voor interactie (Fulcher, 2010).

Het probleem bij het meten van interactionele competentie is dat het lastig is om criteria te definiëren die deze interactionele competentie kunnen meten als een individuele competentie. De individuele bijdrage aan een gesprek is geslaagd als de spreker in staat is om de boodschap of de vraag van de gesprekspartner te interpreteren en hij/zij hierop een passende reactie kan formuleren. Met andere woorden, het gesprek is een gemeenschappelijk product, terwijl de kandidaten later individueel worden beoordeeld. Doordat interactie ontstaat in dialoog met de gesprekspartner(s) worden de individuele bijdragen door bijdragen van anderen beïnvloed. Dit betekent bijvoorbeeld dat een meer vaardige spreker niet al zijn taalcompetentie zou kunnen tonen als de gesprekspartner in de test minder vaardig is (zogenaamde interlocutor-effecten), en dat het problematisch is om de minder vaardige en/of minder dominante spreker op interactie te beoordelen. Vaak wordt daarom gekozen voor het zogenaamde OPI (Oral Proficiency Interview) model in plaats van voor parentests. Dit laat de recente review van Sandlund, Sundqvist en Nyroos (2016) bijvoorbeeld zien. Binnen het OPI model voert de kandidaat een gesprek met de testleider, waardoor beïnvloeding van een mindere spreker geen negatieve invloed kan hebben op de prestaties. Wanneer de testleider daarbij gebruik zou maken van een gescripte, gestandaardiseerde versie van het gesprek, zouden de gesprekken bovendien een hoge mate van uniformiteit hebben. Direct gevolg daarvan is dat OPI

modellen weinig gevarieerde vormen van interactie meten en daarmee staat de authenticiteit van de gemeten interactievaardigheid ter discussie.

Vanwege de evidente voordelen zijn voor deze peiling taken volgens het OPI model in overweging genomen. Daarbij was de verwachting dat in deze peiling vooral het lage competentieniveau van de leerlingen problematisch zou kunnen zijn voor het beoordelen van interactie en in mindere mate de interlocutor-effecten en de complexiteit van het individueel beoordelen van de bijdrage van een leerling in een gezamenlijk product. In haar studies naar de parentests in de First Certificate van Cambridge, merkt Galaczi (2008, 2014) bijvoorbeeld op dat bij de A2 interactietaak de examiner veelvuldig ingrijpt en veel ondersteuning biedt. Dit was voor haar de reden om interactie in deze testvorm te bestuderen vanaf het B1 niveau. In een onderzoek naar de toepassing van parentests bij jonge taalleerders (Butler & Zeng, 2014) was de conclusie dat jongere leerlingen (9-10 jaar) meer moeite hadden met gezamenlijke ontwikkeling van het gesprek en meer vaste formules gebruikten om een beurt te nemen dan oudere leerlingen (11-12 jaar). Het voordeel van parentests, namelijk het uitlokken van meer gevarieerde vormen van interactie, lijkt dus niet optimaal te kunnen worden benut bij jongere leerlingen (van 9-10 jaar). De oudere leerlingen lijken de taak al beter te kunnen uitvoeren.

Zowel OPI modellen als parentests kennen dus hun voordelen en beperkingen. Uiteindelijk is gekozen voor parentests en niet voor meting volgens de OPI systematiek. In het licht van de huidige peiling bij relatief jonge leerlingen kent het OPI-model namelijk duidelijke nadelen. Een nadeel is bijvoorbeeld dat een dergelijke toetsafname niet rechtstreeks aansluit bij de eerdergenoemde communicatieve aanpakken waarin authentieke situaties zoveel mogelijk moeten worden nagebootst (Rose, 2016) en ook niet bij een klassensituatie waarin alle leerlingen gestimuleerd zouden moeten worden om zoveel mogelijk te praten. Het onvoldoende aansluiten bij authentieke situaties, waarbij interactie met een (onbekende) Engelssprekende leeftijdgenoot realistischer wordt geacht dan met een onbekende volwassene, en het niet tegemoetkomen aan de ecologische validiteit die in de huidige peiling prioriteit had waren beslissende factoren in de uiteindelijke keuze voor een opzet waarin gesprekken tussen twee leerlingen plaatsvinden. Gestreefd werd naar taken met zo authentiek mogelijke interacties tussen peers. Naast deze *ecologische factoren* speelden *psychologische factoren* een belangrijke rol: de leerlingen zijn relatief jong (11-12 jaar) en de testleider zou per definitie een onbekende volwassene zijn. Uitgangspunt voor de meting van gespreksvaardigheid was dat de leerlingen zich niet bedreigd of geremd moesten voelen, maar zo ontspannen en spontaan mogelijk meedoen aan een gesprek in de vreemde taal met een klasgenoot, zonder te denken dat ze beoordeeld gingen worden. Er is daarom bewust gekozen voor een opzet waarbij de testleider niet als gesprekspartner deelnam aan het gesprek.

In een recent onderzoek naar mondelinge vaardigheid Engels aan het eind van het voortgezet onderwijs zijn leerlingen in tweetallen getest door een testleider die ook aan het gesprek meedeed (Fasoglio & Tuin, 2017). Een tweede testleider was aanwezig als beoordelaar. Dit is het model dat bijvoorbeeld in de Cambridge toetsen wordt ingezet. In tegenstelling tot deze opzet heeft in de huidige peiling met basisschoolleerlingen de testleider zich beperkt tot het geven van gestandaardiseerde instructie en nam deze alleen de rol op zich van gesprekspartner wanneer een van de leerlingen het gesprek niet en de andere het gesprek wel kon voeren. Dit gebeurde overigens in alle gevallen pas na de eerste taak, waardoor de leerlingen al een beetje aan de situatie (waarin ze met iemand Engels moesten spreken) en de testleider gewend waren. De keuze om de rol van de testleider zoveel mogelijk te beperken was het gevolg van het psychologische uitgangspunt dat

leerlingen zich veilig en op hun gemak moesten voelen. Tenslotte waren er *logistieke factoren* waar rekening mee gehouden moest worden. Door leerlingen gesprekken te laten voeren, zouden zoveel mogelijk leerlingen kunnen worden getest in de beschikbare tijd.

De drie genoemde *ecologische, psychologische* en *logistieke* factoren hadden een hoge prioriteit. Voor het slagen van de gesprekstaking leek het met name van groot belang dat de taak aansloot bij de belevingswereld van leerlingen (ook in vorm) en hun niveau en dat hun welbevinden hoog was. Daarnaast zijn er ook taken ontwikkeld op B1-niveau, omdat er werd aangenomen dat een deel van de leerlingen een hoger beheersingsniveau zou hebben. Zoals hierboven beschreven bieden parentests op B1-niveau juist voordelen voor interactie. De parentests hadden als bijkomend voordeel dat de interactievaardigheid van de leerlingen specifiek in kaart kon worden gebracht, waardoor de gesprekstaking voor twee doeleinden bruikbaar waren: voor individuele (ERK)-beoordeling en voor een parenbeoordeling. Mogelijke interlocutorproblemen zouden daarbij kunnen worden geminimaliseerd door de vorming van tweetallen bestaande uit leerlingen met een vergelijkbaar niveau Engels. In de pilots zijn hiertoe twee varianten uitgetoetst: de vorming van tweetallen op basis van 1) random selectie en 2) niveau Engels. Als indicator van dit niveau Engels is in een prepilot het voorlopige schooladvies en in de pilot een leerkrachtinschatting van het niveau Engels van de leerling gebruikt. Volgens de toetsleiders waren beide indicatoren geen goede voorspeller voor de interactievaardigheid van de leerlingen. Bij gebrek aan een goede indicator voor het niveau Engels van de leerlingen is gekozen voor random selectie van de tweetallen. Vanuit het oogpunt van validiteit had deze keuze voor random selectie ook de voorkeur: we wilden weten wat de gespreksvaardigheid van leerlingen was in een gesprek met een willekeurige klasgenoot en niet alleen in een gesprek met een leerling met een vergelijkbaar niveau

Het gevolg van de keuze voor de parentest was dat de rollenspellen zo geconstrueerd moesten worden dat de leerlingen met minimale talige input (in het Nederlands) gesprekken met identiek verloop en vergelijkbare output konden voeren. De Nederlandse input is geminimaliseerd om te voorkomen dat leerlingen zouden gaan vertalen, in plaats van zelf zinnen te construeren. Input in het Engels was niet wenselijk, hiermee zouden woorden worden “weggegeven”. In de constructie van de rollenspellen is daarom zoveel mogelijk gebruik gemaakt van visueel materiaal.

Tenslotte dient te worden opgemerkt dat productieve taalvaardigheden in vreemde talen moeilijker aan te leren zijn in een schoolcontext dan de receptieve. Vooral het oefenen van interactie in de vreemde taal in de klas vormt een uitdaging voor de leerkracht, niet alleen vanwege tijdgebrek of te grote groepen, maar ook door het gebrek aan een natuurlijke communicatieve situatie die leerlingen ‘dwingt’ om het Engels te gebruiken om met elkaar te praten. De leerkracht moet motiverende activiteiten construeren, die aansluiten bij de wereld van de kinderen, met bijvoorbeeld onvolledige informatie (information gaps), of met een spelelement. Dit alles vraagt specifieke deskundigheid en daarom wordt gespreksvaardigheid vaak amper in de klas geoefend. Veel leerlingen zijn dus niet gewend om in tweetallen spontane gesprekken te voeren. Een meer gebruikte vorm is het laten voorlezen van dialogues, waarbij dus geen sprake is van spontaan taalgebruik en werkelijke interactie.

### **1.5.1 Gespreksvaardigheid: instrumentarium**

Juist vanwege het belang voor leerlingen dat ze mondeling kunnen communiceren in het Engels is gespreksvaardigheid gemeten met behulp van taken die zoveel mogelijk aansloten bij hun

belevingswereld en die gesprekssituaties nabootsen waarin leerlingen mogelijk terechtkomen. De prestaties van de leerlingen zijn beoordeeld op correctheid en fluency en daarnaast zijn ook criteria gedefinieerd om de gesprekken op interactie te beoordelen.

### **Gespreksvaardigheid - 3 gesprekstaken**

De toets gespreksvaardigheid bevatte 3 taken. De taken werden in principe uitgevoerd door twee leerlingen. Aangezien de taken opliepen in moeilijkheid, besloot de testleider na iedere taak of één of beide leerling(en) door mocht/mochten naar de volgende taak. Niet alle leerlingen deden dus alle drie gesprekstaken.

De drie taken kunnen als volgt worden beschreven:

- Taak 1 ('Wie ben ik?') is ontworpen als A1-taak, met A2-elementen. De taak is een rollenspel, waarbij leerlingen ieder een praatplaat over hun fictieve persoon krijgen en dan aan de hand van vraagkaartjes proberen informatie van de ander te krijgen en informatie over hun fictieve persoon te geven. De fictieve personen zijn meisjes en jongens in dezelfde leeftijdscategorie als de geteste leerlingen. Binnen de taak spreken de leerlingen met elkaar over eenvoudige, alledaagse onderwerpen. Taak 1 wordt in principe door beide leerlingen uitgevoerd.
- Taak 2 ('Wat spreken we af?') is ontworpen als A2-taak, met A2+ elementen. De leerlingen blijven dezelfde fictieve persoon als bij Taak 1 en proberen een afspraak te maken (een alledaagse bezigheid). De leerlingen krijgen ieder een eigen agenda en ontvangen beurtelings kaartjes met daarop vragen/antwoorden. Gezien hun volle agenda moeten ze onderhandelen om tot de afspraak te komen en samenhang tussen eenvoudige zinsconstructies weergeven. Taak 2 kan door twee leerlingen worden uitgevoerd of door één leerling en een testleider. (N.B. de andere leerling werkt dan in een voor dit doel samengesteld schrijfboekje met woordpuzzels, woordspelletjes en een schrijfboekje).
- Taak 3 ('Wat vind je ervan?') is ontworpen als A2+/B1-taak. De leerlingen gaan in debat over de vraag of het goed of slecht is om veel te gamen. De ene leerling is voor, de ander tegen en deze posities moeten zij onderbouwen aan de hand van informatie op kaartjes, die door de testleider worden verdeeld. Taak 3 kan door twee leerlingen worden uitgevoerd of door één leerling en een testleider. (N.B. de andere leerling werkt dan in een voor dit doel samengesteld schrijfboekje).

#### **1.5.2 Gespreksvaardigheid: getoetste vaardigheden en subvaardigheden**

Een gesprek is geslaagd als het gaat over een bepaald vastgesteld gespreksonderwerp of uitwisseling van informatie, het gekenmerkt wordt door regelmatige beurtwisseling (beide sprekers hebben voldoende inbreng in het gesprek), het een gestructureerd verloop heeft en er wordt voortgebouwd op wat de gesprekspartner van een ongeveer gelijkwaardig taalvaardigheidsniveau heeft ingebracht. Basis voor de geconstrueerde taken waren de "can do"-statements binnen het ERK (Taalprofielen, 2015). Per taak wordt hieronder beschreven hoe de opdrachten de oplopende ERK-niveaus representeren:



### Taak 1: Wie ben ik?

De eerste opdracht is hoofdzakelijk een A1 taak, met enkele A2 kenmerken. De *can do* descriptors die opgenomen zijn in de toetsmatrijs zijn:

- **A1:**
  - Kan zichzelf en anderen voorstellen en reageren als iemand voorgesteld wordt. Kan op een eenvoudige manier groeten en afscheid nemen.
  - Kan eenvoudige vragen stellen en beantwoorden, en eenvoudige uitspraken doen en beantwoorden op het gebied van primaire behoeften of over zeer vertrouwde, concrete onderwerpen.
  - Kan vragen beantwoorden en stellen over zichzelf en over anderen, waar zij wonen, wie zij kennen, wat zij bezitten.
  - Kan zeggen wat hij of zij leuk of lekker vindt en wat niet.
- **A2:**
  - Kan beperkte informatie uitwisselen over vertrouwde en alledaagse zaken van praktische aard.
  - Kan beperkte informatie uitwisselen over vertrouwde en alledaagse zaken van praktische aard.
  - Kan informatie van persoonlijke aard vragen en geven.

### Taak 2: Wat spreken we af?

De tweede opdracht is een A2-taak die ook A2+ uitingen uitlokt. De volgende *can do* descriptors zijn in de toetsmatrijs opgenomen:

- **A2:**
  - Kan eenvoudige alledaagse beleefdheidsvormen gebruiken om anderen te begroeten en aan te spreken.
  - Kan uitnodigingen doen, op uitnodigingen ingaan of deze afslaan, suggesties opperen.
  - Kan op eenvoudige wijze voorkeur en mening uitdrukken over vertrouwde alledaagse onderwerpen.
  - Kan communiceren over eenvoudige en alledaagse taken in eenvoudige bewoordingen om dingen te vragen en te verschaffen, eenvoudige informatie te verkrijgen en te bespreken wat er vervolgens moet gebeuren.
  - Kan afspraken maken.
  - Kan vragen stellen en beantwoorden over wat men op het werk en in de vrije tijd doet.
- **A2+:**
  - Kan te kennen geven het (on)eens te zijn met anderen.
  - Kan deelnemen aan korte gesprekken over belangwekkende onderwerpen in een alledaagse context.
  - Kan relevante informatie uitwisselen en zijn of haar mening geven over praktische problemen wanneer dat rechtstreeks gevraagd wordt, mits hij of zij enige hulp krijgt bij het formuleren en indien nodig om herhaling van belangrijke punten kan vragen.
  - Kan bespreken wat er vervolgens moet gebeuren, daarbij suggesties doen en beantwoorden, en aanwijzingen vragen en geven.
  - Kan vragen stellen en beantwoorden over gewoonten en routines.

- Kan vragen stellen en beantwoorden over tijdverdrijf en vroegere of toekomstige activiteiten.

### Taak 3: Wat vind jij ervan?

De derde opdracht is een A2+/B1 taak. De volgende *can do* descriptorren zijn in de toetsmatrijs opgenomen:

- **A2+:**
  - Kan te kennen geven het (on)eens te zijn met anderen.
  - Kan deelnemen aan korte gesprekken over belangwekkende onderwerpen in een alledaagse context.
- **B1:**
  - Kan gevoelens uiten en op gevoelens van anderen reageren.
  - Kan persoonlijke standpunten, commentaar en meningen geven of er om vragen in discussies over belangwekkende onderwerpen.
  - Kan zijn of haar meningen en reacties overbrengen met betrekking tot oplossingen voor problemen of praktische vragen.
  - Kan op beleefde wijze een overtuiging, een mening, instemming en afkeuring uitdrukken.
  - Kan een standpunt helder overbrengen, maar heeft er moeite mee om in debat te gaan.
  - Kan deelnemen aan een alledaagse formele discussie over vertrouwde onderwerpen die gaan om de uitwisseling van feitelijke informatie, het ontvangen van en reageren op aanwijzingen of het bespreken van oplossingen voor praktische problemen.
  - Kan zijn of haar mening geven en reacties overbrengen met betrekking tot mogelijke oplossingen van problemen en het nemen van praktische beslissingen, en daarbij beknopte redenen en verklaringen geven.

#### 1.5.3 Fasen in de ontwikkeling van de gesprekstaken

De ontwikkeling van de gesprekstaken kende meerdere fasen. Aanvankelijk werden drie uitgebreide gesprekstaken ontwikkeld. Deze zijn in juni/juli 2017 op vier scholen uitgetoetst in een try-out met 40 leerlingen. Het betrof twee VVTO scholen en twee reguliere basisscholen. De leerlingen die deelnamen aan de try-out hadden VO-adviezen die uiteenliepen van BB/KB tot gymnasium. Aan de hand van de bevindingen bij deze try-outs zijn de opdrachten steeds verder aangescherpt.

Het streven naar authentieke taken leidde in eerste instantie tot zo min mogelijk gescripte taken. De taken bestonden aanvankelijk uit meer open vragen, maar deze werden na de diverse try-outs omgebouwd tot meer gesloten vragen om te voldoen aan de voorwaarde dat zij goed meetbare en vergelijkbare taaluitingen zouden uitlokken. Om te voorkomen dat de taak een vertaalopdracht zou worden werd zoveel mogelijk gebruik gemaakt van plaatjes en emoticons, in plaats van woorden of zinnen. Woorden die gegeven werden waren in het Nederlands.

Tijdens de pretest in september (zie hoofdstuk 4 en 5), die is uitgevoerd op vijf scholen in zes klassen, bleek dat de opdrachten verder ingekort moesten worden in verband met de beperkte afnametijd.

#### 1.5.4 Selectie van leerlingen voor de gesprekstaken

Met behulp van gespreksvaardigheid is getracht de vaardigheid van leerlingen te meten om gezamenlijk een gesprek in het Engels te voeren. Per school werden de gesprekstaken in principe bij

zes leerlingen afgenomen (in tweetallen) en werd de ankertaak (de spreektaak), per duo, door één van de twee leerlingen uitgevoerd.<sup>13</sup> Bij de selectie van leerlingen deden zich twee aandachtspunten voor:

1. Allereerst diende te worden bepaald welke leerlingen zouden moeten worden geselecteerd en hoe. Met name de vraag of selectie op basis van niveau of random zou moeten plaatsvinden was relevant. Een random selectie zou conclusies toelaten over de vraag of leerlingen in staat zijn om met andere leerlingen uit de klas (groep 8) een gesprek te voeren. Grote niveaueverschillen zouden echter kunnen leiden tot de eerdergenoemde interlocutor-effecten, die een adequate meting van het mondelinge vaardigheidsniveau van de leerling zouden kunnen beïnvloeden. Besloten is om in een pretest, uitgevoerd in september, beide varianten toe te passen. Analyses en ervaringen van de testleiders lieten echter geen duidelijke voorkeur voor een selectie op niveau zien, waardoor gekozen is voor een random selectie van de leerlingen.
2. Aangezien het maken van video-opnamen voor beoordelingsdoeleinden noodzakelijk was, waren onderzoekers genoodzaakt actieve toestemming aan ouders te vragen voor de mondelinge taken. Een volledig gerandomiseerde selectie is hierdoor niet mogelijk geweest (in paragraaf 2.5.2 staan de gevolgen van deze selectieprocedure nader verantwoord). Scholen zijn eerst gevraagd of zij bereid waren de ouders te vragen om actieve toestemming voor deze taken (1 school was hiertoe niet bereid met als gevolg dat hier geen mondelinge taken zijn afgenomen). Een gebruiksklare voorbeeldbrief is hiertoe verstrekt. Uiterlijk een week voor de afnamedatum is de school herinnerd aan het verzoek de actieve toestemming te vragen aan ouders. Indien voorafgaand bij de onderzoekers en/of testleider nog niet bekend was welke leerlingen toestemming hadden, heeft de testleider dit op de afnamedag besproken met de leerkracht en heeft hij/zij drie tweetallen random geselecteerd uit de lijst met leerlingen die ouderlijke toestemming hadden. Deze koppels zijn voorgelegd aan de leerkracht, die bezwaar kon maken ingeval van gedragsmatige problematiek (zoals leerlingen die elkaar niet mogen, faalangst). Niveau speelde hierbij geen rol. Ingeval van ingeschatte problemen door de leerkracht, waren vooraf ook al reservekoppels door de testleider geselecteerd. Binnen ieder koppel werd tevens de leerling geselecteerd die de ankertaak (=de spreektaak) ging doen. Ook dit is willekeurig gebeurd, waarbij de testleider er wel rekening mee mocht houden of leerlingen zich op hun gemak voelden. Dit vooral omdat het de eerste taak betrof en het welbevinden van de leerling voorop stond. Ook hier speelde niveau geen rol.

### 1.5.5 Gesprekstaken: afname en procedure

Afname van de gespreksvaardigheidstaken vond plaats in duo's. Per klas namen in principe zes leerlingen deel aan de gesprekstaken, die elk ongeveer 20 minuten duurden. De leerlingen werden tijdens de taken leesvaardigheid, luistervaardigheid en woordenschat door de testleider uit de klas gehaald om in een rustige afgesloten ruimte de taken mondelinge taalvaardigheid uit te voeren. Ten behoeve van de beoordeling zijn video opnamen gemaakt.

---

<sup>13</sup> Dit is op de meeste scholen gerealiseerd. Op scholen met relatief weinig groep-8 leerlingen was dit soms niet het geval omdat er op de dag van de afname minder dan zes geschikte leerlingen aanwezig waren door ziekte en/of het ontbreken van ouderlijke toestemming (zie paragraaf 2.5.2.). Op één school hebben geheel geen gesprekstaken plaatsgevonden omdat de school de toestemmingsbrief niet aan ouders had meegegeven.

Leerlingen zaten – zo veel mogelijk – tegenover elkaar. Om de uniformiteit van de afname te bevorderen is altijd gestart met een gestandaardiseerde instructie door één van de testleiders op film waarin het rollenspel werd uitgelegd. Deze korte instructiefilm werd gevolgd door een kort filmpje waarin twee voorbeeldleerlingen (studenten) de taak uitvoerden, inclusief de geïnstrueerde gedragingen als een leerling het niet weet (elkaar om hulp vragen). Om geen antwoorden weg te geven, is hiervoor een subtaak uit Taak 1 gebruikt die na de pretest is afgefallen. Na afloop van de filmpjes kregen de leerlingen een gesprekskaart en moesten zij elkaar vragen stellen en antwoord geven in het Engels aan de hand van dezelfde vier kaartjes die de testleider hen beiden om beurten gaf. Op deze kaartjes stonden vooral visuele aanwijzingen (plaatjes en pictogrammen), indien nodig aangevuld met minimale talige (Nederlandse) uitingen. In de instructiefilm werd leerlingen gevraagd alleen Engels te spreken, en zo nodig gebaren of andere omschrijvingen te gebruiken. Alleen als de twee leerlingen er samen echt niet uit kwamen schoot de testleider te hulp door bijvoorbeeld te vragen of de leerling het wellicht in andere woorden kon omschrijven. Als een leerling er echt niet uit kwam, mocht hij/zij het in het Nederlands zeggen, zodat de andere leerling kon reageren.

Omdat de taken opliepen in moeilijkheid, diende de testleider na afloop van de eerste gesprekstaak te besluiten of de leerlingen door konden naar de volgende taak of dat één (of beide) leerling(en) moest(en) stoppen, omdat het niveau te hoog lag. Uitgangspunt daarbij was om leerlingen door te laten gaan, *tenzij* er sterke twijfels waren. Twijfels konden ontstaan doordat de leerling a) meermaals erg lang moest nadenken en lange stiltes liet vallen (15 seconden), b) (bijna) alleen verbasterd Nederlands gebruikte, c) zich niet prettig leek te voelen/dichtklapte en/of d) meermaals geen vraag stelde en/of antwoord gaf. Leerlingen die niet doorgingen met de volgende taak konden aan de slag gaan met het activiteitenboekje (het schrijfboekje). Testleiders hebben zich ingespannen om deze leerlingen niet de indruk te geven dat de keuze gebaseerd was op niveau.

Bij de tweede taak gingen de leerlingen door als de fictieve personages uit Taak 1. Ze kregen nu een agenda uitgereikt met daarop de afspraken en verplichtingen die ze hadden. De leerlingen kregen weer om beurten kaartjes aangereikt met daarop vragen of antwoorden. In tegenstelling tot Taak 1 werden niet dezelfde vragen door beide leerlingen gesteld en beantwoord. Na afloop besloot de testleider weer of de leerling(en) de volgende taak konden uitvoeren. De leerlingen die niet verder gingen, gingen werken in het schrijfboekje.

Bij de derde taak moest één van beide leerlingen de stelling innemen dat veel gamen goed is, de ander moest het tegendeel verdedigen. Er werd gewerkt met groene en rode kaartjes om de stellingname te benadrukken. De eerste ('groene') leerling kreeg een kaartje met een argument dat hij/zij zo goed mogelijk moest trachten te verwoorden. De andere ('rode') leerling kreeg dan eerst een kaartje met een reactie en vervolgens een kaartje met tegenargumenten. Op deze manier wisselden zij elkaar af.

De testleiders die de gesprekstakingen afnamen zijn van tevoren een dag getraind en hebben een protocol ontvangen met daarin gestandaardiseerde instructies, richtlijnen en praktische handreikingen. De testleiders zijn allen geselecteerd op ervaring en kennis met het ERK. Het grootste deel van de testleiders had bovendien (ruime) ervaring met kinderen. Daarnaast is zowel in de training als in het protocol benadrukt dat het welzijn van de leerlingen tijdens de taken voorop stond (zorgen voor ongedwongen sfeer, veel positieve bekrachtiging, positieve bejegening bij fouten ('never mind') en voorkomen van lange stiltes (max 15 seconden). Testleiders zijn nadrukkelijk geïnstrueerd geen inhoudelijke hulp in het Engels te geven. Om de druk bij leerlingen te verlagen zijn de taken door de testleiders voorgesteld als oefeningen waarmee wordt nagegaan hoe je Engels in

de klas kunt oefenen en of de leerlingen dit leuk vinden. Na afloop van de taken heeft de testleider bij de leerlingen nagevraagd hoe zij de mondelinge en schrifttaken hebben ervaren.

### 1.5.6 Beoordeling gesprekstaken

Om de gesprekken te beoordelen is gebruik gemaakt van twee beoordelingsformulieren, met ieder een andere doelstelling: 1) Een individuele beoordeling aan de hand van ERK niveaus en 2) een beoordeling van de interactie op paarniveau<sup>14</sup>. Voor informatie over de beoordelingen wordt verwezen naar paragraaf 5.2.3.

## 1.6 Woordenschat: beschrijving en theoretische achtergrond

### 1.6.1 Woordenschat: beschrijving

Woordenschat maakt integraal deel uit van elk van de afzonderlijke taalvaardigheden. In het ERK is woordenschat daarom niet opgenomen als afzonderlijke vaardigheid. Er zijn ook geen specifieke beschrijvingen voor woordenschat. Onderzoek laat echter zien dat woordenschat een belangrijke voorwaarde is voor succesvol oefenen van taalvaardigheid, inclusief mondelinge taalvaardigheid (Rose, 2016). In de lespraktijk binnen het primair onderwijs wordt er veel aandacht aan woordenschat besteed (Geurts & Hemker, 2013). Gezien het centrale belang van woordenschat en de sterke interactie met de andere vaardigheden is deze daarom wel afzonderlijk opgenomen in deze peiling.

De beoordeling van de woordenschat van tweedetaalverwervers kent vele vormen, die afhankelijk is van de definitie van woordkennis. Naast een globale inschatting van de omvang van woordenschat is er in toenemende mate aandacht voor de kwaliteit of niveaus van woordkennis (zie, bijvoorbeeld Laufer & Goldstein, 2004). Lexicale kennis kan variëren van zeer oppervlakkige bekendheid van woorden tot een correct gebruik van woorden in spontane taalproductie (Palmberg, 1987). Een belangrijk onderscheid dat wordt gemaakt in het toetsen van woordenschat is de bekendheid met de vorm van woorden en met de betekenis van woorden. Het verband tussen vorm en betekenis kan direct worden getoetst (Laufer & Nation, 1999), wanneer de leerling de juiste vorm moet produceren bij een gegeven betekenis, maar kan ook indirect worden getoetst aan de hand van associaties (Read, 1993). Daarnaast kan de richting van deze relatie worden onderscheiden – het produceren van de juiste vorm bij een gegeven betekenis, of het produceren van de juiste betekenis bij een gegeven vorm. De richting wordt doorgaans aangegeven met de termen actief (produceren van de juiste woordvorm vanuit de gegeven betekenis) en passief (produceren van de juiste betekenis vanuit de gegeven vorm). De woordenschat die taalgebruikers passief (of *receptief*) tot hun beschikking hebben is aanzienlijk groter dan de woordenschat die zij actief (of *productief*) gebruiken. Vooral binnen het gebruik van de tweede- of vreemde taal is dit een relevant onderscheid (zie Webb, 2008; Laufer, 1998). Het verschil tussen actieve en passieve woordenschat blijft bestaan met toenemende taalvaardigheid en is niet afhankelijk van de frequentie van de woorden (Webb, 2008). Het onderscheid tussen actieve en passieve woordkennis wordt door sommige onderzoekers beschreven als een duidelijk binair verschil (Meara, 1990), maar wordt door anderen gezien als een continuüm (Palmberg, 1987). In het breed geaccepteerde CATTs-model (Laufer & Goldstein, 2004)

---

<sup>14</sup> De beoordelingsformulieren zijn als Bijlage 1 en 2 toegevoegd

wordt een kwalitatieve dimensie toegevoegd aan het onderscheid tussen actief en passief. Zij onderscheiden binnen zowel actieve als passieve kennis het kunnen oproepen van woorden (Recall) en herkennen van woorden (Recognition). Laufer en Goldstein (2004) komen hiermee op vier categorieën van woordenschatverwerving, zoals aangegeven in Figuur 1.1, waarin L1 staat voor de moedertaal en L2 voor de tweede taal die wordt verworven. Paragraaf 1.6.2 bevat voorbeelden van opgaven voor het toetsen van deze niveaus.

	Recall	Recognition
Active (retrieval of form)	Supply the L2 word	Select the L2 word
Passive (retrieval of meaning)	Supply the L1 word	Select the L1 word

Figuur 1.1. Niveaus van woordenschat naar Laufer & Goldstein (2004), p407.

De vier categorieën van woordenschat kunnen worden gekoppeld aan een logische verwervingsvolgorde, waarbij het “passief” herkennen van vormen eerder wordt verworven dan het “actief” oproepen van woorden op basis van betekenis (Laufer & Goldstein, 2004). Hoewel er doorgaans een verband wordt gevonden tussen de omvang van de receptieve woordenschat en die van de productieve woordenschat, blijft het bij het vaststellen van het werkelijke beheersingsniveau van woordkennis zinvol om een onderscheid te maken tussen de vier verschillende categorieën en alle vier te meten. In deze peiling zijn er daarom opgaven uit de vier categorieën opgenomen in de opgaven voor woordenschat.

Hoewel formeel gezien losse woorden niet op de taalvaardigheidsniveaus van het ERK-niveau kunnen worden ingeschaald, is het wel mogelijk om met behulp van frequentielijsten en databases een inschatting te maken van de vocabulaire kennis van leerlingen (in termen van ERK-niveaus). Voor de niveaubepaling van woordenschat is voor deze peiling gebruik gemaakt van het English Vocabulary Profile (EVP). EVP maakt deel uit van English Profile van Cambridge University Press, waarin ERK-niveaus kwantitatief zijn ontleend aan het Cambridge Learner Corpus (CLC), een groeiend corpus van meer dan 100.000 scripts van deelnemers aan de Cambridge Exams. Naast het CLC maakt het EVP gebruik van het omvangrijke Cambridge English Corpus, waarin lesmateriaal voor Engels is opgenomen, gekoppeld aan de beoogde taalvaardigheidsniveaus. In EVP is op grond van deze zeer grote database van tweede-taalgebruikers van het Engels een betrouwbaar woordenschatprofiel afgeleid voor elk van de ERK-niveaus. In deze peiling wordt de beheersing van *productieve woordenschat* gepeild voor items die zijn geconstrueerd op de ERK-niveaus A1, A2 en B1, op basis van English Vocabulary Profile (EVP).

Het EVP maakt in de niveaubepaling geen onderscheid tussen receptieve en productieve vaardigheid. Uit het CLC blijkt op welk moment in het leerproces een woord bekend is en onderdeel gaat uitmaken van de productieve woordenschat van leerders uit het corpus. Om een representatieve beoordeling van de woordenschat mogelijk te maken is het bereik van *receptieve woordenschat* gepeild aan de hand van de band van de eerste 5000 woorden van de Corpus of Contemporary American English (COCA).

### 1.6.2 Woordenschat: instrumentarium

Zoals hierboven betoogd is bij de constructie van de woordenschattoets rekening gehouden met het CATTS-model, in zoverre dat er items uit de vier categorieën opgenomen zijn: (1) Productief Recall, (2) Productief Recognition, (3) Receptief Recall, (4) Receptief Recognition. Daarnaast was het uitgangspunt dat de toets een voldoende en bij de doelgroep aansluitende range in moeilijkheid van de te selecteren woorden had. Op basis van de door SLO beschreven verwachte ERK-uitstroomniveaus voor Engels in het PO is ervoor gekozen items te selecteren voor drie ERK-niveaus: A1, A2, B1.

Beperkende voorwaarde voor de toets was dat er zowel een trendvergelijking met de vorige peiling mogelijk moest zijn (hetgeen inhield dat de taak voldoende ankeritems zou bevatten) als een relatie met de prestaties in het VO moest kunnen worden gelegd (waardoor ook items van het Volgsysteem Voortgezet Onderwijs (VoVo) moesten worden toegevoegd).<sup>15</sup> Hoewel de peiling uitging van de CATTS-indeling, waren wij in de praktische uitvoering van de woordenschattaken op een paar punten genoodzaakt af te wijken van deze CATTS-methode. Twee afwijkingen van deze methode zijn van belang. Allereerst worden bij de officiële CATTS- methode 5 sets van 30 woorden bevraagd, waarbij ieder woord ook in de vier hierboven beschreven modaliteiten wordt voorgelegd (leidend tot in totaal 600 items). In de huidige peiling was voor iedere vaardigheid slechts 30 minuten beschikbaar. Dit hield in dat er maximaal 36 woorden per taak (of 37 in sommige taken) konden worden bevraagd. Er is voor gekozen om per leerling 36 (of 37) unieke woorden te bevragen en in totaal 110 unieke woorden. Ieder uniek woord is slechts in één van de vier categorieën getoetst en niet in alle vier de categorieën. Ten tweede wordt de CATTS-methode volledig in de tweede taal afgenomen. Er is voor deze peiling besloten hiervan af te wijken, omdat de moeilijkheid van de opgave voor beginnende leerlingen niet in de bevraging (de stam of de alternatieven) dient te zitten, maar in het bevraagde woord.

De geselecteerde items zijn verdeeld over zes blokken en iedere toets bestond uit twee blokken met items. Tabel 1.4 laat het aantal gebruikte items zien. Er wordt aangegeven hoeveel ankers en nieuwe items zijn gebruikt, hoeveel van deze ankers en nieuwe items per blok zijn geselecteerd en hoeveel productieve en receptieve items zijn gebruikt. Leerlingen hebben toetsen gemaakt bestaande uit 36 of 37 items, opgebouwd uit twee blokken. Uit Tabel 1.5 kan worden afgeleid dat de verhouding receptieve en productieve items ongeveer 1/3 versus 2/3 was. Er waren 43 items geconstrueerd op A1-niveau, 42 op A2-niveau en 25 op B1-niveau.

---

<sup>15</sup> Deze relatie met prestaties in het VO was geïnitieerd door het consortium, omdat het consortium voor onderzoeksdoeleinden wilde kunnen nagaan in hoeverre de vaardigheidsniveaus Engels van leerlingen in groep 8 gerelateerd zijn aan die van leerlingen in het eerste jaar van het voortgezet onderwijs, daarbij controlerend voor hun voorlopige schooladvies.

Tabel 1.5: Aantallen en typen items in de woordenschattaak

Aantal items	Totaal	Per blok
Ankers uit de vorige peiling	33	4 of 5
VoVO ankers	30	5 of 6
Nieuw	47	8
Totaal	110	
Waarvan:		
<i>Productief</i>	70	
Productive recall	21	
Productive recognition	49	
<i>Receptief</i>	40	
Receptive recall	6	
Receptive recognition	34	

### Voorbeelditems

Door uit te gaan van het CATTs-model zijn vier categorieën bevestigd. Per type categorie worden hieronder de richtlijnen vanuit CATTs beschreven en worden voorbeeld-items getoond.

#### Productive Recall

Voor productive recall geldt de volgende aanwijzing:

In each sentence **one word** is missing.

The meaning of the missing word appears in brackets under each of the sentences.

Fill in the missing word. Use the first letter/s provided for you.

Do not repeat the words given in brackets.

Example:

They have a lot of **t**

(Hint/Meaning: "**hours**")

Answer: **time**

Deze aanwijzing komt terug in de twee onderstaande voorbeelditems (instructie en opgave).



### Voorbeeldopgave 1

Instructie: Write down the correct word.

Schrijf naast elk plaatje het juiste Engelse woord.



### Voorbeeldopgave 2

Instructie: Write down the missing word

Schrijf het goede Engelse woord tussen de twee gegeven woorden.

(tijd) second - \_\_\_\_\_ - hour

### Receptive Recall:

Voor Receptive Recall geldt de volgende aanwijzing:

In each sentence **one word** is underlined.

Complete the sentence with one or more words **in English** in a way that clarifies the meaning of the **underlined word**. Do not use the underlined words in your answer.

Example:

'I have **a lot of time**' means that I

**Possible answers:**

have enough hours, have plenty of minutes, have a long enough period...

Deze aanwijzing komt terug in het onderstaande voorbeelditem (instructie en opgave).

### Voorbeeldopgave 1

**Put in the right order.**

Schrijf de Engelse woorden op in de goede volgorde.

Schrijf op in volgorde van vroeg naar laat (delen van de dag):  
afternoon – evening – morning – night

---

### Productive Recognition

Voor Productive Recognition geldt de volgende aanwijzing:

In each sentence one word is missing.

The meaning of the missing word appears in brackets under each of the sentences.

Choose the word with the closest meaning to that presented in brackets.

Example:

They have a lot of

(Hint/Meaning: "hours")

- a. time
- b. money
- c. work
- d. food

*Productive Recognition* kan op verschillende manieren worden bevraagd. In de peiling zijn verschillende vraagsoorten gebruikt. Hieronder worden er twee gepresenteerd:

### Voorbeeldopgave 1

Which word means the same?

Welk woord heeft ongeveer dezelfde betekenis als het onderstreepte woord? Kies A, B of C

> Will you come to our house?      < Perhaps.

Welk woord heeft ongeveer dezelfde betekenis als perhaps?

- A Absolutely
- B Maybe
- C Of course

### **Voorbeeldopgave 2**

Make the right choice

Welk woord past het beste op de open plek? Kies A, B of C

Welk woord past het best op de open plek?

I can \_\_\_\_\_ (makkelijk) run 5 kilometres, no problem.

- A easily
- B hardly
- C never

### **Receptive Recognition**

In het CATTS wordt de volgende aanwijzing gegeven.

In each sentence one word is underlined.

Choose the best option to complete the sentence in a way that clarifies the meaning of the underlined word.

Example:

I have a lot of time means that

- a. I have many hours
- b. I have a lot of money
- c. I have many opportunities
- d. I have a lot of work

### **Voorbeeldopgave 1**

Make the right choice

Wat is de Nederlandse betekenis van het onderstreepte woord? Kies A, B, C of D

Ruth and Jim will organise the barbecue together

- A graag
- B meteen
- C samen
- D vanavond

### 1.6.3 Woordenschat: afname toets

De toets woordenschat is klassikaal afgenomen. Leerlingen maakten de toets individueel. Het was de laatste taak die de leerlingen hebben gemaakt op de afnamedag. Voor alle klassikale taken is gebruik gemaakt van 1 toetsblad. Testleiders instrueerden de leerlingen conform het protocol (zie hoofdstuk 4). Leerlingen hadden 30 minuten om de toets te maken.

### 1.6.4 Woordenschat: beoordeling

De items die receptieve woordkennis bevroegen waren meerkeuzevragen, waarbij het antwoord goed of fout was. Scoring van de receptieve woordenschatkennis was gebaseerd op het aantal juiste antwoorden. Voor de productieve woordenschatkennis waren twee aspecten van belang. De vraag of een leerling het woord actief kon benoemen en de vraag of de leerling het woord goed kon spellen. Hoewel het de vraag is of een beoordeling van de woordenschattoets, waarbij antwoorden enkel als goed werden gescoord ingeval van correcte spelling, volledig recht doet aan de daadwerkelijke beheersing van het domein door de leerling, is toch gekozen voor deze strikte beoordelingswijze. Onderzoek laat namelijk zien dat de betrouwbaarheid van de meting van de vaardigheid op het gebied van woordenschat bij strikte beoordelingswijzen hoger is (Westhoff, 1981).

## 1.7 Luistervaardigheid: beschrijving en theoretische achtergronden

### 1.7.1 Luistervaardigheid: beschrijving

Luisteren is een complex proces en omvat vaardigheden als 'begrijpen', 'interpreteren' en 'reflecteren' binnen de context waarin luisteren plaatsvindt (Buck, 1991; Damhuis & Litjens, 2003; Krom, Ouborg, & Kamphuis, 2001; Levelt, 1989). Luisteren in een tweede- of vreemde taal vereist bekendheid met alle voor luisteren relevante kennisfactoren in de vreemde taal. Bekendheid met het klanksysteem en kennis over de waarschijnlijkheid van het voorkomen van combinaties van lettergrepen (Safran, 2003) is noodzakelijk voor het decoderen van gesproken taal. Kennis van vocabulaire is noodzakelijk voor het herkennen van betekenis, en cultuurspecifieke informatie is noodzakelijk voor het interpreteren van de gesproken tekst (zie bijvoorbeeld Flowerdew & Miller, 2005).

Luisteren en lezen in de vreemde taal hebben veel gemeenschappelijke kenmerken: in beide gevallen moet de taalleerder decoderen en interpreteren, op basis van kennis van de vreemde taal en van kennis van de wereld. Maar luisteren vraagt meer cognitieve inspanning, omdat het in real-time gebeurt: de luisteraar heeft geen controle op het tempo waarop de informatie verwerkt moet worden en het werkgeheugen speelt een grotere rol (Vandergrift & Baker, 2015). Factoren die het luisteren positief beïnvloeden zijn werkgeheugen, auditieve capaciteit en vocabulairekennis (Vandergrift & Baker, 2018). Multimediale input (video en tekst naast audio) kan een belangrijke bijdrage leveren aan de ontwikkeling van luistervaardigheid (Vandergrift, 2007).

In de literatuur is er overeenstemming over het feit dat het testen van luistervaardigheid weinig is onderzocht en dat de didactiek van luistervaardigheid in de klas een test-format heeft. Leerlingen krijgen luistertaken, de antwoorden worden gecontroleerd, maar ze krijgen weinig hulp in het verbeteren van hun vaardigheid: het product lijkt belangrijker te zijn dan het proces (Vandergrift & Goh, 2012; Graham, Santos & Francis-Brophy, 2014; Harding, Alderson & Brunfaut, 2015).

Luisteren gaat vaak gepaard met kijken, er zijn niet veel situaties waarin de luisteraar alleen op audio aangewezen is (in wezen alleen in telefoongesprekken, radio-uitzendingen of aankondigingen in openbare ruimtes). Voor jonge leerlingen speelt beeld een steeds belangrijkere rol. Taalmethodes en/of leerkrachten maken steeds meer en vaker gebruik van beeld als instructiemateriaal of om de leerstof te introduceren of te verduidelijken. Ook in de lessen 'luisteren' wordt in toenemende mate gebruikgemaakt van beeldmateriaal. Uit onderzoek is bekend dat video in de regel het begrip vergemakkelijkt, omdat gebaren, mimiek en andere elementen die de communicatie ondersteunen zichtbaar worden. Onderzoek naar het gebruik van video ten opzichte van audio in tests geeft daarentegen geen duidelijk beeld van voordelen van video (Batty, 2015).

De combinatie van kijken en luisteren in het Engels is echter meer beschikbaar dan ooit tevoren, en daarnaast worden in Nederland veel Engelstalige programma's en films ondertiteld en niet nagesynchroniseerd. Door dit aanbod en de taalverwantschap tussen Engels en Nederlands, is het niet verrassend dat Nederlandse leerlingen goed op luistertoetsen presteren (bijna 60% van de 15-jarigen zit op B2 niveau volgens de European Survey on Language Competences, 2011). Lopend onderzoek binnen het Orwell project laat verder zien dat bij jonge kinderen buitenschools contact met Engels (via familie of media en games) relatief meer impact heeft op receptieve vocabulairekennis dan bijvoorbeeld motivatie (Van Koert-Hoogervorst, M. J. H., Rispens, J. E., Leona, N. L., Van der Molen, M. W. & Snellings, P. J. F., 2017).

Aan luisteren werd volgens de peiling 2012 de meeste aandacht besteed door leerkrachten (Geurts & Hemker, 2013). Leerlingen zijn dus redelijk gewend aan het oefenen met luistervaardigheid in de klas. De niveaubeschrijving van luistervaardigheid in het ERK richt zich op de mate waarin de leerling een moedertaalspreker als gesprekspartner kan begrijpen, gesprekken tussen moedertaalsprekers kan begrijpen, mededelingen en instructies kan begrijpen en kijk- en luisterteksten in audiovisuele media kan begrijpen.

### **1.7.2 Luistervaardigheid: instrumentarium**

Bij het subdomein luistervaardigheid is gebruik gemaakt van audio(visuele) fragmenten die qua medium en onderwerp zoveel mogelijk aansluiten bij de belevingswereld van leerlingen. Alle opgaven vielen binnen de domeinen (1) dagelijks leven, (2) publiek domein, (3) school/opleiding. Gezien de hierboven genoemde sterke ontwikkelingen in onze maatschappij op audiovisueel gebied, en het streven naar zo authentiek mogelijke taken, bestond het totale instrumentarium uit zowel audio- als audiovisuele opgaven.

In totaal zijn 48 items geconstrueerd. Van die 48 items zijn 20 ankeritems die in de vorige peiling(en) zijn gebruikt. Er zijn 28 nieuwe items geconstrueerd. Conform het design (paragraaf 3.3.2) bevatte ieder van de zes toetsversies twee blokken, met in totaal 16 items. Iedere gemaakte toets bevatte negen audio-items en zeven audiovisuele items. Bij de audio fragmenten werd een foto op het beeldscherm vertoond die verband hield met het item. Bij video ging het om bewegende beelden. Om te voorkomen dat de moeilijkheid van de opdrachten veroorzaakt zou kunnen worden door de vraagstelling, is ervoor gekozen om de vragen in het Nederlands te formuleren.

Voor de nieuwe video-opgaven is gebruik gemaakt van o.a. het Engels jeugdjournaal, documentaires, en twee online vlogs. De nieuwe audio-opgaven bevatten (nieuws- en omroep) berichten over uiteenlopende onderwerpen.

### 1.7.3 Luistervaardigheid: getoetste (sub)vaardigheden

De geconstrueerde toetsen voor luistervaardigheid bevatten items die aansloten bij het A1-, A2- en B1-niveau van het ERK. Er zijn twee sub-vaardigheden gepeild, te weten 'luisteren naar mededelingen en instructie' en 'kijken en luisteren naar audiovisuele media'. De bijbehorende "can do"-statements staan weergegeven in Tabel 1.7.

Tabel 1.7: Gemeten subvaardigheden en *can do* statements voor luisteren

ERK-niveau	Subvaardigheid	<i>can do</i> descriptoren
A1	Luisteren naar mededelingen en instructies	<ul style="list-style-type: none"> <li>• Kan tot hem of haar gerichte vragen en opdrachten verstaan en korte, eenvoudige aanwijzingen volgen.</li> <li>• Kan in korte, duidelijk gesproken teksten, getallen en bekende woorden verstaan die gericht zijn aan de luisteraar.</li> <li>• Kan korte, eenvoudige waarschuwingen begrijpen die gericht zijn aan de luisteraar</li> </ul>
	Kijken en luisteren naar audiovisuele media	<ul style="list-style-type: none"> <li>• Kan het onderwerp bepalen van korte kijk-/luisterteksten</li> </ul>
A2	Luisteren naar mededelingen en instructies	<ul style="list-style-type: none"> <li>• Kan in vertrouwde situaties eenvoudige feitelijke informatie begrijpen.</li> <li>• Kan een korte uitleg begrijpen.</li> <li>• Kan aanwijzingen begrijpen over de werking van vertrouwde apparaten, mits het apparaat voorhanden is.</li> </ul>
	Kijken en luisteren naar audiovisuele media	<ul style="list-style-type: none"> <li>• Kan de wezenlijke informatie verstaan en onderscheiden in korte opgenomen passages over voorspelbare alledaagse zaken.</li> <li>• Kan zich een indruk vormen van de belangrijkste inhoud in een feitelijk nieuwsbericht op televisie.</li> <li>• Kan het belangrijkste punt herkennen van nieuwsberichten op televisie waarin verslag wordt gedaan van gebeurtenissen, ongelukken en dergelijke en waarin het beeld het commentaar ondersteunt (A2+).</li> </ul>
B1	Luisteren naar mededelingen en instructies	<ul style="list-style-type: none"> <li>• Kan eenvoudige technische informatie begrijpen.</li> <li>• Kan gedetailleerde aanwijzingen volgen</li> </ul>
	Kijken en luisteren naar audiovisuele media	<ul style="list-style-type: none"> <li>• Kan de hoofdpunten verstaan van nieuwsberichten op de radio en van eenvoudiger opgenomen materiaal over vertrouwde onderwerpen.</li> <li>• Kan inhoudelijke informatie van het meeste opgenomen of uitgezonden geluidsmateriaal over onderwerpen van persoonlijk belang verstaan (B1+).</li> <li>• Kan de hoofdpunten vatten van televisieprogramma's over vertrouwde onderwerpen wanneer deze betrekkelijk langzaam en helder worden gepresenteerd.</li> <li>• Kan een groot deel begrijpen van veel televisieprogramma's over onderwerpen van persoonlijk belang, zoals vraaggesprekken, korte voordrachten en nieuwsverslagen, wanneer deze betrekkelijk langzaam en helder worden gepresenteerd (B1+).</li> </ul>

#### 1.7.4 Voorbeeldopgaven

In deze sub-paragraaf staan enkele voorbeelditems voor luistervaardigheid weergegeven. In de linkerkolom staat telkens de tekst van het audio-/videofragment weergegeven en in de rechterkolom staat de vraag.

##### ***Fishing in school - nieuwsbericht***

<p>Student: "It's the summer holidays at the moment and we are on an activity day but we'll also be fishing in term time as well. My school in Bolton has its own lake next to the school. They even have their own fishing team.</p> <p>We are fishing for perch, roach and rudd. We have to use maggots as bait. They're pretty disgusting. An angling coach teaches us the skills. How difficult is it to start fishing?"</p> <p>Teacher: "It's not difficult at all. There are plenty of clubs that have coaching. And other organisations like scouts, the cubs, the guides all do fishing as well."</p> <p>Student: "Is it expensive?"</p> <p>Teacher: "No it's not expensive at all. The gear we're using today, probably cost you less than 20 pounds."</p> <p>Student: "I didn't catch any fish. But some of my classmates did, and there are always plenty more fish in the sea. Or lake."</p>	<p>Wat vertelt de visleraar?</p> <p><b>A</b> Dat het heel makkelijk is om te beginnen met vissen</p> <p><b>B</b> Dat je beter kunt vissen als je een dure hengel hebt</p> <p><b>C</b> Dat je het beste lid kunt worden van een visvereniging</p>
--	--

##### ***Bern***

<p>Bern - A man was locked in a phone box by people who saw him robbing a woman in the Swiss capital, Bern. The man stole a wallet out of the woman's handbag while she was making a phone call. The people helped the woman out of the phone box and locked the man in until the police arrived.</p>	<p>Wat gebeurde er bij een telefooncel in Bern?</p> <p><b>A</b> De politie arresteerde er een man die er niet uit wilde</p> <p><b>B</b> Een man werd betrapt toen hij de apparatuur kapot sloeg</p> <p><b>C</b> Mensen hielden er tijdelijk een overvaller in gevangen</p>
---	--

### **Orinoco-rivier**

<p>As the Orinoco flows down onto flat land, the trees begin to thin out and the land becomes easier for people to inhabit.</p> <p>Here the Orinoco is joined by a smaller river, or tributary, and it becomes bigger. It could become even bigger if there's a lot of rain.</p> <p>This ferryboat is passing between rocks that would be covered by the river during the rainy season.</p> <p>But this is the dry season and the river is safe for children to play in.</p>	<p>Wanneer kunnen de kinderen veilig spelen in de <i>Orinoco</i>-rivier?</p> <p><b>A</b> Als de veerboten een andere route nemen</p> <p><b>B</b> Als het lang niet geregend heeft</p> <p><b>C</b> Als ze goed kunnen zwemmen</p>
--	--

#### **1.7.5 Luistervaardigheid: procedure en afname**

Luistervaardigheid werd als derde taak aangeboden op de afnamedag. De toets is klassikaal afgenomen, leerlingen maakten de toets individueel. Audio(-visuele) fragmenten werden via het digibord of videoscherm aangeboden. Audiofragmenten werden daarbij met begeleidende foto's ondersteund.

Na een korte instructie door de testleider conform het protocol (zie hoofdstuk 4) startte de testleider de toets. De daadwerkelijke toets werd voorafgegaan door, nogmaals, een korte standaardinstructie via het digibord en een voorbeeldopgave, zodat leerlingen bekend werden met de afnamewijze. Leerlingen kregen voorafgaande aan ieder fragment ongeveer een halve minuut om alvast de opgave te lezen. Deze tijd werd tevens gebruikt om, indien nog niet gedaan, het juiste antwoord op het beluisterde fragment in te vullen. Fragmenten duurden ongeveer tussen 30 en 50 seconden. De totale afnameduur was 30 minuten.

Luistervaardigheid is gemeten met behulp van meerkeuzevragen, waarbij leerlingen konden kiezen tussen drie antwoordmogelijkheden.

#### **1.8 Leesvaardigheid: beschrijving en theoretische achtergrond**

Het begrijpen van geschreven teksten wordt, net als luisteren, gerekend tot de receptieve vaardigheden. Anders dan de term "passieve kennis", die vaak voor deze vaardigheden wordt gebruikt, doet vermoeden, vereist ook lezen een actief, constructief en complex proces. Het herkennen van letters en combinaties van letters, het koppelen van letters aan het klanksysteem van de taal, het koppelen van klanken aan woorden, het oproepen van de betekenis van woorden en vervolgens de interpretatie van deze betekenis in de context van de zin, de tekst, en de wereld is geen eenvoudige opgave. In het model van Perfetti (1999) wordt leesbegrip in de moedertaal uitgewerkt vanuit de interactie van deze componenten. Leesvaardigheid wordt in dit model beschreven als de interactie van enerzijds het identificeren van woorden, en anderzijds het gebruik van begripsmechanismen waarin de woorden worden gecombineerd tot betekenisvolle boodschappen. Hierin spelen kennis van de fonologie, het lexicon, de morfosyntaxis en algemene (wereld-)kennis op verschillende momenten van het proces een belangrijke rol. Een verstoring in een of meer van deze systemen zal leiden tot een onvolledig begrip. De ontwikkeling van leesvaardigheid



volgt een cyclisch proces: tijdens het lezen wordt de betekenis van de tekst geconstrueerd (Kintsch, 1998). Wanneer deze betekenis eenmaal is geconstrueerd, kan zij worden geïntegreerd en worden samengevoegd met al bestaande kennis, die hiermee ook weer wordt uitgebreid. Net als in het model van Perfetti, blijkt dus de integratie van een breed spectrum van kennis en vaardigheden cruciaal voor een goed tekstbegrip, waarin lexicale kennis een cruciale rol vervult (Perfetti, Landi, & Oakhill, 2008). Leesvaardigheid ontwikkelt zich naarmate de lezer meer leeservaring opdoet.

Het lezen van een vreemde taal kan deels steunen op de vaardigheden die zijn opgedaan in het leesproces in de moedertaal, maar leesvaardigheid in de moedertaal kan pas een rol gaan spelen bij een drempelwaarde van de taalvaardigheid in de tweede taal (Lee & Schallert, 1997). De uitgangspunten voor leesvaardigheid in de tweede taal zijn hetzelfde als die in de moedertaal, maar in elk van de stappen in het proces zijn hiervoor extra uitdagingen opgenomen. Voor het lezen van Engelse teksten komt daar als extra moeilijkheid bij dat de relatie tussen het schrijfsysteem en het klanksysteem verre van consistent is. Deze beperkte transparantie van geschreven taal in het Engels maakt het lezen van Engels ook voor moedertaalsprekers relatief moeilijker dan voor moedertaalsprekers van andere talen (Ziegler et al., 2010). Het is daarom niet verwonderlijk dat lezen in het Engels vaak relatief laag scoort ten opzichte van andere vaardigheden (zie bijvoorbeeld de resultaten van de vorige peiling – Geurts & Hemker, 2013).

Voor Engels in het po lijkt lezen een minder belangrijke vaardigheid. Leerkrachten focussen meer op andere vaardigheden en ook het expertpanel gaf geen prioriteit aan het peilen van leesvaardigheid, maar noemde leesvaardigheid wel relevant in verband met de overgang naar het vo (Rose, 2016).

### **1.8.1 Leesvaardigheid: instrumentarium**

Zoals hierboven beschreven kan lezen gekenmerkt worden als ingewikkeld proces en bij het lezen van Engels is de geringe consistentie tussen klank- en lettersysteem een bemoeilijkende factor. De vaardigheid in het lezen van het Engels is in de peiling gemeten door middel van een toets waarin leerlingen korte en wat langere teksten en berichten moesten lezen. Daarna dienden zij een of meerdere vragen over dit tekstfragment te beantwoorden. Evenals bij de andere taken sloten de teksten zoveel mogelijk aan bij de belevingswereld van leerlingen uit groep 8, zowel qua onderwerpen als medium (bijvoorbeeld sociale-mediaberichten). De tekstjes behandelen de volgende domeinen: (1) dagelijks leven, (2) publiek domein, (3) school/opleiding.

Evenals bij woordenschat zijn voor de meting van lezen verschillende typen items ingezet. In totaal zijn 51 leesitems geconstrueerd. Er zijn 17 ankeropgaven ingezet, 23 VoVo opgaven (dit zijn opgaven uit Volgsysteem voortgezet onderwijs) en 11 nieuw geconstrueerde opgaven. De daadwerkelijke toetsen die leerlingen maakten bevatten 16 of 17 vragen, afhankelijk van de versie. Gezien de koppeling van een of meerdere items aan een tekstfragment was de uniformiteit wat betreft de precieze hoeveelheid vragen per toets ondergeschikt aan de kwaliteit van de items en de samenstelling van verschillende tekstvormen. Evenals bij woordenschat en luisteren is er ook bij lezen voor gekozen om de vragen in het Nederlands te formuleren, om te voorkomen dat de moeilijkheid van de opdrachten veroorzaakt zou kunnen worden door de vraagstelling in plaats van door de tekst.

### 1.8.2 Leesvaardigheid: te meten (sub)vaardigheden

Binnen het ERK worden de niveaus van lezen gedefinieerd aan de hand van tekstkenmerken als woordfrequentie, visuele presentatie, tekstlengte en leesstrategieën (SLO-Taalprofielen, 2015). De peiling concentreerde zich op de ERK-niveaus A1, A2 en B1. Vooral op de eerste twee basale niveaus van taalvaardigheid wordt leesvaardigheid gekenmerkt door een beperkte herkenning van een relatief kleine woordenschat en een beperkte morfosyntactische kennis, gebruikt binnen een vertrouwde context. De niveaubeschrijving in Taalprofielen (2015) richt zich op het begrip van geschreven teksten, onderverdeeld naar het doel van de leestaak: 1) correspondentie lezen, 2) oriënterend lezen, 3) lezen om informatie op te doen en 4) lezen van instructies. In de toets zijn hiertoe diverse tekstgenres in de toets opgenomen, waarbij naar een zo evenredig mogelijke verdeling over de zes toetsblokken en vaardigheidsniveaus is gestreefd. De subvaardigheden en "can do"-descriptoren die zijn getoetst staan weergegeven in Tabel 1.8.

Tabel 1.8: Subvaardigheden en *can do* statements per ERK niveau

ERK-niveau	Subvaardigheid	Can do
A1	Correspondentie lezen	<ul style="list-style-type: none"> <li>• Kan korte, eenvoudige mededelingen begrijpen, bijvoorbeeld via sociale media of op brief- of ansichtkaarten.</li> <li>• Kan voorgedrukte kaarten begrijpen met standaardboodschappen</li> </ul>
	Oriënterend lezen	<p>Kan een korte standaard-mededeling lezen.</p> <p>Kan dingen opzoeken in of kiezen uit een lijst.</p> <p>Kan eenvoudige informatie op een poster, mededelingenbord of in een brochure lezen</p>
	Lezen om informatie op te doen	<p>Kan zich een idee vormen van de inhoud van een korte tekst die waar mogelijk visueel ondersteund wordt.</p> <p>Kan in korte informatieve teksten informatie over personen en plaatsen begrijpen</p>
	Instructies lezen	<p>Kan zeer eenvoudige, korte en goed gestructureerde instructies begrijpen</p>
A2	Correspondentie lezen	<p>Kan een korte, eenvoudige (standaard)brief of e-mail begrijpen.</p> <p><i>Kan elementaire soorten standaardbrieven, faxen (inlichtingen, bestellingen, bevestigingsbrieven, enzovoort) of een algemene kennisgeving over vertrouwde onderwerpen begrijpen (A2+).</i></p>
	Oriënterend lezen	<p>Kan specifieke informatie vinden en begrijpen in eenvoudig, alledaags materiaal.</p> <p>Kan eenvoudige advertenties met weinig afkortingen begrijpen.</p> <p>Kan in lijsten, overzichten en formulieren specifieke informatie vinden en begrijpen.</p> <p>Kan veelvoorkomende borden en mededelingen begrijpen</p>
	Lezen om informatie op te doen	<p>Kan specifieke informatie begrijpen in eenvoudige teksten.</p> <p>Kan de hoofdlijn begrijpen van eenvoudige teksten in een tijdschrift, krant of op een website.</p> <p>Kan korte, beschrijvende teksten over vertrouwde onderwerpen begrijpen.</p> <p>Kan door meelesen eenvoudig audiovisueel materiaal begrijpen.</p>

<b>A2</b>	Instructies lezen	Kan eenvoudige, korte en goed gestructureerde instructies begrijpen. <i>Kan regels en bepalingen begrijpen, wanneer deze in eenvoudige taal zijn gesteld (A2+)</i>
<b>B1</b>	Correspondentie lezen	Kan persoonlijke brieven, e-mails en vormen van sociale media voldoende begrijpen om met iemand te kunnen corresponderen. Kan een eenvoudige formele brief of e-mail voldoende begrijpen om adequaat te kunnen reageren
	Oriënterend lezen	Kan relevante informatie vinden en begrijpen in brochures en korte officiële documenten op internet of in andere media. <i>Kan in langere teksten op internet of in andere media informatie zoeken over thema's binnen het eigen interessegebied (B1+).</i>
	Lezen om informatie op te doen	Kan belangrijke feitelijke informatie begrijpen in korte verslagen en artikelen. Kan eenvoudige jeugdliteratuur lezen. Kan door meelesen alledaagse audiovisueel materiaal begrijpen. <i>Kan hoofdthema en belangrijkste argumenten begrijpen in eenvoudige teksten in tijdschriften, kranten of op internet (B1+).</i>
	Instructies lezen	Kan helder geschreven, ondubbelzinnige instructies begrijpen

### 1.8.3 Voorbeeldopgaven

Hieronder staan enkele voorbeelditems voor leesvaardigheid weergegeven.

#### Correspondentie lezen:

Op een BBC-website voor jongeren staat het onderstaande tekstje.

#### Question

My sister Mary is always getting me into trouble. She does things wrong and then tells my Dad that I did them but I didn't. My parents always believe her. What shall I do?

From: Frances

Wat zegt Frances over haar zus Mary?

- A Mary hoeft nooit iets in het huishouden te doen.
- B Mary klaagt altijd dat ze niets mag van haar ouders.
- C Mary zorgt ervoor dat Frances overal de schuld van krijgt.

#### Oriënterend lezen:



#### First book for young writers

Children aged from 8 to 10 from Brant Broughton Primary School can see their own work in print, after the publication of a new book. The story *A Deer Called Dennis*, was written by the pupils together with Molly Burkett, the children's book author.

Also, pencil drawings by the youngsters illustrate the book.

Congratulations kids!

Wat wordt er gezegd over de kinderen op het plaatje?

- A Iemand heeft een boek over hun schoolklas geschreven.
- B Ze hebben samen met een schrijfster een boek gemaakt.
- C Ze kregen een boek als beloning voor hun prestaties op school.

### Lezen om informatie op te doen:

Hieronder zie je een nieuwsberichtje.

**Panda has operation on his teeth**



Ming is a 17-year old panda who lives in Berlin Zoo. One day keepers saw that it was difficult for Ming to chew his favourite food: bamboo shoots. Dentists repaired his teeth. The old panda had a two-hour operation. He got fillings in three teeth. The operation was so successful that two other pandas will get an operation too.

Waarom werd Ming geopereerd?

- omdat hij al heel oud was
- omdat hij haast niet kon kauwen
- omdat hij veel kiespijn had

#### 1.8.4 Leesvaardigheid: procedure en afname

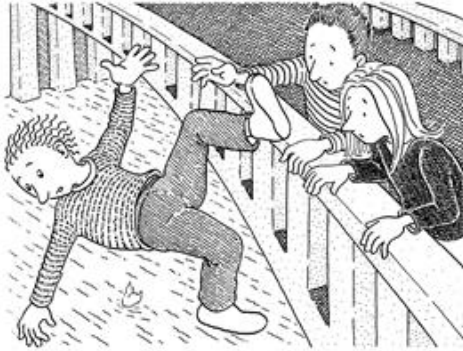
Lezen werd als tweede taak, na het invullen van een leerlingvragenlijst, aangeboden. De toets is klassikaal afgenomen. Leerlingen maakten de toets individueel. Na een korte instructie door de testleider conform het protocol (zie hoofdstuk 4) startte de testleider de toets. Voor alle klassikale taken is gebruik gemaakt van 1 toetsblad. Totale afnameduur was 30 minuten.

Leesvaardigheid is gemeten met behulp van meerkeuzevragen, waarbij leerlingen konden kiezen uit drie of vier antwoorden.

#### 1.9 Spreekvaardigheid

Spreekvaardigheid, het monologisch spreken, wordt niet door experts geprioriteerd als een vaardigheid die gepeild moet worden (Rose, 2016). Om toch een vergelijking te kunnen maken met de mondelinge (spreek)vaardigheid van leerlingen in eerdere peilingen is besloten om een ankertaak mee te nemen. Hiervoor is de taak 'De Brug' geselecteerd die ook onderdeel was van de peilingen in 2006 en 2012. Bij deze taak moesten leerlingen aan de hand van een stripverhaal, bestaande uit de zes plaatjes hieronder, in het Engels vertellen wat er op deze plaatjes gebeurde.

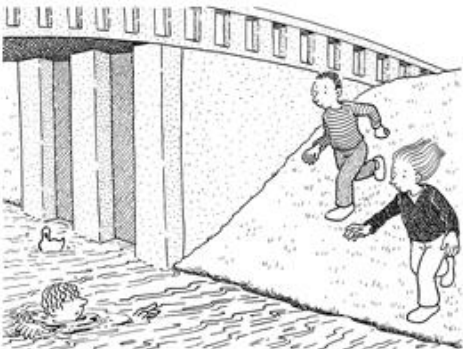
1



2



3



4



5



6



### **Spreekvaardigheid: procedure en afname**

Tijdens de taken lezen, luistervaardigheid en woordenschat werden twee leerlingen door een testleider uit de klas gehaald om de taken voor mondelinge vaardigheid uit te voeren. Dit gebeurde in een afgesloten, rustige ruimte. Na een korte instructie over doel en gang van zaken, werd bij een van de beide leerlingen de spreektaak 'De brug' afgenomen. De andere leerling werd voorafgaand aan het werk gezet met een activiteitenboekje met daarin diverse Engelse activiteiten, zoals puzzels en woordspelletjes. De testleider instrueerde de leerling die de spreektaak ging doen conform het protocol. De instructie was exact dezelfde als tijdens de eerdere peilingen. Testleiders waren voorts geïnstrueerd de leerlingen geen aanwijzingen te geven en een maximale spreektijd van vijf minuten te hanteren. Ten behoeve van de beoordeling zijn video-opnamen gemaakt.



### **Spreekvaardigheid: Beoordeling**

Voor de beoordeling van de spreektaak zijn exact hetzelfde beoordelingsformulier en dezelfde analyses gebruikt als in de vorige peiling (Geurts & Hemker, 2013). Het beoordelingsformulier is weergegeven in Bijlage 3.

## **1.10 De huidige peiling in relatie tot de vorige peiling**

De huidige peiling is gestoeld op de kerndoelen uit 2006 en is deels gebaseerd op dezelfde grondslag en opzet als de vorige peilingen, waardoor een vergelijking van een deel van de resultaten mogelijk is. De aansluiting tussen de vorige peilingen kan voornamelijk worden gevonden in de gerichtheid op het ERK, waar het gaat om de constructie van de klassikale toetsen. De huidige en de vorige peiling zijn daarmee voor de vaardigheden lezen, luisteren en woordenschat gericht op communicatief handelen in authentieke situaties, zoals gedefinieerd in de *can do* descriptors van het ERK. De resultaten werden in de vorige peiling echter niet geïnterpreteerd in termen van de ERK-niveaus én er was destijds geen koppeling met het ERK voor mondelinge vaardigheid (spreken), hetgeen in de huidige peiling wel het geval is (waar het gaat om de constructie en interpretatie van gespreksvaardigheid). Ook in de vorige peiling waren de taken, net als in de huidige peiling, ontwikkeld vanuit deze "can do"-descriptors. De meting van de (achtergrond)kenmerken van leerlingen, leerkrachten, het onderwijsleerproces en de school was in deze peiling bovendien grotendeels gebaseerd op de vragenlijsten die in de vorige peiling zijn gebruikt. Om een trendvergelijking met de vorige peiling mogelijk te maken zijn in de huidige peiling voor een groot deel van de in 2012 gemeten vaardigheden ankeritems toegepast, te weten woordenschat, luistervaardigheid, leesvaardigheid en spreekvaardigheid.

### **1.10.1 Verschillen ten opzichte van de vorige peiling**

De huidige peiling onderscheidt zich echter op de volgende punten van de vorige peiling:

- Conform de uitgangspunten van het ERK en de prioritering van de experts uit het werkveld heeft deze peiling zich gericht op het meten van de gespreksvaardigheid van leerlingen, gekoppeld aan een ERK-interpretatie. Dit is gebeurd door leerlingen met elkaar in gesprek te laten gaan over onderwerpen die aansluiten bij hun belevingswereld en binnen taken die zo goed als mogelijk authentieke situaties voor leerlingen nabootsen. In de peiling van 2012 werd mondelinge vaardigheid enkel gemeten met behulp van taken voor spreekvaardigheid. Eén van deze taken is in de huidige peiling als ankeritem meegenomen om over het niveau van spreekvaardigheid van de leerlingen over de tijd te kunnen rapporteren.
- Deze peiling heeft zich niet alleen wat betreft de ontwikkeling van de taken voor lees-, luister- en gespreksvaardigheid gericht op het ERK, maar ook in de interpretatie van de resultaten van de leerlingen. In deze peiling was de standaardbepaling voor lezen, luisteren en gespreksvaardigheid namelijk gericht op het vaststellen van de A1-, A2 en B1-niveaus op basis waarvan de taken geconstrueerd waren. In de vorige peiling zijn ook standaarden bepaald, maar deze waren vastgesteld voor de minimum, voldoende en gevorderde leerling. Hiermee konden geen conclusies over de beheersing van de ERK-niveaus worden getrokken.

- Naast ankeritems om een trendvergelijking met de eerdere peiling te kunnen maken, zijn in de huidige peiling tevens VoVo-items meegenomen in de taken woordenschat en lezen. Door de toevoeging van deze items aan de instrumenten is het mogelijk de resultaten van groep 8 te koppelen met vo-prestaties. Deze toevoeging van VoVo-items kende de vorige peiling niet.
- Woordenschat is in de huidige peiling zowel receptief als productief gemeten, terwijl in de vorige peiling woordenschat receptief werd gemeten. Echter, in de vorige peiling werd ook de spelvaardigheid van leerlingen gemeten. Onder de noemer spelling speelde productieve woordenschat “weliswaar een rol, maar van een gerichte meting kan niet gesproken worden”(Geurts & Hemker, 2013, p. 125).
- In de vorige peiling is ook de schrijfvaardigheid van leerlingen gepeild, bijvoorbeeld door korte schrijftaken als het schrijven van een ansichtkaart. Van het meten van schrijfvaardigheid is in de huidige peiling afgezien vanwege a) de relatief lage prioritering door de experts uit het werkveld (Rose, 2016) en b) de beperkte tijd die beschikbaar was voor de toetsafnames (maximaal 2 uur).
- De vorige peiling bevatte minder scholen (51 reguliere en 6 VVTO-scholen). De huidige peiling bevat 95 scholen, waarvan 33 VVTO-scholen. Dit bevordert de generaliseerbaarheid van de uitkomsten van deze peiling.
- In de huidige peiling wordt een verdiepend onderzoek uitgevoerd bij de bovengemiddeld presterende scholen. Op deze scholen is met behulp van zowel observaties als interviews de invulling van de lessen Engels en de beweegredenen hiervoor nader onderzocht, met als doel best practices voor het werkveld op te leveren. Een dergelijk nader onderzoek van het onderwijsleerproces op goed presterende scholen behoorde niet tot de vorige peiling.
- Om een goede vergelijking met de resultaten uit de vorige peiling te kunnen maken, was de afname bij de ankeritems hetzelfde als in deze vorige peiling. Twee uitzonderingen moesten worden gemaakt, waarbij gepoogd is de afname zo vergelijkbaar mogelijk te maken:
  - Luistervaardigheid is afgenomen met behulp van luistertoetsen op het digibord. De in de vorige peiling gebruikte afbeeldingen waren in enkele gevallen van dermate slechte beeldkwaliteit dat nieuwe afbeeldingen aan het audiofragment zijn toegevoegd.
  - Spreekvaardigheid is in de vorige peiling afgenomen met behulp van computer en microfoon, terwijl in de huidige peiling de ankertaak voor spreekvaardigheid door een testleider is afgenomen en leerlingen de taak op papier voor zich hadden. Redenen hiervoor waren met name praktisch van aard: de testleider zou nog meer spullen moeten meenemen en correct moeten gebruiken. De inschatting was dat het nog weer verder belasten van de testleider met een extra digitale tool, een adequate afname van de gesprekstaken (met daarbij de inzet van standaardinstructies en video-opnamen) in gevaar zou brengen. Dit terwijl de testleider zich juist vooral moest richten op het welbevinden van de deelnemende leerlingen en een correcte afname van de gesprekstaken. De testleider had de opdracht strikt dezelfde instructie te geven als in de vorige peiling en is door onderzoekers geïnstrueerd zich niet met de output van de leerlingen te bemoeien. Op deze manier is geprobeerd het verschil met de afname in de vorige peiling te minimaliseren.

## Hoofdstuk 2. Steekproef, werving en onderzoeksgroep

De peiling Engels einde basisonderwijs heeft tot doel de vaardigheden Engels van leerlingen van groep 8 in het reguliere basisonderwijs in kaart te brengen. Hiervoor is voorafgaande aan de peiling een steekproef uit de populatie getrokken. In dit hoofdstuk wordt een verantwoording gegeven van de gekozen steekproeftrekking, het steekproefkader en de daarbij spelende overwegingen (paragrafen 2.1 en 2.2). De opzet en het verloop van de werving van scholen is, inclusief de non-respons, beschreven in paragraaf 2.3. We besluiten het hoofdstuk met een overzicht en een representativiteitsanalyse van de gerealiseerde steekproef op school- en leerlingniveau (paragrafen 2.4 en 2.5).

### 2.1 Steekproefopzet

Ten behoeve van de peiling is er een steekproef getrokken van leerlingen die in het schooljaar 2017-2018 in groep 8 van hun basisonderwijs zaten. Het speciaal onderwijs en het speciaal basisonderwijs maakten geen deel uit van dit onderzoek en daarmee ook niet van de onderzoekspopulatie. De steekproef van leerlingen is indirect getrokken door een steekproef van scholen te trekken, en (in principe) alle leerlingen in groep 8 op een geselecteerde school op te nemen in de steekproef.

Aangezien de klas hier als eenheid beschouwd werd, hebben, op verzoek van enkele scholen met combinatieklassen, ook leerlingen uit groep 7 aan de taken mogen meedoen.<sup>16</sup> Deze leerlingen zijn voor de beschrijvingen en berekeningen in dit rapport uit het databestand gehaald. Wel zijn de scholen geïnformeerd over hun prestaties. Van belang voor dit rapport is echter dat alle beschrijvingen in dit rapport alleen leerlingen in groep 8 betreffen.

Bij de steekproeftrekking werden twee stratificatiecriteria gebruikt. Bij een steekproef die volledig conform de populatieverdeling plaats zou vinden, zouden bij een aantal groepen weinig waarnemingen gedaan worden. Als gevolg daarvan zou het maken van goede vergelijkingen tussen verschillende typen scholen lastiger worden. Om die reden dienden er voldoende scholen in de steekproef te worden opgenomen waarbij er sprake was a) van een vroeg startmoment (VVTO-scholen) en b) verschillende percentages van aantallen leerlingen met een formatiegewicht op de school.

#### 2.1.1 Stratificatiecriteria

##### *Eerste stratificatie-criterium: startmoment Engels*

Het eerste stratificatie-criterium was het startmoment Engels. Er werd een onderscheid gemaakt tussen 'reguliere scholen' en scholen met vroeg vreemdetalenonderwijs (de 'VVTO-scholen'). Dit onderscheid hield in dat scholen die niet binnen de hieronder beschreven definitie van VVTO-scholen vielen, behoorden tot de reguliere scholen. Deze laatste groep is diffuus en kent verschillende startmomenten, omdat zij zowel de 'EIBO-scholen' bevat, waar Engels vanaf groep 7 wordt gegeven, als de 'vervroegde EIBO-scholen', waar Engels vanaf groep 5 of 6 wordt gegeven.

---

<sup>16</sup> Op andere scholen met combinatieklassen zijn de groep-7 leerlingen bewust buiten de afname gehouden.



Bij de steekproef is een oververtegenwoordiging van scholen getrokken uit het VVTO. In principe zijn VVTO-scholen scholen die hun leerlingen vanaf groep 1 onderwijs in het Engels aanbieden. Voor de peiling is deze definitie echter, in overleg met de opdrachtgever, enigszins aangepast: scholen die vanaf 2013-2014 Engels hebben aangeboden aan de leerlingen die ten tijde van de peiling in groep 8 zitten, worden binnen dit peilingsonderzoek gedefinieerd als VVTO-scholen. Er is derhalve gewerkt met een ondergrens die gelegd is op het schooljaar 2013-2014, het vierde leerjaar van de groep 8 leerlingen in de peiling<sup>17</sup>. Als de grens bij een eerder jaar gekozen werd, dus vereiste dat de VVTO-leerlingen in de peiling minimaal vanaf groep 3 Engels hadden gekregen, waren er te weinig scholen om een goede steekproef te trekken. Bij een later jaar zouden de leerlingen in groep 8 relatief weinig van het VVTO-karakter van de school hebben meegekregen. VVTO-scholen worden in dit peilingsonderzoek daarom als volgt gedefinieerd:

*Scholen die vanaf schooljaar 2013-2014 of eerder Engels aanboden (dus waar leerlingen in ieder geval vanaf groep 4 structureel Engels hebben gehad) worden beschouwd als VVTO-scholen. De reguliere scholen zijn alle scholen die niet als VVTO-school geteld worden.*

Naast EIBO en VVTO-scholen wordt in de regel nog een derde groep scholen onderscheiden, de vervroegd EIBO scholen die vanaf groep 5 of 6 Engels aanbieden. Op voorhand was echter geen informatie beschikbaar over de mogelijke omvang van deze groep scholen: er zijn geen bronnen beschikbaar waarin dit type scholen geregistreerd staat. Derhalve kon hier geen steekproef voor getrokken worden. In de intakegesprekken is overigens wel aandacht besteed aan de onderwijstijd van leerlingen en de indicatie die scholen zichzelf geven. Op basis hiervan kan een schatting worden gedaan over het aantal deelnemende 'Vervroegd EIBO-scholen'. De geïnteresseerde lezer wordt hiervoor naar paragraaf 6.1.1 verwezen.

#### **Tweede stratificatie-criterium: percentage leerlingen met formatiegewicht op school**

De tweede stratificatie vond plaats op basis van het percentage leerlingen met een formatiegewicht ('gewichtleerlingen') op een school. Dit criterium is voor alle scholen in de onderzoekspopulatie (het regulier basisonderwijs) bekend. Het was de wens van de opdrachtgever om hierbij vier gewichtenstrata<sup>18</sup> te gebruiken. De definitie van deze gewichtenstrata zijn gegeven in Tabel 2.1.

Tabel 2.1: Betekenis van de gewichtenstrata

Gewichtenstratum		Omschrijving	
Label	Afkorting	Volledig	Verkort
Stratum 1	S4o1	0% gewichtenleerlingen (geheel geen gewichtenleerlingen)	0% gew.
Stratum 2	S4o2	>0%-10% gewichtenleerlingen	>0-10%
Stratum 3	S4o3	>10%-25% gewichtenleerlingen	>10-25%
Stratum 4	S4o4	> 25% gewichtenleerlingen	>25%

<sup>17</sup> Voorwaarde was dat de huidige leerlingen van groep 8 vanaf schooljaar 2013-2014 onderwijs in Engels hadden ontvangen. Een VVTO-certificering was geen voorwaarde.

<sup>18</sup> De afkorting S4o verwijst hier naar de gewichtenstratum vierdeling van de onderwijsinspectie

In deze verdeling wordt er geen verschil gemaakt tussen leerlingen met een laag gewicht (0,30) en een hoog gewicht (1,20). Dit is een gebruikelijke keuze<sup>19</sup>.

### 2.1.2 Impliciete stratificatiecriteria

Bij het trekken van de steekproef moest tevens rekening worden gehouden met twee impliciete stratificatiecriteria, te weten *grootte van de gemeente* (geoperationaliseerd als 'mate van stedelijkheid') en *grootte van de school*. Door gebruikmaking van deze twee impliciete stratificatiecriteria is ernaar gestreefd de representativiteit van de steekproef te vergroten.

### 2.1.3 Exclusiecriteria

#### *Exclusie criterium op schoolniveau*

Het gehanteerde exclusie criterium op schoolniveau betrof het (verwachte) aantal leerlingen in groep 8 in schooljaar 2017-2018. Er is hier gewerkt met verwachte aantallen, aangezien op het moment van de steekproeftrekking de werkelijke gegevens niet bekend waren. Het vereiste minimum aantal leerlingen in groep 8 in het peilingsjaar was zes. Dit betekende dat een school met vijf of minder leerlingen in groep 8 buiten het steekproefkader viel. Het exclusie criterium had een praktische reden: het aantal leerlingen dat mee zou moeten doen aan de mondelinge taken, waaronder de gesprekstaken, was minstens zes (zoals beschreven in paragraaf 1.5.1.4). Dit exclusie criterium is geoperationaliseerd door uit te gaan van het aantal leerlingen in leerjaar 7 in het jaar voor de peiling (2016-2017). Deze leerlingen zouden namelijk naar alle waarschijnlijkheid het jaar daarop in leerjaar 8 zitten – het jaar van het peilingsonderzoek. Het uitsluiten van deze scholen heeft een minimale impact op het steekproefkader (zie paragraaf 2.2.2). Het aantal groep 8 leerlingen is bij de werving van scholen gecontroleerd. Op alle benaderde scholen was het beoogde minimum aantal leerlingen aanwezig.

#### *Exclusiecriteria op leerlingniveau*

In het peilingsonderzoek zijn drie exclusiecriteria op leerlingniveau gehanteerd. Leerlingen met een gediagnosticeerde taalontwikkelingsstoornis (TOS) of een visuele beperking werden uitgesloten van deelname. Slechthorende kinderen hoefden niet deel te nemen aan de luistertoets, maar namen wel deel aan de overige taken. Tijdens het intakegesprek is met behulp van de leerlingenlijst nagegaan of bepaalde leerlingen in de klas(sen) moesten worden uitgesloten van deelname. In het bestand van aangemelde leerlingen zitten drie leerlingen met een gediagnosticeerde Taalontwikkelingsstoornis. Twee daarvan hebben niet deelgenomen. De derde leerling heeft op verzoek van de school wel deelgenomen, maar alleen aan de klassikale taken. Er zijn uiteindelijk geen leerlingen als gevolg van de exclusiecriteria 'visuele beperking' of 'slechthorendheid' van deelname (van één of enkele onderdelen) uitgesloten. Leerlingen met gedragsstoornissen en dyslexie konden gewoon met de afname meedoen.

---

<sup>19</sup> Voor de definitie van de gewichten wordt onder andere verwezen naar <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2006/04/26/nieuwe-gewichtenregeling-basisonderwijs/brochuregewichtenregeling.pdf>

## 2.2 Bronnen en definitief steekproefkader

### 2.2.1 Bronnen: beschikbaarheid van gegevens

Om een goede steekproef te kunnen trekken moest eerst het steekproefkader gedefinieerd worden. Zoals gezegd was de doelpopulatie groep 8 op EIBO en VVTO-scholen in schooljaar 2017-2018. Scholen met Tweetalig Primair Onderwijs ('TPO-scholen') maakten geen onderdeel uit van deze doelpopulatie. Het volmaakte steekproefkader was echter niet beschikbaar aangezien op het moment van de steekproeftrekking nog niet bekend was hoeveel leerlingen per school in 2017-2018 in groep 8 zouden zitten. Voor het inschatten van leerlingaantallen is gebruik gemaakt van DUO-gegevens. Hieronder staat weergegeven welke gegevens benodigd waren.

#### *Aantal leerlingen in groep 8 in 2017-2018*

Op logische gronden kon gesteld worden dat het aantal groep 8-leerlingen het beste geschat kon worden door de leerlingen die in 2016-2017 in groep 7 zaten, als benadering te gebruiken voor het aantal leerlingen dat in het daaropvolgende jaar in groep 8 zou zitten. Om na te gaan of deze redenering correct was is een controle uitgevoerd door de correlatie te berekenen van het aantal leerlingen in groep 7 in schooljaar 2015-2016 en het aantal leerlingen in groep 8 in schooljaar 2016-2017. Deze correlatie ligt tussen 0,97 en 0,99, afhankelijk van de aannamen<sup>20</sup>. Bij meer dan 90% van de scholen wordt een verschil gevonden van 3 of minder leerlingen. Het aantal leerlingen in groep 7 op een school in 2016-2017 leek daarmee een adequate indicatie voor het aantal leerlingen in groep 8 op een school in 2017-2018.

#### *Percentage gewichtenleerlingen in 2017-2018*

Ook het stratificatiecriterium percentage gewichtenleerlingen in 2017-2018 was ten tijde van de steekproeftrekking nog niet geheel bekend. Ook hier zijn de getallen van 2016-2017 gebruikt, omdat zij de best mogelijke beschikbare gegevens waren. Voor de meeste gewichtenstrata maakte dit weinig uit, maar doordat het onderscheid tussen S4o1 en S4o2 een onderscheid was tussen 0 leerlingen en minstens 1 leerling op school, is de kans op een verandering voor het toewijzen van een school in een van deze gewichtenstrata wat gevoeliger.

#### *Exclusiecriterium op schoolniveau – minimum aantal leerlingen in groep 8*

De onzekerheid van de aantallen leerlingen in het schooljaar 2017-2018 speelde ook een rol bij het exclusiecriterium *verwachte groepsgrootte* (6 leerlingen of minder) in groep 8 in 2016/2017. Ook hier is het aantal leerlingen in groep 7 in 2016-2017 gebruikt. Zoals gesteld, is om te verifiëren of deze inschatting correct was, aan het begin van het wervingsgesprek navraag gedaan naar het aantal leerlingen in groep 8. Gevraagd is of het (verwachte) aantal leerlingen in groep 8 van schooljaar 2017-2018 lager was dan zes leerlingen. In dat geval viel de school alsnog af en is gebruik gemaakt van de extra back up-school die naast de reservescholen getrokken is.

---

<sup>20</sup> In opvolgende jaren vinden er bij scholen administratieve verschuivingen plaats, bijvoorbeeld op vestigingsniveau. Leerlingen die eerst bij de ene vestiging ingeschreven staan, komen bijvoorbeeld bij een andere vestiging terecht. Als hier geen rekening mee gehouden wordt dan is de correlatie tussen twee jaren .971. Als de verschuivingen van meer dan 30 leerlingen niet meegenomen worden stijgt deze naar 0.984. Als een maximaal verschil van 10 leerlingen als acceptabel wordt beschouwd is de correlatie 0.994.

## Startmoment

Voor het eerste stratificatiecriterium waren gegevens nodig om te bepalen of een school als VVTO-school kon worden geclassificeerd. Hiervoor zijn meerdere bronnen geraadpleegd, die uiteenlopende informatie bevatten (variërend van onder de 400 tot meer dan 1000 Nederlandse scholen). In overleg met de opdrachtgever is uiteindelijk besloten gebruik te maken van twee van Nuffic afkomstige bestanden die via de opdrachtgever waren verkregen<sup>21</sup>. Bijkomende moeilijkheid was dat het onderzoek leerlingen betrof die in groep 8 zaten en de VVTO-kwalificatie slechts betekenis heeft als de leerlingen een aanzienlijk aantal jaren Engels hebben gehad. Met de opdrachtgever is afgesproken dat er sprake zou zijn van een VVTO-school indien de school vanaf schooljaar 2013-2014 Engels onderwijs aanbood. Dit gegeven is gehanteerd in de steekproeftrekking. Om na te gaan of een school die volgens de steekproef een VVTO-school zou moeten zijn ook daadwerkelijk een VVTO-school was, zijn aan het begin van het wervingsgesprek bij VVTO-scholen twee vragen gesteld: 'Vanaf welk leerjaar geeft de school Engels?' en, als het antwoord op deze vraag leerjaar 1 was: 'In welk leerjaar is de school daarmee begonnen?'.

### 2.2.2 Steekproefkader

In deze paragraaf worden relevante data gepresenteerd die het steekproefkader beschrijven. Allereerst zijn de scholen (en leerlingen) die op basis van de geformuleerde exclusiecriteria af zouden moeten vallen bepaald. Tabel 2.2 laat zien dat 252 scholen (897 leerlingen) op voorhand niet in de steekproeftrekking konden worden meegenomen vanwege het hierboven genoemde exclusiecriteria op schoolniveau. Op schoolniveau betreft dit bijna 4% van de populatie en op leerlingniveau een half procent. Dit percentage ligt ruim onder de in de aanvraag gestelde grens van 5% exclusie op leerlingniveau.

Tabel 2.2: Exclusiegegevens (aantallen en percentages)

Exclusie	Aantal		Percentage	
	Scholen	Leerlingen LJ7	Scholen	Leerlingen LJ7
Exclusie: te weinig verwachte leerlingen in groep 8	252	897	3.89	0.51
Geen exclusie	6225	176116	96.11	99.49
Totaal	6477	177013	100.0	100.0

In de hiernavolgende beschrijving van het steekproefkader worden alleen resultaten getoond zonder deze exclusiegegevens. De aantallen scholen en leerlingen worden hieronder dus beschreven ten opzichte van 6225 scholen en 176.116 leerlingen in groep 7 (als schatting van het verwachte aantal leerlingen in groep 8 van het reguliere basisonderwijs in schooljaar 2017-2018).

Uit Tabel 2.3 kan worden afgeleid hoeveel scholen en leerlingen binnen ieder gewichtenstratum als EIBO en als VVTO konden worden geclassificeerd. Bepaling van de grens op schooljaar 2013-2014 is zinvol om er in ieder geval zorg voor te dragen dat er voldoende scholen in de steekproef zitten met leerlingen die (in ieder geval een groot deel van) de gehele 8 jaar Engels hebben gehad. Tabel 2.3 laat tevens de verdeling van de beschikbare scholen per gewichtenstratum

<sup>21</sup> Adressen vvto-scholen gestart voor 2011.xlsx (mail: 16/5/2017) en een kopie van Excellijst vvto-scholen Engels v.a. groep 1 start 2012.xlsx (mail:18/5/2017)

zien, waarbij de VVTO-scholen verder zijn onderverdeeld naar scholen die in schooljaar 2012-2013 met VVTO zijn gestart en scholen die zijn gestart in 2011 of eerder ('-2011'). Duidelijk wordt dat gewichtenstratum S4o2 (>0%-10% gewichtenleerlingen) het vaakst in de populatie voorkomt, zowel bij de EIBO als bij de VVTO-scholen.

Tabel 2.3: Expliciete stratificatie startmoment en percentage gewichtenleerlingen

aantal scholen	VVTO (voor 2014)				Totaal
	Niet	2012-2013	-2011	VVTO totaal	
S4o1 (0% gew.lIn)	347	28	39	67	414
S4o2 (>0-10%)	3447	173	236	409	3856
S4o3 (>10-25%)	1244	55	91	146	1390
S4o3 (>25%)	531	13	21	34	565
Totaal	5569	269	387	656	6225

aantal leerlingen L17	VVTO (voor 2014)				Totaal
	niet	2012-2013	-2011	VVTO totaal	
S4o1 (0% gew.lIn)	8614	688	1030	1718	10332
S4o2 (>0-10%)	105268	4790	7458	12248	117516
S4o3 (>10-25%)	31282	1234	1960	3194	34476
S4o3 (>25%)	12994	323	475	798	13792
Totaal	158158	7035	10923	17958	176116

De verdeling van aantallen scholen en leerlingen in groep 7 over de twee impliciete criteria 'mate van verstedelijking' en 'grootte van de school' wordt weergegeven in Tabel 2.4 en Tabel 2.5, respectievelijk.

Tabel 2.4: Verdeling naar mate van verstedelijking van scholen en leerlingen verwacht in leerjaar 8

Mate van verstedelijking	Aantal		Percentage	
	Scholen	Leerlingen	Scholen	Leerlingen
Zeer sterk stedelijk	794	28812	12.8	16.4
Sterk stedelijk	1456	49436	23.4	28.1
Matig stedelijk	1252	37444	20.1	21.3
Weinig stedelijk	1619	39444	26.0	22.4
Niet stedelijk	1104	20980	17.7	11.9
Totaal	6225	176116	100.0	100.0

Tabel 2.5: Verdeling naar schoolgrootte van scholen en leerlingen verwacht in leerjaar 8

Schoolgrootte categorie (aantal leerlingen op school)	Aantal		Percentage	
	Scholen	Leerlingen	Scholen	Leerlingen
1-100	990	10112	15.9	5.7
101-200	2118	41221	34.0	23.4
201-300	1680	50746	27.0	28.8
301-400	816	34889	13.1	19.8
> 400	621	39148	10.0	22.2
Totaal	6225	176116	100.0	100.0

Op basis van de twee expliciete stratificatiecriteria kunnen in totaal acht combinatie-strata worden vastgesteld. Deze volgen uit de combinatie van het startmoment (EIBO/VVTO) en het percentage gewichtenleerlingen op school (S4o1-S4o4). Voor deze acht combinatie-strata zijn tevens de gemiddelde aantallen berekend van het aantal leerlingen op school en van het aantal leerlingen in groep 7 in 2016-2017. In Tabel 2.6 staat per stratum de verdeling van gemiddelde aantallen leerlingen op de scholen en in leerjaar 7 weergegeven voor respectievelijk EIBO (Niet VVTO/na 2013 VVTO) en VVTO (2013 of eerder).

Tabel 2.6: Verdeling van verwacht aantal leerlingen op school en leerjaar 8 in de combinatie-strata

Gewichtenstratum \ VVTO school	Gemiddelde aantal leerlingen			
	Op school		Verwacht in LJ8	
	EIBO	VVTO	EIBO	VVTO
S4o1 (0% gew.lIn)	197	211	25	26
S4o2 (>0-10%)	241	244	31	30
S4o3 (>10-25%)	203	173	25	22
S4o3 (>25%)	207	190	24	23

### 2.2.3 De steekproef

Voorafgaand aan de uiteindelijke steekproeftrekking is een aantal aspecten gedefinieerd. De steekproefgrootte was gedefinieerd in het aantal te trekken scholen, namelijk 100. De eenheid “school” was gedefinieerd door de combinatie van BRIN-code en vestigingsnummer. De verdeling van deze scholen in de steekproef was evenredig gegeven het steekproefkader, maar kende een oververtegenwoordiging van VVTO-scholen.

Er waren meerdere redenen om geen representatieve steekproef te trekken. Allereerst diende er na afloop van het hoofdonderzoek een vervolgonderzoek plaats te vinden, waarin alle gewichtenstrata vertegenwoordigd dienden te zijn. Een representatieve steekproef zou tot gevolg hebben dat het aantal scholen in de kleinere gewichtenstrata (zoals VVTO-S4o1 en VVTO-S4o4) zodanig laag zouden liggen dat het aantal benodigde scholen voor het vervolgonderzoek in een stratum niet gehaald kon worden of dat vrijwel alle scholen hiervoor geselecteerd zouden worden – hetgeen onwenselijk was. Daarnaast zou de power om onderscheid te maken tussen de combinatie-strata dan ook een stuk lager zijn, waardoor werkelijke verschillen tussen de strata moeilijker te identificeren zijn.

Om ervoor te zorgen dat maximaal een op twee scholen geselecteerd zou kunnen worden voor het vervolgonderzoek en om een voldoende aantal leerlingen te hebben per combinatie-stratum worden er minimaal 8 scholen uit een combinatiestratum getrokken. Dat heeft tot gevolg dat bij een trekking van 100 scholen er sprake is van een niet proportionele steekproef. Het heeft tot gevolg dat voor 64 van de 100 te trekken scholen vastligt uit welk combinatiestratum deze komen (8 bij ieder van de combinatie-strata). De overige 32 zijn proportioneel verdeeld naar het aantal beschikbare scholen in de strata. In dit geval was er bij 6 van de 8 strata al sprake van een oververtegenwoordiging. Alleen uit de strata EIBO-S4o2 en EIBO-S4o3 zijn additionele scholen getrokken, waarna deze strata overigens nog steeds ondervertegenwoordigd zijn in vergelijking met de populatieverdeling<sup>22</sup>.

<sup>22</sup> Om twee redenen is er niet gekozen voor een geheel gelijke verdeling van 100 scholen over de 8 combinatie-strata (12 a 13 scholen per stratum). Ten eerste zou een dergelijke oververtegenwoordiging een vergrote toets-druk op de kleinere gewichtenstrata opleveren, hetgeen gevolgen zou kunnen hebben voor de bereidheid tot deelname van de scholen met

Wanneer we de verdeling van scholen vergelijken met de gegevens in Tabel 2.3, dan valt al op dat S4o4 en S4o1 inderdaad oververtegenwoordigd zijn. In Tabel 2.7 wordt per cel aangegeven welk deel van de steekproef in de populatie zit<sup>23</sup>.

Tabel 2.7: Aantal scholen in de steekproef (links) en percentage van populatie per cel (rechts)

Aantal scholen in enkelvoudige steekproef	VVTO			Percentage steekproef t.o.v. totale groep	VVTO		
	niet	wel			niet	wel	
S4o1 (0%)	8	8	16	S4o1 (0%)	2,3	11,9	3,9
Gewichten S4o2 (>0-10%)	38	8	45	Gewichten S4o2 (>0-10%)	1,1	2,0	1,2
stratum S4o3 (>10-25%)	14	8	23	stratum S4o3 (>10-25%)	1,1	5,5	1,7
S4o4 (>25%)	8	8	16	S4o4 (>25%)	1,5	23,5	2,8
	68	32	100		1,2	4,9	1,6

De mate van over of onder-representatie is af te leiden uit het percentage in de steekproef ten opzichte van de totale groep. Bij een steekproef van 100 scholen uit een steekproefkader van 6225 scholen, betekent dit dat 1,6% van de scholen in de steekproef zit. Bij een evenredige trekking zou dit percentage bij alle combinatie-strata worden getrokken. Bij de huidige trekking is dat duidelijk niet het geval. Bij strata waar het percentage lager ligt dan 1,6% is er sprake van ondervertegenwoordiging in de steekproef. Bij percentages boven de 1,6% is er sprake van oververtegenwoordiging. Zo geldt voor VVTO-S4o4 dat met acht van de 34 scholen 23,5% van de scholen uit dat deel van de populatie in de steekproef zat. Bij de steekproef is er ook rekening gehouden met het startjaar waarbij er een onderscheid gemaakt wordt tussen de scholen die in de periode tussen schooljaar 2010-2011 tot en met 2013-2014 gestart zijn. Dit om ervoor te zorgen dat er voldoende leerlingen tussen zitten die mogelijk de volle 8 jaar VVTO hebben gevolgd.

#### *Back up van de steekproef*

Bij de steekproeftrekking zijn voor iedere getrokken school in ieder van de 8 stratumgroepen back up steekproef-scholen getrokken worden voor het geval van non-response: voor iedere school in de steekproef zijn er drie specifieke “reserve”-scholen die direct bij de steekproeftrekking bekend waren, en geordend waren als eerste, tweede en derde reserveschool. Dit waren scholen uit hetzelfde combinatie-stratum, die niet alleen vergelijkbaar waren op de expliciete stratificatiecriteria (startmoment en percentage gewichtenleerlingen), maar ook op de impliciete criteria. Hierdoor was de reserveschool vergelijkbaar met de school in de steekproef waaraan ze gekoppeld waren. In het geval scholen na eerdere toezegging tot deelname zich onverhoopt afmelden, dan was voor iedere school duidelijk welke andere school benaderd kon worden.

Bij een eerdere peiling bleek dat er soms scholen uitvallen omdat de BRIN-informatie niet correct is: scholen zijn bijvoorbeeld gefuseerd of opgeheven. Dit wordt aangemerkt als technische non-response. Voor deze gevallen is er voor 92 van de steekproefscholen, naast de drie back-up scholen, ook op voorhand een additionele vierde school getrokken. Deze additionele school kwam alleen in aanmerking voor werving op het moment dat er een school uitviel wegens incorrecte informatie voor 2017-2018 waar op het moment van de steekproeftrekking geen rekening mee

---

relatief veel gewichtenleerlingen. Ten tweede was er, gezien de gestelde voorwaarde dat uit ieder van de acht cellen van het steekproefkader (de combinatie van startmoment en 4 gewichtenstrata) minstens 8 scholen zouden moeten worden getrokken, al sprake van een oververtegenwoordiging van de kleinere gewichtenstrata.

<sup>23</sup> Zoals uit de hiernavolgende tekst kan worden afgeleid zijn uiteindelijk vier steekproeven getrokken.



gehouden kon worden<sup>24</sup>. De oververtegenwoordiging van de VVTO-S4o4 scholen was zo groot dat het niet mogelijk was om voor deze groep een additionele school te trekken: vrijwel alle scholen in die groep zaten al in de steekproef (geselecteerd of als 'reserve'school).

### **Samengevat**

In totaal is dus een steekproef getrokken van 100 scholen, 300 reservescholen en een set van 92 additionele scholen, die alleen in uitzonderlijke gevallen gebruikt kon worden. Additionele scholen ontbraken voor het stratum VVTO S4o4. In het databestand is per school in de steekproef aangegeven wat de eerste tot en met vierde steekproef school was, en wat de (vijfde) additionele school was van de getrokken school. Er is gestreefd naar een zo hoog mogelijk percentage toezeggingen binnen de steekproef.

### **Verantwoording van aantallen leerlingen in de steekproef**

Het verwachte aantal leerlingen in de steekproef lag op 2400 leerlingen. Dit aantal was gebaseerd op DUO gegevens, waaruit blijkt dat klassen gemiddeld uit 24 leerlingen bestaan. Om na te gaan of dit verwachte aantal leerlingen op basis van bovenstaande steekproefopzet zou kunnen worden bereikt, is het aantal leerlingen in de eerste steekproef bekeken, alsmede het totaal aantal leerlingen. Inspectie van zowel de eerste steekproef als ook het gemiddelde van de steekproeven (tussen de 2600-2700 leerlingen) liet zien dat in de meeste gewichtenstrata het aantal leerlingen in groep 8 hoger lag dan verwacht. Onder aanname van 100 deelnemende scholen zou het verwachte aantal van 2400 leerlingen nog steeds worden gehaald als gemiddeld ongeveer 7% van de leerlingen zou afvallen.

Door in het ontwerp voor over- en ondervertegenwoordiging van specifieke categorieën te kiezen, is bij de analyses achteraf gewogen naar de populatie. Meer hierover is te vinden in de paragraaf 8.1.

## **2.3 Opzet en uitvoering van de werving**

Een goed steekproefkader en een correcte steekproeftrekking zijn van belang voor een goed beeld van het Engels in het primair onderwijs is. Daarnaast is het natuurlijk noodzakelijk dat er voldoende scholen aan de peiling willen deelnemen. Om aan deze voorwaarde te voldoen is begin juni 2017 een gerichte wervingscampagne uitgezet. Hieronder wordt een toelichting gegeven op de aankondiging van de peiling bij scholen, de inzet van ambassadeurs en het verloop van de wervingsperiode.

### **2.3.1 Informeren beoogde scholen**

Schoolleiders van 68 EIBO-scholen ('reguliere' basisscholen) en 32 VVTO-scholen uit de (eerste) steekproef hebben begin juni 2017 een e-mail bericht ontvangen met daarin een algemene beschrijving van het belang van Engels in het basisonderwijs en een link naar een wervingsfilm. In die film was er aandacht voor Engels als (toekomstig) communicatiemiddel, voor het belang van goed onderwijs om Engelse vaardigheden aan te leren, voor de opzet van het onderzoek, en voor de opbrengsten van het onderzoek voor de school. In het filmpje werd ook de verdere voortgang van het peilingsonderzoek aangekondigd. In de e-mail is tevens aangekondigd dat de scholen persoonlijk benaderd zouden worden door zogenaamde ambassadeurs. Deze persoonlijke benadering zou

---

<sup>24</sup> Wanneer de back-up school vervolgens zou weigeren, zou de eerste en vervolgens tweede reserve-scholen in beeld komen voor benadering. Dat is niet voorgekomen.



binnen twee weken na ontvangst van het bericht plaatsvinden. De e-mail met de link naar het filmpje is opgenomen als bijlage 5.

De scholen van de eerste en tweede reservesteekproef zijn één respectievelijk twee weken later met een vergelijkbare e-mail benaderd. Omdat in de meeste gevallen nog niet bekend was of de eerste school zou deelnemen, is de wervingstekst in het bericht enigszins aangepast van daadwerkelijke benadering naar 'mogelijke benadering' door ambassadeurs.

De besturen van alle (300) scholen in de drie steekproeven zijn in week twee van de wervingscampagne geïnformeerd met een brief ondertekend door zowel het consortium als de plaatsvervangend Inspecteur-generaal. Naast een inhoudelijke toelichting van de peiling en een aankondiging dat we de directeuren van de betreffende scholen (mogelijk) zouden benaderen, had de brief vooral tot doel dat het bestuur de scholen zou motiveren aan de peiling deel te nemen. Voorafgaand aan de informatieronde is een website ([www.peil-engels.nl](http://www.peil-engels.nl)) ingericht met alle informatie over het onderzoek en eventuele deelname. In de e-mails en brieven werd naar deze website verwezen. De website bevatte eveneens de wervingsfilm.

### 2.3.2 Inzet ambassadeurs

Om scholen te motiveren voor deelname aan onderzoek, lijkt niets beter te werken dan een persoonlijke benadering. De 100 scholen uit de eerste steekproeftranche zijn daarom vanaf half juni benaderd en deels 'geworven' door een team van tien ambassadeurs. Deze ambassadeurs hadden allen een sterke persoonlijke band met Engels in het basisonderwijs. Zij waren met name (oud) vakleerkracht, dan wel directeur op een basisschool of werkzaam (geweest) als vakdocent of taalwetenschapper bij de Hogeschool van Amsterdam of de Rijksuniversiteit Groningen. In het contact met de scholen fungeerden de ambassadeurs als een soort 'visitekaartje' voor het onderzoek. In eerste instantie verliep de benadering via telefonisch contact en e-mail. Indien de school dat wenste, kon de ambassadeur ook een persoonlijke afspraak op de school maken om het onderzoek nader toe te lichten. Ter ondersteuning van de werving is een 'draaiboek werving' opgesteld en gedeeld met de ambassadeurs. De ambassadeurs zijn feitelijk meteen aan de slag gegaan en werden begeleid door de onderzoekers bij vragen over de onderzoeksopzet of bij specifieke vragen van schoolleiders. In augustus 2017 heeft een ambassadeursdag plaatsgevonden waarop de ambassadeurs ervaringen en tips hebben uitgewisseld.

De ambassadeurs ontvingen verder per toegewezen school een logboek waarin o.a. de gegevens van de beoogde school en bijbehorende reservescholen waren opgenomen. De toegewezen scholen waren waar mogelijk verdeeld o.b.v. de voorkeursregio van de ambassadeurs (gegeven het feit dat er mogelijk schoolbezoeken nodig waren voor de werving). Zoals beschreven in paragraaf 2.2.3 was voor elk steekproefnummer op voorhand bekend welke reservescholen aan dat nummer gekoppeld waren. De reservescholen voldeden aan dezelfde expliciete steekproefcriteria als de eerste school, maar waren niet geografisch geclusterd. Uiteindelijk is er in beperkte mate sprake geweest van schoolbezoeken in de gehele wervingsfase. De meeste correspondentie vond op verzoek van de scholen plaats via telefoon en e-mail, eventueel met tussenkomst van de onderzoekers (die nadere toelichting bij vragen gaven). In de laatste fase van de wervingsperiode (oktober-november 2017) heeft een interne onderzoeker de rol van ambassadeur overgenomen. Met die inzet is het beoogde totaal aantal scholen bereikt.

### 2.3.3 Verloop van de werving

De werving startte in juni 2017 en liep door tot begin december 2017. Het streven was dat 100 getrokken scholen deel zouden nemen, waarvan zoveel mogelijk scholen uit de eerste steekproef.

Ondanks de late start in juni had het ambassadeursteam eind juli 2017 reeds van 35 scholen een toezegging mee te doen aan de peiling. Het merendeel betrof scholen uit de eerste steekproef.

De werving voor de zomervakantie kende echter een aantal lastige/vertragende aspecten, aangezien:

- de directie de wervings e-mail in ongeveer een derde van de gevallen niet had gelezen;
- veel schoolleiders niet of slecht bereikbaar of veelvuldig afwezig waren;
- de directie moest kiezen tussen concurrerende verzoeken voor onderzoek;
- relatief veel scholen de beslissing graag 'over de zomervakantie heen tilden';
- er sprake was van een kleine technische non-respons. Twee van de scholen uit de eerste 100 waren of zouden worden opgeheven. Na de zomervakantie zijn hiervoor de additionele scholen benaderd (zie paragraaf 2.2.3).

Begin september was ongeveer 90% van de 100 scholen uit de eerste steekproef benaderd en bereikt, ook al leidde dit niet altijd tot een onmiddellijk besluit. Bij een definitieve afwijzing benaderde de ambassadeur de eerstvolgende reserveschool op de lijst. Met veel reservescholen werd na de zomervakantie voor het eerst concreet contact gelegd. Hoewel eind september/begin oktober zo'n twee-derde van de beoogde 100 scholen was geworven, was de werving niet eenvoudig. Zo zagen na de vakantie vier scholen af van deelname vanwege onvoldoende draagkracht bij de leerkrachten (2 scholen), miscommunicatie over het toetsen van combinatiegroepen (1 school) en verwachte problemen met 'ouders en privacy' (1 school). Voor deze scholen zijn direct reservescholen benaderd. Als gevolg van uitval van ambassadeurs, zijn er begin oktober twee nieuwe ambassadeurs geworven en getraind en heeft een hoofdtestleider meegeholpen bij de werving. De inzet van 'nieuwe' ambassadeurs is goedgekeurd door de opdrachtgever. Dit geldt ook voor een aanpassing van de wervingstactiek, die het wervingsproces zou versnellen: Besloten is om vanaf begin oktober simultaan *de derde en vierde school achter een volgnummer te benaderen, ook wanneer de tweede school nog niet had beslist*. Voor de zekerheid werd vanaf dat moment ook de 5e (back-up school) alvast ingelicht per e-mail (voor het geval de reguliere reserves allen zouden weigeren). In drie gevallen is uiteindelijk ook deze back-up school benaderd (en vastgelegd voor deelname).

Half november stond het aantal aangemelde scholen op 88, waarvan 43 eerste keuze scholen. Voor een aantal 'openstaande' volgnummers bleek geen mogelijkheid tot deelname meer, omdat geen van de vier/vijf benaderde scholen op het volgnummer deel wilde nemen aan de peiling. Met de opdrachtgever is half november afgesproken dat volgnummers ook mochten worden 'ingevuld' door scholen die nog niet waren benaderd (omdat de 1<sup>e</sup> of 2<sup>e</sup> school meedeed) maar wel voldeden aan de twee expliciete steekproefcriteria (dus in hetzelfde stratum zaten). Op deze manier zijn acht extra scholen uit andere regels in het steekproefbestand geworven. Daarnaast is het gelukt om eind november 2017 nog vijf extra scholen te werven op de reguliere manier.

### 2.3.4 Bruto wervingsresultaat en non-respons

Met de hierboven toegelichte initiële opzet van de werving en de aanpassingen van de wervingsstrategieën is uiteindelijk een bruto steekproef van 101 scholen gerealiseerd (Tabel 2.8). In totaal is van 230 scholen een beslissing voor het wel of niet deelnemen ontvangen.

Tabel 2.8: Bruto wervingsresultaat november 2017

		Persoonlijk contact	Aanmelding	Definitief geen deelname
Steekproef	Benaderd	98	41	57
	Back-up scholen <sup>25</sup>	2	2	0
Reserves	Benaderd	130	58	72
	1 <sup>e</sup> reserve (2xxx)	63	26	37
	2 <sup>e</sup> reserve(3xxx)	36	15	21
	3 <sup>e</sup> reserve (4xxx)	28	14	14
	Backupscholen (5xxx)	3	3	0
Totaal		230	101	129

Tabel 2.8 laat zien dat van alle in de wervingsperiode benaderde scholen er 129 hebben aangegeven niet aan het onderzoek te willen deelnemen. Dat betekent dat de algemene bereidwilligheid 44 procent bedroeg. Voor de eerste steekproef van 100 scholen gold dat ruim twee vijfde (41 scholen) deelname had toegezegd. Het aandeel 'eerste keuze' scholen bedraagt 43 procent wanneer de twee backup-up scholen worden meegerekend die direct zijn benaderd vanwege technische non-respons bij de school die als eerste was getrokken, bijvoorbeeld in het geval van gefuseerde scholen. Binnen die groep van 43 scholen bevonden zich 28 EIBO Scholen (van de beoogde 68) en 15 VVTO scholen (van de beoogde 32). De bereidwilligheid tot deelname was derhalve in de eerste steekproef van 100 scholen onder VVTO scholen iets groter dan onder reguliere scholen (47% versus 41%).

#### Non-respons: aantallen

Van de 230 benaderde scholen hadden er 101 toegezegd deel te nemen, terwijl er 129 negatief reageerden op het verzoek. Tabel 2.9 geeft de verdeling van de non-respons naar stratum weer. De non-respons is het laagste in het EIBO met 0%-stratum (33%) en het hoogste in het VVTO met >25%- stratum (68%). Uit de tabel valt op te maken dat het percentage non-respons voor beide 'typen' scholen een oplopend patroon kent naar gewichtenstrata. Hoe meer gewichtenleerlingen op schoolniveau des te kleiner het percentage 'succes' in de werving.

Tabel 2.9: Non-respons naar stratum

Cel	Totaal benaderd	Geen aanmelding	Aangemeld	Percentage Non-respons
EIBO S1	12	4	8	33 %
EIBO S2	90	52	38	58 %
EIBO S3	34	20	14	59 %
EIBO S4	21	13	8	62 %
VVTO S1	14	6	8	43 %
VVTO S2	17	8	9	47 %
VVTO S3	17	9	8	53 %
VVTO S4	25	17	8	68 %
Totaal	230	129	101	56 %

<sup>25</sup> Zoals eerder vermeld zijn twee scholen uit de eerste steekproef door technische non-respons (school opgeheven) niet benaderd. Hiervoor is de back-up school (school 5xxx in de steekproefregel) benaderd in plaats van de reguliere reservescholen. Beide scholen hebben deelgenomen waardoor benadering van de reguliere reserves niet meer nodig was.

De non-respons had met deze verdeling van 101 scholen geen consequenties voor de generaliseerbaarheid van de resultaten van dit peilingsonderzoek naar de expliciete stratificatiecriteria, aangezien de reservescholen altijd behoorden tot hetzelfde stratum en type school als de beoogde 'eerste' school. Ook door de toevoeging van acht scholen uit andere steekproefregels (uit dezelfde strata als de beoogde scholen van een steekproefregel waar alle scholen weigerden) is dit niet in gevaar gekomen.

### *Non-respons: redenen*

Ambassadeurs hadden de opdracht hun benaderingspogingen en wervingsresultaten bij te houden in zogenoemde logboeken. Deze logboeken zijn achteraf geanalyseerd om de non-respons in kaart te brengen. Tabel 2.10 laat de in de logboeken genoteerde redenen zien, gerubriceerd naar vijf categorieën. Scholen konden natuurlijk meerdere redenen opgeven, waardoor het aantal geturfde redenen (134 iets hoger is dan het totaal aantal weigerende scholen (129)).

De toetskalender wordt door bijna 25% van de benaderde scholen (57 keer) genoemd en is daarmee de meest genoemde reden om niet mee te doen. Sommige scholen voorzagen logistieke problemen omdat er al ander (extern) onderzoek in de afnameperiode was gepland. Ook werd de afnameperiode (januari-maart) niet ideaal bevonden, aangezien deze maanden in het teken staan van de standaard Cito volgroetsen (M8), die in groep 8 op veel scholen essentieel zijn als het gaat om het schooladvies. Daarnaast gaven veel scholen meer in algemene zin aan dat men behoorlijk 'onderzoeks- en toetsmoe' was, en derhalve geen extra toetsen in genoemde periode wensten.

Onvoldoende draagvlak voor dit onderzoek (bij leerkrachten en/of directie) is een tweede reden om deelname te weigeren. Scholen hebben specifiek geen belangstelling voor onderzoek op het gebied van Engels. Deze reden wordt vaker aangegeven bij EIBO scholen dan bij VVTO scholen (18% van de benaderde EIBO scholen versus 14% van de benaderde VVTO scholen). De ambassadeurs geven aan dat de angst om leerkrachten en leerlingen bloot te stellen aan iets wat ze nog niet voldoende beheersen, hier mogelijk een rol speelt. Zo werd er onder andere door scholen genoemd dat "het Engels nog een verbeterslag nodig heeft", of "nog te veel in de kinderschoenen staat". Een andere voorbeeld in dit opzicht is "we zijn ontevreden over de methode, en dan krijgen we het verkeerde beeld als we de resultaten ontvangen".

Een derde categorie redenen is dat scholen direct, zonder duidelijke opgave van redenen hebben geweigerd ('expliciete weigering', 7%). Grotendeels gaat het om scholen die elk extern onderzoek afhouden, en daadwerkelijk 'nergens aan meedoen', maar het is mogelijk dat enkele scholen dit aangaven om zodoende het verzoek maar helemaal niet in overweging te hoeven nemen.

Onder de vierde categorie redenen, gelabeld als 'andere toetsingsvorm', vallen scholen die Engels via eigen toetsen of landelijke toetsen (b.v. Cito-toets Engels) in kaart brengen. Nog een keer dezelfde vaardigheden toetsen was voor deze scholen niet opportuun. Deze categorie is vaker gehoord bij de VVTO scholen dan bij EIBO scholen (12% versus 5%). In 2% van de non-respons is er sprake van (verwachte) technische non-respons: scholen waren bij benadering reeds gefuseerd of opgeheven of gaven aan dat deze verandering in het lopende schooljaar zou plaatsvinden (twee steekproefscholen, één reserveschool). Deze groep bevat verder twee reservescholen die bij benadering door de ambassadeur niet bleken te voldoen aan het beoogde type VVTO of EIBO. Deze scholen konden derhalve niet worden geselecteerd voor de steekproef, en in dat geval is meteen de eerstvolgende reserveschool benaderd.

Tabel 2.10 bevestigt verder het oplopende patroon van weigering naar de gewichtenstrata uit tabel 2.9. Alleen ‘expliciete weigering’ van het verzoek zien we iets vaker in de S3 en S4 strata, maar de verschillen zijn niet groot.

Tabel 2.10: redenen om niet deel te nemen, in aantal en in percentages van benaderde scholen

Niet aangemeld, omdat:	Alle benaderde scholen			Benaderd naar stratum			
	EIBO n=157	VVTO n=73	Totaal N= 230	S1 n=26	S2 n=107	S3 n=51	S4 n=46
<i>Totaal genoemde redenen</i>	90	44	134	10	64	30	30
Conflicterend met toetskalender	38 (24%)	19 (26%)	57 (25%)	5 (19%)	28 (27%)	12 (24%)	12 (26%)
Onvoldoende draagvlak voor (opzet) onderzoek	29 (18%)	10 (14%)	39 (17%)	2 (6%)	18 (18%)	10 (20%)	8 (17%)
Expliciete weigering	11 (7%)	5 (7%)	16 (7%)	1 (4%)	7 (7%)	4 (8%)	4 (9%)
Andere toetsingsvorm	8 (5%)	9 (12%)	17 (7%)	1 (4%)	9 (8%)	3 (6%)	4 (9%)
(verwachte) Technische non-respons	4 (3%)	1 (1%)	5 (2%)	1 (4%)	2 (2%)	1 (2%)	1 (2%)
Totaal scholen dat deelname weigert	89 (57%)	40 (55%)	129 (56%)	10 (38%)	60 (56%)	29 (57%)	30 (65%)

### Non-respons: ten slotte

In de logboeken is uiteindelijk vaak kort de voornaamste reden uit de gesprekken met schoolleiders genoteerd. Het is mogelijk dat er meerdere redenen zijn genoemd die niet allemaal zijn genoteerd. De logboeken van ambassadeurs bevestigen het beeld dat het iets lastiger was scholen uit de hogere gewichtenstrata te overtuigen van deelname (zie Tabellen 2.9 en 2.10 hierboven). Dit betekent in ieder geval dat de belangstelling naar onderzoek naar Engels onder dergelijke scholen iets lager is. Of de onzekerheid over (de verwachte resultaten van) het vak Engels bij leerkrachten sterker speelt in deze scholen is onvoldoende duidelijk: in de logboeken wordt maar één keer de leerlingpopulatie specifiek benoemd als reden om niet deel te nemen: geen belangstelling want “we durven het niet aan om deze zwakke leerlingen te toetsen op Engels” (EIBO-school, 10-25%). Zoals eerder opgemerkt heeft dit voor de representativiteit van de resultaten geen gevolg: er zijn immers voldoende scholen uit elk steekproefstratum in de gerealiseerde steekproef opgenomen.

## 2.4 Gerealiseerde steekproef: scholen

Hoewel het bruto wervingsresultaat 101 scholen bedroeg, ligt het nettoresultaat lager. Uiteindelijk zijn zes scholen uitgevallen, ondanks toezeggingen. De redenen hiervoor waren divers: reorganisatie/nieuwe directeur die niet akkoord ging met de reeds gemaakte afspraken (1 school); onrust in de (nieuwe) groep 8 (1 school); langdurige ziekte van de vaste groepsleerkracht (2 scholen), onvoldoende draagvlak voor het onderzoek onder ouders (1 school); en zwangerschapsverlof (waarbij vervangende leerkracht niet wilde deelnemen, 1 school).

Met de uitval van deze zes scholen bedroeg het netto wervingsresultaat **95 scholen**. Tabel 2.12 laat de gerealiseerde steekproef zien. De deelnemende scholen, 32 VVTO scholen en 63 EIBO scholen, zijn min of meer verdeeld over de steekproefcellen zoals beoogd. Ten opzichte van de

steekproefverdeling misten dus vijf EIBO scholen. Vier van deze vijf scholen zitten in het stratum 'EIBO-S4o2 (reguliere school met >0-10% gewichtenleerlingen). In aantal blijft dit met 34 scholen wel de grootste groep van de acht combinatiestrata.

Zoals in paragraaf 2.2.1 is beschreven, waren de aanwezige bronnen onvoldoende eenduidig waar het gaat om het aantal VVTO-scholen in Nederland. Om na te gaan of een school, die volgens de steekproef een VVTO-school zou zijn dit ook daadwerkelijk was, is in het eerste persoonlijke wervingscontact aan de hand van twee vragen gecheckt of de leerlingen in groep 8 vanaf schooljaar 2013-2014 (of eerder) Engels hebben gehad (zie paragraaf 2.2.1). Dat is bij alle beoogde VVTO scholen gebeurd; bij EIBO-scholen kwam dit gaandeweg het gesprek soms aan de orde, maar is dit niet systematisch bevraagd. In paragraaf 2.3.4 wordt reeds beschreven dat tijdens de werving twee scholen al niet bleken te voldoen aan de steekproefcriteria en daarom op voorhand niet zijn geworven.

Met scholen die hadden toegezegd te willen deelnemen is in het najaar een afspraak gemaakt voor een schoolbezoek door ofwel de ambassadeur ofwel de hoofdtestleider, het zogenaamde intakegesprek, dat in paragraaf 4.4 nader wordt toegelicht. Tijdens dit gesprek werd onder meer een korte schoolvragenlijst afgenomen. Deze vragenlijst bevatte onder andere vragen over de profielkeuze van de school (EIBO, vervroegd EIBO, VVTO) en over de onderwijstijd Engels in de leerjaren 1-8 van de deelnemende groep-8 leerlingen.

Gedurende de dataverwerking, na afloop van alle afnamen binnen het hoofdonderzoek, bleek uit de intakegegevens dat voor enkele scholen de profielkeuze en/of de gerapporteerde onderwijstijd niet overeenkwam(en) met het beoogde steekproefstratum. Via telefonisch contact met deze scholen zijn de onduidelijkheden voorgelegd en is nagegaan hoe de school daadwerkelijke getypeerd moest worden. Dit heeft geleid tot een verschuiving in de uiteindelijke verdeling naar strata binnen de gerealiseerde steekproef. Specifiek gaat het om vier beoogde EIBO scholen die, na verificatie, alsnog aan het VVTO criterium bleken te voldoen en om drie beoogde VVTO scholen die niet voldeden aan de harde eis van het inruimen van onderwijstijd Engels vanaf schooljaar 2013-2014 (of eerder), maar getypeerd moesten worden met het profiel 'vervroegd EIBO' (bijvoorbeeld Engels vanaf leerjaar 5). Het uiteindelijke databestand bevat derhalve gegevens van 62 EIBO scholen en 33 VVTO scholen, zoals blijkt uit de kolom 'Deelname na correctie' in Tabel 2.11.

Tabel 2.11: Aanmelding en daadwerkelijke deelname naar stratum

Cel	Steekproef	Beoogd percentage	Aangemeld	Gerealiseerde deelname	Percentage realisatie	Deelname na correctie	Percentage na correctie
EIBO S1	8	8.0 %	8	8	8.4 %	7	7.4 %
EIBO S2	38	38.0 %	38	34	35.8 %	33	34.7 %
EIBO S3	14	14 %	14	14	14.7 %	15	15.8 %
EIBO S4	8	8.0 %	8	7	7.4 %	7	7.4 %
VVTO S1	8	8.0 %	8	8	8.4 %	9	9.5 %
VVTO S2	8	8.0 %	9	9	9.5 %	10	10.5 %
VVTO S3	8	8.0 %	8	7	7.4 %	6	6.3 %
VVTO S4	8	8.0 %	8	8	8.4 %	8	8.4 %
Totaal	100	100 %	101	95	100 %	95	100 %

Bij de evaluatie op de hoofdcategorie 'type school' is te zien dat de EIBO-scholen nu 65,3% van de gerealiseerde steekproef uitmaken (van de beoogde 68%); het percentage VVTO-scholen steeg van 32 naar 34,7%. Hiermee zijn de VVTO scholen (sterker) oververtegenwoordigd ten opzichte van de steekproefopzet. Door in de verdere analyses in dit rapport terug te wegen naar de populatie wordt in de analyseresultaten deze oververtegenwoordiging tenietgedaan. Bij de gewichtenstrata zijn de mutaties in de gerealiseerde steekproef na deze correctie kleiner, in alle gevallen verschilt het percentage ongeveer 1 procentpunt. In het vervolg van dit rapport wordt gebruik gemaakt van deze gecorrigeerde versie van de indeling van scholen.

#### 2.4.1 Representativiteit op schoolniveau

Om de representativiteit van de gerealiseerde steekproef na te gaan is een aantal achtergrondkenmerken van de totale set deelnemende scholen vergeleken met die van de landelijke populaties basisscholen.

Tabel 2.12 bevat de verdeling naar regio, mate van stedelijkheid, schoolgrootte, percentage gewichtenleerlingen op schoolniveau, toezichtarrangement en denominatie binnen de steekproef vergeleken met de totale populatie basisscholen. De eenheid 'school' is in deze peiling gedefinieerd door de combinatie van BRIN-code en vestigingsnummer. Ook nevenvestigingen zijn dus, net als bij de steekproef, betrokken in de vergelijking. In tegenstelling tot de steekproef (schooljaar 2016-2017) kon bij deze vergelijking wel gebruik gemaakt worden van de landelijke DUO-schoolgegevens van het schooljaar van de afname (2017-2018). Er is gebruikt gemaakt van de DUO-bestanden '02.leerlingen-bo-in-bron-2017-2018.xlsx', en '03.leerlingen-bo-sbo-leerjaar-geslacht-2017-2018.csv' die verkrijgbaar zijn via de website van DUO. Het eerstgenoemde bestand kent 6426 schoolvestigingen basisonderwijs (brinnummers), het tweede bestand bevat 6403 schoolvestigingen. Het DUO-bestand bevat geen directe gegevens over de mate van stedelijkheid. Hiervoor is gebruik gemaakt van het CBS-bestand '161010-Kenmerken-postcode-mw.xlsx'. Daarin is de mate van stedelijkheid gekoppeld aan alle Nederlandse postcodes.

De scholen zijn ingedeeld naar percentage gewichtenleerlingen (één van de twee expliciete criteria). Voor het percentage gewichtenleerlingen op schoolniveau zijn dezelfde gewichtenstrata gehanteerd als bij het vaststellen van de steekproef (zie paragraaf 2.1). Omdat we in de steekproef met opzet rekening hielden met een strikte indeling van het percentage gewichtenleerlingen in vier strata (0%, >0-10%, >10-25%, >25%), blijkt de verdeling van de groepen in de steekproef (95 scholen)

niet volledig overeen te komen met de populatie. Daarnaast kunnen sommige scholen van gewichtenstratum zijn verschoven ten opzichte van de steekproef omdat een ander gegevensjaar is gebruikt. In paragraaf 2.2.1 is dit punt reeds aangehaald. Er is sprake van significante verschillen in de verdeling van de scholen over de vier strata ten opzichte van de populatie:  $\chi^2(3; N = 6400) = 13,21$ ,  $p = 0,004$ .

Voor het impliciete criterium 'stedelijkheid' geldt dat er geen statistisch significante verschillen bestaan wanneer de steekproef wordt vergeleken met de totale populatie basisscholen  $\chi^2(4; N = 6413) = 0,34$ ,  $p = 0,987$ . Ook voor het impliciete criterium 'schoolgrootte' geldt dat er geen significante verschillen tussen de steekproef (95 scholen) en de populatie zijn:  $\chi^2(4; N = 6426) = 5,11$ ,  $p = 0,276$ . De uitsluiting in de steekproefopzet van de zeer kleine scholen (minder dan zes te verwachten leerlingen in groep 8) heeft dus niet geleid tot ernstige ondervertegenwoordiging van kleine scholen.

Door de uitsluiting van scholen met een (verwacht) aantal van 6 groep 8-leerlingen of lager, zien we ten opzichte van de gehele populatie vooral een afwijking in de onderste groep (<6 leerlingen), als het gaat om de omvang van groep 8 per school (geen scholen in de steekproef, ten opzichte van 4.2% in de populatie). Bij de overige groepen, die qua aantal leerlingen globaal kunnen worden omschreven als 'maximaal halve klas (6 tot 15)', 'normale klas' (16-30), 'twee klassen (31-60) 'en meer dan twee klassen', zien we dat er vooral een verschil is waar te nemen in de middencategorie (16 tot 30 leerlingen). Ten opzichte van de populatie hebben relatief meer scholen aan de peiling meegedaan waar, in termen van het leerlingenaantal, sprake is van een 'normale klas.

Wanneer gekeken wordt naar de geografische ligging van de scholen, dan worden vier regio's onderscheiden: Noord (provincies Groningen, Drenthe en Friesland), Oost (provincies Gelderland en Overijssel), Midden (provincies Flevoland, Noord-Holland, Zuid-Holland en Utrecht) en Zuid (provincies Noord-Brabant, Limburg en Zeeland). Er blijkt een oververtegenwoordiging van de steekproefscholen uit de oostelijke provincies aan het onderzoek te hebben meegedaan. Voor regio Zuid is er juist sprake van een ondervertegenwoordiging. De gevonden verschillen tussen steekproefscholen en populatie zijn significant met een drempel van  $p < 0,05$ :  $\chi^2(3; N = 6426) = 9,01$ ,  $p = 0,029$ .

Zoals verwacht staat het overgrote deel van de scholen onder basistoezicht (Inspectie van het Onderwijs, peildatum 1 april 2017). Er hebben geen 'zwakke scholen' deelgenomen aan het onderzoek. De verschillen ten opzichte van de populatie zijn statistisch niet significant:  $p = .240$ , *Fisher's exact test*.

Het DUO-bestand bevat zeventien categorieën als het gaat om denominatie. Voor de overzichtelijkheid zijn deze groepen gereduceerd tot vier hoofdgroepen. De categorieën 'algemeen bijzonder', 'openbaar' en 'rooms-katholiek' zijn integraal overgenomen uit het DUO-bestand. Algemeen bijzonder omvat scholen die niet op basis van levensbeschouwing zijn in te delen, maar bijvoorbeeld behoren tot het traditioneel vernieuwingsonderwijs (Jenaplan, Montessori et cetera). Aan de groep protestants-christelijke scholen zijn naast de scholen die zich zo afficheren ook de Evangelische, de Gereformeerd vrijgemaakte en de Reformatorische scholen toegevoegd. In de categorie 'overig levensbeschouwend' vallen alle scholen van religieuze minderheden (islamitische, joodse, hindoeïstische scholen), confessioneel samenwerkende scholen (bijvoorbeeld rooms-katholiek en protestants-christelijk), 'interconfessionele' scholen en antroposofische scholen. De vergelijking laat zien dat er geen sprake is van een statistisch significant verschil tussen de gerealiseerde steekproef en populatie:  $\chi^2(4, N = 6426) = 3,43$ ,  $p = 0,489$ .



Tabel 2.12: Deelname scholen naar achtergrondkenmerken ten opzichte van landelijke populatie

	Steekproef (N=95)		Landelijke populatie (N=6426) <sup>26</sup>	
	n	%	n	%
<i>Regio (N=6426)</i>				
Noord	13	13.7 %	940	14.6 %
Oost	32	33.7 %	1413	22.0 %
Midden	37	38.9 %	2688	41.8 %
Zuid	13	13.7 %	1385	21.6 %
<i>Stedelijkheid (N=6413)</i>				
Zeer sterk stedelijk	14	14.7 %	877	13.7 %
Sterk stedelijk	18	18.9 %	1357	21.2 %
Matig stedelijk	18	18.9 %	1187	18.5 %
Weinig stedelijk	19	20.0 %	1287	20.1 %
Niet stedelijk	26	27.4 %	1705	26.6 %
<i>Schoolgrootte (N=6400)</i>				
1-100 lln.	18	18.9 %	1198	18.7 %
101-200 lln.	30	31.6 %	2125	33.2 %
201-300 lln.	27	28.4 %	1639	25.6 %
301-400 lln.	16	16.8 %	797	12.5 %
401 en meer lln.	4	4.2 %	641	10.0 %
<i>Aantal leerlingen in leerjaar 8 (N=6403)</i>				
Minder dan 6	0	0.0%	272	4.2 %
6 tot 15	19	20.0 %	1447	22.6 %
16 tot 30	47	49.5 %	2606	40.7 %
31 tot 60	27	28.4 %	1797	28.1 %
Meer dan 60	2	2.1 %	282	4.4 %
<i>Percentage gewichtenleerlingen (N=6400)</i>				
0% leerlingen	14	14.7 %	456	7.1 %
> 0%-10% leerlingen	45	47.4 %	3986	62.0 %
>10%-25% leerlingen	24	25.3 %	1419	22.2 %
> 25% leerlingen	12	12.6 %	539	8.4 %
<i>Toezichtarrangement (N=6263)*</i>				
Basis	93	98.9%	6163	98.5 %
Zwak	0	0.0 %	88	1.4 %
Zeer zwak	1	1.1	12	0.2 %
<i>Denominatie (N=6426)</i>				
Algemeen bijzonder	8	8.4 %	324	5.0 %
Openbaar	31	32.6 %	2060	32.1 %
Rooms-katholiek	24	25.3 %	1928	30.0 %
Protestants-christelijk	30	31.6 %	1905	29.6 %
Overig levensbeschouwelijk	2	2.1 %	209	3.3 %

\* het toezichtarrangement ontbreekt voor 1 (steekproef) en respectievelijk 162 (populatie) gevallen

<sup>26</sup> In de genoemde aantallen zijn ook de vestigingen opgenomen waar minder dan 6 groep-8 leerlingen waren ingeschreven, terwijl dergelijke scholen in de steekproef waren uitgesloten. In schooljaar 2017/2018 waren er 272 (4.2%) bo scholen met geen tot en met vijf leerlingen in leerjaar 8. Voor een volledige vergelijking is daarom ook gekeken naar het aantal leerlingen in leerjaar 8.

Ten aanzien van de achtergrondvariabelen kan geconcludeerd worden dat de basisscholen uit de steekproef (95 scholen) voldoende representatief zijn voor de populatie op het ‘percentage gewichtenleerlingen’ en op de variabele ‘regio’ na. Bij gebruik van de data (op schoolniveau) kan er derhalve voor gekozen worden voor deze variabele te corrigeren.

## 2.5 Gerealiseerde steekproef: leerlingen

In de opzet van de peiling was het de bedoeling dat alle leerlingen in leerjaar 8 van de school aan het onderzoek zouden deelnemen. Dit had tot gevolg dat er acht scholen in de gerealiseerde steekproef zitten waar twee of meer groepen 8 hebben deelgenomen. Op verzoek van de school was het verder mogelijk dat, in het geval van gecombineerde groepen (7-8 of 6-7-8), ook de leerlingen in leerjaar 7 aan de afname meededen. Voor dergelijke leerlingen heeft weliswaar een terugkoppeling naar de scholen plaatsgevonden, maar deze zijn niet in de databestanden opgenomen. In de voorbereiding van de afname hebben de 95 scholen de gegevens van 2342 groep-8 leerlingen aangeleverd. In Tabel 2.13 zijn de aantallen *aangemelde* leerlingen uitgesplitst naar stratum. Voorafgaand aan of net na de intakegesprekken zijn bij de contactpersoon op de school de namen en achtergrondgegevens van de beoogde deelnemers opgevraagd<sup>27</sup>. De achtergrondgegevens worden gepresenteerd in paragraaf 2.5.1. Aan elke leerling is een vragenlijstnummer en een antwoordblad (met barcode) toegewezen. Voor elke klas of deel van de klas (bij combinatieklassen) is verder telkens één reservevragenlijst en één reserve-antwoordblad aangemaakt.

Tabel 2.13: Aanmeldingen leerlingen naar stratum

Cel	Scholen	Aantal leerjaar 8	Percentage
EIBO-S4o1 (0%)	7	183	7.8 %
EIBO-S4o2 (>0-10%)	33	843	36.0 %
EIBO-S4o3 (>10-25%)	15	321	13.7 %
EIBO-S4o4 (>25%)	7	226	9.6 %
VVTO-S4o1 (0%)	9	257	11.0 %
VVTO-S4o2 (>0-10%)	10	268	11.4 %
VVTO-S4o3 (>10-25%)	6	134	5.7 %
VVTO-S4o4 (>25%)	8	110	4.7 %
Totaal	95	2342	100

### *Uitval op leerlingniveau*

Tussen half januari en eind maart hebben de afnames plaatsgevonden. In de meeste gevallen betekent dit dat op de school met de beoogde klas(en) alle taken zijn afgenomen. Uitzonderingen op schoolniveau zijn: één school waarvan alleen klassikale taken beschikbaar zijn (want geen actieve toestemming van ouders voor de gesprekstaken) en uitval van één van twee klassen (21 leerlingen) vanwege ziekte van de leerkracht (en het ontbreken van een geschikt vervangingsmoment in de afnameperiode). Op alle 95 scholen heeft derhalve een (groot) deel van de beoogde leerlingen

<sup>27</sup> Enkele scholen kozen er voor geen voor en/of achternamen aan te leveren. In die gevallen is met de school, ten behoeve van de terugkoppeling, een voor de school herkenbaar volgnummer aan de leerlingen toegewezen.

deelgenomen. Op leerling-niveau is er uitval vanwege a) ziekte of afwezigheid vanwege een andere reden en b) het ontbreken van ouderlijke toestemming voor het gehele onderzoek<sup>28</sup>. Twee leerlingen met TOS zijn uitgesloten van deelname (exclusie criterium), terwijl een andere TOS-leerling uiteindelijk toch heeft meegedaan aan de klassikale toetsen, op verzoek van de school. Het reservemateriaal is door negen (9) additionele leerlingen gebruikt, leerlingen die ten tijde van de verzameling van leerlingenlijsten nog niet waren aangemeld<sup>29</sup>. In Tabel 2.14 is de uitval op leerlingniveau ingedeeld naar stratum.

Tabel 2.14: Gerealiseerde aantallen leerlingen en uitval naar stratum

	Aantal Scholen	Aan gemeld	Ziek, afwezig		Geen toestemming		Totale uitval	Extra Lln	Netto
	n	n	n	%	n	%	%	n	n
EIBO-S4o1 (0%)	7	183	17	9.3	2	1.1	10.4		164
EIBO-S4o2 (>0-10%)	33	844	30	3.6	49	5.8	9.4	4	765
EIBO-S4o3 (>10-25%)	15	321	16	5.0	17	5.3	16.8*		267*
EIBO-S4o4 (>25%)	7	226	14	6.2	29	12.8	19.0		183
VVTO-S4o1 (0%)	9	257	11	4.3	11	4.3	8.6	1	235
VVTO-S4o2 (>0-10%)	10	268	14	5.2	11	4.1	9.1	2	243
VVTO-S4o3 (>10-25%)	6	134	8	6.0	2	1.5	7.5	1	124
VVTO-S4o4 (>25%)	8	110	9	8.2	3	2.7	10.9	1	98
<b>Totaal (%)</b>	<b>95</b>	<b>2343</b>	<b>119</b>	<b>(5.1)</b>	<b>124</b>	<b>(5.3)</b>	<b>(11.3)</b>	<b>9</b>	<b>2088</b>

\* inclusief de uitval van een gehele klas van 21 leerlingen

Binnen het peilingsonderzoek is gebruik gemaakt van informed-consent van de ouders. Ouders hebben hiertoe een informatiebrief ontvangen met de mogelijkheid om van deelname af te zien. Ongeveer 95% van de aangemelde leerlingen (2219) had toestemming om aan het onderzoek mee te doen; 124 leerlingen hadden geen ouderlijke toestemming (5.3%). Van de groep leerlingen met toestemming was op de afnamedag vijf procent afwezig vanwege ziekte of andere redenen. Hierboven is al genoemd dat een hele klas van 21 leerlingen (0,9%) uit het onderzoek is weggevalen vanwege ziekte van de leerkracht. In totaal bedraagt de uitval ten opzichte van het totaal aantal aangeleverde leerlingen 11.3%.

Vermoed wordt dat de extra informatieronde naar ouders toe, waarin actieve toestemming voor de gesprekstaken werd gevraagd, (mede) debet is aan het percentage leerlingen dat helemaal geen toestemming heeft gekregen. Onder EIBO scholen met meer dan 25% gewichtenleerlingen is de uitval om die reden relatief gezien het grootst, namelijk 12.8%. Nadere inspectie van de toestemming per school laat zien dat op het overgrote deel (53, 56%) van de scholen alle kinderen toestemming hadden om mee te doen (geen enkele ouder tekende bezwaar aan). Op 37 scholen (39%) gaat het steeds om één tot drie leerlingen per klas en blijft de uitval (sterk) onder de 10 procent. Op vijf scholen bedraagt de uitval vanwege deze reden echter tussen de 23 en 43 procent. Het gaat specifiek

<sup>28</sup> Voorafgaand aan het onderzoek is ouders om toestemming voor deelname gevraagd. Ouders ontvingen hiertoe een brief over deelname aan het onderzoeken en de mogelijkheid om deelname van hun kind te weigeren. Voor deze leerlingen was bij de leerkracht een antwoordstrook met deelnameweigering retour gekomen. Deze leerlingen kregen tijdens de afname andere taken, meestal buiten het klaslokaal.

<sup>29</sup> Omdat er actieve toestemming gevraagd is voor het deelnemen aan de gesprekstaken hebben de ouders van deze leerlingen natuurlijk ook de toestemmingsbrief ontvangen.

om twee van 33 scholen in stratum EIBO-S4o2 (>0-10%); twee van zeven scholen in stratum EIBO-S4o4 (>25%) en één school van de 9 scholen in stratum VVTO-S4o1 (0%).

Voor 2088 leerlingen geldt dat er minimaal één toets-onderdeel of een vragenlijst beschikbaar is. Dat wil zeggen dat de meeste van deze leerlingen de vragenlijst en de drie klassikale taken hebben gemaakt en dat steeds zes leerlingen per klas de gesprekstaken hebben gedaan en daardoor niet alle klassikale taken. Voor enkele leerlingen geldt dat zij pas later in de afname aanschoven in de klas of juist tijdens de afname weg moesten (vanwege bijvoorbeeld medische afspraken). Om die reden is er dan alleen een vragenlijst of één of enkele toetsen beschikbaar. Op twee scholen is de luistertoets niet afgenomen, vanwege technische problemen. In de volgende paragraaf worden de achtergrondkenmerken van de 2088 leerlingen beschreven.

### 2.5.1 Representativiteit op leerlingniveau

Om de representativiteit van de deelnemende leerlingen na te gaan is een aantal van hun achtergrondkenmerken vergeleken met die van de landelijke populatie basisscholieren uit groep 8. Voor die vergelijking is gebruik gemaakt van landelijke DUO-gegevens die in bewerkte vorm via de Inspectie van het Onderwijs zijn verkregen. De via de Inspectie van het Onderwijs verkregen gegevens hebben betrekking op het schooljaar 2017/2018, het jaar van de afname van de peiling.

In de overzichten hieronder zijn de 2088 leerlingen van groep 8 meegenomen die minimaal aan één onderdeel van onderzoeksdesign hebben deelgenomen. Tabel 2.15 bevat de verdeling van de leerlingen naar sekse, leeftijd, en leerlinggewicht voor de steekproefscholen en de populatie. Zoals uit de tabel kan worden afgeleid, is het percentage jongens en meisjes in de peiling redelijk in lijn met de verwachtingen. Verschil met de populatie is niet significant  $\chi^2(1, N = 175070) = 1.74, p = 0,187$ . Wat de verdeling naar leeftijd betreft geldt dat voor de bepaling van de representativiteit is uitgegaan van de leeftijd van de leerlingen uit de steekproef op 01-10-2017. De gegevens van de populatie hebben betrekking op groep 8 leerlingen in het schooljaar 2017/2018 en daarbij is dezelfde peildatum van de leeftijd gehanteerd. Uit de tabel valt op te maken dat de leerlingen gemiddeld gezien iets ouder lijken dan in de landelijke populatie. De verschillen zijn echter niet significant,  $\chi^2(4, N = 175061) = 7.68, p = 0,105$ . Bij analyses naar leeftijd (op een specifieke peildatum) hoeft dus niet te worden teruggewogen. Verder blijkt voor de variabele leerlinggewicht dat de leerlingen met name zijn oververtegenwoordigd in de categorie 1.2. Bij een gehanteerde  $p$ -waarde van  $<0.01$  zijn de verschillen tussen leerlingen van de steekproef en populatie zoals verwacht statistisch significant  $\chi^2(2, N = 175061) = 22.56, p = 0,000$ . Hierbij moet worden opgemerkt dat voor een relatief groot deel van de deelnemende leerlingen (14.1%) geen gewicht is aangeleverd door de scholen. De oververtegenwoordiging van het percentage kan (mede) verklaard worden doordat de steekproef een oververtegenwoordiging van scholen met een hoog percentage gewichtenleerlingen bevat.

Tabel 2.15: Deelname leerlingen naar achtergrondkenmerken ten opzichte van landelijke populatie

	Steekproef (95 scholen)		Landelijke populatie (6426 scholen)	
	n	%	n	%
<i>Sekse</i>				
Jongen	1008	48.6	87.555	50.01
Meisje	1068	51.4	87.515	49.99
<i>Totaal</i>	2076		175.070	
Niet bekend	12			
<i>Leeftijd op 1-10-2017</i>				
9 jaar oud	0	0	193	0.0
10 jaar oud	240	12.5	20.322	11.6
11 jaar oud	1414	73.9	131.177	74.9
12 jaar oud	248	13.0	22.802	13.0
13 jaar oud	11	0.6	567	0.0
<i>Totaal</i>	1913		175.061	
Niet bekend	175		9	
<i>Gemiddelde leeftijd</i>	11,2 jaar		10,98 jaar	
<i>Leerlinggewicht</i>				
Geen gewicht	1577	88.2	159.123	90.9
Gewicht 0,3	92	5.1	8.251	4.7
Gewicht 1,2	119	6.7	7.696	4.4
<i>Totaal</i>	1788		175.070	
Niet bekend	300			

## 2.5.2 Representativiteit leerlingen gesprekstaken

### *Beschikbare leerlingen voor de gesprekstaken*

Niet alleen de representativiteit van de steekproef op leerlingniveau is relevant. Voor dit onderzoek is het ook belangrijk na te gaan in hoeverre de groep leerlingen, die toestemming hadden voor de gesprekstaken, representatief was. Zoals in de hoofdstukken 1 en 4 is beschreven, was het voor de afname van de gesprekstaken, die op video werden opgenomen, noodzakelijk om actieve toestemming van de ouders te verkrijgen. Van de 2088 groep 8-leerlingen is voor 339 van de leerlingen (16.3%) geen actieve toestemming van de ouders ontvangen voor de afname van de gesprekstaken. Dat is inclusief 20 leerlingen waarbij de schoolleider weigerde de actieve toestemming bij ouders op te vragen. Van de overige 94 scholen is op 24 scholen (26% van deze 94 scholen) van alle ouders actieve toestemming van de ouders ontvangen. Op de overige scholen (74% van de 94 scholen waar actieve oudertoestemming is gevraagd) varieerde het percentage leerlingen waarvoor toestemming is verkregen van 52% tot en met 98% (mediaan=80%). Al met al waren er 1749 leerlingen op 94 scholen beschikbaar om de gesprekstaken uit te voeren. In principe zouden in ieder geval 6 leerlingen per school de gesprekstaken uitvoeren. Op acht scholen lijkt dit niet gelukt te zijn. Hiervan waren er op vier scholen geen zes leerlingen beschikbaar, omdat er ten tijde van de

afname geen 6 leerlingen in groep 8 aanwezig waren (2 scholen<sup>30</sup>) of omdat we niet voor 6 leerlingen actieve toestemming hadden ontvangen (ook 2 scholen). Voor vier scholen is niet meer na te gaan waarom er geen zes leerlingen zijn gefilmd. Mogelijk was hier sprake van een mislukte opname, waardoor er slechts van vier leerlingen opnames waren, of is er om een andere onverklaarbare reden één gesprek niet afgenomen (bijvoorbeeld op verzoek van de school). Op negen scholen zijn bij meer dan 6 leerlingen de gesprekstaken afgenomen. Hiervan zijn op vijf scholen door twaalf leerlingen gesprekstaken uitgevoerd. Daarnaast zijn er in totaal nog drie scholen waar bij tien, negen of zeven leerlingen de gesprekstaken zijn afgenomen.

### **Representativiteit**

Om de representativiteit van de leerlingen die de gesprekstaken mochten uitvoeren, na te gaan, zijn de resultaten van beide groepen leerlingen (met en zonder toestemming voor de gesprekstaken) op de toetsen Lezen, Luisteren en Woordenschat met elkaar vergeleken. Toetsing van de verschillen in de gemiddelde scores met behulp van de Independent Samples T-test laat zien dat de prestaties op Lezen, Luisteren niet significant van elkaar verschillen; bij Woordenschat was dit wel het geval maar lijkt het geringe absolute verschil niet heel relevant. De resultaten bij Lezen van de groep leerlingen met actieve toestemming ( $M=.42$ ,  $S.D.=0.45$ ) verschilden niet significant van die van de groep zonder deze toestemming ( $M=0.36$ ,  $S.D.=0.43$ ;  $t(df=1858)=1.91$ ,  $p=0.06$ ). Ook de resultaten bij Luisteren van de groep leerlingen met actieve toestemming ( $M=.31$ ,  $S.D.=0.42$ ) verschilden niet significant van die van de groep zonder deze toestemming ( $M=0.28$ ,  $S.D.=0.42$ ;  $t(df=1817)=1.14$ ,  $p=0.25$ ). De resultaten van beide groepen verschilden wel op woordenschat. De gemiddelde score van de groep leerlingen met actieve toestemming ( $M=0.38$ ,  $S.D.=0.44$ ) is significant hoger dan die van de groep leerlingen zonder actieve toestemming ( $M=0.31$ ,  $S.D.=0.42$ ;  $t(df=1801)=2.67$ ,  $p=0.01$ ). Het geconstateerde verschil in gemiddelden is 0.07. Gezien de behaalde range aan scores op de vaardigheidsschaal woordenschat van 2.67 lijkt dit geringe absolute verschil niet heel relevant te zijn.

---

<sup>30</sup> Ook al is er bij de werving systematisch nagegaan of er per school minimaal 6 leerlingen in groep 8 zouden zitten, het aantal leerlingen per school dat deelnam lag op 2 scholen lager dan 6.



## Hoofdstuk 3. Design

In dit hoofdstuk wordt het design van de peiling beschreven. In paragraaf 3.1 gaan we allereerst in op de uitgangspunten voor het design. In paragraaf 3.2 wordt in algemene termen besproken welke taken de leerlingen hebben uitgevoerd. In paragraaf 3.3 wordt per kenmerk aangegeven hoe het toetsdesign er uitzag. De manier waarop vragenlijsten, toetsen en spreek- en gespreksopdrachten gecombineerd worden staat beschreven in paragraaf 3.4. Ten slotte komen in paragraaf 3.5 de aantallen observaties ter sprake.

### 3.1 Uitgangspunten

De peiling Engels had tot doel om een breed beeld van het prestatieniveau van groep 8 leerlingen in het reguliere basisonderwijs in kaart te brengen. Dit is gedaan door bij de leerlingen een vragenlijst af te nemen en vier vaardigheden te toetsen. Om ervoor te zorgen dat alle relevante aspecten binnen deze vaardigheden werden gemeten, is er gewerkt met een design voor afname en toetsing waarin iedere leerling slechts een deel van alle opdrachten hoefde uit te voeren. In dit hoofdstuk geven we een verantwoording van de overwegingen en keuzes die gemaakt zijn wat betreft het onderzoeksdesign, de selectie van de taken, het toetsdesign en de combinatie van taken binnen dat toetsdesign.

Bij opstellen van het toetsdesign werd met een viertal randvoorwaarden rekening gehouden: (1) de maximale tijd die de leerling deelneemt; (2) de vergelijkbaarheid van de resultaten; (3) de dekking van de vaardigheid door de opgaven; en (4) het aantal observaties per opgave en in totaal. De keuzen die omtrent deze randvoorwaarden zijn gemaakt leverden een viertal uitgangspunten die leidend waren in het design.

De keuze voor de bestede toetstijd per leerling is van logistieke aard: om de belasting voor uitvoering van het onderzoek voor leerlingen en scholen beperkt te houden, was het eerste uitgangspunt dat alle leerlingen in totaal maximaal twee uur aan het invullen van vragenlijsten, het maken van opgaven of het uitvoeren van mondelinge taken mochten besteden. Het tweede uitgangspunt was dat de resultaten van de leerlingen per vaardigheid met elkaar vergelijkbaar moesten zijn. Dit betekent dat binnen de peiling de resultaten van leerlingen voor bijvoorbeeld de vaardigheid *lezen* met elkaar vergeleken moesten kunnen worden. Daar waar mogelijk worden per vaardigheid tevens vergelijkingen met de resultaten van leerlingen in eerdere peilingen gemaakt, en zijn ook vergelijkingen met leerlingen in het voortgezet onderwijs mogelijk. Hiervoor is gebruik gemaakt van designs met ankeropgaven. Het derde uitgangspunt was dat er voldoende opgaven werden afgenomen, zodat (enige) generalisatie over de vaardigheid toelaatbaar is. Het gekozen aantal varieerde per vaardigheid. Het vierde uitgangspunt was dat er voldoende observaties per opgave en per vaardigheid zijn om betrouwbare uitspraken te kunnen doen. Om aan deze vier uitgangspunten te kunnen voldoen is gekozen voor een onvolledig design, waarin leerlingen een deel van alle opdrachten hebben uitgevoerd.



## 3.2 Algemene opzet

Binnen de peiling zijn in totaal vijf hoofdkenmerken gemeten. Het eerste hoofdkenmerk had te maken met de ervaring met en attitude ten opzichte van (het vak) Engels zoals deze door de leerlingen zelf wordt gerapporteerd. Dit is gemeten met een vragenlijst, die in principe door alle leerlingen is gemaakt. De overige vier kenmerken waren de vaardigheden: Luisteren, Lezen, Woordenschat en mondelinge vaardigheden. Deze laatste werd vooral gemeten als gespreksvaardigheid. Het invullen van de vragenlijst en het meten van ieder van deze vaardigheden nam telkens een 30 minuten in beslag. Uitgaande van een maximale toetsduur van in totaal twee uur en een meting van zoveel mogelijk verschillende vaardigheden bij zoveel mogelijk leerlingen betekende dit dat bij iedere leerling minstens drie van de vier vaardigheden zijn gemeten.

## 3.3 Toetsdesign per design-onderdeel

### 3.3.1 Toetsdesign voor de vragenlijst

Voor de vragenlijst is een eenvoudig design opgesteld: alle kandidaten krijgen alle vragen voorgelegd. De vergelijkbaarheid van de kandidaten is op de (sub)testcores bepaald.

### 3.3.2 Toetsdesign voor de vaardigheden Luisteren, Lezen en Woordenschat

Voor de vaardigheden Luisteren, Lezen en Woordenschat zijn designs gebruikt die sterk op elkaar lijken. Alle opgaven zijn op papier aangeboden. Er is overwogen om een gedeelte digitaal en gedeelte op papier aan te bieden, maar aangezien niet uitgegaan kon worden van voldoende betrouwbare internetverbindingen op alle scholen (die het aan zouden kunnen dat circa 24 leerlingen gelijktijdig een digitale toets maken) is hiervan afgezien. Voor ieder van de drie vaardigheden zijn zes varianten van papieren toetsen ("toetsboekjes") gemaakt, waarbij er voldoende overlap was in de gebruikte items in de verschillende varianten.

In Figuur 3.1 wordt de algemene opzet gegeven van het design voor de drie vaardigheden. Als voorbeeld wordt de vaardigheid Lezen genomen, maar dezelfde opzet geldt voor Woordenschat en Luisteren. De gehele verzameling opgaven Lezen die in de peiling aan de leerlingen werd aangeboden is in zes groepen verdeeld: Lezen blok 1 tot en met Lezen blok 6. Ieder blok bevatte een verzameling opgaven die in een kwartier kon worden gemaakt en een toets bestond telkens uit twee blokken. Ieder (item) blok (b1, b2, b3, etc.) kwam een keer aan het begin en een keer aan het einde van een toets voor om voor een deel rekening te houden met volgorde effecten door vermoeidheid. Elk van de zes gemaakte toetsen per vaardigheid bestond daarmee uit een specifieke combinatie van blokken in een vooraf opgestelde volgorde. Zo is **Le P1** (toets 1 bij Lezen) bijvoorbeeld te beschrijven als **Le P1 b1 I (blok 1 bij toets 1 bij Lezen, als onderdeel 1 (I) van de toets)** en **Le P1 b2 II (blok 2 bij toets 1 bij Lezen, als onderdeel 2 (II) van de toets)** en **Le P6** (toets 6 bij Lezen) als **Le P6 b6 I** en **Le P1 b1 II**. Daarbij was het streven dat ieder van de blokken (Lezen blok 1 tot en met Lezen blok 6) ongeveer evenveel opgaven zou bevatten. De blokken rouleerden bovendien over de zes versies, zodat er altijd sprake was van een overlap bestaande uit één blok.

Het aantal opgaven in een blok hing af van de hoeveel tijd die er nodig was per opgave. Dit aantal kon binnen een vaardigheid verschillen, bijvoorbeeld in het geval van matrix-vragen, waarin leerlingen meerdere subvragen per item moeten beantwoorden, zoals bij een leestekst met meerdere opgaven. Uitgangspunt bij de samenstelling van een blok was dat ieder blok een min of meer representatieve set van opgaven bevatte: daar waar mogelijk bestond elk blok uit zowel

nieuwe als ankeritems. De clustering van vragen bij Lezen en Luisteren bemoeilijkte dit, maar daar waar mogelijk zijn representatieve sets van opgaven gebruikt. Het aantal vragen in één blok verschilde per vaardigheid.

Op basis van de pilot is nagegaan of de hoeveelheid opgaven per vaardigheid geschikt was voor de 30 minuten die een toets per vaardigheid mocht duren. Naar aanleiding van de pilot zijn de aantallen opgaven niet gewijzigd en ook de aantallen nieuwe en ankeropgaven bleven onveranderd. Het totaal aantal opgaven bij Woordenschat was 110, bij Lezen 51 en bij Luisteren 48.

Figuur 3.1: Basis design voor de papieren toetsen (voorbeeld Lezen)

Vaardigheid lezen	Item sets (blokken) papier					
Toets	Lezen blok 1	Lezen blok 2	Lezen blok 3	Lezen blok 4	Lezen blok 5	Lezen blok 6
Papier 1 (Lezen P1)	Le P1 b1 I	Le P1 b2 II				
Papier 2 (Lezen P2)		Le P2 b2 I	Le P2 b3 II			
Papier 3 (Lezen P3)			Le P3 b3 I	Le P3 b4 II		
Papier 4 (Lezen P4)				Le P4 b4 I	Le P4 b5 II	
Papier 5 (Lezen P6)					Le P5 b5 I	Le P5 b6 II
Papier 6 (Lezen P6)	Le P6 b1 II					Le P6 b6 I

Het toetsdesign voor de drie vaardigheden Lezen, Luisteren en Woordenschat was hetzelfde. Het samengevatte toetsdesign voor deze vaardigheden is gegeven in Figuur 3.2.

Figuur 3.2: Toetsen Luisteren, Woordenschat en Lezen

Toetsen	Boekje Lezen (LE)		Boekje Lu (LU)		BoekjeWoordenschat (WS)	
Papier 1 (Lu P1)	LeP1 b1 I	LeP1 b2 II	LuP1 b1 I	LuP1 b2 II	WsP1 b1 I	WsP1 b2 II
Papier 2 (Lu P2)	LeP2 b2 I	LeP3 b2 II	LuP2 b2 I	LuP2 b3 II	WsP2 b2 I	WsP2 b3 II
Papier 3 (Lu P3)	LeP3 b3 I	LeP4 b3 II	LuP3 b3 I	LuP3 b4 II	WsP3 b3 I	WsP3 b4 II
Papier 4 (Lu P4)	LeP4 b4 I	LeP5 b4 II	LuP4 b4 I	LuP4 b5 II	WsP4 b4 I	WsP4 b5 II
Papier 5 (Lu P5)	LeP5 b5 I	LeP6 b5 II	LuP5 b5 I	LuP5 b6 II	WsP5 b5 I	WsP5 b6 II
Papier 6 (Lu P6)	LeP6 b6 I	LeP1 b1 II	LuP6 b6 I	LuP6 b1 II	WsP6 b6 I	WsP6 b1 II

In ieder blok zit  $1/6^e$  van de opgaven. Doordat iedere leerling bij wie de vaardigheid gemeten is twee blokken maakte, maakte iedere leerling  $1/3^e$  van het totaal aantal opgaven per vaardigheid, dus 36 of 37 bij woordenschat, 16 of 17 bij lezen en 16 bij luisteren. Evenzo gold ook dat iedere opgave is gemaakt door een op de drie leerlingen bij wie de vaardigheid gemeten wordt.

In oorsprong werd gestreefd naar combinatie-toetsen voor Lezen en Woordenschat (conform bovenstaand design, waarbij een toets bestond uit een blok Lezen en een blok Woordenschat). Hiervoor waren twee redenen: 1) op deze manier herkennen de leerlingen de toets niet direct als toetsen voor de twee vaardigheden en 2) de ervaring leert dat leerlingen toetsen waarin enkel lezen wordt gemeten relatief zwaar vinden. Echter, de resultaten van de pilotanalyses lieten zien dat er bij Woordenschat een hoog aantal ontbrekende waarden was, waardoor via de combinatie-toetsen de werkelijke vaardigheid voor Lezen en Woordenschat onvoldoende nauwkeurig gemeten werd. Het was namelijk niet duidelijk of een lage score voor woordenschat het gevolg was van moeite met Woordenschat of met Lezen (de leerling kon te veel tijd hebben besteed aan lezen). Beide situaties

waren onwenselijk, want ook de resultaten voor Lezen zouden niet bruikbaar zijn als leerlingen langer dan de ingeschatte duur aan de toets besteedden. Op basis van deze resultaten is besloten aparte toetsen voor lezen en woordenschat af te nemen – zoals weergegeven in Figuur 3.2.

### 3.3.3 Toetsdesign voor de spreek- en gespreksvaardigheid

De vaardigheden spreken en gespreksvaardigheden zijn gemeten met vier praktische taken. De eerste opdracht was een individueel uit te voeren spreektaak (“De Brug”). Deze taak diende tevens als anker, aangezien dit een taak was die in de vorige peiling ook is afgenomen. De overige taken waren gesprekstakingen, waarin een tweetal leerlingen een (gestructureerd) gesprek voerden. De eerste gesprekstaak was zo gemaakt dat deze te doen zou zijn voor leerlingen op de ERK-niveaus A1/A2, de tweede zo dat die geschikt was voor leerlingen met een wat hoger ERK-niveau (A2/A2+) en de derde zo dat die geschikt was voor leerlingen die met ERK-niveau B1. In het geval de leerlingen zeer duidelijk veel moeite hadden met de eerste/tweede gesprekstaak dan was het niet zinvol een volgende gesprekstaak af te nemen. Dit zou alleen tot frustratie van de leerling(en) leiden. Zoals in paragraaf 1.5.5 staat beschreven diende de testleider na afloop van de eerste gesprekstaak te besluiten of de leerlingen door konden naar de volgende taak of dat één (of beide) leerling(en) moest(en) stoppen, omdat het niveau te hoog lag. De gespreksleider voerde in dat laatste geval de volgende gesprekstaak met de leerling die het niveau nog aan zou kunnen. Testleiders zijn geïnstrueerd wanneer hiertoe kon worden overgegaan en hoe zij een dergelijk gesprek konden voeren.

In paragraaf 1.5.4 is weergegeven hoe op basis van de pilot is besloten voor een zo willekeurig mogelijke selectie van leerlingen. Doordat actieve toestemming van ouders noodzakelijk was, is volledig gerandomiseerde selectie niet mogelijk geweest.

## 3.4 Toets-design over vaardigheden heen

In Figuur 3.3 is het toetsdesign weergegeven over de vaardigheden heen, in volgorde van afname. Op alle scholen kende de afname dezelfde volgorde. Er werd gestart met de afname van de vragenlijst bij alle leerlingen, waarna respectievelijk de toetsen Lezen, Luisteren en Woordenschat werden afgenomen. De vragenlijst is op alle scholen als eerste afgenomen, aangezien het de bedoeling was dat de leerlingen, ongeacht hun gepercipieerde prestaties op de toetsen, een mening zouden geven over hun attitude ten opzichte van en ervaring met het Engels. In principe was er tussen het tweede en het derde onderdeel een langere pauze gepland, terwijl er een kort pauzemoment was tussen de overige taken (respectievelijk tussen eerste en tweede en derde en vierde onderdeel). Alleen in verband met praktische overwegingen op een school werd hiervan afgeweken en is een ander pauzemoment gekozen, bij voorkeur tussen het derde en het vierde onderdeel.

Figuur 3.3: Samenstelling van de papieren afnamen

TOETSSET	Vragenlijst	Lezen	Luisteren	Woordenschat
SET 1	VL	Le P1	Lu P1	Ws P1
SET 2	VL	Le P2	Lu P2	Ws P2
SET 3	VL	Le P3	Lu P3	Ws P3
SET 4	VL	Le P4	Lu P4	Ws P4
SET 5	VL	Le P5	Lu P5	Ws P5
SET 6	VL	Le P6	Lu P6	Ws P6

Een school heeft in principe altijd één toets-set toegewezen gekregen. In het geval van meerdere klassen is de volgende procedure gevolgd: bij afname in parallelklassen op dezelfde dag is dezelfde toets-set bij alle leerlingen afgenomen, bij afname op meerdere dagen zijn verschillende toets-sets bij de verschillende parallelklassen afgenomen. Op deze manier kon de inhoud van de toets niet binnen de school verspreid worden. Voornaamste reden voor toewijzing van toets-sets op school (en in sommige gevallen klasniveau) was dat de luistertoets klassikaal getoond werd via het digibord of een videofilm en leerlingen derhalve dezelfde opgaven moesten beantwoorden (met als gevolg: dezelfde toetsen maakten). De leerlingen op school maakten dus dezelfde papieren opgaven voor alle klassikaal getoetste vaardigheden. Echter, niet alle leerlingen op een school maakten *alle* papieren toetsen, omdat een deel van de leerlingen ook mondelinge taken uitvoerde op het moment dat de rest van de klas een papieren onderdeel deed. Deze opzet was ongeacht de toets-set die een school kreeg toegewezen. Dit leverde het design op dat is weergegeven in Figuur 3.4.

In Figuur 3.4 worden vier verschillende mogelijkheden onderscheiden voor de toetsen die aan de leerlingen werden toegekend. Twee leerlingen in de klas startten met de vragenlijst, deden daarna de mondelinge taken (buiten de klas), dan de toets Luisteren en ten slotte de toets Woordenschat (SET # s1). Twee andere leerlingen in de klas startten met de vragenlijst, maakten daarna de toets Lezen, voerden daarna de mondelinge taken uit en eindigden met de toets Woordenschat (SET # s2). Nog weer twee andere leerlingen in de klas startten ook met de vragenlijst, maakten daarna respectievelijk de toets Lezen en Luisteren, en voerden tenslotte de mondelinge taken uit (SET # s3). De overige leerlingen in de klas startten ook met de vragenlijst en maakten daarna allemaal de toetsen Lezen, Luisteren en Woordenschat (SET # \_\_). In Figuur 4 staat SET # voor de toets-set, waarbij de *hashtag*(#) 1 tot en met 6 kan zijn, afhankelijk van de versietoekenning op school-/klasniveau.

Figuur 3.4: Design binnen een school (gebaseerd op Figuur 3.2)

Binnen School	Vragenlijst	Lezen/ Mondeling	Luisteren/ Mondeling	Woordenschat/ Mondeling	aantal
SET # s1	VL	Mondeling	Lu P#	Ws P#	2 leerlingen
SET # s2	VL	Le P#	Mondeling	Ws P#	2 leerlingen
SET # s3	VL	Le P#	Lu P#	Mondeling	2 leerlingen
SET # __	VL	Le P#	Lu P#	Ws P#	overige leerlingen

### 3.5 Aantallen Leerlingen

Binnen het voorgestelde design zou iedere school één van de zes toets-sets toegewezen krijgen. Op basis van de 101 scholen die zich in eerste instantie hadden aangemeld betekende dit dat een toets-set op 16 of 17 scholen aangeboden zou worden. Bij de toewijzing van de toets-sets is er zo veel mogelijk rekening mee gehouden dat in een stratum alle zes de toets-sets werden afgenomen, waardoor iedere toets-set dus in ieder geval waarnemingen bevatte uit alle acht combinatie-strata die de basis voor de streekproeftrekking waren. Er is daarnaast zoveel mogelijk geprobeerd om bij de toewijzing van de toets-sets aan de scholen ervoor te zorgen dat het aantal waarnemingen per toets-set vrijwel even groot zou zijn.

Bij het opstellen van het design is voorts uitgegaan van DUO-gegevens om schattingen te maken over aantallen te toetsen leerlingen. Uitgaande van 100 scholen was de inschatting dat het

totaal aantal leerlingen op 2400 zou liggen. Dit hield in dat de vragenlijst gemaakt zou worden door 2400 leerlingen. De toets-sets zijn ook gebundeld en er zijn ook 2400 toets-sets over de scholen verdeeld. Het aantal observaties per taak (lezen, luisteren of woordenschat) was by-design echter lager dan deze 2400, aangezien op voorhand al besloten was dat er tijdens iedere taakafname twee leerlingen een gesprekstaak zouden gaan maken.

De mondelinge taken zouden worden gemaakt door 600 leerlingen: 100 scholen met zes leerlingen per school, zij het dat het aantal observaties op de tweede en derde wat moeilijker gesprekstaken mogelijk lager kon liggen. De inschatting voor het aantal observaties voor de andere drie vaardigheden (Lezen, Luisteren en Woordenschat) zou per vaardigheid gelijk zijn aan 2200 (2400 – 100\*2 leerlingen). Aangezien de leerlingen een derde van de opgaven zouden moeten maken, zou iedere opgave door ruim 700 leerlingen gemaakt worden.

In de praktijk waren er enige afwijkingen van de aantallen. Doordat zes scholen naderhand van deelname hebben afgezien, is het aantal waarnemingen per toets-set niet meer volledig vergelijkbaar en is ook de verdeling van het aantal waarnemingen over de strata niet meer even groot. De verdeling van het aantal varianten dat bij de combinatie-strata terecht kwam is gegeven in Tabel 3.1. De gegevens in Tabel 3.1 zijn gebaseerd op de strata-verdeling. Doordat de afzegging van de scholen na toewijzing aan de scholen plaatsvond, was het niet meer mogelijk om tijdige wijzigingen in de versie-verdeling te maken.

Tabel 3.1: verdeling van de varianten over de combinatie-strata

Combi code 8: VVTO (2) - Stratum (4)	versie_afname						Totaal
	1	2	3	4	5	6	
Regulier - S1(0)	10	11	32	45	66	23	187
Regulier - S2(>0-10)	198	123	94	107	163	174	859
Regulier - S3(>10-25)	48	26	65	75	77	38	329
Regulier - S4(>25)	58	26	39	28	30	27	208
VVTO - S1(0)	35	29	22	60	47	69	262
VVTO - S2(>0-10)	46	27	50	93	16	28	260
VVTO - S3(>10-25)	47	0	5	31	21	32	136
VVTO - S4(>25)	36	15	21	40	9	16	137
	478	257	328	479	429	407	2378

Als we kijken naar hoe de versies bij scholen terecht zijn gekomen dan is te zien dat bij 90 van de scholen inderdaad een versie afgenomen is, en bij 5 scholen twee verschillende versies gebruikt zijn. Dit zijn scholen die ook met minstens twee klassen meededen. Het aantal school-versie combinaties komt daarmee op 100. Versie 2 kwam, zoals te verwachten als gevolg van de weigering van zes scholen, op minder scholen terecht (14), terwijl versie 1 op relatief veel scholen terecht kwam (19). De andere versies kwamen op 16 tot 17 scholen terecht. Aangezien ieder item in twee versies zat, lag het aantal scholen waar de items terecht kwamen tussen de 30 en 36.

Bij de aantallen versies moet opgemerkt worden dat er ook reserveboekjes meegestuurd zijn. Niet alle versies die naar de scholen gestuurd zijn, zijn ingevuld. Dit kwam deels doordat het reserve-versies waren voor als er toch meer leerlingen waren dan verwacht, deels doordat er gelijktijdig ook mondelinge taken uitgevoerd waren, en deels door ziekte, dokters-/tandartsbezoek of andere redenen voor afwezigheid. Het gemiddelde aantal observaties per item

lag daardoor met 631 wel wat lager dan de beoogde 733, maar is voor de meest toepassingen ruim te noemen, zeker bij het toepassen van een IRT-model dat gebruik maakt van twee itemparameters (zie de COTAN richtlijnen<sup>31</sup>). De exacte gegevens met betrekking tot de uiteindelijke aantallen observaties zijn gegeven in de paragrafen 2.4 en 2.5.

---

<sup>31</sup> COTAN Beoordelingssysteem voor de kwaliteit van tests (Evers, Lucassen, Meijer & Sijtsma, 2010; <https://www.psynip.nl/wp-content/uploads/2016/07/COTAN-Beoordelingssysteem-2010.pdf>)



## Hoofdstuk 4. Voorbereiding en uitvoering

Een grootschalige nationale peiling in het basisonderwijs vraagt om een gedegen voorbereiding. In dit hoofdstuk beschrijven we de verschillende stappen in de voorbereiding van de afnames op de scholen. De voorbereiding heeft plaatsgevonden in de periode van augustus tot en met december 2017. In de eerste plaats wordt in dit hoofdstuk een beschrijving gegeven van het afnameteam en hoe de testleiders in deze teams zijn geworven, hoe zij zijn getraind en welke taken zij hadden. Daarna wordt een beschrijving gegeven van de opzet en de globale aanpassingen naar aanleiding van de pilotafnames die in september-begin oktober 2017 en november-december 2017 hebben plaatsgevonden. Tenslotte wordt een beschrijving gegeven van de intakegesprekken die voorafgaande aan de afname op de scholen plaats hebben gevonden als ook een beschrijving van de logistieke voorbereiding.

### 4.1 Werving en selectie van (hoofd)testleiders

De uitvoering van de peiling in de praktijk is uitgevoerd door twaalf "hoofdttestleiders" en 21 "reguliere" testleiders die allen werden aangestuurd door de onderzoekers. De hoofdttestleiders hadden een centrale rol in de peiling als het aanspreekpunt voor schoolleiders en leerkrachten, als gesprekspartner in de intakegesprekken en als gespreksleider in de afname van de gesprekstaken. Voor de rol van hoofdttestleider zijn personen geselecteerd met een ruime ervaring, in het vakgebied van de Engelse taalverwerving en/of in het werken met kinderen uit de leeftijdsgroep. Een viertal van de hoofdttestleiders was eerder in het project betrokken als ambassadeur. De overige hoofdttestleiders zijn geworven in de periode juni tot oktober 2017 binnen de netwerken van de consortiumpartners. Aangezien de hoofdttestleiders de mondelinge taken in een aparte ruimte bij de leerlingen afnamen, is de hoofdttestleiders verzocht een Verklaring Omtrent Gedrag (VOG-onderwijs) aan te vragen. Alle hoofdttestleiders hebben een VOG ontvangen<sup>32</sup>.

Reguliere testleiders waren in het onderzoek primair verantwoordelijk voor de gestandaardiseerde afname van de klassikale taken. Ze dienden de verschillende onderdelen op systematische wijze af te nemen, waren verantwoordelijk voor een ordelijk verloop tijdens en tussen de onderdelen in en hadden de taak de leerlingen volgens protocol te begeleiden bij het invullen van de vragenlijst en het antwoordformulier van de toetsen. De testleiders werden geworven in de periode juli-november 2017 onder studenten, recent afgestudeerden en promovendi van de diverse betrokken opleidingen (Engels en Applied linguistics RuG, lerarenopleiding HvA, Academische Opleiding leraar basisonderwijs RuG). Een deel van de testleiders is gevonden via de pool van testleiders die eerder werkzaam waren geweest voor de peiling Bewegingsonderwijs (Timmermans e.a., 2017). Specifiek ging het om zes afgestudeerde bewegingswetenschappers met lesbevoegdheid. Na aanmelding heeft met elke beoogde testleider een kennismakingsgesprek plaatsgevonden. Naast de feitelijke beschikbaarheid voor training, pilotonderzoeken en afnames in het hoofdonderzoek, zijn de testleiders geselecteerd op affiniteit met de Engelse taal, hun ervaring met (klassikaal) lesgeven of het begeleiden van (jonge) kinderen.

---

<sup>32</sup> Overigens hebben ook de beoordelaars van de mondelinge taken, die op video waren opgenomen, een VOG aangevraagd en ontvangen.



## 4.2 Training van (hoofd)testleiders

In de voorbereiding op de training is voor alle betrokken testleiders een algemeen handboek ontwikkeld. In dit handboek staat allereerst een korte introductie op de peiling Engels. Deze introductie was bedoeld om de testleiders meer zicht te geven op de context waarbinnen de peiling wordt uitgevoerd, de resultaten die de peiling beoogde op te leveren en een beknopt overzicht van de taken die in de peiling gedaan moesten worden. Het tweede deel van het handboek bestaat uit protocollen en beschrijvingen van de verschillende onderdelen (vragenlijst, toetsen, testen gespreksvaardigheid) die in de peiling afgenomen zouden worden. Per opdracht is beschreven welke materialen nodig waren voor de opdracht, welke procedure gevolgd moest worden, welke instructies aan de leerlingen moesten worden gegeven en hoe testleiders om moesten gaan met (veelvoorkomende) vragen. Dit geheel was aangevuld met een overzicht van eventuele aandachtspunten en tips.

Voor de hoofdtestleiders, die belast zouden worden met de leiding van de afnames en de afname van de mondelinge taken, is een supplement op het handboek gemaakt waarin zowel de ankertaak spreken als de drie gesprekstaken zijn beschreven en voor elke afzonderlijke taak een afnameprotocol is opgenomen. Het supplement bevatte verder secties over de voorbereiding (materialen thuis), de afstemming met de leerkracht voor de start van het eerste onderdeel (o.a. voorstellen van de reguliere testleider, globaal overzicht van de taken, gespreksruimte), en het protocol voor het ophalen van de leerlingen tussen de klassikale taken door. Alhoewel alle hoofdtestleiders ervaring hadden met het begeleiden van kinderen was er in het supplement nog bijzondere aandacht voor het welbevinden van de leerlingen en voor een veilig en prettig verloop van de gesprekstaken.

Om de betrouwbaarheid van de metingen te vergroten zijn alle testleiders, voorafgaand aan de (pilot)metingen, getraind, zodat zij de testen en gesprekstaken volgens de opgestelde protocollen zouden afnemen. Het trainingstraject betrof voor de reguliere testleiders twee dagen, terwijl de hoofdtestleiders nog een extra trainingsdag hebben doorlopen. Hieronder wordt de opzet van de trainingsdagen weergegeven.

- Dag 1, ochtend:
  - Algemene voorlichting omtrent de peiling Engels.
  - Toelichting op organisatorische aspecten van de afname.

Dag 1, middag:

- Reguliere testleiders werden getraind in de afname van de klassikale taken (vragenlijst, leesvaardigheidstoets, luistertoets, woordenschattoets). De hoofdtestleiders werden getraind in de afname van de gesprekstaken. Naast de protocollen voor de taken werd ook aandacht geschonken aan: gedrag op school, afstemming met de leerkracht, protocollen voor vragen, tijdstippen van pauzes en invulling van de pauzes (bijvoorbeeld de inzet van "energizers" of korte spelletjes wanneer gewacht moest worden op de terugkomst van leerlingen die de gesprekstaken deden). De testleiders werden getraind met behulp van de (pilot)instrumenten en via rollenspellen. Elke testleider heeft tijdens de training minimaal de uitvoering van één klassikale taak geoefend en ontving daarop feedback van de onderzoekers. Hoofdtestleiders kregen eerst een overzicht van de

mondelijke taken, de doelstellingen van de taken en de te meten vaardigheidsniveaus. Daarna werden zij door de onderzoekers van de HvA en RuG uitgebreid getraind in de daadwerkelijke afname van de drie gesprekstaken. Hiervoor zijn audio en video-opnames van taken uit de kleinschalige testperiode (van Cito) gebruikt, maar de taken zijn vooral via simulatiegesprekken tussen onderzoeker (in de rol van leerling) en hoofdtestleider geoefend.

- Dag 2: proefafname op één van de zes klassen in de pilotscholen of op één van de drie scholen waar een generale repetitie heeft plaatsgevonden. (zie onder).
- Dag 3: nadere training van de uitvoering van de gesprekstaken (alleen voor hoofdtestleiders). In deze training is vooral gebruik gemaakt van videobeelden uit beide pilotafnames en is de afname van gesprekstaken (en begeleiding van de leerling) gefinetuned en, herhaaldelijk, via rollenspel geoefend.

Vanwege de doorloop van de werving van ‘reguliere’ testleiders hebben niet alle testleiders op hetzelfde moment de training ontvangen. De eerste aanvullende pool testleiders is getraind voorafgaande aan de generale repetitie (zie onder). Zij hebben derhalve in die dress rehearsals hun tweede trainingsdag gehad. Vanwege uitval van enkele testleiders zijn voorts begin december nog drie extra testleiders getraind en omdat zij niet konden meedraaien in de geplande dress rehearsals, hebben zij, met een beperkt actieve rol, ‘meegelopen’ bij de eerste afnames in januari 2018.

### 4.3 Pilotonderzoek

Voorafgaand aan de afnames van de peiling zijn zes pilotafnames (op vijf scholen) en drie zogenoemde “dress rehearsals” uitgevoerd. De pilotafnames vonden plaats in september en begin oktober 2017 en de dress rehearsals in november en begin december 2017. Bij alle proefafnames waren telkens onderzoekers van GION en/of de HvA aanwezig. Alle scholen die aan de proefafnames deelnamen waren scholen uit de netwerken van consortiumpartners. Bij de werving van scholen is rekening gehouden met een representatieve verdeling over Nederland en met verschillen in startmoment Engels en de leerlingpopulatie (conform de steekproefcriteria). Dit betekende dat in beide testperiodes zowel een VVTO- als een EIBO-school betrokken was en dat zowel een school zonder (of een zeer gering aantal) als een school met relatief veel (>20%) gewichtenleerlingen heeft deelgenomen.

#### 4.3.1 September 2017

De pilotafnames hadden tot doel om alle items die voor deze peiling waren ontwikkeld dan wel als ankeritems waren aangeduid uit te testen, om te bepalen of de doorlooptijden van de onderdelen globaal gezien correct waren en om te bepalen of de ontwikkelde protocollen voldoende uitvoerbaar en eenduidig waren. Er waren in totaal zes toetsversies geconstrueerd. Elke versie bestond daarbij uit twee blokken en elk blok zou in twee versies voorkomen (hetgeen betekent dat er overlap in de blokken over de versies heen zat). Om zo alle items te testen zijn de zes beoogde versies van de klassikale taken uitgetest. Daarnaast hadden de proefafnames tot doel de pilotversies van de gesprekstaken als ook het beoogde protocol en duur van de afname van de gesprekken uit te proberen. De proefafname was inclusief het uittesten van de *volgorde* van de verschillende taken, en

het uittesten van het werken met digitale hulpmiddelen (powerpoints met instructies en luistertoets via internet film, usb en/of dvd). Ook het afnemen van de leerlingvragenlijst was opgenomen in de proefafname, zodat we konden nagaan of voor alle leerlingen het beantwoorden van de vragen binnen de beschikbare tijd haalbaar was, en of er geen lastige of multi-interpretabele items in de vragenlijst waren opgenomen. Dit laatste is gedaan middels klassikale bespreking van de vragenlijst: leerlingen konden aangeven welke vragen ze niet snaptten of anderszins vreemd vonden. Ten slotte is geoeftend met de rollen van hoofdtestleider en testleiders.

Na afloop van de proefafnames hebben er verschillende evaluaties en aanpassingen plaatsgevonden en zijn zowel de instrumenten zelf als ook de protocollen voor de afnames uitgebreid aangepast en 'gefinetuned'. Zo bleek gesprekstaak 1 te lang, waardoor de afname van Taak 2 en Taak 3 in het gedrang kwam. Taak 1 bevatte oorspronkelijk meerdere onderdelen op A2-niveau. Aangezien A2 ook met Taak 2 werd gemeten zijn A2-items verwijderd. Taak 1 bevatte uiteindelijk nog één onderdeel waarmee A2 gemeten werd, zodat de testleider kon inschatten of de leerlingen de volgende taak aan zouden kunnen. Daarnaast liet het videomateriaal van de pilot zien dat de instructies en begeleiding van de verschillende testleiders nog niet uniform waren. De afname-instructies zijn in het protocol verder aangescherpt, waarbij de instructies meer gescript zijn en het 'speelse en leuke' aspect van de gesprekjes sterker werd benadrukt (om te voorkomen dat leerlingen zenuwachtig zouden worden). Ook is een standaardinstructie op film gemaakt en een filmpje waarin een voorbeeld wordt gegeven van wat de bedoeling is (zodat de leerlingen wisten wat van ze werd verwacht). Deze werden aan het begin van de afname van de gesprekstakingen getoond. De aangescherpte afnamecondities zijn in de volgende training voor de testleiders sterker benadrukt. Ze zijn niet alleen besproken, maar ook extra getraind m.b.v. de video-opnamen. Tijdens de training zijn daarnaast heldere regels opgesteld over de beslissing om wel of niet met één of beide leerlingen door te gegaan met de volgende (moeilijker) taak. Voor de afname van de gesprekstakingen werd daarnaast een extra onderdeel geïntroduceerd, namelijk het zogenaamde "activity-boekje" met schrijf- en puzzel-opdrachten. Deze kon worden ingezet om de leerling die niet deelnam aan de anker- of gesprekstaak actief aan het werk te zetten. Dit activity-boekje is geen onderdeel van de te meten vaardigheden Engels. Bij de luistertoets bleek bovendien dat het zinvol zou zijn wanneer leerlingen met een oefenvraag konden beginnen, zodat zij aan het testformat en de procedure konden wennen: de introductie van de luistertoets werd aangepast en de oefenvraag werd in het hoofdonderzoek vanzelfsprekend niet gescoord. Meer informatie over de aanpassingen van de instrumenten naar aanleiding van de pilots staat beschreven in de paragrafen 5.1.1 en 5.2.1. Deze aanpassingen zijn eind november ook verwerkt in een nieuwe versie van het handboek en het supplement voor hoofdtestleiders en toegepast in de trainingen voorafgaande aan de dress rehearsals.

#### **4.3.2 November en december 2017**

In november en december 2017 werd een generale repetitie uitgevoerd op drie scholen. Doelstelling van deze afnames was derhalve om de gewijzigde opzet van de afname in de praktijk te oefenen en om te controleren of de instrumenten en protocollen naar behoren waren aangepast en aangescherpt. Dat betekent dat op drie scholen de definitieve vragenlijst, en drie definitieve versies van de leesvaardigheidstoets, woordenschattoets en luistertoets zijn afgenomen alsof het een reguliere afname betrof. In de dress rehearsal zijn verder de klassikale instructies (powerpoints) gebruikt die in de daadwerkelijke afnamecyclus ook zijn ingezet en is gebruik gemaakt van een

gestandaardiseerde afnamevolgorde. Deze volgorde schreef voor dat op alle scholen in het hoofdonderzoek zou worden gestart met de vragenlijst waarna volgde: een korte pauze; 30 minuten leestoets; een langere pauze; 30 minuten luistertoets; een korte pauze; 30 minuten woordenschattoets; en een afsluiting.

Naar aanleiding van de dress rehearsals kon worden vastgesteld dat zowel de instrumenten als de protocollen en doorlooptijd van de klassikale taken niet meer hoefden te worden aangepast. De generale repetitie verliep, in de klas, zoals beoogd. Ook de meer gestandaardiseerde afname van de (herziene) gesprekstaken verliep naar verwachting. De ervaringen van deze afnames leverde nog wel wat extra tips en trucs op. Die zijn met alle hoofdtestleiders uitgewisseld.

De protocollen, het handboek en het supplement voor hoofdtestleiders (versie eind november 2017) zijn in ongewijzigde vorm ingezet in het hoofdonderzoek. Testleiders die wel in de pilots, maar niet in de dress rehearsals betrokken waren hebben deze documenten begin december ontvangen en waar nodig zijn de veranderingen in de instrumenten en afnameprotocollen persoonlijk aan hen toegelicht.

## 4.4 Voorbereiding van de scholen

### *Intakegesprekken*

Voorafgaand aan de daadwerkelijke peilingsafnames op de scholen hebben de hoofdtestleiders een intakegesprek op de scholen afgenomen. Hierbij waren de schoolleider, één of meerdere groepsleerkracht(en) en eventueel de vakleerkracht Engels aanwezig. Het intakegesprek had naast een nadere kennismaking vooral tot doel om enige informatie te verzamelen over het aanbod Engels op de scholen. Ook werd er in het gesprek een toelichting op het toetsmateriaal en het afnameprotocol gegeven en werden een aantal organisatorische zaken met de scholen kortgesloten. De meeste intakegesprekken hebben op school plaatsgevonden. Uitzonderingen betreffen twee skypegesprekken en zes telefonische gesprekken (op verzoek van de school).

Ter ondersteuning van de intakegesprekken is een protocol ontwikkeld met de onderwerpen die tijdens het gesprek aan de orde zouden moeten komen. De onderwerpen in het protocol voor het intakegesprek waren als volgt:

1. Uitwisselen van contactgegevens van hoofdtestleider aan contactpersonen op de school en vice versa. De hoofdtestleider was het vaste aanspreekpunt voor de school.
2. Toelichting geven op de opzet van de afname en de inhoud van de verschillende klassikale taken; voorbeelden tonen van vragenlijst en toetsboekjes.
3. Toelichting geven op de opzet van de afname en de inhoud van de taken gespreksvaardigheid; nut en noodzaak van video-opnamen toelichten.
4. Tonen van de leerkrachtvragenlijst en afspraak maken dat deze op de afnamedag zou worden ingevuld.
5. Toelichten van de afnamecondities (o.a. leerkracht aanwezig in de klas tijdens afname).
6. Toelichten van het toestemmingsverzoek aan de ouders van de leerlingen (zie paragraaf hieronder).
7. Overleg over eventuele exclusie van individuele leerlingen ((audio)visuele beperkingen).
8. Behandelen van de vragen/topics over het aanbod Engels.
9. Toelichting op de schoolrapportages die de school na afloop van het onderzoek ontvangt.

10. Nagaan of de benodigde administratieve leerlinggegevens (al) naar de onderzoekers zijn gestuurd. Eventueel navragen (waarom) welke gegevens niet of wel beschikbaar kunnen worden gesteld.
11. Datum en dagdeel voor de peiling inplannen (tussen 9 januari en 31 maart 2018). Bij voorkeur op een moment waarop de groep 8-leerlingen normaal gesproken Engelse les krijgen.
12. Aankondiging van het verdiepende onderzoek.
13. Vraag naar en eventueel bezichtiging van de ruimte waar de gesprekstaken tijdens de afname zouden plaatsvinden.

### *Toelichting op de ouderlijke toestemming*

De ouders van leerlingen is gevraagd naar toestemming voor deelname aan het peilingsonderzoek. Dit betrof in eerste instantie het passief toestemming vragen via een (uniforme) brief en antwoordstrook die door de school wordt verstuurd. Deze brief kon zowel digitaal als op papier aangeleverd worden afhankelijk van de wens van de school. Passief betekent praktisch gezien 'ouders stemmen in met deelname aan het onderzoek, tenzij ze zelf geen toestemming geven voor a) het gehele onderzoek of b) deelname aan op video geobserveerde gesprekken'. Indien ouders niet toestemmen met deelname van hun zoon/dochter aan de peiling konden ze dit aan de contactpersoon (vak/groepsleerkracht) op school kenbaar maken via de antwoordstrook.

In de loop van de periode van intakes werden de onderzoekers door de (ethische commissie Faculteit Gedrags- en maatschappijwetenschappen van de Rijksuniversiteit Groningen) verzocht om *actieve* toestemming voor de mondelinge taken aan alle ouders te vragen. Deze actieve toestemming was vereist, aangezien er bij de mondelinge taken sprake was van het gebruik van video-opnamen, (die noodzakelijk waren voor een adequate beoordeling). Dit verzoek is via de scholen, met een aangepaste brief en antwoordformulier, uitgezet aan a) alle ouders van scholen die de intake nog niet hadden gehad en b) ook aan alle ouders van scholen, die in eerste instantie de passieve versie van de ouderbrief hadden verstuurd. Aan alle scholen is het verzoek voorgelegd of zij bereid waren deze actieve toestemming voor de mondelinge taken aan de ouders te vragen. Dat was immers een extra belasting voor de scholen. Uiteindelijk heeft slechts één school aangegeven geen actieve toestemming te willen vragen. Op deze school zijn dan ook geen gesprekstaken afgenomen/gesprekken op video vastgelegd. In paragraaf 1.5.1 staat beschreven hoe de procedure rond het verzamelen van actieve toestemming is verlopen.

### *Voorbereiding leerlingmaterialen*

Om enige achtergrondkenmerken van de te toetsen leerlingen te verzamelen (voor latere analyses) en om de afnames goed te kunnen voorbereiden zijn scholen door de ambassadeurs bij de werving reeds gewezen op het aanleveren van leerlinggegevens. Bij het maken van de afspraak voor het intakegesprek is door de hooftestleiders een formeel verzoek neergelegd om voorafgaande of tijdens het intakegesprek leerling kenmerken en achtergrondgegevens aan te leveren via een gestandaardiseerd excel-formaat of via een uitdraai van het schooladministratiesysteem. Specifiek ging het om:

- voornaam, tussenvoegsel, achternaam
- sekse (**M**eisje of **J**ongen)
- geboortedatum (maand, dag, geboortjaar)

- naam van de klas waarin de leerling zit
- leerlinggewicht/formatiegewicht (0; 0,3; 1,2)
- de scores voor begrijpend lezen in groep 7 (eindtoets groep 7 in het leerlingvolgsysteem, E7)
- het schooladvies of voorlopig schooladvies (begin groep 8)
- eventueel: indicatie van gediagnosticeerde dyslexie

In het begin van het verzamelen van deze gegevens werd naast bovenstaande gegevens ook aan leerkrachten gevraagd in de excel een inschatting te noteren van het algehele niveau Engels van de leerling. Dit was bedoeld om (beter) een paarsgewijze selectie te maken van leerlingen voor de gesprekstaken op drie niveaus: laagpresteerders, middengroep en hoogpresteerders. Nadat uit de pilots bleek dat de selectiemethode voor de gesprekstaken in het hoofdonderzoek op random wijze – en dus niet op (ingeschat) niveau - zou plaatsvinden is deze variabele niet meer actief verzameld (ook om de leerkracht te ontlasten). De niveau-inschattingen die wel zijn verkregen zijn echter ongewijzigd opgenomen in de dataset en in hoofdstuk 8 wordt deze inschatting gerelateerd aan de leerlingprestaties.

Met de verkregen gegevens zijn leerlingnamen gekoppeld aan barcodes ten behoeve van de antwoordformulieren en volgnummers voor de vragenlijst. De barcodes bestaan uit zeven cijfers: de eerste vier corresponderen met het schoolvolgnummer en de laatste drie met het volgnummer in de (op alfabet gesorteerde) leerlingenlijst. Voorafgaand aan de afname zijn vragenlijsten bedrukt met de volgnummers 1001 tot 3600. Ze werden toegewezen aan de leerlingen per school in de volgorde van het ontvangen van de leerlingenlijst. Leerling 1001 is zodoende de eerst aangemelde leerling, waarvan achtergrondgegevens zijn ontvangen. Het vragenlijstnummer is als uitgangspunt genomen voor de gesprekstaken: in plaats van het opnoemen van een school en/of de eigen naam werd in de video alleen de (vragenlijst)volgnummers van de leerlingen genoemd. Met de barcodes en vragenlijstnummers werd het tevens mogelijk voor de school, wanneer zij dat wenste, om op enigszins geanonimiseerde manier deel te nemen aan de peiling. Aan de barcodes en vragenlijstnummers werd dan in plaats van voor- en achternaam een voor de school herkenbare code gekoppeld (bijvoorbeeld het interne leerlingnummer van de school), in combinatie met de achtergrondkenmerken zoals hierboven genoemd.

#### 4.5 Organisatie van de afnames

De planning en de organisatie van de afnames werden vanuit het GION gedaan, waar twee veldwerkcoördinatoren in contact stonden met hoofdtestleiders, testleiders en de contactpersonen op de scholen. Scholen gaven bij de intake aan de hoofdtestleiders door op welke dag en welke tijd de afname plaats zou kunnen vinden, dit in overleg met de agenda van de hoofdtestleider die in principe ook de afname zou verzorgen. De veldwerkcoördinator zocht het aantal benodigde reguliere testleiders voor de afname die tezamen met de hoofdtestleider de school zou bezoeken. Bij meerdere groepen 8 zocht de coördinator ook een tweede hoofdtestleider zodat beide groepen op dezelfde dag getoetst konden worden. In een enkele instantie moest de coördinator door ziekte van de beoogde hoofdtestleider een andere hoofdtestleider inschakelen.

Per klas werd in principe een team van één hoofdtestleider en één reguliere testleider ingeroosterd voor de afname. Vier scholen die deelnamen met meerdere klassen zijn meerdere

malen bezocht. Als een klas uit meer dan 26 leerlingen bestond werd een extra 'reguliere' testleider gezocht voor de afname. Dat was ook het geval bij enkele scholen met samengestelde groepen groter dan 26 (bijvoorbeeld alle groep-8 leerlingen uit meerdere (6)7-8 combinatiegroepen). Op die manier kon de afname soepel en vlot verlopen.

Zodra een afnameteam compleet was ontvingen alle betrokken testleiders een bevestigingse-mail met daarin a) de afspraken voor de afname en het adres van de school, b) de contactinformatie van alle betrokkenen (telefoonnummers, emailadres) en c) de PowerPoint van de versie die zou worden afgenomen (inclusief de juiste link naar de luistertoets en bijbehorend wachtwoord). Ook de contactpersoon op de school ontving, in ieder geval een week van te voren, nog per e-mail een bevestiging van de afnamedatum en de namen van de personen die de school zouden bezoeken.

Ter voorbereiding op een peilingsafname zorgde de hoofdtestleider voor het inpakken van de juiste materialen. Hoofdtestleiders ontvingen per post, voorafgaand aan de afname, van de veldwerk coördinator per klas het volgende:

- Een namenlijst van de leerlingen (ter controle en voor de random selectie van de leerlingen die de gesprekstaken zouden voeren);
- Toetsboekjes in de variant 1-6, voor Lezen, Luisteren en Woordenschat en per klas twee setjes reservematerialen. NB: Indien er sprake was van meerdere klassen op een school en er was sprake van *verschillende* afnamedagen, dan werden twee verschillende toetsversies gebruikt;
- Antwoordformulieren (versie 1-6) voor alle leerlingen, en een reserve-exemplaar. De formulieren waren voorzien van voornaam, achternaam en volgnummer;
- Activity-boekjes voor de leerlingen die tijdens de mondelinge taken *niet* de ankertaak spreken zouden maken of niet alle gesprekstaken zouden uitvoeren;
- Een genummerd exemplaar van de leerkrachtvragenlijst per leerkracht (bij deeltijdbanen of meerdere deelnemende klassen).

Alle betrokken hoofdtestleiders waren verder in het bezit van de volgende algemene materialen:

- Een camera met statief en lege sd-kaarten voor het opnemen van de mondelinge taken;
- Gelamineerde opdrachtkaarten van de gesprekstaken (zie Hoofdstuk 1 voor een beschrijving van de gesprekstaken);
- Een tablet met daarop de video-instructies voor de gesprekstaken;
- Een USB met daarop alle powerpoints van de instructies per taak en alle luistertoetsen (videobestanden);
- Een DVD met de (videos van de) luistertoetsen;
- Voldoende potloden en gummen (de portloden waren bedrukt en er was sprake van kleurige gummetjes in verschillende vormpjes. Leerlingen mochten deze beide na de afname houden);
- Puntenslijpers, verlengsnoeren, handleiding en camera.
- 

#### 4.6 Verloop van een gemiddelde afnamedag

De regie op de afnamedag zelf lag bij de hoofdtestleiders. De hoofdtestleider zorgde voor een goed verloop van de peilingstaken en een goede communicatie met de school. Na aankomst van het team op de school voerde de hoofdtestleider het introductiegesprek met de leerkracht (en/of schoolleider), stelde zijn testleiders voor, deed (in principe) de eerste klassikale introductie voor de leerlingen (voorafgaand aan de vragenlijst,) stuurde de testleiders aan en zorgde er voor dat alle materialen aan het einde van de afname weer mee naar huis werden genomen. Ook de toetsboekjes werden elke keer door de hoofdtestleider meegenomen: deze konden hergebruikt worden op scholen waar dezelfde versie afgenomen diende te worden. Voor de overzichtelijkheid hadden de toetsboekjes voor de te toetsen vaardigheden verschillende kleuren.

Alle afnamedagen verliepen in principe via hetzelfde patroon. Dit is weergegeven in het Handboek voor testleiders<sup>33</sup>.

---

<sup>33</sup> Dit handboek is beschikbaar bij de onderzoekers.





## Hoofdstuk 5. Kwaliteit van de meetinstrumenten, vragenlijsten en schalen

Dit hoofdstuk beschrijft de kwaliteit van de meetinstrumenten die gebruikt zijn om eigenschappen van leerlingen te meten. Dit betreft de beschrijving van de toetsen (lezen, luisteren, woordenschat) en taken (gesprekken), maar ook de gebruikte leerlingvragenlijst. Daarnaast wordt ingegaan op de gebruikte leerkracht- en schoolvragenlijst. De kwaliteit van de meetinstrumenten voor het meten van leerlingeigenschappen wordt ten eerste beschreven in termen van de klassieke testtheorie (KTT), voorafgegaan door een korte beschrijving van de resultaten van de pilotafnamen. In paragraaf 5.1 wordt dit gedaan voor de vaardigheden Lezen, Luisteren en Woordenschat, die zijn gemeten met behulp van klassikale toetsen. De kwaliteit van de praktische taken voor mondelinge vaardigheid staat centraal in paragraaf 5.2, inclusief een beschrijving van de gebruikte beoordelingsmodellen. In paragraaf 5.3 worden de analyses en schaling besproken die zijn uitgevoerd voor de vragenlijst die de leerlingen hebben ingevuld. Ook de schoolvragenlijst en de leerkrachtvragenlijst worden in deze paragraaf besproken. De resultaten van toetsen en taken zijn omgezet naar scores op vaardigheidsschalen. De kwaliteit van deze vaardigheidsschalen wordt in paragraaf 5.4 beschreven. In paragraaf 5.4.4. wordt daarnaast nog ingegaan op de schaling van de klassikale taken op een enkele schaal voor 'Engelse vaardigheid', ten behoeve van de selectie van scholen voor vervolgonderzoek. Voor dit vervolgonderzoek dienden schoolbezoeken op de bovenmatig presterende scholen te worden afgelegd (zie hoofdstuk 10 voor de opzet en resultaten van dit onderzoek).

### 5.1 Toetsen: Lezen, Luisteren en Woordenschat

De vaardigheden lezen en luisteren en de competentie woordenschat zijn allen met toetsen gemeten. Het merendeel van de vragen in de toetsen waren meerkeuzevragen. Uitzondering hierop waren enkele vragen in de woordenschattoetsen. Voordat de toetsen in het hoofdonderzoek zijn afgenomen, heeft in september/begin oktober 2017 eerst een pilotonderzoek plaatsgevonden.

#### 5.1.1 Pilotonderzoek toetsen

Om een indruk te krijgen van de kwaliteit van de items en toetsen is een pilot uitgevoerd, waarbij vrijwel alle opgaven van het hoofdonderzoek zijn afgenomen. De pilot had als belangrijkste doel om te achterhalen of de verschillende peilingsonderdelen opgaven of elementen bevatten die voor de daadwerkelijke afname aangepast of geheel verwijderd zouden moeten worden, omdat deze niet voldeden.

#### *Pilot lezen en woordenschat*

Bij 155 leerlingen van 5 zeer verschillende scholen (met 6 klassen) zijn hiertoe gegevens verzameld bij drie toets-versies, die ieder bij twee scholen/klassen werden afgenomen. Iedere toets-versie bestond uit drie blokken: twee gecombineerde lees- en woordenschat-blokken en een luister-blok. Een gecombineerd lees-woordenschat-blok bevatte opgaven zowel voor woordenschat als voor lezen. De aantallen opgaven voor lezen en woordenschat en leerlingen zijn gegeven in Tabel 5.1.

Tabel 5.1: Kenmerken afname toetsen pilot lezen en woordenschat

	aantal items			aantal leerlingen		Leerlingen
	totaal	Lezen	woordenschat	totaal	definitief	% uit
Set 1	29	10	19	34	29	15%
Set 2	29	10	19	30	23	23%
Set 3	29	10	19	47	45	4%
Set 4	30	11	19	47	47	0%
Set 5	29	10	19	50	50	0%
Set 6	30	11	19	49	46	6%
Totaal	176	62	114	257	240	7%

Vanuit de gedachte dat iedere leerling in principe twee blokken maakte voor lezen en woordenschat, zou de verwachting zijn dat het totaal aantal afnames uitkomt op 310 (2 x 155 leerlingen). Dit was echter niet het geval, omdat er ook leerlingen waren die spreek- en gesprekstakingen uitvoerden tijdens de afname. Daarnaast zijn andere leerlingen die 100% ontbrekende waarden (bijvoorbeeld als gevolg van tandartsbezoek) hadden niet in deze aantallen opgenomen.

Het pilotonderzoek was met name bedoeld om meer over de opgaven te weten te komen. Om die reden zijn niet alleen de leerlingen met 100% ontbrekende waarden niet opgenomen in het onderzoek, maar ook de leerlingen bij wie een reeks van vijf of meer opgaven op rij ontbreekt zijn niet in dit onderzoek opgenomen. In de meeste gevallen geeft het ontbreken van opgaven op rij namelijk meer informatie over leerlingen (bijvoorbeeld dat ze een blad hebben overgeslagen of niet tot het einde van de toets kwamen) dan specifiek over opgaven.

Iedere opgave zat slechts in één opgaven-set en de pilot was opgezet bij een beperkt aantal leerlingen (n=155). Dit aantal leerlingen lag te laag om goed item-response-theorie (IRT) analyses uit te voeren.

De pilot leverde ondanks het geringe aantal observaties inzichten op die invloed hadden op de definitieve samenstelling van de toetsen in het onderzoek. De combinatie-toetsen werden heroverwogen, omdat zij geen voldoende nauwkeurige meting van de vaardigheden woordenschat en lezen toestonden (zie paragraaf 3.3.2 voor meer informatie). Op basis van de resultaten van de pilot is besloten om voor lezen en woordenschat aparte toetsen te maken.

Een ander resultaat van de pilot bij woordenschat was de beslissing hoe de open opgaven te beoordelen. Bij woordenschat was op de pilotgegevens een tweetal analyses uitgevoerd, aangezien de open opgaven op strenge en milde wijze geanalyseerd konden worden. Bij de strenge beoordelingswijze wordt een opgave alleen goedgekeurd bij correcte spelling, terwijl dit bij de milde beoordelingswijze ook het geval is bij fonetisch geschreven goede antwoorden. Beide wijzen geven een indicatie van de lexicale beheersing. Vanwege de hogere betrouwbaarheid van strenge beoordelingen is uiteindelijk gekozen voor deze strenge beoordelingsvorm (Westhoff, 1981, zie ook paragraaf 1.6.4). Deze strenge beoordeling is ook in de voorgaande peiling gehanteerd.

Opgaven waarvan de psychometrische kenmerken niet heel goed waren, bijvoorbeeld extreme p-waarden of negatieve item-rest-correlaties<sup>34</sup> zijn aangepast of in de definitieve afname in het geheel niet opgenomen. Ankeritems die afweken zijn, na inhoudelijke evaluatie onveranderd in

<sup>34</sup> Extreme p-waarden geven weer dat de opgaven heel makkelijk of heel moeilijk waren. In beide gevallen is inhoudelijke evaluatie nodig om na te gaan of de opgaven qua moeilijkheid geschikt zijn voor opname in het peilingsonderzoek. Negatieve item-rest correlaties komen naar voren als laagscorerende leerlingen het item wel goed hebben gemaakt, maar de hoogscorerende leerlingen niet. Ook dit vraagt om inhoudelijke evaluatie.

het onderzoek opgenomen, aangezien zij in de vorige peiling goed functioneerden. Het aantal overgebleven opgaven was groot genoeg voor de definitieve afname. De pilot gaf ook genoeg vertrouwen dat de definitieve toetsen betrouwbaar genoeg zouden zijn.

### **Pilot luisteren**

De blokken bij luisteren zaten anders in elkaar dan bij lezen en woordenschat. Ten eerste was er maar één blok per leerling en een blok bevatte alleen luisteropgaven. Er waren in totaal vier sets van acht opgaven waardoor er 32 verschillende items in de pilot zaten. Iedere toets bevatte twee sets van acht opgaven (16 items per toets). Bij de drie toets-versies leverde dit de volgende verdeling op: de eerste toets bestond uit set 1 en 2 (gemaakt door 37 leerlingen), de tweede uit 3 en 4 (50 leerlingen) en de derde uit 1 en 3 (46 leerlingen). Set 1 en 3 zijn hierdoor vaker afgenomen (respectievelijk 83 en 96 keer) dan set 2 en 4 (37 en 50 keer). Er waren geen leerlingen die vijf opgaven op een rij niet hadden ingevuld. Bij luistertoetsen kunnen leerlingen hun tijd niet zelf indelen: doordat de opgaven per fragment geordend zijn maken de leerlingen allen op hetzelfde moment dezelfde opgaven. De pilot leidde niet tot heel nieuwe inzichten wat betreft de afname. Negatieve *Rir*-waarden werden gevonden bij 7 luisteropgaven. Na inhoudelijke evaluatie zijn deze vervangen of aangepast. Eén audio-item is verwijderd vanwege een te hoge *p*-waarde en om tot gelijke aantallen in de toetsboekjes te komen. Ook hier bleven ankeritems ongewijzigd.

### **5.1.2 Definitieve toetsen: Lezen, Luisteren en Woordenschat**

De vaardigheden lezen, luisteren en woordenschat zijn ieder met een eigen toets afgenomen. Van iedere toets zijn zes versies afgenomen, waarbij iedere opgave in twee toetsen zat. De afnametijd per toets was 30 minuten. Meer informatie over de samenstelling van de toetsen is te vinden in paragraaf 3.3.

In de hiernavolgende beschrijving van de resultaten wordt onderscheid gemaakt tussen de kwaliteit van de opgaven en die van de toetsen als geheel. Ten behoeve van de KTT-analyses zijn als eerste stap de leerlingen met 100% ontbrekende waarden uit alle analyses naar de kwaliteit van de toetsen gelaten. Daarna is bij de KTT-analyses onderscheid gemaakt tussen de analyses met *en* zonder de leerlingen die vijf opgaven op een rij misten. De aanname is dat bij de leerlingen met meer dan vijf missende waarden op rij vaak meer aan de hand is dan een lagere vaardigheid, zeker als deze reeks in het midden van de toets zit (denk aan een overgeslagen bladzijde), maar ook als deze op het eind zit: de toetsen zijn namelijk kort genoeg om tot (bijna) het einde te komen. De serie waarbij de leerlingen uitvallen wanneer ze 5 opgaven op een rij missen wordt de *x*-serie genoemd. Bij de leerlingen zonder vijf ontbrekende waarden op rij is het hoogste gevonden percentage ontbrekende waarden 41% bij Lezen (bij Luisteren en Woordenschat is dat 38% en 36%). Alle leerlingen met meer dan 41% ontbrekende waarden in de dataset zijn daarom automatisch verwijderd, omdat zij vijf ontbrekende opgaven op een rij hadden. Bij de leerlingen die vijf opgaven op rij niet hadden ingevuld (en minder dan 100% missende waarden hadden), was het laagste percentage ontbrekende waarden 14% bij Woordenschat, en 29% en 31% bij Lezen en Luisteren. Meestal lagen die percentages echter hoger. De verschillen in de KTT-analyses voor de groep leerlingen die *iets* gedaan had (dat wil zeggen, die geen 100% ontbrekende waarden hebben) en de groep leerlingen die uitgesloten werd bij vijf ontbrekende waarden op rij zijn zeer gering. Bij toetsing worden er geen significante verschillen gevonden tussen de twee keuzes. Gekozen is derhalve om zoveel mogelijk leerlingdata te gebruiken bij de bepaling van de kwaliteit van de opgaven en toetsen. De hieronder gerapporteerde gegevens

betreffen daarom de groep die 'iets' gedaan heeft, de groep die dus geen 100% ontbrekende waarden heeft (de zogenaamde y-serie).

Merk op dat de KTT-resultaten voor de kwaliteit van de toetsen in alle gevallen ongewogen resultaten betreffen. Door de over-representatie van de scholen met relatief veel gewichten-leerlingen zouden de resultaten onderschat kunnen zijn, maar aangezien ook de VVTO-scholen oververtegenwoordigd zijn, zouden de waarden ook overschattingen kunnen zijn voor de gehele populatie. De KTT-resultaten zullen niet afzonderlijk worden teruggewogen naar de populatieresultaten. Dat wordt binnen de IRT-resultaten bij de beantwoording van onderzoeksvragen wel gedaan (zie hoofdstuk 8).

### *Kwaliteit van de opgaven*

In Tabel 5.2 wordt over alle toetsen heen informatie over de opgaven gegeven. Zo is te zien dat er een item was met een negatieve item-rest-correlatie ( $R_{ir}$ ), maar dat item heeft in een ander boekje een positieve  $R_{ir}$ , die echter met een waarde van 0,14 nog steeds laag is. Weliswaar is dat in het kader van een peiling nog steeds interessante informatie. De gemiddelde  $R_{ir}$ -waarden zijn goed te noemen en zijn bij Woordenschat (WS) en Lezen (LE) bij ruim 90% van de gevallen boven de 0.20, met andere woorden boven de minimale streefwaarde. Dit betekent dat de toets voor het overgrote deel (meer dan 90% van de items) bestaat uit items die een goed onderscheid maken tussen leerlingen met een hoge en lage toetsscore. Bij Luisteren (LU) geldt dat voor ruim 75% van de items.

Uit Tabel 5.2 kan tevens worden afgeleid dat het aantal ontbrekende waarden per item laag ligt. Alleen bij Woordenschat (WS) is het percentage wat hoger. Dat is te wijten aan de open korte antwoordopgaven waar de percentages ontbrekend relatief hoog zijn: het gemiddelde percentage ontbrekende waarden bij de 84 meerkeuze-opgaven Woordenschat ligt, evenals bij Lezen en Luisteren, onder de 1%, terwijl dat percentage bij de 26 open opgaven boven de 12% ligt. Al met al zien de gegevens voor de opgaven er goed uit.

Tabel 5.2: Itemgegevens over alle toetsen heen

	Percentage ontbrekend			p-waarden (*100)			Item Rest correlaties (*100)		
	LE	LU	WS	LE	LU	WS	LE	LU	WS
Gemiddelde	0.2	0.2	3.5	69.5	67.3	66.7	36.9	28.4	37.9
Minimum	0	0	0	38	25	15	11	7	-2
P10	0	0	0	52	39	39	25	12	22
P25	0	0	0	61	54	54	30	23	30
Mediaan	0	0	1	71	73	69	37	28	39
P25	0	0	3	78	80	80	44	34	46
P90	1	1	13	84	89	87	49	40	53
Maximum	1	1	36	90	94	97	59	50	64

### *Kenmerken van de toetsen: Betrouwbaarheid van de meetinstrumenten*

In Tabel 5.3 worden de kenmerken van de toetsen gegeven. Gegeven worden het aantal items per toets (N items) en het aantal personen per toets (N pers) dat een toets gemaakt heeft. Daarna volgen de itemeigenschappen per toets. Opgemerkt moet worden dat de gemiddelde p-waarden (gem. p) hier - in tegenstelling tot Tabel 5.2 - per toets zijn uitgerekend. In plaats van de  $R_{ir}$ -waarden zijn hier de (gemiddelde) item-test-correlaties ( $R_{it}$ ) gegeven zodat ook over de  $R_{it}$ -waarden informatie

gegeven wordt. Voor de beoogde hoogte van de  $R_{it}$ -waarden heeft de COTAN richtlijnen gegeven<sup>35</sup>, waar veruit de meeste items in deze toetsen aan voldoen<sup>36</sup>. De gemiddelde  $R_{it}$ -waarden zijn goed te noemen.

Na de gemiddelde itemeigenschappen worden de toets-kenmerken gegeven. Dit zijn achtereenvolgens de gemiddelde score, de standaardafwijking en de standaardmeetfout (SEM; volgens een KTT-berekening). Tot slot wordt de betrouwbaarheid gerapporteerd door middel van de klassieke interitembetrouwbaarheidsmaten Cronbach's Alpha en Guttman's Lambda-2. Dit zijn beiden onderschattingen van de werkelijke betrouwbaarheid, waarbij de Lambda-2-waarde nooit onder die van de waarde van Alpha kan liggen. Uit Tabel 5.3 kan worden afgeleid dat de verschillen tussen deze twee maten zeer klein zijn. Bij de waarden van Alpha zijn ook de standaardfouten weergegeven.

Tabel 5.3: Kenmerken van de toetsen

VDH	versie	N items	N pers	Item eigenschappen		Toetseigenschappen (score)			Betrouwbaarheid		
				gem. p	gem. Rit	Gemiddelde	sd	SEM	Alpha	SE(a)	Lambda2
LE	v1	17	402	73.2	0.51	12.4	3.7	1.4	0.82	0.01	0.82
LE	v2	17	211	68.3	0.51	11.6	3.8	1.7	0.81	0.02	0.82
LE	v3	17	273	66.7	0.51	11.3	3.9	1.7	0.82	0.02	0.82
LE	v4	17	411	67.7	0.47	11.5	3.6	1.5	0.77	0.02	0.78
LE	v5	17	313	71.2	0.46	12.1	3.5	1.4	0.77	0.02	0.77
LE	v6	17	314	71.1	0.50	12.1	3.7	1.4	0.81	0.02	0.81
LU	v1	16	395	64.7	0.42	10.4	3.0	1.4	0.68	0.02	0.69
LU	v2	16	206	66.2	0.48	10.6	3.3	1.6	0.77	0.02	0.78
LU	v3	16	248	70.8	0.40	11.3	2.6	1.6	0.64	0.03	0.65
LU	v4	16	422	66.4	0.41	10.6	2.8	1.5	0.66	0.02	0.67
LU	v5	16	295	69.9	0.44	11.2	3.0	1.5	0.71	0.02	0.72
LU	v6	16	300	66.8	0.42	10.7	2.9	1.5	0.68	0.03	0.69
WS	v1	36	396	70.3	0.48	25.3	7.4	2.3	0.90	0.01	0.91
WS	v2	36	206	70.4	0.47	25.3	7.0	2.2	0.90	0.01	0.90
WS	v3	37	248	69.4	0.44	25.7	6.8	2.3	0.88	0.01	0.89
WS	v4	37	421	60.7	0.44	22.5	7.4	1.9	0.88	0.01	0.88
WS	v5	37	313	65.0	0.40	24.0	6.4	2.0	0.84	0.01	0.85
WS	v6	36	302	66.1	0.43	23.8	6.9	1.7	0.87	0.01	0.88

De gemiddelde betrouwbaarheid bij Luisteren (LU) is het laagst met een gemiddelde Lambda-2 van 0,70, met een minimum waarde van 0,65. Bij Lezen (LE) is deze waarde gemiddelde 0,80 (en minimaal 0,77) en bij Woordenschat is dat bijna 0,90 met een minimum van 0,85. Zoals vermeld gaan we bij de beoordeling van de betrouwbaarheid uit van de COTAN richtlijnen voor het vergelijken van leerlingen op groepsniveau, zoals het geval is bij systeemevaluatie en terugrapportage op schoolniveau. Een betrouwbaarheid van 0,60 wordt daarbij als voldoende beschouwd en een betrouwbaarheid vanaf 0,70 als goed. De gevonden betrouwbaarheden zijn in alle gevallen

<sup>35</sup> Voor een beoordeling goed:  $R_{it} > .30$ ; voor een beoordeling voldoende  $R_{it} > .20$ . Zie Evers, A., Lucassen, W., Meijer, R.R. & Sijtsma, K. (2010). COTAN Beoordelingssysteem voor de Kwaliteit van Test (p.40). COTAN ( <https://www.psynip.nl/wp-content/uploads/2016/07/COTAN-Beoordelingssysteem-2010.pdf> )

<sup>36</sup> De waarden van  $R_{it}$  zijn altijd hoger dan die van de  $R_{it}$ -waarden die gegeven worden in Tabel 5.1.

voldoende tot (zeer) goed voor rapportage op groepsniveau. Bij de selectie van scholen is ook gebruik gemaakt van een totaalscore waarbij de vaardigheden Lezen, Luisteren en Woordenschat zijn samengenomen. Die samengestelde score was (uiteraard) ook betrouwbaar genoeg.

## 5.2 Taken: Spreekvaardigheid en Gespreksvaardigheid

### 5.2.1 Pilotonderzoek gesprekstaken

Voor de trendvergelijking voor mondelinge vaardigheid is een spreektaak gebruikt ('De Brug'). Vanwege de vergelijkbaarheid met de vorige peiling dienden de toets en het scoreformulier voor deze spreektaak identiek te zijn aan die in de vorige peiling. Herziening op basis van pilotresultaten was daarom niet aan de orde en er heeft geen pilotafname plaatsgevonden. De ontwikkelde taken voor gespreksvaardigheid zijn daarentegen wel gepilot. Gespreksvaardigheid is gemeten met behulp van drie gesprekstaken. Deze zijn in de pilot onderzocht bij 40 leerlingen, verdeeld over 20 paren. De drie te onderzoeken taken liepen op qua moeilijkheid: de eerste was gericht op het meten van het A1-niveau binnen het ERK, de tweede op het A2-niveau en de laatste op het B1-niveau. In principe konden alle drie de taken door alle leerlingen gemaakt worden, maar de afname was zo opgezet dat, als men zwak presteerde op een taak of te lang over een taak deed de volgende taak niet meer afgenomen zou worden.

De gesprekstaken zijn op video opgenomen. Ieder paar is minstens door één beoordelaar beoordeeld, die de prestaties van beide leerlingen codeerde in alle gesprekken die dit paar voerde. De meeste paren zijn ook door een tweede beoordelaar beoordeeld, die ook de prestaties van beide leerlingen codeerde. Het leerling-bestand bestond zodoende uit 80 records: 20 gesprekken \* 2 leerlingen \* 2 beoordelaars. In ieder record kunnen gegevens staan van drie gesprekken. Naast een codering van de leerlingenresponsen zijn ook de gesprekken als geheel gecodeerd (parenoordeel) door dezelfde beoordelaars die ook de prestaties van de leerlingen hebben gecodeerd in het gesprek. Dat leverde een bestand met 40 records (20 gesprekken \* 2 beoordelaars), ook weer met ruimte voor drie gesprekken per record.

De gesprekstaken waren allen nieuw, en zodoende ook de beoordelingsmodellen voor de individuen en de paren<sup>37</sup>. Hieronder volgen de belangrijkste gevolgen van de pilot voor de definitieve afname.

Taak 1 bestond in het pilot-format uit een tiental beoordelingspunten waarop iedere leerling beoordeeld werd. Bij die punten kreeg de leerling een score op een driepuntsschaal (scores 0,1 of 2) voor drie elementen van het gesprek: inhoud, realisatie kwaliteit en realisatie interactie en vloeiendheid (*fluency*). De scorering liep zodoende van 0 tot en met 60 in de pilot. In de pilot sloegen beoordelaars relatief veel opgaven over. Dit is bij dergelijke taken lastig in de verwerking, omdat een ontbrekende opgave niet noodzakelijkerwijs aan de leerling ligt, maar ook aan de beoordelaar kan liggen. De pilot leek er ook op te wijzen dat sommige beoordelaars vaker beoordelingspunten oversloegen dan anderen. Om die reden is de instructie aangescherpt en is de beoordelaars op het hart gedrukt dat zij geen beoordelingen mogen overslaan.

Er zijn verschillende vormen van beoordelaarsbetrouwbaarheid. De beoordelaarsbetrouwbaarheid op itemniveau kan enig inzicht geven in welke opgaven lastiger te beoordelen zijn. Hiertoe is het percentage overeenstemming bepaald voor de opgaven. Er was een aantal opgaven

---

<sup>37</sup> De verwerking van de pilot-resultaten per taak en voor de paarbeoordeling is beschikbaar via de onderzoekers.

waarbij een negatieve waarde voor Cohen's Kappa gevonden was. Dit is onwenselijk, omdat dit op een lage overeenstemming duidt. Door de lage aantallen observaties was het lastig te zeggen of er op basis hiervan items verwijderd moesten worden of aangepast wat betreft de scoring.

Op basis van de pilot en op inhoudelijke gronden is dit later in de definitieve versie van de taak aangepast naar 18 beoordelingspunten, ook weer met trichotome scores op drie elementen (range: 0-108 – 18 items\*3 elementen\*minimumscore0/maximumscore2)<sup>38</sup>. Over het algemeen was de overeenstemming in orde: de correlatie op totaalscoreniveau tussen twee beoordelaars waarbij geen ontbrekende waarden gevonden waren, was 0,79. De interitembetrouwbaarheid was met waarden boven 0,90 nog hoger.

Een verschil tussen taak 1 en de andere twee taken (Taak 2 en Taak 3) is dat bij de eerste taak alle leerlingen op dezelfde punten beoordeeld worden. Taak 2 en Taak 3 zijn eerder dialogen waarbij iedere leerling een net iets andere rol heeft en daarmee op (inhoudelijk) net iets andere beoordelingspunten beoordeeld wordt. Deze zijn wel zo verdeeld dat de scores uiteindelijk equivalent zijn, maar in het scoreproces op net iets andere momenten beoordeeld worden. De opzet van het scoreformulier had in de pilot bij Taak 2 en 3 echter hetzelfde format als bij Taak 1, waarbij de beoordelaar per leerling een serie "missing-by-design" zou hebben. Dit bleek niet goed te functioneren en relatief veel invulfouten op te leveren, met scores op niet verwachte posities en ontbrekende waarden op andere plekken. Op basis van de pilot is de opzet van het scoreformulier aangepast, waardoor de beoordeling duidelijk meer per leerling geordend zijn.

Door het lage aantal observaties bij Taak 2 en 3, te weten 15 en 9 leerlingen, was het lastig de psychometrische resultaten goed te kunnen duiden. Op zich waren de resultaten bij Taak 2 bemoedigend, en bij Taak 3 iets minder, maar er was behalve de aanpassing van het invulformulier geen reden veel aanpassingen te doen. Wel was het vermoeden dat bij de werkelijke afname er erg weinig resultaten voor Taak 3 zouden zijn, omdat er weinig leerlingen aan de taak zouden toekomen.

Naast de beoordelingen van de individuele leerlingen was er ook sprake van een beoordeling van itemparen. Het oorspronkelijke beoordelingsmodel voor de interactievaardigheid was gericht op aspecten als het vragen van hulp, het bieden van hulp, het al dan niet reageren, het gebruikmaken van non-verbale communicatie, betrokkenheid in het gesprek en het begrip van dat wat door de ander gezegd is – dit alles in relatie tot het gebruik van Engels/Nederlands. Dit model bestond uit twee onderdelen: onderdeel A met twee vragen en onderdeel B met tien vragen. Bij onderdeel A werd bevraagd of het gesprekstempo goed en het gesprek zonder interventies verliep. Bij dit onderdeel bleek per beoordelaar de correlatie tussen de twee vragen 1,0 te zijn, maar tussen de beoordelaars was de kappa met een waarde van 0,26 weer laag (overeenstemming van 71%). Bij de vragen bij onderdeel B naar de geobserveerde interactie-elementen, die deels afhankelijk waren van deel A, waren de overeenstemmingsmaten nog lager. Er was ook weinig variatie in scores: op de driepuntschaal ( $X_i = 0, 1, \text{ of } 2$ ) werd in 84% van de gevallen een score 0 gegeven. Op basis van die resultaten is er voor gekozen deze paarsgewijze beoordeling geheel aan te passen en voor een nieuwe benadering te kiezen. De volgende procedure is daarbij gevolgd. In overleg met de beoordelaars is besloten minder aspecten te scoren, die specifiekere waren geformuleerd. Ook is nader gedefinieerd wat we onder interactie verstonden. De volgende stap was het verzamelen van extra informatie over dit aangepaste model en de bijbehorende scoringsinstructie. Dit is gedaan tijdens een college binnen de master Applied Linguistics van de RuG, waarin studenten met het instrument hebben gewerkt aan de hand van twee video-opnamen van leerlinggesprekjes uit de

---

<sup>38</sup> Ook de taak zelf was overigens enigszins aangepast. De taak bleek namelijk teveel tijd in beslag te nemen (zie § 4.3.1).



pilot. De aldus verzamelde feedback leerde ons dat het problematisch was de constructen ‘begrip’ en ‘betrokkenheid’ te observeren en dat het model niet overzichtelijk was. De scope van het instrument werd derhalve verder ingeperkt, opdat antwoord kon worden gegeven op de vraag of er a) sprake was van een vorm van interactie en b) of de interactie in het Engels was. Deze vragen leken het meest relevant voor het meten van interactie in het Engels op een laag competentieniveau. Tenslotte zijn wijzigingen ten behoeve van de volledigheid en logica van het instrument doorgevoerd.

Ervaringen tijdens de pilot leidden ook tot wijzigingen in de beoordelingsprocedure. In tegenstelling tot het pilotonderzoek, waar de individuele en paarbeoordeling voor een specifiek paar door dezelfde beoordelaar werd uitgevoerd, zijn de individuele en paarbeoordelingen in het hoofdonderzoek door verschillende beoordelaars uitgevoerd. Dit om te voorkomen dat het parenoordeel beïnvloed zou worden door de individuele oordelen over de Engelse vaardigheid van leerlingen. Daarnaast is een digitale invoerversie ontwikkeld, waarin beoordelaars met behulp van checks werden gewezen op eventuele invoerfouten (zoals teveel scores of ontbrekende scores). De beoordelaars zijn in januari 2018 getraind aan de hand van videomateriaal.

### 5.2.2 Definitieve taken: Spreekvaardigheid

Als anker is een enkele taak meegenomen. Dat is de taak “De Brug” waarbij de leerling zes plaatjes moet beschrijven. De eerste beoordeling die de beoordelaar moet uitvoeren is of de taak überhaupt te beoordelen is. De beoordelaar kan aangeven dat de taak niet te beoordelen is, wanneer de opnamekwaliteit te slecht is of de leerling bij geen van de plaatjes een te beoordelen antwoord geeft. Dat is allebei een keer het geval. Daarnaast wordt ook de beoordeelbaarheid per plaatje geëvalueerd: uiteindelijk blijken nog eens 9 leerlingen op geen enkel plaatje te beoordelen te zijn, ondanks dat de beoordelaar aan het begin van de taak niet had aangevinkt dat de taak niet te beoordelen was. In totaal vallen daardoor 11 leerlingen af zonder enige beoordeling. Het aantal beoordelingen met scores is daarom 300. Daarvan zijn er 28 dubbele beoordelingen van in totaal 14 leerlingen, waardoor er voor 286 leerlingen taken minstens een taak beoordeeld is. Voor de verdere analyses is alleen gewerkt met een beoordeling per leerling, die op minstens een taak te beoordelen is.

Tabel 5.4: Percentages leerlingen met beoordelingen

<i>aantal plaatjes</i>	<i>per aantal</i>	<i>cumulatief</i>	<i>Plaatje</i>	<i>per plaatje</i>
6	52.1%	52.1%	Plaatje 1	95.8%
5	25.2%	77.3%	Plaatje 2	94.8%
4	8.7%	86.0%	Plaatje 3	94.1%
3	5.9%	92.0%	Plaatje 4	91.3%
2	3.8%	95.8%	Plaatje 5	89.5%
1	4.2%	100.0%	Plaatje 6	85.0%

Uit Tabel 5.4 valt af te leiden dat hoe verder het plaatje in de taak voorkomt, hoe minder personen te beoordelen zijn op het plaatje. Het is echter niet zo dat het ontbreken van een beoordeling van een plaatje betekent dat alle volgende plaatjes ook ontbreken.

De kwaliteit van de beoordelingen wordt eerst per plaatje geanalyseerd, waarbij alleen leerlingen worden meegenomen waarvoor het plaatje te beoordelen was. Het valt daarbij op dat,

wanneer de taak per plaatje wordt bekeken, de kwaliteit te wensen overlaat. De gevonden betrouwbaarheden zijn laag en minder dan de helft van de opgaven heeft een rit-waarde boven de 0,20 (zie Tabel 5.5).

Tabel 5.5: Beoordelingen per plaatje

	plaatje 1	plaatje 2	plaatje 3	plaatje 4	plaatje 5	plaatje 6
<b>N leerlingen</b>	274	271	243	261	269	256
<b>Lambda-2</b>	0.37	0.67	0.47	0.43	0.44	0.25
<b>Cronbach's alpha</b>	0.34	0.64	0.42	0.36	0.39	0.20
<b>N items</b>	5	8	6	6	9	6
<b>N items rit &gt; .00</b>	5	8	4	5	8	4
<b>N items rit &gt; .20</b>	2	7	3	2	4	1

Als we nu naar de somscore kijken, kunnen we de keuze maken om de plaatjes die niet te beoordelen zijn een score 0 mee te geven, of alleen te kijken naar de leerlingen die op alle plaatjes te beoordelen zijn. Aangezien er in dat laatste geval net de helft van de leerlingen overblijft, kijken we naar de variant waar de score 0 gegeven wordt wanneer een plaatje overgeslagen wordt. In dat geval is de betrouwbaarheid 0,80 (en Cronbachs alpha 0,78). Er zijn dan twee opgaven met een negatieve rit en 23 opgaven met rit-waarden boven de 0,20 en 15 opgaven met een rit tussen de 0,00 en 0,20. Deze uitkomsten kunnen mogelijk verklaard worden doordat de uitspraak op het laagste niveau veel variatie bevat, waardoor het onderscheidende vermogen van items relatief laag kan zijn. Dit in ogenschouw nemend lijkt de kwaliteit van de taak een trendvergelijking op stelselniveau toe te staan.

De spreektaak is beoordeeld door drie beoordelaars: een met 15 beoordelingen, een met 107 en een met 189 beoordelingen. Dubbele beoordelingen zijn gedaan door de personen met meer dan 100 beoordelingen. Na de randomtrekking van dubbele beoordelingen zijn de resultaten bepaald op 101 beoordelingen van de 107 en 181 van de 189 beoordelingen. De beoordelaars-overeenstemming per opgave was voor 28 opgaven te berekenen en bleek 0,65. Bij de overige opgaven was geen variatie in scores bij minstens een van de beoordelaars. Bij 23 opgaven was dat een 100% overeenstemming tussen de twee beoordelaars, bij 10 opgaven had een van de twee beoordelaars een andere score. De overeenstemming is dus redelijk, zij het dat dit op zeer beperkte data is bepaald. Informatie over de overeenstemming tussen beoordelaars in 2012 op deze specifieke taak is onbekend.

### 5.2.3 Definitieve taken: Gesprekstaken 1, 2 en 3

#### *Modellen voor Individuele beoordelingen gesprekstaken*

Voor de individuele beoordelingen is per gesprekstaak een beoordelingsformulier geconstrueerd, bestaande uit 3 beoordelingsmodellen, voor iedere taak één. Voor iedere leerling werd voorts genoteerd of de testleider de instructie volgens de richtlijnen had uitgevoerd, als indicatie van de uniformiteit van de afnames. Ook werd, ná beoordeling van de uitgevoerde gesprekstaken, gevraagd naar het ingeschatte ERK-niveau van de leerling. Deze informatie is voor validatiedoeleinden gebruikt.

Elke taak kende een verschillend aantal beoordelingspunten en was opgedeeld in onderdelen. Deze onderdelen zijn beoordeeld op drie criteria: 1) inhoud, 2) realisatie kwaliteit en 3)

realisatie interactie en vloeiendheid (*fluency*). Op deze manier kon antwoord worden gegeven op de volgende vragen:

*Wat is het niveau van de Engelse gespreksvaardigheid van individuele leerlingen in groep 8?*

- *In hoeverre zijn leerlingen in groep 8 in staat om de gespreks(sub)taken volledig in het Engels uit te voeren?*
- *Wat is de kwaliteit van de bijdragen aan de gespreks(sub)taken van leerlingen in groep 8?*
- *Wat is de vloeiendheid en interactievaardigheid van de bijdragen aan de gespreks(sub)taken van leerlingen in groep 8?*

Voor de eerste vraag werd het criterium *Inhoud* gebruikt. Scores liepen van 0 (onvoldoende), 1 (voldoende) tot 2 (goed). Onvoldoende betekende dat het gevraagde niet in het Engels werd gerealiseerd en dat de communicatieve boodschap helemaal niet overkwam. Bij voldoende kon de leerling de communicatieve boodschap nog niet volledig overbrengen omdat hij/zij a) niet alles vertelde wat op het kaartje stond en/of b) deels in het Nederlands, deels in het Engels sprak. Goed betekende dat de leerling het ERK-niveau waarop de taak is gebaseerd liet zien. De leerling was in staat de communicatieve boodschap volledig en in het Engels over te brengen, op het niveau van de taak. Dit betekende dat bij Taak 1 score '2' staat voor A1-niveau, bij Taak 2 voor A2-niveau en bij Taak 3 voor B1-niveau.

De tweede en derde vraag konden respectievelijk met *Kwaliteit* en *Interactie & Fluency* (I&F) worden beantwoord. Voor de beoordeling werd gebruikgemaakt van de ERK-beschrijvingen. Kwaliteit werd vastgesteld op basis van woordenschat, grammatica en samenhang. Bij I&F was het tempo, de snelheid, een belangrijke indicator. Ook hier werd onvoldoende, voldoende en goed gescoord. Bij beide categorieën gold de stelregel dat voldoende indicatief is voor het ERK-niveau van de taak. Een score voldoende stond bij alle items/subtaken van Taak 1 voor A1-niveau, van Taak 2 voor A2-niveau en van Taak 3 voor B1-niveau. De score '2' (goed) gaf weer dat een leerling meer liet zien dan bij het ERK-niveau wordt verwacht.

Om de uniformiteit tussen de beoordelaars te bevorderen zijn er op inhoudelijke gronden algemene scoringsregels vastgesteld waar elke beoordelaar zich aan diende te houden. Deze stonden weergegeven in het scoringsprotocol. Tevens is daarin opgenomen hoe om te gaan met missende waarden. Tijdens de trainingen zijn bovendien specifieke afspraken gemaakt voor bepaalde korte uitingen in Taak 1 en Taak 3, die weliswaar de communicatieve boodschap uitdrukten, maar waarmee niet werd aangetoond dat een leerling kwalitatief meer kon dan het aangegeven niveau. Een voorbeeld hiervan is het enkel noemen van een getal om aan te geven hoe oud je bent.

Een groep van 12 beoordelaars, allen ERK-deskundigen, is in januari 2018 twee dagen getraind met behulp van calibratiesessies (aan de hand van videomateriaal). Na afloop van de tweede sessie hebben bleken de resultaten onvoldoende betrouwbaar. Hierop zijn aanvullende instructies gegeven, bestaande uit 1) nog specifiekere richtlijnen voor scoring volgens het vastgestelde scoringsmodel, 2) de werkwijze van de 'meesterbeoordelaar' (een beoordelaar met veel ervaring op het gebied van het beoordelen van gespreksvaardigheid en die het meest consistent over de taken was gebleken) en 3) algemene richtlijnen t.b.v. de consistentie.

### **Beschikbare data gesprekstaken**

De ontvangen gegevens van gesprekken betroffen 732 bruikbare records. Deze waren afkomstig van 583 unieke leerling-nummers. Deze 583 unieke leerling-nummers behoorden tot 288 unieke leerling-

paren<sup>39</sup>. Van de 583 leerlingen is er bij twee leerlingen geen beoordeling, bij 447 een enkele beoordeling, bij 124 beoordelingen van twee beoordelaars, en bij tien beoordelingen van drie beoordelaars. De gegevens waren afkomstig van 11 verschillende beoordelaars, die gemiddeld 73 werken beoordeeld hebben. Het laagste aantal beoordeelde werken door een beoordelaar is 30, terwijl het hoogste aantal 119 is.

### *Kwaliteit van de opgaven - gesprekstaken*

Bij de beoordeling zijn er in totaal 120 items gebruikt. Dit waren 54 items bij Taak 1, 30 bij Taak 2 en 36 bij Taak 3. Een individuele leerling kon maximaal op 87 items beoordeeld zijn: weer 54 bij Taak 1, maar bij Taak 2 en Taak 3 telkens maar de helft door de verschillende rollen binnen die taak. De helft van de items betreft de rol Lisa/Leo of de rol Rosa/Roy bij 2, met 15 opgaven per rol. Bij Taak 3 betreft de helft van de opgaven een "Voor-argument"-rol en de helft een "Tegen-argument"-rol, met 18 opgaven per rol. Er is daarnaast nog sprake van een ander type opdeling van opgaven, namelijk die naar Inhoud (I), Fluency (F) en Kwaliteit (K). Ieder beoordeeld element in de taak wordt op alle drie de kenmerken beoordeeld. Dit betekent dat bij Taak 1 een leerling beoordeeld is op  $(54/3=)$  18 elementen, bij Taak 2 op  $(30/2/3=)$  5 elementen en bij Taak 3 op  $(36/2/3=)$  6 elementen.

Bij de opgaven bij Taak 1 zijn 674 tot en met 726 beoordelingen per opgave in de dataset (gemiddelde 716), bij Taak 2 was dat 295 tot en met 312, en bij Taak 3 was dat 117 tot en met 154. Bij deze laatste taak waren er meer observaties horende bij de rol Voor (147-154) dan Tegen (117-120)<sup>40</sup>. De verschillen in aantallen tussen Taak 1, 2 en 3 worden veroorzaakt, doordat men pas aan een volgende taak kon beginnen na (kwalitatief redelijke) afronding van de vorige taak. In het geval van Taak 3 was er ook vaak sprake van tijdnood, waarbij tijdnood als indicator van niveau kan worden beschouwd. Tijdnood duidt er namelijk op dat leerlingen de juiste uitdrukkingen in het Engels niet kunnen vinden, veel haperen en zichzelf herstellen of dat ze door gebrekkige lexicale of grammaticale kennis op zoek moeten gaan naar omschrijvingen – kortom, het duidt op moeite met de inhoud van de taak. De verschillen binnen Taak 3 worden veroorzaakt doordat niet altijd beide leerlingen doorgingen naar Taak 3. Er wordt geen verschil gevonden tussen de I, K en F-items voor wat betreft het aantal observaties.

Iedere opgave is een trichotome opgave met mogelijke score 0, 1 en 2. Bij ieder van de 120 opgaven is iedere score wel eens waargenomen, behalve bij een inhoudsbeoordeling binnen taak 1 (item *A11aal* – een I-type opgave bij taak 1). De p-waarden variëren van 23 (F-type opgave bij taak 3) tot en met 99 (I-type opgave bij taak 1). Het is duidelijk dat de opgaven bij taak 1 gemakkelijker zijn met een gemiddelde p-waarde van 73. De gemiddelde p-waarde bij Taak 2 en 3 is 43. De p-waarden bij Taak 3 liggen iets hoger dan bij taak 2 (44 tegenover 42), maar dat is mede veroorzaakt doordat minder goede leerlingen niet aan Taak 3 toekwamen. In de schaling – zie paragraaf 5.4.2 – is hier rekening mee gehouden. Binnen de taken zijn de I-typen opgaven over het algemeen iets gemakkelijker dan de F-type (p-waarde 63 over alle taken heen tegen over 60). De moeilijker opgaven zijn de K-type, met een p-waarde van 47.

Binnen de opgaven is er geen extra onderzoek uitgevoerd naar de beoordelaars-overeenstemming per opgave. Wel is er op somscore-niveau onderzoek uitgevoerd. Meer hierover bij de evaluatie van de kwaliteit van de toetsen.

---

<sup>39</sup> De overige zeven leerlingen hadden enkel met een testleider gesproken of waren niet identificeerbaar als paar

<sup>40</sup> Waarschijnlijk komt dit doordat testleiders leerlingen vaak lieten kiezen welke rol de leerling wilde spelen op het moment dat de testleider de taak met de leerling uitvoerde. De meeste leerlingen wilden het wel met de stelling eens zijn.

### Kwaliteit van de toetsen – individuele oordelen gesprekstaken

Voor de evaluatie van de toetsen is onderscheid gemaakt tussen de eerste en de tweede of derde beoordelaar. Eerst is de beoordelaarsovereenstemming tussen de beoordelaars per taak geëvalueerd, waarna de betrouwbaarheid van de taakscores beoordeeld is. Deze tweede evaluatie is altijd gedaan op maar één beoordeling per leerling, dus bij dubbel beoordeelde leerlingen op basis van een random getrokken beoordeling.

De taakscores worden gevormd door de ongewogen som te nemen van de itemscores. Bij Taak 1 levert dit per leerling mogelijke scores van 0 tot en met 108 op, bij taak 2 van 0 tot en met 30 en bij Taak 3 van 0 tot en met 36. Als we kijken naar de interitembetrouwbaarheid zoals gemeten met Guttman's Lambda-2, dan is te zien dat deze per taak zeer hoog is. Dat is voor een deel te danken aan de hoge samenhang per element tussen de I, K en F items. Kijken we naar de betrouwbaarheid van de sets items met alleen de I, K, of F items, dan is de interitembetrouwbaarheid nog steeds hoog. De gevonden betrouwbaarheden staan in Tabel 5.6. Merk op dat de betrouwbaarheden voor Taak 2 en Taak 3 per rol zijn berekend, omdat de oordelen per rol gegeven zijn (RR=Roy/Rosa, LL=Leo/Lisa, V=Voor en T=Tegen). De hoogte van de betrouwbaarheid per taak is zeer hoog te noemen.

Tabel 5.6: betrouwbaarheid per taak

	Taak 1	Taak 2 RR	Taak 2 LL	Taak 3 V	Taak 3 T
<b>Alle items</b>	<b>0.96</b>	<b>0.95</b>	<b>0.95</b>	<b>0.94</b>	<b>0.95</b>
<i>Inhoud items</i>	0.87	0.84	0.85	0.80	0.85
<i>Kwaliteit items</i>	0.89	0.88	0.85	0.86	0.85
<i>Fluency items</i>	0.88	0.87	0.84	0.85	0.86

Naast de interitembetrouwbaarheid is er ook de beoordelaarsbetrouwbaarheid. De dubbel beoordeelde leerlingen zijn door zo veel mogelijk verschillende beoordelaarsparen beoordeeld. Bij 11 beoordelaars zijn er 55 mogelijke paren. Daarvan zijn er vier paren nooit geobserveerd; alle 51 andere koppels hebben gemiddeld 3 leerlingen dubbel beoordeeld<sup>41</sup>. De reden voor het ontbreken van vier paren is gelegen in onverwachte tussentijdse uitval van één van de beoordelaars, waardoor deze beoordelaar niet met iedere beoordelaar een beoordelaarspaar vormde.

Als twee beoordelaars of drie beoordelaars hebben beoordeeld en de laagste scorende beoordelaar wordt gezien als de eerste en de hoogst scorende als de tweede dan zijn de interbeoordelaars-betrouwbaarheden 0,88 (Taak 1), 0,92 (Taak 2) en 0,83 (Taak 3). Het gemiddelde (absolute) verschil tussen de twee beoordelaars is dan wel het grootst, respectievelijk 8,3 punt, 4,3 en 7,1. In ons geval trekken we echter een random beoordelaar als eerste en als tweede beoordelaar. Omdat verschillende random trekkingen verschillende resultaten op zouden kunnen leveren, zijn er voor deze evaluatie 200 trekkingen gedaan. De resultaten hiervan staan in Tabel 5.7 weergegeven.

Tabel 5.7: Resultaten na 200 trekkingen

	Correlatie			Verschil			Absoluut verschil		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
<b>Gemiddelde</b>	<b>0.80</b>	<b>0.81</b>	<b>0.65</b>	<b>0.05</b>	<b>-0.05</b>	<b>-0.09</b>	<b>0.81</b>	<b>0.37</b>	<b>0.95</b>

<sup>41</sup> Merk op dat als een leerling door 3 beoordelaars beoordeeld wordt dit 3 verschillende paren oplevert.

<i>Mediaan</i>	0.80	0.81	0.65	0.11	-0.07	-0.19	0.67	0.29	0.77
----------------	------	------	------	------	-------	-------	------	------	------

De resultaten wat betreft de correlatie tussen de eerst en tweede random getrokken beoordelaar waren ook al gevonden na 50 trekkingen, hetgeen aangeeft dat meer dan 200 trekking niet nodig lijken en dat deze resultaten stabiel genoeg zijn. De verschillen in de scores liggen in absolute waarden onder een punt.

Al met al geeft dit per taak vertrouwen dat de taakscores goed beoordeeld zijn. Met interbeoordelaarsbetrouwbaarheden van 0,80 is dit zelfs goed te noemen, terwijl bij een correlatie deze resultaten voldoende zijn, volgens de COTAN richtlijnen voor metingen van groepen zoals bij een peiling. Een mogelijke reden voor de lagere betrouwbaarheid bij Taak 3 kan zijn dat de spreiding in vaardigheid bij deze taak lager ligt dan bij de andere twee taken: de matig presterende leerlingen komen niet toe aan taak drie, terwijl de volledige spreiding aan vaardigheden nog te vinden is bij Taak 1. Tot slot kan opgemerkt worden dat de definitieve beoordeling niet gedaan wordt op een enkele taak maar op de combinatie van taken, waardoor de interbeoordelaarsbetrouwbaarheid nog hoger zal liggen.

### *Beoordelingsmodel voor interactie op paarniveau - gesprekstaken*

Er vanuitgaande dat interactie het product is van gedragingen tussen twee leerlingen en een interactie niet geslaagd is als slechts één van beide personen actief deelneemt, is in het beoordelingsformulier een score per paar toegekend en niet per persoon apart. Voor de beoordeling per paar is gekeken naar de interactie tussen twee leerlingen op Taak 1, Taak 2 en Taak 3. Iedere taak was opgedeeld in clusters, waarvoor telkens één paarscore is gegeven. Deze clusters waren dusdanig samengesteld dat beide leerlingen minimaal één beurt hebben en zowel initiator als reactor zijn. Een parenoordeel is daarbij alleen gegeven wanneer *twee leerlingen met elkaar* de taak uitvoerden. Tussentijdse interactie met een testleider of volledige uitvoering van de taak met een testleider zijn niet gescoord.

Met behulp van scores op paarniveau wordt een antwoord gegeven op de volgende vragen:

*Interacteren leerlingen tijdens de gesprekstaken en zo ja, op welke manier?*

- *In hoeverre is er sprake van interactie per tweetal per gespreks(sub)taak?*
- *In hoeverre is er sprake van Engelstalige interactie in leerlingparen uit groep 8?*

Interactie is hierbij gedefinieerd als “het geheel van verbaal en nonverbaal gedrag dat de gesprekspartners tijdens een gesprekstaak vertonen”. Hierbij is het feit dat ze Engels gebruiken belangrijk, maar het was ook relevant om te observeren in hoeverre de gesprekspartners contact met elkaar hebben, via gesproken taal en lichaamstaal. De gedragingen die als relevant werden gezien in het gespreksbeoordelingsmodel zijn:

- Het paar spreekt wel/niet Engels
- Het paar spreek wel/niet Nederlands
- Het paar communiceert wel/niet nonverbaal

Uitgaande van bovenstaande definitie kan verbaal gedrag dus zowel Engels als Nederlands zijn, terwijl nonverbaal gedrag alleen dan is gescoord als er sprake is van compensatoir nonverbaal gedrag. Hieronder werd het nonverbale gedrag (zoals gebaren) verstaan dat de leerling gebruikte ter compensatie voor het feit dat hij/zij zich niet volledig verbaal kon uitdrukken. Op het moment dat de leerling een uiting volledig in het Engels kon realiseren en gebaren aan deze uiting toevoegde zonder dat zij betekenis toevoegde, werd alleen het verbale gedrag gescoord.

De drie genoemde gedragingen konden ook in combinatie met elkaar worden geobserveerd. Om de boodschap over te brengen waren de volgende 8 vormen van interactie onderscheidbaar:

- (1) Volledig Engelstalig
- (2) Deels Engelstalig, deels nonverbaal
- (3) Deels Engelstalig, deels Nederlandstalig
- (4) Deels Engelstalig, deels Nederlandstalig, deels nonverbaal
- (5) Deels Nederlandstalig, deels nonverbaal
- (6) Volledig nonverbaal
- (7) Volledig Nederlandstalig
- (8) Geheel afwezig (niet communicatief: stil; amper beweging)

Deze lijst was voldoende uitputtend voor het onderzoek. De beoordelaar vinkte altijd per cluster aan van welke van de acht vormen van interactie in dat cluster sprake was en deze interactievorm werd als '1' gescoord. Wanneer er bijvoorbeeld 4 clusters in een taak waren, is er dus bij elk cluster een interactietype aangevinkt. Bij Taak 1, bestaande uit 4 clusters, leverde dit per interactievorm een score op van 0-4; bij Taak 1 was de somscore van de somscores van de acht typen interactie altijd 4. In het gebruikte digitale invoerbestand zijn de beoordelaars m.b.v. checks gewezen op deze maximale scores.

Voorts bevatte de rangordening een hiërarchie. Zij kon geïnterpreteerd worden als een waardering, waarbij de hoogste waardering gegeven werd aan de interactievorm (1 = volledig Engelstalig) en de laagste voor (8 = geheel afwezig). Onze eerste vraag was of paren überhaupt interacteren. Indien dat niet het geval was, werd per cluster interactievorm 8 ingevuld, die een score '0' voor dat cluster opleverde. Daarnaast was de vraag of paren in staat zijn om enkel in het Engels te spreken. Dit is de score die volgt uit formule 2, waarin alleen punten worden toegekend aan de clusters die volledig in het Engels zijn uitgevoerd. In een mildere variant (formule 3) worden daarnaast ook interactievormen 2, 3 en 4 als 'een uiting van interactie in het Engels' beschouwd. De waardering kwam tot uiting in de weging die gehanteerd is om tot een definitief oordeel te komen (waarbij bij de milde variant het volledig in het Engels spreken zwaarder woog dan bijvoorbeeld het gedeeltelijk spreken van Engels/Nederlands):

1. Interactiescore:  $1*G1(1) + 1*G1(2) + 1*G1(3) + 1*G1(4) + 1*G1(5) + 1*G1(6) + 1*G1(7) + 0*G1(8)$
2. Interactie\_Engels\_strikt:  
 $1*G1(1) + 0*G1(2) + 0*G1(3) + 0*G1(4) + 0*G1(5) + 0*G1(6) + 0*G1(7) + 0*G1(8)$
3. Interactie\_Engels\_mild:  
 $2*G1(1) + 1*G1(2) + 1*G1(3) + 1*G1(4) + 0*G1(5) + 0*G1(6) + 0*G1(7) + 0*G1(8)$

### **Beschikbare data leerlingparen - gesprekstaken**

Aangezien de gesprekstaken conform design bij 600 leerlingen zouden worden afgenomen, was de verwachting dat er 300 paren zouden worden beoordeeld (voor Taak 1; voor Taak 2 en zeker voor Taak 3 was de verwachting dat minder paren zouden worden beoordeeld, omdat niet alle leerlingen door zouden gaan naar Taak 2 en/of 3). Het totaal unieke paren dat uiteindelijk is beoordeeld, was 287. Van deze 287 beoordelingen is uiteindelijk voor 283 paren een bruikbaar oordeel gegeven. Drie opnamen bleken niet bruikbaar (want mislukt) en bij één opname was een leerling niet zichtbaar, waardoor beoordeling van nonverbaal gedrag niet mogelijk was. Het overzicht van de aantallen paren waarbij 1, 2 of 3 taken zijn beoordeeld wordt gegeven in Tabel 5.8. Daarin is ook aangegeven



welke taken beoordeeld zijn. Op basis van Tabel 5.8 is af te leiden dat bij Taak 1 er 283 unieke paren zijn beoordeeld (98%), bij Taak 2 zijn dat er 224 (78%) en bij Taak 3 zijn dat 79 paren (28%).

Tabel 5.8: Aantal paren met beoordeelde taken.

Aantal	Aantal	%	Welke	Aantal	%
0 Taken	4	1.4	Geen	4	1.01
1 Taak	59	20.2	Taak 1	57	19.86
			Taak 2	1 <sup>42</sup>	0.00
2 Taken	147	50.3	Taak 1 & 2	143	50.83
			Taak 1 & 3	2	0.01
			Taak 2 & 3	1	0.00
3 Taken	82	28.1	Taak 1, 2 & 3	79	27.53

### Beoordelaarsbetrouwbaarheid (paarbeoordeling)

De procedure voor het bepalen van de beoordelaarsbetrouwbaarheid is vergelijkbaar met die bij de enkele beoordeling. Bij een beperkt deel van de paren zijn de interactiebeoordelingen dubbel beoordeeld. Het betrof 14 paren waarvan met zekerheid gesteld kon worden dat deze dubbel beoordeeld waren. Bij alle 14 was Taak 1 dubbel beoordeeld; bij negen paren was Taak 2 ook door beide leerlingen gedaan; en bij slechts drie paren was Taak 3 door allebei de leerlingen gedaan.

Gegeven de eigenschappen van de Interactie-score is het duidelijk dat als bijna alle paren de maximum score krijgen dit ook in de Beoordelaarsbetrouwbaarheid geen variatie oplevert. Een interbeoordelaarsbetrouwbaarheid levert dit dan ook niet op: wel een 100% overeenstemming.

De betrouwbaarheid over de Milde en de Strikte score verschilden binnen een Taak niet veel. Er is bij Taak 1 bij een 50-tal combinaties van paren<sup>43</sup> bekeken en daar was de correlatie ,70. Dit was overigens al bereikt bij 10 combinaties van paren. Bij Taak 2 werd een interbeoordelaarsbetrouwbaarheid van 0.68 gevonden. Deze was een honderdste lager dan het resultaat na 10 combinaties van paren. Bij Taak 3 waren er maar drie paren en zijn alle combinaties bekeken. Deze leverde een interbeoordelaarsbetrouwbaarheid van 0.95.

## 5.3 Vragenlijst leerlingen, scholen en leerkrachten

### 5.3.1 Analyses voor de leerlingvragenlijst in de pilot

De vragenlijst voor de leerlingen was grotendeels gebaseerd op de vragenlijst die in de vorige peiling is gebruikt. Dit betekent dat er voor een groot deel van de vragen al veel bekend was over deze opgaven. Desalniettemin was de vragenlijst in de pilot bij 70 leerlingen afgenomen. Op een leerling na die te veel ontbrekende waarden had, waren deze pilotresultaten bruikbaar voor analyse. De pilot kon goed gebruikt worden om verder te onderzoeken hoe de gegeven antwoorden konden worden omgezet om ervoor te zorgen dat de hoogste score de meest positieve attitude dan wel meeste ervaring met Engels weergeeft, en de laagste score de minst positieve attitude dan wel de minste ervaring.

<sup>42</sup> Voor in totaal 4 paren waren er slechts interactieoordelen voor taak 2 (en niet taak 1), taak 1 en 3 en taak 2, en 3. Reden voor deze missing waren opnieuw opnameproblemen (uitval camera bij één specifieke taak).

<sup>43</sup> Per combinatie kan over de paren verschillen wie de eerste en wie de tweede beoordelaar was.



In de pilot bleek een aantal vragen niet sterk samen te hangen met de overige opgaven. Van deze vragen is de vraag “Ben je wel eens op vakantie geweest in een Engelstalig land?” daarnaast ook om inhoudelijke redenen niet in de uiteindelijke vragenlijst opgenomen. De vraag “Met wie spreek je het vaakst Engels buiten school?” met zes sub-vragen is aangepast naar zes losse vragen omdat deze in de opzet van de pilot geen bruikbare resultaten opleverde. De open vragen zijn wel in de vragenlijst gebleven, maar leverden geen scorebare informatie op. Het ging hierbij om vragen naar de specifieke taal die met vader, moeder of verzorgende werd gesproken. Dit leverde een dusdanig grote hoeveelheid unieke talen op dat die moeilijk te classificeren was; nadere analyse was weinig zinvol.

### 5.3.2 Analyses voor de leerlingvragenlijst in de definitieve afname

De leerling-vragenlijst wordt in principe bij alle leerlingen afgenomen. Er is dus geen sprake van een incompleet design. Er waren 2117 kandidatengegevens verzameld bij de vragenlijst, maar na exclusie van de groep 7-leerlingen (die mochten meedoen), bleven daar 2075 van over. Dat betekent dat slechts bij 13 van de 2088 leerlingen in de steekproef geen gegevens beschikbaar zijn: zes leerlingen hadden niets ingevuld tijdens de afname, en van zeven leerlingen is er geen vragenlijstformulier ontvangen.

De opgaven in de vragenlijst zijn genummerd van 1 tot en met 30, maar een aantal vragen bestaat uit deelvragen waardoor het aantal in te vullen vragen voor de leerlingen uitkomt op 54. Daarvan zijn er twee open opgaven waar leerlingen alleen iets in hoefden te vullen als de belangrijkste verzorgers van de leerling een andere taal dan Nederlands zouden spreken. Al met al leverde dat 52 vragen op waarbij ontbrekende waarden ook echt overgeslagen vragen konden zijn. Bij 1724 leerlingen (83% van de leerlingen waar gegevens over bekend zijn) is er geen enkele van de deze opgaven ontbrekend. Van 9% en bijna 3% van de leerlingen ontbreekt één dan wel twee antwoorden, waardoor van 95% van de leerlingen zeer bruikbare gegevens aanwezig zijn. Van 1% van de leerlingen ontbreekt meer dan een kwart van de antwoorden en bij in totaal vijf leerlingen mist meer dan de helft. Het ontbreken van de antwoorden lijkt niet samen te hangen met de vaardigheid van de leerlingen. Alleen bij luisteren is er sprake van een significante samenhang (bij een alpha van 0,01), maar met een correlatie van -0,06 is deze samenhang verwaarloosbaar. De opgaven die in de schalen voor de vragenlijst voor de leerlingen voorkomen, staan weergegeven in Bijlage 9.

#### *Kwaliteit van de items in de leerlingvragenlijst*

Voordat de antwoorden gebruikt konden worden moesten de ingevulde vragenlijst-antwoorden omgezet worden. In een aantal gevallen betrof dat een omzetting van scores om ervoor te zorgen dat de hogere scores een positievere attitude geven. In een aantal gevallen leverde een enkel antwoord een aantal verschillende responsen op zoals vraag 10 (“Wat vind je leuk tijdens Engelse les?”) waar meer vakjes ingevuld konden worden. In totaal waren er 62 scores met betrekking tot de attitude en ervaring met het Engels, met daarnaast twee open vragen, en een vraag over de verwachte doorstroom naar het voortgezet onderwijs (van praktijkonderwijs naar vwo). Van deze 62 vragen zijn er 36 dichotoom (scores 0 en 1) en 18 trichotoom (0, 1, 2). Van de overige acht opgaven hebben er zes scores lopend van 0 tot 3 en twee van 0 tot en met 4.

Bij de evaluatie is ook onderscheid gemaakt tussen de vragen naar specifieke eigen ervaring met Engels, en naar de thuistaal en talen die de ouders spreken. Deze laatste vragen (vraag 23 tot en met 27) zijn als achtergrondvariabelen geïnterpreteerd en niet als onderdeel van een mogelijke

attitude-/ervaringsschaal. Dit geldt ook voor twee andere vragen die als achtergrondvragen gezien werden, omdat ze niet over Engels gingen. Het ging om de vraag over de verwachte doorstroom naar het voortgezet onderwijs en de vraag “Vind jij taal (Nederlands) op school gemakkelijk”.

Bij goede opgaven, na de omzetting van de scores, hangt een hogere score samen met een hogere totaalscore. Bij goede opgaven geldt dit voor iedere opvolgende categorie: het gevonden gemiddelde op de overige items bij score  $x_i$  zou lager moeten liggen dan bij  $x_{i+1}$ . Dit is onderzocht voor alle opgaven, na de logische inhoudelijke hercodering. Bij veruit de meeste opgaven die mogelijk onderdeel van de attitude/ervaringsschaal zouden kunnen zijn was te zien dat de opgaven aan deze eisen voldeden.

Uitzonderingen waren<sup>44</sup>:

- 10.4. Wat vind je leuk tijdens Engels: luisteren (liedjes/verhalen)
- 10.5. Wat vind je leuk tijdens Engels: liedjes zingen
- 10.6. Wat vind je leuk tijdens Engels: niets
- 14.4. Op tv, dvd, internet of in de bioscoop Engelstalige films, series of andere programma's kijken met Nederlandse ondertiteling

Vragen 10.4 tot en met 10.6 waren dichotome vragen. Bij 10.4 en 10.5 zou een positief antwoord inhoudelijk eigenlijk een positieve attitude of ervaring moeten representeren. Dat bleek niet het geval. Echter, het hercoderen van scores was inhoudelijk niet logisch. Dat is anders bij vraag 10.6: als een leerling niets leuk vindt bij Engels zou het gegeven antwoord “ja” oorspronkelijk een score 1 opleveren. Het is echter eerder indicatief voor een negatieve houding ten opzichte van Engels. Deze opgave is daarom omgezet naar 10.6R. Opgave 14.4 is een opgave met 4 scorecategorieën, waarbij de leerlingen die nooit met ondertiteling kijken een fractie hoger scoren dan leerlingen die dat soms doen. Het verschil was echter minimaal waardoor er niet voor een hercodering is gekozen<sup>45</sup>.

Tot slot behoeft vraag 17 speciale aandacht. De vraag luidt:

*Welke uitspraak klopt voor jouw gevoel? Mijn klasgenootjes....*

- luisteren minder vaak naar Engelse liedjes dan ik.
- luisteren net zo vaak naar Engelse liedjes als ik.
- luisteren veel vaker dan ik naar Engelse liedjes.
- Ik weet niet hoe vaak mijn klasgenootjes naar Engelse muziek luisteren.

Het merendeel (62% van de leerlingen die de vraag niet overslaan) van de kandidaten geeft aan niet te weten hoeveel klasgenootjes naar Engels liedjes luisteren. Deze categorie kan als een ontbrekende waarde gebruikt worden, maar dat betekent dat de somscore met deze opgave in totaal veel

---

<sup>44</sup> Andere uitzonderingen waren een aantal vragen dat gesteld werd bij de opgaven 23 tot en met 27 en vraag 30, maar deze zijn geen onderdeel van de te vormen schalen.

<sup>45</sup> Het is wel te vatten dat leerlingen die veel liever in het Engels dan in het Nederlands communiceren “nooit” kiezen, maar dat is niet de enige reden waarom men *nooit* boven *soms* zou selecteren. Inhoudelijk is de ordening van *nooit* naar *altijd* ook beter te onderbouwen.

ontbrekende waarden op gaat leveren. De gemiddelde score op de rest van de opgaven is bij het antwoord “weet niet” hoger dan bij het antwoord “Klasgenoten luisteren vaker” maar iets lager dan bij “Klasgenoten luisteren even vaak” en veel lager dan bij “Klasgenoten luisteren minder vaak”. Er is voor gekozen om de “weet niet”-categorie samen te voegen met de “even vaak” –categorie omdat er anders erg veel ontbrekende waarden bij deze vraag zouden zijn. Dit levert de variabele 17T<sup>46</sup> op.

### **Kwaliteit van de schalen in de leerlingvragenlijst<sup>47</sup>**

Op inhoudelijke gronden was er een aantal schalen te onderscheiden namelijk *Attitude*, *Linguistic Self Confidence (LSC)*<sup>48</sup>, *Buitenschools Gebruik (BsG)* en *Inschatte Vaardigheid (IV)*. De laatste was ook op te delen in vier subschalen: Lezen, Luisteren, Spreken en Schrijven. Deze schalen worden hieronder beschreven.

### **Schaal Attitude (leerlingvragenlijst)**

De opgaven die inhoudelijk bij deze schaal zouden horen zijn opgaven 1, 3, 4, 8 en 10, waarbij 10 opgedeeld is in zes subvragen (10.1 tot en met 10.6). Bij deze schaal werkten we met de 10.6R. Bij een factoranalyse<sup>49</sup> kwamen hier drie dimensies uit, waarvan één alleen item 8 betrof en één de items 10.4 en 10.5. De overige items zaten allen in een schaal. Bij de schaalanalyse bleek ook dat de  $R_{it}$ -waarde van item 10.4 negatief was, en die van 10.5 zeer laag ( $<.10$ ). Item 8 had de op twee na laagste  $R_{it}$ -waarde. De betrouwbaarheid van de schaal was met alle opgaven 0,60 (Gutman's Lambda-2). Als de drie afwijkende opgaven, die ook inhoudelijk afwijkend zijn, verwijderd werden steeg de betrouwbaarheid naar 0,69. De attitudeschaal zonder deze drie opgaven is de definitieve attitudeschaal. Alle zeven opgaven op de schaal hebben een  $R_{it}$ -waarde hoger dan 0,20 (COTAN-grens), met een gemiddelde  $R_{it}$ -waarde van 0,38.

De variabele die deze schaalscore weergeeft is *VL\_Attitude*. Deze variabele is samengesteld uit items 1, 3, 4, 10.1, 10.2, 10.3 en 10.6R en heeft een scorering van 0 tot en met 10. Deze scorering komt tot stand doordat de maximale score op items 1, 3 en 4 telkens 2 is en op de overige items 1. De schaal heeft een gemiddelde van 6,63 en een standaardafwijking van 1,95. Er zijn 2065 leerlingen voor wie deze score te berekenen is. Dat zijn de leerlingen die op geen van de samenstellende opgaven een ontbrekende waarde hebben.

### **Schaal Linguistic self confidence (leerlingvragenlijst)**

De opgaven die horen bij deze schaal zijn 2, 5, 6, 7 en 11. Een factoranalyse leverde een een-factor-oplossing op. De betrouwbaarheid (lambda-2) van de schaal is 0,71. De laagste (corrected)  $R_{it}$ -waarde is groter dan 0,30, en de  $R_{it}$ -waarden hebben een gemiddelde van 0,48. De schaalscore wordt gegeven als *VL\_LSC*. De scorering van deze variabele is van 0 tot en met 12. Deze scorering van de items 2, 5, 6 en 7 is '0-2' en van item 11 '0-4'. De schaal heeft een gemiddelde van 8,13 en een

---

<sup>46</sup> T voor trichotoom, omdat die variant nu drie scorecategorieën heeft. Er zijn ook alternatieven onderzocht, zoals “weet niet” als aparte categorie iets onder “even veel”, maar die functioneerde minder goed.

<sup>47</sup> De verdelingseigenschappen die hier gegeven worden zijn gewogen gemiddelden en standaarddeviaties waarbij rekening gehouden is met de overrepresentatie van VVTO-scholen en gewichtenstrata. Meer hierover is te vinden in hoofdstuk 2 waar op de verdelingseigenschappen wordt ingegaan.

<sup>48</sup> Wat onder Attitude en Linguistic Self Confidence wordt verstaan staat beschreven in paragraaf 6.5.2.

<sup>49</sup> Principale componenten op de correlatiematrix met varimax rotatie.

standaardafwijking van 2,50. Er zijn 2035 leerlingen uit groep 8 voor wie geen enkele ontbrekende waarde gevonden is op de vijf samenstellende items, en waarvoor deze variabele berekend is.

### *Schaal Buitenschools gebruik (leerlingvragenlijst)*

De opgaven die horen bij deze schaal zijn de vraag 12 tot en met 22, waarbij vraag 14 in totaal uit zes items bestaat. In de analyses bleek dat 17T duidelijk de laagste  $R_{it}$ -waarde had (0,131) die onder de beoogde 0,20 ligt. Ook in de factoranalyse bleek dit item opvallend. Deze opgave is verwijderd uit de schaal. De resulterende schaal heeft een betrouwbaarheid van 0,85 ( $\text{Lambda-2}$ )<sup>50</sup>. Op deze schaal zijn twee opgaven met  $R_{it}$ -waarden onder 0,40, te weten 14.4 en 14.5 met  $R_{it}$ -waarden 0,33 en 0,25. Deze opgaven konden in de schaal blijven. Het verwijderen van deze twee opgaven zou de schaal wel betrouwbaarder maken (0,86), en bij meer leerlingen is een score te berekenen<sup>51</sup>. Alle overige opgaven hebben een  $R_{it}$ -waarde boven 0,40, met een gemiddelde van 0,52. Dit is een behoorlijk homogene schaal.

Voor beide versies van de schaal is een variant aangemaakt, te weten VL\_BsG15 met 15 items, dus inclusief 14.4 en 14.5, en VL\_BsG13. In deze rapportage is gebruik gemaakt van VL\_BsG15. Deze is berekend bij 1956 leerlingen, heeft een gemiddelde van 18,96 en een standaardafwijking van 6,92. De scorering loopt van 0 tot en met 38. Deze scorering komt als volgt tot stand: de scorering van items 12, 13, 16-22 (minus 17) loopt van '0-2', de scorering van 14-1/6 loopt van '0-3' en van 15 loopt van '0-4'.

### *Schalen Ingeschatte vaardigheid: Lezen, Luisteren, Spreken en Schrijven (leerlingvragenlijst)*

Vraag 28 en 29 vragen naar de ingeschatte vaardigheid (IV) van de leerlingen. Vraag 28 bestaat uit 9 items waarbij de leerling bij 4 leesactiviteiten en 5 luisteractiviteiten aangeeft wat hij of zij meent te kunnen. Vraag 29 doet dat voor spreken (5 activiteiten) en schrijven (5 activiteiten).

Er is een exploratieve factoranalyse<sup>52</sup> uitgevoerd. Deze leverde niet een vier maar een drie-factoroplossing, waarbij twee opgaven minder goed pasten, namelijk spreekopgave 29.3 ("Als mensen langzaam spreken en dingen herhalen, kan ik deelnemen aan een gesprek in het Engels") en schrijfopgave 29.9 ("Ik kan mijn naam, adres en nationaliteit invullen op een formulier"). De eerste van deze twee is meer een 'ontvouwingsopgave': mensen die heel slecht of heel goed zijn in Engels zullen hier negatief op kunnen reageren. Bij de tweede vraag kan men bediscussiëren of deze vraag werkelijk over Engels schrijven gaat. De eerste factor bevatte de helft van de leesopgaven, de helft van de overgebleven spreekopgaven, en alle overgebleven schrijfopgaven. De tweede factor bevatte de andere helft van de lees- en spreekopgaven en twee van de luisteropgaven. De laatste factor bevatte drie luisteropgaven<sup>53</sup>.

De schalen zijn dus niet evident per vaardigheid geordend. Als direct naar de correlatiematrix gekeken wordt is ook te zien dat de correlaties tussen opgaven die over dezelfde vaardigheid gaan gemiddeld niet hoger onderling correleren dan opgaven die over verschillende vaardigheden gaan. Om inhoudelijke redenen zijn echter wel de vier IV-schalen per vaardigheid geformeerd. Bij deze

<sup>50</sup> De gevonden betrouwbaarheid is een fractie hoger dan met opgave 17T erbij. Andere varianten van opgave 17 leveren geen betere resultaten op.

<sup>51</sup> Wanneer leerlingen op een van de opgaven geen antwoord gegeven hebben, kan er geen score gegeven worden. Er wordt niet met geïmputeerde waarden gewerkt.

<sup>52</sup> Principale componenten op de correlatiematrix met varimax rotatie; eigenwaarden groter dan 1.

<sup>53</sup> De factoren: Factor 1: LE(28.01 & 03), SP(29.01 & 05) en SC(29.06-08 & 10); Factor 2: LE(28.02 & 04), LU(28.05 & 07) en SP(29.02&04); Factor 3: LU(28.06&08-09); de twee afgevalen opgaven passen het best bij Factor 3.

schalen viel opgave 29.9 weer op door een lagere  $R_{it}$ -waarde, die als enige onder 0,20 lag. Deze opgave is dan ook uit deze schaal verwijderd –hetgeen ook nog een hogere betrouwbaarheid van de schaal opleverde – aangezien er nu, naast de inhoudelijke evaluatie, twee indicaties zijn dat deze vraag niet functioneert zoals bedoeld. Opgave 29.3 blijft wel in de schaal aangezien deze een redelijke  $R_{it}$ -waarde heeft en het verwijderen de schaal ook minder betrouwbaar zou maken. De overige kenmerken van de vier schalen worden gegeven in Tabel 5.9.

Tabel 5.9: Kenmerken van de vier schalen Ingeschatte vaardigheden

Vaardigheidsschaal		Aantal			Betr.b.h*	Scores (gewogen)			
Label		Items	Observaties	Ontbrekend	Lambda-2	Range	P'	Gemiddelde	St.Afw.
VL_IV_LE	IV Lezen	4	2029	46	0.55	0 – 4	76	3.05	1.03
VL_IV_LE	IV Luisteren	5	2011	64	0.57	0 – 5	74	3.70	1.25
VL_IV_SP	IV Spreken	5	1986	89	0.59	0 – 5	66	3.30	1.19
VL_IV_SC	IV Schrijven	4	1991	84	0.61	0 – 4	55	2.18	1.47

\* betrouwbaarheid ; bij ontbrekend zijn niet de 13 leerlingen meegerekend bij wie 100% ontbreekt

De onderlinge samenhang tussen de vier schalen is hoog te noemen, zeker gerelateerd aan de geschatte betrouwbaarheden van de schalen. Deze correlaties zijn gegeven in Tabel 5.10. Er zijn maar twee correlaties die na attenuatiecorrectie niet gelijk aan 1 zijn en dat zijn die tussen Lezen en Schrijven (0,93) en Luisteren en Schrijven (0,82).

Tabel 5.10: Samenhang van de 4 schalen Ingeschatte vaardigheden

Pearson correlaties	VL_IV_LE	VL_IV_LU	VL_IV_SP	VL_IV_SC
VL_IV_LE		0.63	0.62	0.54
VL_IV_LU	0.63		0.60	0.48
VL_IV_SP	0.62	0.60		0.60
VL_IV_SC	0.54	0.48	0.60	

De hoge onderlinge (latente) correlaties en de resultaten van de factoranalyses geven aan dat naast de subschalen er ook een algemene schaal Ingeschatte vaardigheid gemaakt kan worden. Deze bestaat uit de som van de 4 subschalen, en is dus exclusief item 29.9. De betrouwbaarheid van die schaal is 0,86, en op een opgave na<sup>54</sup> hebben alle opgaven een  $R_{it}$ -waarde van boven de 0,30, met een gemiddelde  $R_{it}$ -waarde van 0,45. Deze schaal score is te berekenen voor 1913 leerlingen, heeft een scorereange van 0 tot en met 18, een (gewogen) gemiddelde 12,24 van en een standaardafwijking van 4,05

### *Samenhang tussen de schalen (leerlingvragenlijst)*

Tot slot is gekeken naar de samenhang tussen de vier belangrijkste schalen die uit de vragenlijst komen. Dat zijn Attitude, Linguistic Self Confidence, Buitenschools gebruik<sup>55</sup> en Ingeschatte vaardigheid. In Tabel 5.11 zijn de eigenschappen gegeven, inclusief de correlaties tussen de schalen. Een overzicht van de opgaven in deze schalen is gegeven in Bijlage 9.

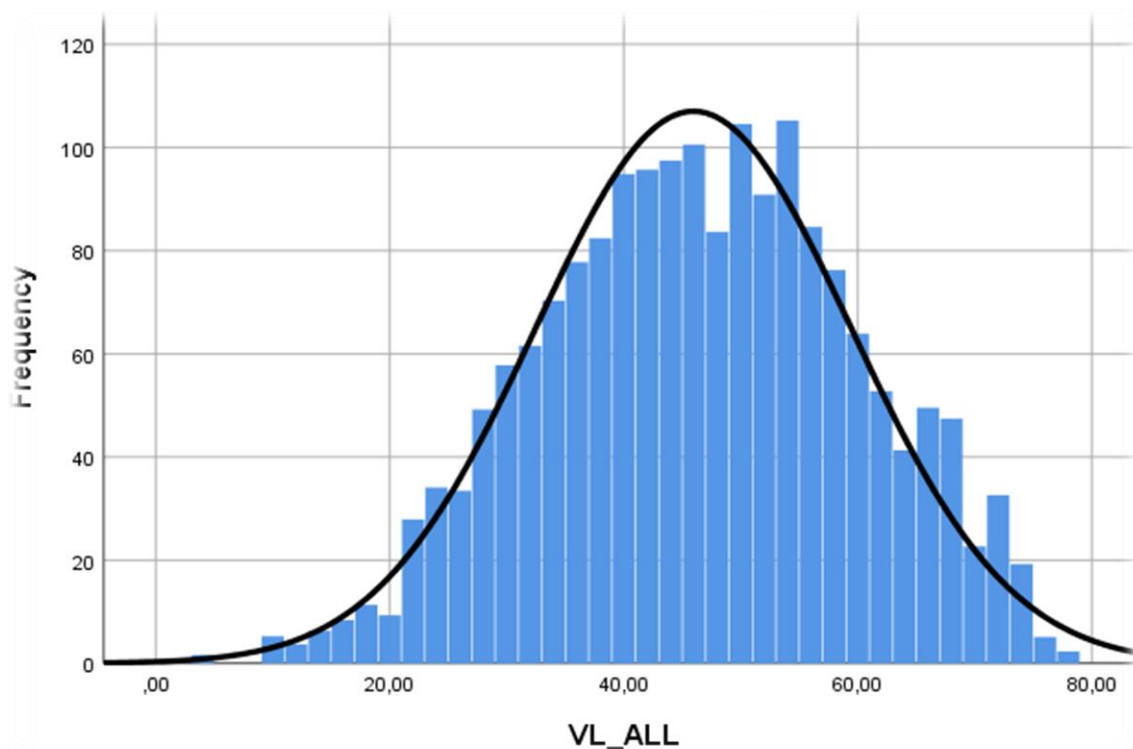
<sup>54</sup> Items 29.4: Spreken: "Ik kan eenvoudige vragen stellen en korte antwoorden geven" met een  $R_{it}$ -waarde van 0,29. Ook deze opgave zou gezien kunnen worden als een ontvouwingsopgave.

<sup>55</sup> Variant met 15 opgaven

Tabel 5.11: Kenmerken van de Vragenlijstschalen

aantal		Scores				Correlaties			
Schaal	items	Max	Gem.	Standaard afwijking	Lambda-2	Attitude	LSC	BsG15	IV_ALL
VL_Attitude	7	10	6.63	1.95	0.6		0.58	0.55	0.56
VL_LSC	5	12	8.13	2.50	0.71	0.58		0.67	0.73
VL_BsG15	15	38	18.96	6.92	0.85	0.55	0.67		0.68
VL_IV_ALL	18	18	12.24	4.05	0.86	0.56	0.73	0.68	

De correlaties tussen de schalen is redelijk en liggen na attenuatiecorrectie tussen 0,77 en 0,93, met een gemiddelde van 0,84. Het is daarmee duidelijk dat de schalen niet exact hetzelfde meten, maar wel sterk samenhangen. Het is ook mogelijk een enkele samengestelde vragenlijst score te maken die een positieve houding en ervaring met het Engels weergeeft. De schaal heeft een betrouwbaarheid van 0,93 en alle opgaven hebben een  $R_{it}$ -waarde boven 0,20 (gemiddelde  $r_{it}$ : 0,45).



Figuur 5.1 Gevonden score verdeling voor de “vragenlijst-schaal”

Deze schaal bestaat uit 45 opgaven met een theoretische scorering van 0 tot en met 78. De score van de variabele VL\_ALL is uit te rekenen voor 1799 personen, waarbij de laagst gevonden waarde 4 is<sup>56</sup> en de hoogst gevonden waarde 77. De gemiddelde score is 45,95, met een standaardafwijking van 13,50. De gevonden score verdeling is gegeven in Figuur 5.1.

<sup>56</sup> De minimum score is een *outlier*: de een-na-laagste waarde is 10

### Schoolvragenlijst

De basis voor de schoolvragenlijst lag in de korte vragenlijst die bij de vorige peiling is afgenomen bij de scholen (Geurts & Hemker, 2013). De vragenlijst bevatte zeven vragen, waarvan drie vragen sub-items bevatten. De directie is gevraagd naar: a) de aard van het aanbod (regulier EIBO met Engels in jaargroep 7 en 8, vervroegd EIBO met Engels vanaf jaargroep 5 of 6, of vroeg vreemdetalenonderwijs met Engels vanaf jaargroep 1 (VVTO)), b) het startmoment Engels voor de huidige groep 8, c) de geprogrammeerde onderwijstijd in elk van de groepen waarin Engels wordt gegeven, d) de aanwezigheid van een coördinator of werkgroep voor Engels, e) de faciliteiten die de school voor Engels biedt aan leerkrachten, f) een inschatting van de (talige) thuissituatie van de leerlingen op school en g) de overdracht van leerlinggegevens aan het voortgezet onderwijs. Met deze vragen wilden we inventariseren op welke manier scholen het onderwijs Engels op schoolniveau hadden georganiseerd en vormgegeven. Vanwege de inventariserende aard van deze vragen zijn er geen nadere analyses voor kwaliteit en schaling uitgevoerd.

### Leerkrachtvragenlijst

Evenals de schoolvragenlijst was de leerkrachtvragenlijst grotendeels gebaseerd op de vragenlijst uit de vorige peiling (Geurts & Hemker, 2013). Over de achtergrondkenmerken van de leerkracht, het aanbod voor Engelse taalvaardigheid en -didactiek in de genoten opleiding en het eigen gebruik van het Engels is aanvullende informatie bevraagd. Met de meeste vragen werd beoogd een beschrijving te kunnen geven van achtergrondgegevens van de leerkrachten en met name van de manier waarop zijn het onderwijsleerproces bij Engels vormgegeven. Voorbeelden zijn het aantal jaren leservaring of ervaring met het geven van Engels, gebruikte activiteiten in lessen Engels, gebruikte doelen of de gebruikte methode. Deze vragen waren inventariserend van aard en de uitkomsten worden in de paragrafen 6.2-6.4 beschreven.

De vragenlijst is uitgezet bij alle leerkrachten in de groepen 8 van de deelnemende scholen. In totaal bevatte het bestand 112 ingevulde vragenlijsten, waarvan 1 vragenlijst volledig blanco bleek. Bij inspectie van de ingevulde gegevens bleek verder dat één leerkracht alleen les gaf aan groep 6-7 en dat één leerkracht expliciet aangaf geen rol te spelen in de verzorging van Engels in de klas. Het netto aantal leerkrachten met bruikbare gegevens bedraagt derhalve 109. Tien van deze leerkrachten had meer dan 10% aan missende waarden (9.17%). De leerkrachten zijn werkzaam op 90 verschillende scholen, hetgeen betekent dat we van vijf scholen geen leerkrachtgegevens hebben. Op 15 scholen hebben twee leerkrachten de vragenlijst ingevuld, op twee scholen drie leerkrachten. Voor de analyses aangaande de leerkrachtvragenlijst zijn de antwoorden van alle 109 leerkrachten meegenomen, behalve bij de vraag naar de gebruikte methode. Hier is gekozen voor analyse op schoolniveau (89 scholen).

Voorafgaand aan de analyses zijn, waar nodig, de scores aangepast, zodat alle vragen indicatief zijn voor een positieve attitude dan wel meer (buitenschoolse) ervaring met het Engels. Die hercodering zorgt ervoor dat de hoogste score de meest positieve attitude dan wel meeste ervaring met Engels weergeeft, en de laagste score de minst positieve attitude dan wel de minste ervaring.

De leerkrachtvragenlijst kende twee vragen waarvoor schalen zijn opgesteld: buitenschools gebruik Engels en de inschatting van het eigen Engelse vaardigheidsniveau. De kwaliteit en schaling wordt hieronder beschreven.

### *Schaal buitenschools gebruik Engels (leerkrachtvragenlijst)*

De vragenlijst bevatte zeven vragen over de mate waarin de leerkracht het Engels buitenschools gebruikt. De leerkracht diende telkens één antwoord te kiezen op een schaal van (bijna) nooit (1) tot en met dagelijks (5). Een factoranalyse leverde een een-factor oplossing op. De schaalscore<sup>57</sup> bedraagt het gemiddelde buitenschoolse gebruik over de 7 vragen, de score range loopt derhalve van 1 tot en met 5. De betrouwbaarheid is goed te noemen: Cronbach's  $\alpha = .79$

### *Schaal eigen inschatting vaardigheid Engels (leerkrachtvragenlijst)*

Een inschatting van de eigen vaardigheid in termen van de ERK-niveaus A2, B1, B2 en C2 is bevestigd voor vier vaardigheidsdomeinen: 1) lezen, 2) luisteren, 3) schrijven en 4) spreken. Per domein kregen leerkrachten vier beschrijvingen voor de vier genoemde ERK-niveaus voorgelegd en konden zij aankruisen of ze het niveau nog niet (score '0') of juist wel beheersten (score '1'). Om vergelijking met uitkomsten uit de vorige peiling (Geurts en Hemker, 2013) mogelijk te maken is de vraagstelling exact hetzelfde gehouden. In tegenstelling tot de vorige peiling, waarin de uitkomsten enkel beschrijvend zijn gepresenteerd, is nagegaan of de items geschaald konden worden<sup>58</sup>. Aangezien er sprake was van oplopende moeilijkheid in de beschrijvingen is er gebruik gemaakt van Mokken-analyses om te komen tot cumulatieve schalen (Sijtsma & Molenaar, 2002). Cumulatieve schalen zijn gemaakt voor respectievelijk de receptieve (lezen, luisteren) en productieve schalen (spreken, schrijven). De Mokkenschaalanalyses laten zien dat er voor alle vier de schalen sprake is van sterke schalen ( $H=0.69$  voor spreken,  $H=0.75$  voor schrijven,  $H=0.82$  voor luisteren en  $H=1.00$  voor lezen; Sijtsma & Molenaar, 2002). De schaling kent echter wel een probleem: bij lezen en luisteren was er nauwelijks variantie in de items op A2 en B1-niveau (bijna iedereen gaf aan de beschrijving te beheersen). Dit betekent dat de schalen eigenlijk slechts op twee items gebaseerd zijn (B2 en C1 niveau), die wel oplopend zijn. De schalen voor spreken en schrijven geven wel weer dat de vier descriptorren oplopend in moeilijkheid zijn. De schaalscore per vaardigheid is gebaseerd op de somscore van de vier beschrijvingen – dus ook voor de schalen luisteren en lezen.

### *Geconstrueerde variabelen voor de meerniveau-analyses*

Ten behoeve van de meerniveau-analyses zijn enkele vragen in de leerkrachtvragenlijst aangepast. In Bijlage 11 wordt een overzicht van de samenstelling van de variabelen gegeven die in de meerniveau-analyses zijn meegenomen. Naast variabelen op leerkracht- en leerkrachtniveau (waarvan de informatie afkomstig is uit de leerkrachtvragenlijst) worden tevens variabelen op leerlingniveau weergegeven.

## **5.4 Constructie van de vaardigheidsschalen**

Voor de getoetste vaardigheden zijn vaardigheidsschalen geconstrueerd. Bij de constructie van deze vaardigheidsschalen is er onderscheid gemaakt tussen drie typen schalen. Het eerste type zijn vaardigheidsschalen behorend bij de vaardigheden die gemeten zijn met toetsen. Aangezien niet alle leerlingen alle opgaven hebben gemaakt wordt bij deze schalen gewerkt met Item Response Theorie (IRT). De tweede type schalen zijn de mondelinge taalvaardigheden die gemeten zijn met praktische opdrachten (de gesprekstaken). De kandidaten die deze taken hebben uitgevoerd konden in principe

<sup>57</sup> Principale componentenmatrix op de correlatiematrix; 1 component.

<sup>58</sup> In paragraaf 6.2 staan zowel de schaalscores als de frequentieverdeling op de losse items beschreven.



alle opgaven maken. Hier is begonnen met een totaalscoreschaal per schaal, maar deze zijn ook omgezet naar een latente vaardigheid voor gesprekken. Het derde type schaal of schalen zijn die schaal/schalen waarbij meerdere vaardigheden samen genomen worden. Dat was ten eerste een schaal waarbij is geprobeerd een score te leveren voor woordenschat, lezen en luisteren, en een schaal voor de mondelinge taalvaardigheden. Of deze schalen mogelijk waren hing ook van data af. Dat was helemaal het geval wat betreft het samenvoegen van de met toetsen gemeten vaardigheden en de praktische mondelinge taalvaardigheden. In de volgende paragrafen wordt beschreven welke stappen er in de analyses genomen zijn. De scores op de vaardigheidsschalen worden geleverd in termen van Weighted Maximum Likelihood (WML)-vaardigheden. Dit betreft de verdeling na weging op de acht combinatie-strata, om de overrepresentatie in de steekproef van de hogere gewicht-strata en VVTO-scholen te compenseren. Hierdoor betreft dit de populatieverdeling.

Naast de vaardigheidsschattingen zijn de gegevens ook vertaald naar ERK-niveaus. Hoe dit gedaan is, is beschreven in het hoofdstuk over standaardbepaling (hoofdstuk 7). Hoewel voor woordenschat geen ERK-niveaus bestaan, is in hoofdstuk 7 ook een alternatief beschreven voor de vraag hoe de vaardigheid voor woordenschat meer betekenis kan worden gegeven.

#### **5.4.1 Schalen op basis van de toetsen binnen de peiling 2018**

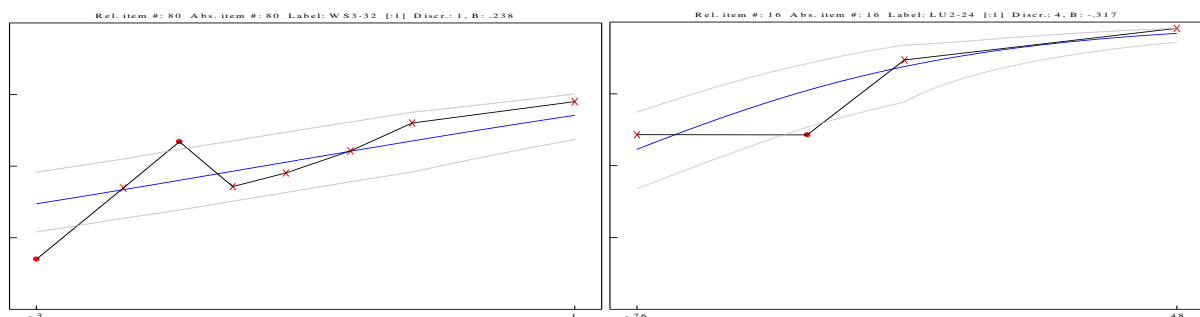
Bij de schalen voor woordenschat, leesvaardigheid, en luistervaardigheid zijn de gegevens verzameld met een incompleet design. Om deze reden zijn IRT-methoden gebruikt om de scores op de varianten van de toets op één gezamenlijke schaal te brengen. Bij alle drie de vaardigheden ging het om gegevens van zes toetsen, waarvan opgaven altijd in twee verschillende toetsen voorkwamen. Hoeveel items in een itemblok zaten verschilde per vaardigheid. Bij lezen was dat acht of negen items per itemblok, bij luisteren acht items en bij woordenschat 18 of 19 items. Hierdoor bevatten de toetsen respectievelijk 16/17, 16 en 36/37 items. Meer over de toetsdesigns is beschreven in hoofdstuk 3.

De leerlingsscores op de afzonderlijke items zijn geanalyseerd met behulp van het One Parameter Logistic Model (OPLM; Verhelst en Glas, 1995). In dit model wordt de kans dat een bepaalde leerling een bepaald item goed beantwoordt, uitgedrukt als functie van de latente (“ware”) vaardigheid van de leerling enerzijds, en de moeilijkheid en het onderscheidende vermogen van het item anderzijds. De interpretatie van het OPLM is vergelijkbaar met die van het Two Parameter Logistic Model (2PLM), omdat beide modellen rekening houden met variatie in de moeilijkheid én het onderscheidende vermogen van verschillende items (in tegenstelling tot het eenvoudigere Rasch-model, dat alleen rekening houdt met de moeilijkheid). Een praktisch verschil tussen het OPLM en het 2PLM is dat in het OPLM de discriminatieparameters van de items a-priori als constanten worden ingevoerd. Door deze keuze kan de (gewogen) ruwe score van een leerling op de toets direct uit de data berekend worden, zodat het mogelijk wordt om de itemparameters te schatten met conditional maximum likelihood (CML).

De mogelijkheid om CML-schatting te gebruiken is een belangrijk voordeel van het OPLM in het huidige onderzoek, omdat CML-schatting van de itemparameters niet gepaard gaat met assumpties over de vaardigheidsverdeling in de populatie, over de wijze van steekproeftrekking of over de toewijzing van toetsvarianten in een incompleet design (Eggen, 2004). Het heeft zodoende géén nadelige invloed op CML-schattingen van de itemmoeilijkheden (het schalen van de opgaven) dat de steekproef waarmee we werken niet representatief is. Een ander voordeel is dat de modelpassingsmaten binnen CML beter ontwikkeld zijn.

Voor de modelschatting gebruiken we het computerprogramma OPLM (Verhelst, Glas & Verstralen, 1995), dat tevens een aantal statistische toetsen uitvoert op grond waarvan we kunnen bepalen of het model een adequate beschrijving geeft van de data. Belangrijk zijn de zogenaamde itemgeoriënteerde *S*-toets en de overall *R1c*-toets. De *S*-toets is gebaseerd op de verschillen tussen de geobserveerde en verwachte proporties antwoorden in homogene scoregroepen. Een uniforme verdeling van *p*-waarden voor de *S*-toetsen in het interval [0,1] pleit voor adequate passing van het model (zie Verhelst, Glas & Verstralen, 1995).

Het evalueren van de passing kan onder andere visueel gebeuren. Daartoe worden de geschatte item-response functies (IRF) vergeleken met de geobserveerde *p*-waarden op een aantal punten op de schaal. In Figuur 5.2 is een geschatte IRF weergegeven als een vloeiende blauwe lijn, en zijn de *p*-waarden weergegeven als een rood kruis, of een rode bol. De weergave is een rood kruis als de *p*-waarde binnen het betrouwbaarheidsinterval (weergegeven door de grijze vloeiende lijnen) van de schatte IRF ligt; een rode bol geeft aan dat deze buiten het betrouwbaarheidsinterval ligt. Zoals gesteld kunnen de afwijkingen van geschatte IRF ook weergegeven worden door middel van de *S*-*statistic* die per opgave gegeven kan worden. Volgens de *S*-*statistics* is opgave 80 van de Woordenschaatschaal (WS3-32) het slechts passende item. Visuele inspectie laat zien dat aan de onderkant van de schaal de passing inderdaad niet geheel correct is, maar het model kan dergelijke schendingen wel aan: het model is robuust genoeg. Het een-na slechts passende item is het 16<sup>e</sup> item van de Luisteren-schaal (LU2-24). Bij visuele inspectie is duidelijk dat de afwijking daar nog minder erg is. Beide opgaven zijn afgebeeld in Figuur 5.2.



Figuur 5.2: Passing van items WS3-32 (links) en LU2-24 (rechts)

Voor alle drie de vaardigheden is te zien dat over de hele schaal heen de gemiddelde *a*-parameter gelijk aan 3 is en de gemiddelde *b*-parameter 0, maar deze waarden zijn het gevolg van gemaakte keuzen. Bij het bepalen van de *a*-parameter is voor dit gemiddelde gekozen en aangezien de *b*-parameters vrij geschat konden worden is het gemiddelde hier 0. De *a*-parameters variëren van 1 tot en met 5 bij alle schalen: er zijn dus geen zeer extreme waarden gevonden. Het percentage items met een *a*-parameter van 1 (vlakke, weinig onderscheidende item response functies) ligt op 5%, met weinig verschillen over de vaardigheden en geen van de *a*-parameters met een negatieve  $\text{Gam}_i$ , wat indicatief zou zijn voor een dalende item response functie. In het geval van peilingen hoeven dergelijke opgaven niet verwijderd te worden omdat het juist interessant kan zijn bij welke opgaven vaardige en minder vaardige leerlingen weinig van elkaar verschillen.

Het gemiddelde aantal observaties per opgave bij alle drie de vaardigheden ligt boven de 600. Bij een opgave is het aantal observaties 383 (WS1-14), maar bij alle ander opgaven ligt dat boven de 450. De gemiddelde SE(B) ligt onder de 0,04 en de gemiddelde bijdrage aan de *R1c* ligt op 1,15.

Kijken we naar de passing van de gehele schaal dan zien we dat de verdeling van de  $S_i$ -statistics (Tabel 5.12) in grote lijnen een uniforme verdeling van p-waarden voor de S-toetsen oplevert in het interval [0,1]. Dit pleit voor adequate passing van het model. De resultaten voor de R1c-statistics zijn iets minder fraai. Bij Luisteren is de p-waarde van de R1c toets 0,0008, wat nog steeds een significante afwijking betekent. Deze afwijking is echter niet dramatisch.

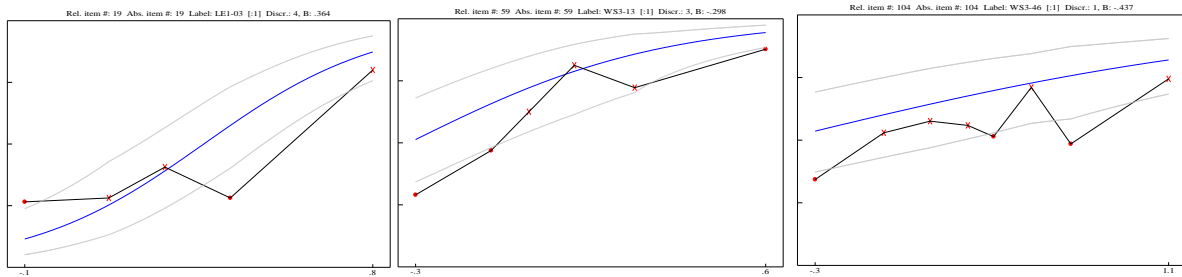
Tabel 5.12: verdeling van p-waarden voor de S-toetsen

	0,01	0,05	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00
LE (N items=51)	0	3	2	5	5	6	7	6	4	5	5	3
LU (N items=48)	0	4	4	7	6	2	3	6	6	4	3	3
WS (N items=110)	1	7	4	11	13	15	12	13	10	5	9	10
LE	0%	6%	10%	20%	29%	41%	55%	67%	75%	84%	94%	100%
LU	0%	8%	17%	31%	44%	48%	54%	67%	79%	88%	94%	100%
WS	1%	7%	11%	21%	33%	46%	57%	69%	78%	83%	91%	100%

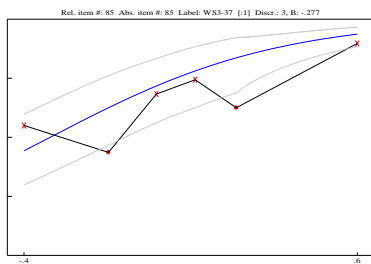
Om de stelling te onderbouwen dat het model robuust genoeg is en om na te gaan of er, ondanks kleinere schendingen van het model, zonder problemen met het model kan worden gewerkt is er een controle uitgevoerd. Voor alle schalen en items is een inschatting gemaakt hoe goed de geobserveerde waarden – zoals de p-waarden of de  $R_{it}$ -waarden, of de scoreverdeling bij de verschillende toetsen – teruggeschat kunnen worden met het model. Er is geen standaardrichtlijn om hier te stellen dat het model robuust genoeg is, maar in de beschrijving van de resultaten kon wel een kwalitatieve analyse gemaakt worden of het model robuust genoeg is om te kunnen gebruiken. Deze leverde goede resultaten. Op toets-niveau is het grootst gevonden verschil in gemiddelde scores 0,1 bij Woordenschat-toets 4 (geschat 22,8 versus geobserveerd 22,7 bij 37 opgaven). De standaardafwijkingen worden iets overschat (2%) en de betrouwbaarheden daarmee ook iets, zij het dat dit onder de 0,01 blijft. Op itemniveau zijn de afwijkingen ook niet heel erg groot: in absolute waarde zijn de gemiddelde p-waarde-verschillen kleiner dan 0,02 (mediaan 0,016).

Een ander onderzoek betreft het onderzoek naar vraagpartijdigheid (Differential Item Functioning; afgekort als DIF). In dergelijk onderzoek wordt onderzocht of de geschatte parameters ook voor verschillende subgroepen van de populatie geldig zijn. De afstanden tussen de opgaven op de latente schaal zouden namelijk voor iedere groep hetzelfde moeten zijn. Hier is onderzoek gedaan naar de tweedeling op basis van de scholen (EIBO – VVTO) en op basis van geslacht (jongen – meisjes). Als we hierbij kijken naar de passingsmaten zien we dat er weinig verschil is tussen de passing (met gefixeerde parameters) als alle leerlingen samengenomen worden en als deze per subgroep geanalyseerd worden. Enige uitzondering is hier woordenschat waarbij we een iets slechtere passing hebben bij de opdeling naar geslacht. Dit is overigens een niet ongebruikelijk resultaat: er zijn woorden die jongens meer aanspreken en woorden die meisjes meer aanspreken. Het verwijderen van opgaven met DIF op basis van geslacht verarmt de meting van het concept.

De slechtst passende opgaven zijn bij de data met alleen de jongens de onderstaande opgaven in Figuur 5.3a. De slechts passende opgave bij data met alleen meisjes is gegeven in Figuur 5.3b. Dit zijn de opgaven die een sterk significant afwijkende s-statistic hebben ( $p=.000$ ). Als we visueel deze opgaven bekijken, valt het gebrek aan passing mee. Op basis van de p-waarden zien we verschillen van maximaal 0,20 punten (0,57 jongens, 0,77 bij meisjes bij het woord “shelf” – WS3-46).



Figuur 5.3a: Opgaven alleen bij jongens met de slechtste passing (LE1-03; WS3-13; WS3-46)



Figuur 5.3b: Opgaven alleen bij meisjes met de slechtste passing (WS3-37)

Al met al lijkt het erop dat ook bij woordenschat het gebrek aan passing meevalt en dat er uitgegaan kan worden van één schaal voor lezen, luisteren en woordenschat.

#### 5.4.2 Schalen op basis van de gesprekstaken binnen de peiling 2018

De gespreksvaardigheden zijn gemeten met behulp van praktische opdrachten. Een aselect gekozen deel van de leerlingen maakte deze praktische opdrachten. Hoewel de gesprekken in paren zijn afgenomen kregen de leerlingen individuele vaardigheidsscores. Hiervoor is eerst de vaardigheidsschaal gemaakt.

De vaardigheidsschaal is ook gemaakt met behulp van OPLM. In plaats van CML-schattingen, zoals bij de toetsen, is hierbij MML gebruikt. Dat is gedaan omdat het design bij deze vaardigheden anders was dan het design bij de toetsen. Bij de toetsen lag vooraf vast welke opgaven de leerling zeker niet ging maken. Bij de gesprekstaken was dat niet van tevoren duidelijk, omdat dit mede afhing van de resultaten tijdens de afname. De taken hadden voor een deel het karakter van een multi-stage toets, namelijk dat de prestatie op de vorige toets invloed had op de taak die hierna voorgelegd zou worden. Als de leerling te zwak presteerde, kreeg deze de volgende toets niet aangeboden. Een verschil met een standaard multi-stage toets is dat dit besluit niet strikt afhing van de score op de eerdere toets. Bij een multi-stage toets is het gebruikelijk dat er per toets of fase (*stage*) scoregrenzen bepaald zijn waarmee het toetsverloop bepaald wordt. Deze scoregrenzen bepalen dus of de leerling een nieuwe toets krijgt en zo ja welke. In ons geval waren die grenzen er niet. De enige beslissing die hier genomen moest worden was of de volgende toets afgenomen zou worden of niet en deze beslissing werd genomen op basis van de inschatting van de toetsleider en de tijd die er nog beschikbaar was. Als het een zuivere multi-stage toets zou zijn, zou een MML-schatting strikt noodzakelijk zijn. MML is wel de schattingsmethode die hier toegepast is, maar er is ook een controle gedaan met CML-schattingen, vanwege de eerder beschreven voordelen. Het bleek dat de resultaten amper van elkaar verschilden.

In de OPLM analyse zijn er zeven verschillende combinaties van toetsen geanalyseerd. Ten eerste betrof dat de set leerlingen die alleen Taak 1 gedaan hebben, ten tweede de twee varianten (Roy/Rosa, en Leo/Lisa) van Taak 2 samen met Taak 1, en tot slot de vier mogelijke combinaties van Taak 2 en Taak 3 (Voor/Tegen). Als we kijken naar de passingsmaten dan zien die er minder fraai uit dan bij de toetsen (Tabel 5.13). We verwachten ongeveer 2 of 3 (1% van 240) van de s-statistics van 0,01 of lager, maar dat zijn er in dit geval 13 (5%); bij 0,05 en 0,10 is dat verwacht bij 12 (5% van 240 opgaven) en 24 opgaven (10% van 240 opgaven), maar gevonden bij 32 en 43 opgaven. Er zijn dus meer slecht passende opgaven dan verwacht.

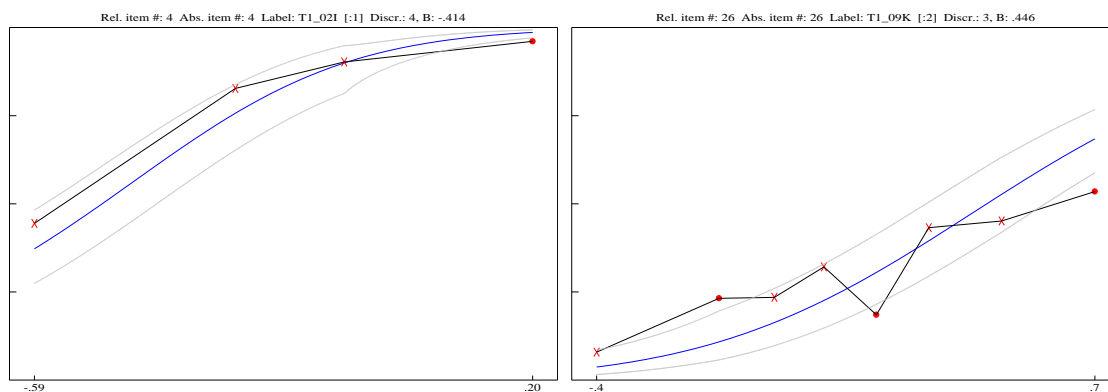
Tabel 5.13: verdeling van p-waarden voor de S-toetsen

	0,01	0,05	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00
Gesprekken (N items=240)	13	19	11	17	20	22	18	18	12	29	17	44
Gesprekken	5%	13%	18%	25%	33%	43%	50%	58%	63%	75%	82%	100%

Hier zijn enkele opmerkingen op zijn plaats. Ten eerste zijn de in Tabel 5.13 gepresenteerde passingsmaten niet heel zuiver omdat dit de maten zijn, nadat de parameters gefixeerd zijn. De CML-passing is nu gedaan op de met MML-geschatte waarden. Het betreft hier dus niet een echte, onafhankelijke toetsing<sup>59</sup>. Een ander punt is dat het hier polytome opgaven betreft, waarbij de parameters voor twee itemstappen<sup>60</sup> binnen een opgave geschat worden. De a-parameter wordt daarbij gelijk verondersteld bij beide stappen binnen een opgave. Daarin lijkt dit model binnen opgaven enigszins op een 1-parameter model. Dat blijkt bij sommige opgaven een wat strenge eis, waardoor bij polytome opgaven de modellen vaak minder goed passen. Om een indruk te geven hoe erg het gebrek aan passing is, worden de figuren getoond van de twee itemstappen met de slechtste passing, te weten de eerste stap (van score 0 naar score 1) van item 4, en de tweede stap (van score 1 naar score 2) van item 26 (Figuur 5.4). Bij item 4 (linker plaatje) is de eerste stap wat stijler dan de tweede stap wat voor een minder passing zorgt. Bij item 26 (rechter plaatje) is de vorm van de geobserveerde item-stap-response-functie wat vlakker dan de eerste, en deze is ook minder monotoon. Robuustheidsonderzoek laat ook hier zien dat de impact van de niet perfecte passing niet groot is; deze komt nergens boven 1% per scorepunt uit (maximum verschil bij boekje 6: taak 1+taak2'L/L'+taak3'Voor': 0,007% per scorepunt).

<sup>59</sup> Alternatief zou zijn de dataset in tweeën op te delen en de ene helft te gebruiken voor de schatting en de tweede voor de toetsing, maar deze dataset is wat klein en daardoor zouden bij een aantal items de parameterschattingen voor de toetsing gebaseerd zijn op heel weinig observaties.

<sup>60</sup> Er zijn verschillende definities van itemstappen mogelijk; de definitie die hier gebruikt is, is de Partial Credit (of, Adjacent Categories) definitie; zie Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.



Figuur 5.4: Twee itemstappen met minste passing bij de schaal voor gesprekstaken

### 5.4.3 Het maken van vaardigheidsscores

Nadat de vaardigheidsschaal gemaakt is, is het mogelijk om de schattingen van de vaardigheden van de leerlingen te verkrijgen. Aangezien we hier met gewogen scores werken, is dat gedaan met WML-schattingen. Deze zijn berekend met het programma OPLAT, waarbij de variant met data gebruikt is.

### 5.4.4 Het combineren van vaardigheidsscores

Voor vervolgonderzoek moest op basis van de vaardigheden een selectie van de beste scholen per stratum worden gemaakt. Wanneer deze selectie op basis van de meer verfijnde achtdeling zou worden gemaakt, zou dit een beperking hebben opgeleverd, omdat de kans dan zeer reëel was dat in een relatief slecht presterend stratum de beste scholen duidelijk nog slechter zouden kunnen zijn dan redelijk presterende scholen in een goed presterend stratum. Gekozen is daarom voor de CC4-definitie van stratum waarbij de gewichtenstrata gedichotomiseerd zijn. De S4o1 en S4o2 zijn samen genomen, hetgeen een categorie opleverde van scholen met 10% of minder gewichten-leerlingen, en de categorieën S4o3 en S4o4 zijn samengevoegd tot de categorie van scholen met meer dan 10% gewichtenleerlingen. De opdeling VVTO tegenover Eibo blijft bestaan. Dit leverde in combinatie met de samengevoegde gewichtenstrata vier groepen op (Eibo, 0- 10% gewichtenleerlingen; Eibo, >10%; VVTO 0-10% en VVTO >10% gewichtenleerlingen).

De gecombineerde score moest worden gemaakt voor de klassikale taken. Om de scores samen te kunnen voegen moesten deze klassikale taken samenhangen. Daartoe is de correlatie genomen tussen de verschillende vaardigheidsschattingen van de leerlingen<sup>61</sup>. Deze correlaties zijn gegeven in Tabel 5.14. De correlaties zijn redelijk hoog, en hoog genoeg om voor de selectie van de scholen een samengestelde score te kunnen maken. Deze samengestelde score heeft alleen betekenis voor de selectie van scholen voor het vervolgonderzoek.

5.14: correlaties tussen de vaardigheidsscores

	LE	LU	WS	GESPREK
LE		0.64	0.76	0.65
LU	0.64		0.64	0.56
WS	0.76	0.64		0.75
GESPREK	0.65	0.56	0.75	

<sup>61</sup> Puur ter informatie zijn ook de correlaties met gespreksvaardigheid weergegeven. De gespreksvaardigheid maakte geen deel uit van de gecombineerde score.

De opgaven voor de toetsen zijn met OPLM goed op een schaal te brengen. We zien dat de correlaties tussen de toetsen met waarden groter dan 0,60 redelijk hoog is, zeker als we rekening houden met de betrouwbaarheid van de toetsen: dan is de latente correlatie tussen de toetsen bijna 0,90. Deze waarden zijn hoger dan in de pilot gevonden was. Het lijkt dus mogelijk de toetsen op een schaal te brengen.

Als we de opgaven op een schaal brengen dan is door de overdaad aan woordenschatopgaven (110, goed voor ruim 50% van het totaal aantal opgaven) de impact van die vaardigheid erg hoog. Wanneer dan ook nog eens een programma gebruikt wordt om de a-parameters in te schatten kan die impact nog hoger worden als de woordenschatopgaven meer onderling samenhangen. Om die reden zijn in drie afzonderlijke sessies voor de drie vaardigheden de a-parameters ingeschat: Bij lezen (51 opgaven) met een gemiddelde a-parameter van 4,10, bij luisteren (48 opgaven) met een gemiddelde a-parameter van 4,36 en bij woordenschat (110 opgaven) met een gemiddelde van 1,90. Hierdoor loopt het theoretisch gewogen scorebereik per vaardigheid van 0 tot en met 209 en weegt ieder van de vaardigheden even zwaar mee, terwijl binnen de vaardigheid verschillen tussen opgaven wat betreft onderscheidend vermogen meegenomen kunnen worden.

Hierbij moet wel opgemerkt worden dat de passing statistisch gezien niet zeer goed is. De woordenschat-opgaven worden in vergelijking met de andere opgaven in de schaal als te weinig onderscheidend ingeschat met deze gemiddelden. Dat is echter doelbewust gedaan om deze vaardigheid niet te dominant te laten zijn in de inschatting van de gehele vaardigheidsschatting. De samengestelde vaardigheid wordt in dit onderzoek alleen gebruikt voor de selectie van scholen voor het verdiepend onderzoek. In hoofdstuk 10 wordt verder beschreven hoe de selectie van de scholen voor het verdiepend onderzoek heeft plaats gevonden.

## Hoofdstuk 6. Kenmerken deelnemers

Naast het vaststellen van de vaardigheden Engels van leerlingen in groep 8 heeft de peiling tot doel het onderwijsleerproces op de scholen in kaart te brengen. Zowel voor de directie als voor leerkrachten zijn vragenlijsten opgesteld die informatie hieromtrent op respectievelijk school- en klasniveau verzamelen. Met de leerkrachtvragenlijst zijn tevens achtergrondkenmerken van de leerkrachten vastgelegd. Daarnaast is een vragenlijst bij leerlingen afgenomen. In principe is deze vragenlijst, als eerste onderdeel van de afname, door alle aanwezige leerlingen ingevuld.

Met behulp van de resultaten van de school-, leerkracht- en leerlingvragenlijsten worden de kenmerken van de deelnemende scholen, leerkrachten/klassen en leerlingen beschreven. In respectievelijk in paragraaf 6.1, 6.2 en 6.3 staan kenmerken van de scholen, de leerkrachten en het onderwijsleerproces centraal, terwijl dit in paragraaf 6.4 voor de leerlingen het geval is. De bijbehorende vragenlijsten zijn opgenomen als Bijlage 6 tot en met 8. Bij de presentatie van de onderzoeksbevindingen worden naast resultaten van de totale onderzoekspopulatie veelvuldig resultaten op basis van groepen scholen weergegeven. Er wordt daarbij een onderscheid gemaakt tussen EIBO- scholen en VVTO-scholen. Daarnaast komen eventuele verschillen aan bod tussen EIBO-scholen met 0-10% gewichtenleerlingen, EIBO-scholen met meer dan 10% gewichtenleerlingen, VVTO-scholen met 0-10% gewichtenleerlingen en VVTO-scholen met meer dan 10% gewichtenleerlingen. Op deze manier wordt niet alleen recht gedaan aan mogelijke verschillen als gevolg van startmoment, maar wordt daarbij ook het percentage gewichtenleerlingen op een school in aanmerking genomen. In dit hoofdstuk worden resultaten op basis van ongewogen analyses gepresenteerd.

### 6.1 Kenmerken van de deelnemende scholen

Voorafgaand aan de afname van het hoofdonderzoek hebben bij alle deelnemende scholen intakegesprekken plaatsgevonden met de schooldirectie en, indien mogelijk, de groepsleerkracht(en) van de deelnemende klas(sen). Tijdens deze intakegesprekken is een aantal vragen voorgelegd om de organisatie en inrichting van het aanbod Engels op de scholen in kaart te brengen. De vragen waren onderdeel van het intakeformulier dat werd ingevuld door de schooldirecteur en/of de contactpersoon. Ook kwam het voor dat de vragen door de hoofdtestleider werden gesteld en werden ingevuld. Op de scholen waar een face-to-face intakegesprek plaatsvond werden de vragenlijsten direct weer meegenomen door de hoofdtestleiders. Op zestien scholen vond op verzoek van de school een telefonische intake of een intake via skype plaats. Deze scholen dienden de vragen in het intakeformulier na afloop in te vullen en digitaal op te sturen. Ondanks herhaalde verzoeken hebben zes scholen hier uiteindelijk geen gehoor aan gegeven. In totaal zijn gegevens van 89 van de 95 scholen (94%) via de vragenlijst verkregen en opgenomen in de analyses. Om die reden tellen niet alle tabellen over het schoolproces in het vervolg van dit hoofdstuk op tot 95.

De schoolvragenlijst bevatte zeven vragen gericht op de wijze waarop scholen het onderwijs in Engels op schoolniveau hebben georganiseerd en vormgegeven. Hieronder gaan we in afzonderlijke paragrafen in op de resultaten.



### Startmoment Engels – aanbod profiel

Scholen zijn vanaf 1986 verplicht om Engels vanaf groep 7 aan te bieden en de tendens is dat steeds meer scholen op een eerder moment starten met dit aanbod. Scholen is gevraagd aan te geven tot welke categorie scholen zij behoorden: 1) vroeg vreemdetalenonderwijs, 2) vervroegd EIBO (Engels vanaf groep 5 of groep 6) en 3) 'regulier' EIBO (Engels vanaf groep 7)<sup>62</sup>. Voor een categorisering als VVTO-school is daarbij leidend geweest of de school vanaf het schooljaar 2013-2014 Engels aan de groep 8-leerlingen heeft aangeboden, conform de in dit onderzoek gehanteerde definitie. De gehele set van 95 scholen is te verdelen in 62 EIBO-scholen en 33 VVTO-scholen. De definitieve verdeling van de profielen is weergegeven in Tabel 6.1. Informatie over het specifieke EIBO-profiel (vervroegd-regulier) is onbekend bij vijf van de zes scholen waarvan het intakeformulier niet is ontvangen. Wel is vastgesteld op basis van aanwezige informatie uit de vragenlijst en nagezochte internetinformatie dat deze scholen definitief tot het EIBO-profiel behoren<sup>63</sup>: ze zijn dan ook als EIBO-school gecategoriseerd. Het blijkt dat minimaal de helft van de 62 EIBO scholen een vervroegd aanbod profiel heeft.

Tabel 6.1: aanmelding en deelname naar profiel

Profiel	Beoogde Steekproef	Deelname
EIBO	68	62
- <i>Regulier EIBO</i>		- 26
- <i>Profiel EIBO onbekend</i>		- 5
- <i>Vervroegd EIBO</i>		- 31
VVTO	32	33
Totaal	100	95

### Onderwijstijd

De hoeveelheid onderwijstijd die aan Engels wordt besteed in de huidige groep 8 is opgevraagd in termen van het aantal weken en het gemiddeld aantal lesminuten per week. Tabel 6.2 geeft het totaal aantal lessen per jaar aan. Van de 89 scholen met (intake)gegevens is van een viertal scholen het aantal weken of de minuten per week niet bekend. Gemiddeld worden 31,5 uren lestijd besteed aan het geven van Engels. Vergeleken met de vorige peiling (Geurts en Hemker, 2013) blijkt dat de totale tijd in groep 8 licht is gedaald<sup>64</sup>. In de vorige peiling was de ingeschatte onderwijstijd gemiddeld 32 uur per jaar volgens de leerkrachten en 34 uur per jaar volgens de schoolleiders. De informatie over het aantal uren per jaar is nu verkregen in het intakegesprek waar soms geen leerkracht bij aanwezig was.

<sup>62</sup> De vraag naar het startmoment voor de huidige groep is gebruikt ter verificatie van de steekproefindeling van de scholen. Met deze vraag is nogmaals nagegaan of de school, op basis van de aanwezige informatie voor steekproeftrekking, daadwerkelijk aan het juiste stratum was toegekend wat betreft startmoment voor de huidige groep 8 (vanaf schooljaar 2013-2014). Zeven scholen bleken tot een incorrect stratum te behoren; dit is gewijzigd in het databestand. Zie voor meer informatie paragraaf 2.4.

<sup>63</sup> De zesde school waarvan geen intakeformulier was ontvangen was een VVTO school en bleek dat na nader onderzoek ook te zijn.

<sup>64</sup> Bij gebrek aan data van de peiling 2012 kan er slechts in beschrijvende zin een vergelijking worden gemaakt met de vorige peiling. Eventuele verschillen konden niet op significantie worden getoetst.

Uit de tabel wordt duidelijk dat de onderwijstijd in groep 8 op VVTO-scholen hoger is. Gemiddeld ontvangen VVTO-leerlingen per jaar ruim 6 uur meer Engels onderwijs in groep 8 dan leerlingen op EIBO scholen. Wat verder opvalt aan Tabel 6.2 is dat, ook al zijn de aantallen klein, de EIBO-scholen met veel gewichtenleerlingen meer tijd aan Engels besteden dan de EIBO-scholen met weinig gewichtenleerlingen, terwijl dit andersom is voor de VVTO-scholen: Op VVTO-scholen met veel gewichtenleerlingen is er minder tijd voor Engels dan op VVTO-scholen met weinig gewichtenleerlingen.

Tabel 6.2: gemiddelde aantal uren Engels per schooljaar in groep 8, naar strata (N=85)

Aantal uren Engels in groep 8	n	M	S.D.	Mediaan	Minimum	Maximum
EIBO < 10	36	27.86	10.70	30	6	40.
EIBO > 10	19	31.97	6.90	30	20	40
VVTO <10	17	37.71	18.86	30	22.50	100
VVTO > 10	13	32.90	7.34	33.33	15	40
Totaal EIBO	55	<b>29.28</b>	9.69	30	6	40.08
Totaal VVTO	30	<b>35.62</b>	14.98	31.67	15	100
Totaal	85	<b>31.52</b>	12.13	30	6.00	100.00

Tabel 6.3 bevat de onderwijstijd in gemiddeld aantal minuten per week, aantal weken en totaal aantal uren. Veruit de meeste leerlingen ontvangen 31 tot 60 minuten Engelse les per week (samen 84% van de scholen). Lesminuten per week boven de 60 zijn uitzonderlijk, terwijl lestijden onder de 30 minuten alleen voorkomen bij EIBO-scholen. Bijna viervijfde van de scholen biedt Engels gedurende het gehele schooljaar aan (36-40 weken), terwijl in 10 procent van de gevallen 0-20 schoolweken worden benut. De totale tijdsbesteding per schooljaar ligt zoals Tabel 6.2 laat zien gemiddeld op 31,5 uren. Ongeveer 18% (15 scholen) zit daar beduidend onder (tot 20 uren per jaar).

Tabel 6.3: aantal uren en weken Engels per schooljaar in groep 8

Minuten per week	Aantal scholen	%	Aantal weken	Aantal scholen	%	Totaal uren	Aantal scholen	%
0-30	9	10%	0-20	9	10%	Tot 20 uur	15	18%
31-45	38	44%	21-25	0	0%	21-35 uur	40	47%
46-60	35	40%	26-30	3	3%	36-40 uur	26	31%
61-80	2	2%	31-35	6	7%	40-45 uur	1	1%
81-100	2	2%	36-40	68	79%	46-60 uur	2	2%
Meer dan 100	1	1%	40 of meer	0	0%	61 uur of meer	1	1%
Totaal	87			86			85	

### 6.1.2 Organisatie van het vak Engels op school

Het vak Engels is op de deelnemende scholen op verschillende wijze ingebed en georganiseerd. Veelal kan en moet de leerkracht van groep 8, behoudens wellicht de keuze voor de gehanteerde hoofdmethode, de lessen grotendeels individueel invullen. In ruim twee vijfde van de deelnemende scholen wordt het vak echter aangestuurd via een coördinator Engels en/of of een werkgroep Engels (Tabel 6.4). Het blijkt dat een dergelijke coördinator of werkgroep, zoals zou kunnen worden verwacht, relatief vaker aanwezig is op de VVTO scholen (72 % versus 25% in de EIBO groep). Acht VVTO scholen hebben zowel een coördinator als een werkgroep (25%) Engels; slechts één van de 56 EIBO scholen geeft aan beide te hebben.

Tabel 6.4: aantallen scholen met een coördinator en/of werkgroep voor Engels, naar strata (N=88)

	Aantal	Aanwezig		Niet aanwezig	
		N	%	n	%
EIBO	56	14	25 %	42	75 %
VVTO	32	23	72 %	9	28 %
Totaal	88	37	42 %	51	58 %

Scholen kennen verschillende faciliteiten om de lessen Engels in groep 8 goed te kunnen organiseren. Scholen hebben bijvoorbeeld een budget voor aanvullende materialen, laten leerkrachten en leerlingen deelnemen aan uitwisselingen, of stimuleren met doeltaal-voertaal-activiteiten buiten de reguliere les het aanbod van Engels (bijvoorbeeld Engelstalige gasten, volgen van Engelse schooltelevisie, Engelse lunch). Tabel 6.5 geeft aan in welke mate scholen dergelijke zaken (kunnen) faciliteren. Op VVTO scholen zijn, naar verwachting, deze faciliteiten relatief meer aanwezig dan op EIBO scholen. Opvallend is dat ruim de helft van de EIBO scholen aangeeft geen enkele faciliteit te hebben voor het vak Engels. NB: scholen konden meerdere faciliteiten aanvinken. Onder 'andere faciliteiten' worden overigens uiteenlopende zaken verstaan, variërend van 'de mogelijkheid bijscholing te volgen' tot organiseren van 'Engelse week' of 'deelname aan de Dag van de Talen'.

Tabel 6.5: Facilitering voor aanbod Engels (N= 89)

	Geldt voor	Naar startmoment	
		EIBO (n=57)	VVTO (n=32)
Budget voor aanvullende of ondersteunende materialen voor Engels	31 (35%)	13 (23%)	18 (56%)
Stimuleren van deelname aan internationale partner- en uitwisselingsprojecten	12 (14%)	3 (5%)	9 (28%)
Stimuleren van activiteiten waarbij Engels voertaal is	15 (17%)	4 (7%)	11 (34%)
Anders	22 (25%)	15 (47%)	7 (22%)
GEEN faciliteiten	38 (43%)	30 (53%)	8 (25%)

### Professionalisering

Om het Engels onderwijs te professionaliseren en te optimaliseren kunnen scholen verder bijscholing voor hun leerkrachten organiseren en financieel ondersteunen. Ruim 70 procent van de 89 scholen geeft aan dat er in het jaar voorafgaande aan de peiling geen professionaliseringsactiviteiten hebben plaatsgevonden. Op ongeveer dertig procent (25 scholen) hebben leerkrachten juist wel deelgenomen aan cursussen of trajecten als het gaat om de verbetering van de lespraktijk voor het vak Engels. Daarbij waren 182 leerkrachten betrokken, variërend van 1 leerkracht op de school tot gehele teams van 16 tot 26 leerkrachten. In de meeste gevallen ging dit om een cursus die zowel de taalvaardigheid van de leerkracht(en) als de didactische vaardigheden van de betrokken leerkracht(en) trachtte te verbeteren (Tabel 6.6). Professionalisering van het onderwijs Engels lijkt relatief meer plaats te vinden in de VVTO scholen dan in EIBO scholen: van de 182 leerkrachten waren er 127 werkzaam op VVTO scholen en 55 op EIBO scholen.

Tabel 6.6: Insteek professionaliseringsactiviteiten (N=89)

Insteek activiteit	Totaal		EIBO Scholen	VVTO scholen
	n	%	n	n
Taalvaardigheid	2	2 %	1	1
Didactiek	4	4 %	4	0
Taalvaardigheid en didactiek	14	16 %	4	10
Anders	1	1 %	0	1
Meerdere combinaties	4	4 %	2	2
Totaal	25	28 %	11	15
Geen activiteiten	64	72 %	47	17

### Overdracht

Naast Engels als belangrijke 21<sup>e</sup> -eeuwse vaardigheid is het vak in het basisonderwijs voor een belangrijk deel opgenomen in de wettelijke opdracht voor basisscholen om de overstap naar het voortgezet onderwijs te vergemakkelijken (soepele po-vo overgang, doorgaande leerlijnen). De scholen zijn daarom gevraagd in hoeverre en op welke manier zij de vaardigheden in de Engelse taal overdragen naar het secundair onderwijs. De resultaten zijn weergegeven in Tabel 6.7. Het blijkt dat in ruim de helft van de gevallen helemaal geen overdracht van vaardigheids- of toetscores plaatsvindt. In de lijn der verwachting blijkt echter dat VVTO scholen meer dan EIBO scholen actief de prestaties naar het VO doorspelen.

Tabel 6.7: Overdracht van Vaardigheden Engels richting voortgezet onderwijs (N=87).

Type overdracht	Totaal		EIBO		VVTO	
	Aantal	%	Aantal	%	Aantal	%
Geen overdracht	46	53 %	34	61 %	12	39 %
Rapportcijfers voor Engels	13	15 %	8	14 %	5	16 %
Overdrachtdossier	4	5 %	4	7 %	0	-
Mondelinge overdracht po-vo	6	7 %	4	7 %	2	6 %
Anders	6	7 %	1	2 %	5	16 %
Meerdere opties	12	14 %	5	9 %	7	23 %

## 6.2 Kenmerken van de deelnemende leerkrachten

Onderdeel van het onderzoek was de afname van een vragenlijst bij de leerkrachten van groep 8. De vragenlijst bevat vier onderdelen met vragen over 1) achtergrondkenmerken, 2) het onderwijsleerproces in de klas, 3) attitude over Engels in het PO en 4) inschatting van de eigen Engelse vaardigheid en die van de leerlingen in de klas. Onder achtergrondkenmerken vallen, naast individuele kenmerken, tevens het genoten aanbod Engels in de opleiding en het buitenschools gebruik van het Engels.

### Onderzoeksgroep

De vragenlijst is uitgezet bij alle leerkrachten in de groepen 8 van de deelnemende scholen. Vragenlijsten zijn op papier door de leerkracht ingevuld tijdens de afname van de klassikale taken. In geval van duo-leerkrachten heeft de testleider een vragenlijst achtergelaten, die op een later moment door de duo-partner kon worden ingevuld en opgestuurd aan de onderzoekers. Voor 109 leerkrachten, werkzaam op 90 scholen, waren bruikbare gegevens beschikbaar. Voor de analyses zijn de antwoorden van alle 109 leerkrachten meegenomen, behalve bij de vraag naar de gebruikte methode. Hier is gekozen voor analyse op schoolniveau (89 scholen).

### Achtergrondkenmerken

De eerste vragen betroffen achtergrondinformatie, waarin eveneens naar ervaring en taalachtergrond is gevraagd. In Tabel 6.8 zijn de achtergrondkenmerken van de leerkrachten opgenomen. Van de 109 leerkrachten met bruikbare gegevens zijn 98 leerkrachten (90%) de groepsleerkracht van de deelnemende groep 8. Tien leerkrachten zijn dat niet, maar bijvoorbeeld de vakleerkracht en één respondent heeft bij deze vraag niets aangekruist. Van deze respondent is ook het geslacht en de leeftijd niet bekend. Van de 108 leerkrachten is 33% man en 67% vrouw. Analoog aan de steekproefstrata is ruim twee derde van de leerkrachten (74 in totaal) werkzaam op of voor een EIBO-school en is een derde (35 leerkrachten) werkzaam op of voor een VVTO school. Meer dan 96% van de leerkrachten is Nederlands en beschouwt het Nederlands ook als moedertaal. Ongeveer 30% van de leerkrachten geeft aan met regelmaat in het Nederlands en Engels te communiceren in situaties buiten de school.

Tabel 6.8: Achtergrondkenmerken van de leerkrachten

	N		n	Percentage <sup>65</sup>
Geslacht	108	Man	36	33.3 %
		Vrouw	72	66.7 %
Leeftijd	108	20-30 jaar	27	25.0 %
		30-40 jaar	37	34.3 %
		40-50 jaar	22	20.4 %
		50-60 jaar	15	13.9 %
		Ouder dan 60	7	6.5 %
Groepsleerkracht	108	Ja	98	90.1 %
		Nee	10	8.9 %
In Nederland geboren	108	Ja	104	96.3 %
		Nee	4	3.7 %
Nederlands wordt als moedertaal beschouwd	107	Ja	104	97.2 %
		Nee	3	2.8 %
Engels wordt als moedertaal beschouwd	103	Ja	1	1.0 %
		Nee	102	99.0 %
Leerkracht communiceert regelmatig in	99	Nederlands	64	65.3 %
		Nederlands en Engels	29	29.6 %
		Nederlands en anders	2	2.0 %
		Anders	3	3.1 %
Leerkracht is werkzaam op ....	109	EIBO school <10	51	46.8 %
		EIBO school >10	23	21.1 %
		VVTO school <10	22	20.2 %
		VVTO school >10	13	11.9 %

Het overgrote deel van de deelnemende leerkrachten (79%) heeft een reguliere Pabo-opleiding gevolgd, maar ook de verkorte pabo en de academische pabo komen voor (respectievelijk 10% en 9%). Twee procent van de leerkrachten zegt de opleiding niet te hebben afgemaakt. De werkervaring van de leerkrachten is gevarieerd, waarbij de verdeling van leerkrachten over de categorieën min of meer vergelijkbaar is - alleen de categorie 16-20 jaar ervaring is relatief lager (15% van de leerkrachten). Ruim 40% van de leerkrachten heeft vijf jaar of minder ervaring met het geven van Engelse les, terwijl ruim een derde van de leerkrachten meer dan 11 jaar ervaring heeft met het geven van Engelse lessen.

<sup>65</sup> Evenals in de hiernavolgende tabellen geeft het genoemde percentage het percentage weer dat gebaseerd is op het aantal respondenten dat de vraag heeft beantwoord.

Tabel 6.9: Ervaring en opleiding respondenten

	<b>N</b>		<b>n</b>	<b>Percentage</b>
Werkervaring in het PO	108	0-5 jaar	25	23.1 %
		6-10 jaar	21	19.4 %
		11-15 jaar	24	22.2 %
		16-20 jaar	16	14.8 %
		Meer dan 21 jaar	22	20.4 %
Ervaring met geven van Engels	108	0-5 jaar	47	43.5 %
		6-10 jaar	23	21.3 %
		11-15 jaar	21	19.4 %
		16-20 jaar	6	5.6 %
		Meer dan 21 jaar	11	10.2 %
Behaalde diploma	103	PABO	82	79.6 %
		Verkorte PABO	10	9.7 %
		Academische PABO	9	8.7 %
		Diploma nog niet behaald	2	1.9 %

Hoewel het belang van Engels op de Pabo toeneemt, was het vroeger niet altijd een verplichte module en ook de academische Pabo's kennen niet allemaal een verplichte module Engels. Eenentwintig procent van de leerkrachten geeft aan geen verplichte module Engels op de (academische) Pabo te hebben gevolgd (Tabel 6.10). Zes procent van de leerkrachten rapporteert een keuzemodule Engels op de Pabo te hebben gevolgd. Gevraagd naar de inhoud van de scholing op de Pabo geeft 42% van de leerkrachten aan dat deze op zowel taal als didactiek gericht was. Slechts 19% zegt tevreden te zijn over het aanbod op de Pabo. Van de 97 leerkrachten die de vraag hebben beantwoord, geeft slechts 14% aan dat zij na de opleiding na- of bijscholing op het gebied van Engels heeft ontvangen. Deze nascholing was gericht op taal (53%) of taal *en* didactiek (47%). Opvallend is dat één leerkracht heeft aangegeven geen nascholing te hebben gehad, maar wel heeft aangegeven waar de nascholing op gericht was.

Tabel 6.10: Scholing van de leerkrachten op het gebied van Engels

Scholing Engels....		N		Aantal	Percentage
Op de (academische) PABO	Engels was verplichte module	97	Ja	76	78.4 %
			Nee	21	21.6 %
	Engels was keuzemodule	85	Ja	5	5.9 %
			Nee	80	94.1 %
	Nadruk in lessen	97	Taal en didactiek	41	42.3 %
			Alleen Taal	13	13.4 %
			Alleen didactiek	18	18.6 %
			n.v.t.	25	25.8 %
Tevredenheid invulling en inhoud	84	Ja	16	19.0 %	
		Nee	68	81.0 %	
Na-/bijscholing	Nascholing gehad?	97	Ja	14	14.4 %
			Nee	83	85.6 %
	Nadruk in na-/bijscholing	15	Taal- en didactiek	7	46.7 %
			Alleen taal	8	53.3 %

### Buitenschools gebruik Engels

Door het Engels in het dagelijks leven passief of actief te gebruiken kan de leerkracht de eigen Engelse taalvaardigheid bijhouden en verder ontwikkelen. Dit is van belang, aangezien de taalvaardigheid van de leerkracht de prestaties van leerlingen in het mondelinge domein bevordert (Unsworth, et al., 2014). De vragenlijst bevatte zeven vragen over de mate waarin de leerkracht het Engels buitenschools gebruikt. De leerkracht diende telkens één antwoord te kiezen op een schaal van (bijna) nooit (1) tot en met dagelijks (5). Figuur 6.1 toont een voorbeeld-item.

<p><b>Uw eigen gebruik van het Engels in uw dagelijkse leven</b>  <i>U kunt onderscheid maken tussen:</i></p>				
1 = (bijna) nooit	3 = minstens 1 keer per maand	5 = dagelijks		
2 = minstens 1 keer per jaar	4 = minstens 1 keer per week			
<p>Hoe vaak schrijft u een tekst in het Engels?                  (social media, chat, e-mail, brief, kaart)</p>				
<p style="text-align: center;"> <input type="checkbox"/> 1      <input type="checkbox"/> 2      <input type="checkbox"/> 3      <input type="checkbox"/> 4      <input type="checkbox"/> 5                 </p>				

Figuur 6.1: voorbeelditem uit schaal 'buitenschools gebruik Engels'

In Tabel 6.11 zijn de schaalscores voor buitenschools gebruik Engels gegeven voor het totaal en voor de verdeling naar startmoment en stratum. Deze schaalscore geeft de gemiddelde score op de zeven vragen weer. Zoals kan worden afgeleid uit Tabel 6.11 gebruiken leerkrachten in het dagelijkse leven gemiddeld meer dan 1 keer per maand, maar minder dan 1 keer per week actief of passief het Engels. Wanneer we de antwoorden per stratum bekijken zien we dat het gemiddelde gebruik, meer dan 1 keer per maand, voor alle groepen geldt. Leerkrachten op VVTO-scholen en EIBO-scholen verschillen hierin niet van elkaar ( $t(df=106)=0.04, p=0.97$ ).



Tabel 6.11: Schaalscores buitenschools gebruik Engels door leerkrachten

	N	Gem.	SD	Range
EIBO totaal	73	3.42	.73	1-5
EIBO < 10	50	3.49	.75	1-5
EIBO > 10	23	3.28	.66	2-4
VVTO totaal	35	3.43	.78	2-5
VVTO <10	22	3.53	.78	2-5
VVTO > 10	13	3.25	.79	2-5
Totaal	108	3.42	.74	1-5

### *Eigen inschatting vaardigheid Engels*

Voor het geven van Engels in het basisonderwijs wordt sinds 2018 een taalvaardigheidsniveau van minimaal B2 vereist<sup>66</sup>. Het belang van de vaardigheid Engels van de leerkracht voor de Engelse vaardigheid van leerlingen is reeds in eerder onderzoek aangetoond (Unsworth, et al, 2014). In de leerkrachtvragenlijst is aan de leerkrachten gevraagd hun eigen vaardigheid in vier vaardigheidsdomeinen aan te geven: 1) lezen, 2) luisteren, 3) schrijven en 4) spreken. Per vaardigheid werden vier ERK-beschrijvingen gegeven, corresponderend met ERK-niveaus A2, B1, B2 en C1, waarbij leerkrachten konden aankruisen of ze het niveau nog niet (score '0') of juist wel beheersten (score '1').

In de vorige peiling is bij de beschrijving van de uitkomsten geen rekening gehouden met het type school waarop de leerkrachten lesgeven (EIBO of VVTO). Dat is in de huidige peiling wel gedaan, met inachtneming van het percentage gewichtenleerlingen op de school. Daarnaast zijn in de huidige peiling de resultaten op de verschillende vaardigheidsdomeinen geschaald. In Tabel 6.12 staat de eigen inschatting van de leerkracht per domein weergegeven. Omdat de descriptoren opliepen in moeilijkheid en er gebruik is gemaakt van items variërend van A2-C1 (met theoretische schaalscores van '0-5') per vaardigheid kunnen de resultaten in Tabel 6.12 als volgt worden geïnterpreteerd: de eigen inschatting van de leerkrachten voor de receptieve vaardigheden lezen en luisteren ligt op B2-niveau (en voor luisteren zelfs iets daar boven) en voor de productieve vaardigheden ligt die inschatting wat lager; leerkrachten schatten in het B1-niveau te beheersen, maar gemiddeld verwachten zij het B2-niveau nog niet te halen. Zoals Tabel 6.12 laat zien, verschillen de inschattingen van de leerkrachten op de vier onderscheiden typen scholen over het algemeen niet duidelijk van elkaar. Bij spreken lijkt er echter meer onderscheid te zijn, met gemiddelden variërend van 2.31 tot 2.90. Een Anova-analyse laat echter geen significante verschillen zien in ingeschatte spreekvaardigheid tussen de leerkrachten van de verschillende groepen scholen ( $F=1.27$ ,  $df=3$ ,  $p=0.29$ ).

<sup>66</sup> Zie [https://www.10voordeleraar.nl/documents/kennisbases\\_pabo/kennisbasispabo.pdf](https://www.10voordeleraar.nl/documents/kennisbases_pabo/kennisbasispabo.pdf).

Tabel 6.12: Leerkracht inschatting van eigen vaardigheden

Inschatting lezen						
	N	Gem.	SD	Min	Max	Mediaan
EIBO totaal	71	2.97	.79	2	4	3
EIBO < 10	49	3.00	.76	2	4	3
EIBO > 10	22	2.91	.87	2	4	3
VVTO totaal	33	3.03	.85	1	4	3
VVTO <10	21	3.00	.89	1	4	3
VVTO > 10	12	3.08	.79	2	4	3
Totaal	104	2.99	.81	1	4	3
Inschatting luisteren						
EIBO totaal	68	3.21	.89	1	4	3.5
EIBO < 10	45	3.22	.88	1	4	3
EIBO > 10	23	3.17	.94	2	4	4
VVTO totaal	34	3.15	.86	2	4	3
VVTO <10	21	3.14	.91	2	4	4
VVTO > 10	13	3.15	.80	2	4	4
Totaal	102	3.19	.88	1	4	4
Inschatting spreken						
EIBO totaal	71	2.52	1.17	0	4	2
EIBO < 10	49	2.41	1.12	0	4	2
EIBO > 10	22	2.77	1.27	0	4	3
VVTO totaal	33	2.67	1.22	0	4	3
VVTO <10	20	2.90	1.12	0	4	3
VVTO > 10	13	2.31	1.32	0	4	2
Totaal	104	2.57	1.18	0	4	2
Inschatting schrijven						
EIBO totaal	71	2.54	.97	1	4	2
EIBO < 10	49	2.53	.92	1	4	2
EIBO > 10	22	2.55	1.10	1	4	2
VVTO totaal	35	2.54	1.12	0	4	2
VVTO <10	22	2.50	1.37	0	4	2.5
VVTO > 10	13	2.62	.87	2	4	2
Totaal	106	2.54	1.03	0	4	2

### 6.3 Aanbod van Engels in de les

Bij het in kaart brengen van het onderwijsaanbod is gekeken naar een aantal verschillende aspecten. Hieronder wordt ingegaan op: de doelstellingen waaraan gewerkt wordt, de gebruikte methoden en methodieken, de activiteiten in de klas, de manier van evalueren en de overdracht tussen groep 7 en 8.

#### Doelstellingen

De leerkrachten is gevraagd welke doelen zij nastreven voor Engels. Van de 104 leerkrachten die de vraag heeft beantwoord, noemen 31 leerkrachten slechts één doel, terwijl de overige leerkrachten

twee doelen noemen. Zesenvestig (bijna de helft van de) leerkrachten geeft aan zowel de kerndoelen als de doelen die in de methode worden aangegeven, na te streven. VVTO-leerkrachten gaan relatief vaker uit van doelen die zij zelf vaststellen of die binnen de school zijn bepaald.

Tabel 6.13: Doelstellingen van leerkrachten

	Totaal		EIBO		VVTO	
	n	%	n	%	n	%
De kerndoelen Engels voor het po	67	64 %	50	70 %	17	51 %
Doelen die zijn beschreven in het ERK	2	2 %	1	1 %	1	3 %
Doelen die binnen school/schoolbestuur bepaald zijn	13	13 %	7	10 %	6	18 %
Doelen die ikzelf vaststel.	15	7 %	6	8 %	9	27 %
Daarbij gebruik ik uiteenlopende bronnen						
Doelen die in de gebruikte methode worden aangegeven	72	44 %	54	76 %	18	55 %
Totaal	104	100	71	100	33	100

### Methoden en additionele leermiddelen

Voor zowel groep 7 als groep 8 is leerkrachten gevraagd of zij uitsluitend de methode gebruiken of dat zij ook gebruik maken van andere leermiddelen. Van 109 leerkrachten die de vraag beantwoordden voor groep 8 zegt ruim de helft (51.4%) dat zij uitsluitend de methode gebruiken bij de lessen Engels, terwijl 53 leerkrachten (48.6%) ook andere leermiddelen gebruiken. 104 leerkrachten hebben deze vraag ook beantwoord voor groep 7. Bijna drie vijfde (56.7%) rapporteert dat in groep 7 uitsluitend de methode is gebruikt, terwijl volgens 34 leerkrachten (32.7%) ook andere leermiddelen in die groep zijn gebruikt. Ruim 10 procent van de leerkrachten weet niet wat in groep 7 is gebruikt. Het gebruik van de leermiddelen is volgens 74 leerkrachten (68.5%) hetzelfde in groep 8 als in groep 7 en volgens 16 leerkrachten (14.8%) verschillend. Van de 108 leerkrachten die deze vraag hebben beantwoord, weten 18 leerkrachten dit niet.

Van 89 scholen is bekend welke methode in groep 8 wordt gebruikt (Tabel 6.14). Er is een grote diversiteit aan door leerkrachten gerapporteerde gebruikte methoden, waarbij *Take it easy* (34%), *Groove me* (25%) en *Step to* (8%) de meest gebruikte zijn. De overige methoden worden slechts op enkele scholen ingezet. Op vijf VVTO-scholen wordt geen specifieke methode gebruikt, maar worden de lessen door de leerkracht zelfontworpen (soms met input vanuit verschillende methodes). *Take it easy* en *Groove me* worden op respectievelijk 35 procent en 28 procent van de reguliere EIBO-scholen ingezet, terwijl zij op respectievelijk 31 procent en 17 procent van de VVTO scholen worden gebruikt. Daarmee zijn zij in beide groepen scholen het meest gebruikt.

Met behulp van vijf stellingen is leerkrachten gevraagd hoe zij hun methode waarderen. De vijf antwoordmogelijkheden liepen oorspronkelijk van helemaal mee eens (1) tot helemaal mee oneens (5). Hecodering heeft plaatsgevonden om ervoor te zorgen dat een hogere score een hogere attitude weergeeft. Een voorbeeld van de gegeven stellingen is:

	Helemaal eens	Eens	Neutraal	Oneens	Helemaal oneens
De leeractiviteiten in de methode spreken mijn leerlingen aan	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

Figuur 6.2 Voorbeeld stelling over gebruikte methode

Tabel 6.14 Gebruikte methoden, schoolbreed (N=89)

Methode	Totaal		EIBO (N=60)		VVTO (N=29)	
	N	%	n	%	n	%
Take it easy (ThiemeMeulenhoff)	30	34 %	21	35 %	9	31 %
Groove me (Blink)	22	25 %	17	28 %	5	17 %
Step to (Anglia)	7	8 %	4	12 %	3	10 %
Just do it (ThiemeMeulenhoff)	5	6 %	4	8 %	1	3 %
Hello World (Malmberg)	3	3 %	3	5 %	0	-
Real English (ThiemeMeulenhoff)	3	3 %	3	5 %	0	-
Alles apart (Alles in 1)	2	2 %	1	3 %	1	3 %
Backpack (Pearson)	2	2 %	0	-	2	7 %
Our discovery island (Pearson)	2	2 %	2	3 %	0	-
Duolingo (online uitgever)	1	1 %	1	2 %	0	-
Happy Earth (Practicum Educatief)	1	1 %	0	-	1	3 %
Happy House (Practicum Educatief)	1	1 %	0	-	1	3 %
Join in (Malmberg)	1	1 %	1	2 %	0	-
My name is Tom (Bas educatie)	1	1 %	0	-	1	3 %
Rainbow Library	1	1 %	1	2 %	0	-
Such Fun (Early English program)	1	1 %	1	2 %	0	-
The team (Educatheek)	1	1 %	1	2 %	0	-
Geen/andere methode	5	6 %	0	-	5	17 %

De uitkomsten in Tabel 6.15 laten zien dat de leerkrachten neutraal tot positief zijn over verschillende aspecten van de methode (aanspreekkracht, aansluiting bij niveau, relevante aspecten en onderwerpen, leerzaamheid), maar dat ze neutraal tot negatief zijn over de hoeveelheid differentiatie-handreikingen die de methode biedt. Deze bevinding is vergelijkbaar met die van de vorige peiling (en eerder onderzoek van Thijs,2011): leerkrachten schatten de methode als voldoende in qua niveau, maar missen mogelijkheden tot differentiatie. Dat laatste geldt voor zowel leerkrachten in EIBO als VVTO scholen. De scores van leerkrachten op EIBO-scholen zijn gemiddeld hoger m.b.t. 'aanspreekkracht', 'relevante aspecten' en 'leerzaamheid' dan de scores van leerkrachten op VVTO-scholen. De verschillen zijn echter niet significant ( $p < 0.05$ ).

Tabel 6.15: Waardering methode door leerkrachten

	Totaal N=110			EIBO n=74	VVTO n=35
	Gem	Mediaan	SD	Gem	Gem
methode spreekt In aan	3,35	4	1,00	3.46	3.06
activiteiten in methode sluiten aan bij niveau leerlingen	3,50	4	0,83	3.49	3.51
methode besteedt aandacht aan relevante aspecten	3,62	4	0,81	3.68	3.49
leerlingen leren veel door methode	3,38	3	0,88	3.45	3.23
methode biedt mogelijkheden voor differentiatie	2,57	2	0,96	2.55	2.57

Gevraagd naar de ingezette leermiddelen geven 53 leerkrachten (48.6%) aan ook additionele leermiddelen buiten de schoolmethode te gebruiken. Voorbeelden van additioneel materiaal zijn Engelstalige (woordenschat)spelletjes, Squala Engels, flashcards, zelfgemaakte worksheets, werkbladen van internet, Duolingo, Engelstalige Sesamstraat (via Youtube) of school TV, een skype-marathon, 'Energizers' (combinatie (bewegings)spel en lesinhoud) en gebruik van muziek. Ook noemt een enkele leerkracht dat zij Engels inzetten bij projecten of kerst. De inzet van websites van Early Bird, Anglia, British Council, Aduis, Twinkle en Newswise wordt telkens door één leerkracht genoemd. In het kader van gespreksvaardigheid noemt één leerkracht de inzet van VO-leerlingen voor de wekelijkse les gespreksvaardigheid. Ook voor schrijfvaardigheid worden additionele middelen en activiteiten ingezet: Penfriend project, posters maken, interview opstellen en het schrijven van korte tekstjes buiten de methode om.

Om meer inzicht te krijgen in de aard en omvang van de inzet van eventuele andere leermiddelen is gevraagd in hoeverre additioneel materiaal ook authentieke bronnen bevatten en hoe vaak deze worden ingezet (zie Tabel 6.16). Dit was opgenomen als een aparte vraag. Opvallend is dat het totaal aantal leerkrachten dat aangeeft 'af en toe' en 'vaak' additionele middelen in te zetten, hoger is dan het aantal dat in algemene zin aangeeft ook buiten de methode materiaal in te zetten (72 versus 53 van 108 leerkrachten).

Tabel 6.16: inzet additionele leermiddelen (naast de methode)

Inzet van.....		Totaal N=108		EIBO n=73		VVTO n=35	
		N	%	N	%	n	%
authentiek Engels beeld- en geluidsmateriaal	(Bijna) nooit	36	33	29	40	7	20
	Af en toe	50	46	33	45	17	49
	Vaak	22	20	11	15	11	31
authentiek tekstmateriaal	(Bijna) nooit	41	38	30	41	11	31
	Af en toe	54	50	37	51	17	49
	Vaak	13	12	6	8	7	20

### *Het lesprogramma: vaardigheden en activiteiten*

Leerkrachten is gevraagd in welke mate zij aan zeven (deel)vaardigheden aandacht besteden door aan de afzonderlijke vaardigheden een prioritering te geven voor de mate van aandacht die aan

deze vaardigheid wordt besteed. Deze prioritering was een 7-punts likert-schaal, lopend van 1 (minste aandacht) tot 7 (meeste aandacht). Van 107 leerkrachten is informatie ontvangen. Echter hebben drie VVTO leerkrachten niet geantwoord zoals beoogd, maar hebben zij bijvoorbeeld voor elke vaardigheid ‘gemiddeld’ ingevuld. Deze leerkrachten zijn uit de analyse van deze vraag weggelaten. Tabel 6.17 laat zien dat 104 leerkrachten een vaardigheid heeft aangekruist waar zij de meeste tijd aan besteden. Woordenschat en luisteren, respectievelijk 32.7% en 28.8%, zijn daarbij het vaakst aangekruist. Opvallend is dat slechts vijf leerkrachten rapporteren dat zij aan gespreksvaardigheid, m.a.w. de interactie tussen leerlingen en/of met de leerkracht, de meeste aandacht besteden, terwijl 13 leerkrachten de meeste aandacht besteden aan monologisch spreken. Nog opvallender is dat bijna alle leerkrachten die aan spreken en gespreksvaardigheden de meeste aandacht zeggen te besteden, werkzaam zijn op EIBO scholen (17 EIBO leerkrachten versus 1 VVTO-leerkracht).

Tabel 6.17: Vaardigheden waaraan de meeste tijd wordt besteed<sup>67</sup>

(Deel)vaardigheid	Totaal		EIBO		VVTO		EIBO	EIBO	VVTO	VVTO
	N	%	n	%	n	%	< 10	> 10	< 10	> 10
Lezen	9	8.7	6	8.3	3	9.4	4	2	3	-
Luisteren	30	28.8	16	22.2	14	43.8	10	6	9	5
Spreken (monoloog)	13	12.5	12	16.7	1	3.1	10	2	1	-
Schrijven	7	6.7	5	6.9	2	6.3	2	3	1	1
Woordenschat	34	32.7	25	34.7	9	28.1	20	5	4	5
Grammatica	6	5.8	3	4.2	3	9.4	2	1	3	-
Gespreksvaardigheid (interactie)	5	4.8	5	6.9	-	-	2	3	-	-
Totaal	104	100	72	100	32	100	50	22	21	11

De verdeling van de vaardigheden waar *het minste* aandacht aan wordt besteed (Tabel 6.18) laat verder zien dat zes VVTO-leerkrachten (17% van de VVTO leerkrachten) aangeven in groep 8 de minste aandacht te besteden aan gespreksvaardigheid en vier VVTO-leerkrachten (11%) de minste aandacht besteden aan spreken. Aan grammatica wordt, overall gezien, de minste aandacht geschonken: bijna twee vijfde van de scholen in de Peiling. Grammatica lijkt daarbij, relatief gezien, minder van belang te zijn voor VVTO-leerkrachten dan voor EIBO-leerkrachten.

<sup>67</sup> De in Tabel 6.17 gepresenteerde percentages onder EIBO en VVTO geven de percentages van de subgroepen (EIBO en VVTO) weer. Dit geldt ook voor de hiernavolgende tabellen.

Tabel 6.18: Vaardigheden waaraan de minste tijd wordt besteed

(Deel)vaardigheid	Totaal		EIBO		VVTO		EIBO	EIBO	VVTO	VVTO
	N	%	n	%	n	%	< 10	> 10	< 10	> 10
Lezen	6	5.6	5	6.9	1	2.8	4	1	-	1
Luisteren	9	8.4	5	6.9	4	11.4	5	-	4	-
Spreken (monoloog)	21	19.6	17	23.6	4	11.4	10	7	2	2
Schrijven	13	12.1	10	13.9	3	8.6	6	4	3	-
Woordenschat	7	6.5	7	9.7	-	-	6	1	-	-
Grammatica	42	39.3	25	34.7	17	48.6	18	7	11	6
Gespreksvaardigheid (interactie)	9	8.4	3	4.2	6	17.1	2	1	3	3
Totaal	107	100	72	100	35	100	51	21	23	12

In de voorgaande peiling werd het onderscheid tussen spreken en gespreksvaardigheid niet gemaakt. In 2012 gaf 22% van de leerkrachten aan de meeste aandacht aan spreken te besteden. De 17.3% van de huidige leerkrachten die de meeste aandacht aan mondelinge vaardigheid (spreken of gesprekken) zegt te besteden is daarmee lager dan in 2012. Dit is een opvallende uitkomst in relatie tot de prioritering van vaardigheden die is vastgesteld door experts (Rose, 2016). De expertgroep, die in voorbereiding van de huidige peiling de te peilen vaardigheden heeft geprioriteerd, gaf aan dat mondelinge vaardigheid, en met name gespreksvaardigheid, de meest belangrijke vaardigheid in het po zou moeten zijn. Ruime aandacht voor dit onderdeel zou dan in de lijn der verwachting liggen.

Gezien het belang dat wordt gehecht aan gespreksvaardigheid is deze prioritering voor gespreksvaardigheid nader uitgewerkt. Uit Tabel 6.19 kan worden opgemaakt dat de groep leerkrachten die de meeste én de minste aandacht besteden aan gespreksvaardigheid het kleinst is. Ongeveer evenveel leerkrachten besteden vrij veel tot veel aandacht aan gespreksvaardigheid (33.2%) als vrij weinig tot weinig aandacht (35.1%).

Tabel 6.19: mate van aandacht besteed aan gespreksvaardigheid

Mate	Totaal		EIBO		VVTO		EIBO	EIBO	VVTO	VVTO
	n	%	n	%	n	%	< 10	> 10	< 10	> 10
Meeste aandacht	5	4.7	5	7.0	-	-	2	3	-	-
Veel aandacht	14	13.3	12	16.9	2	5.9	10	2	2	-
Vrij veel aandacht	22	21.0	14	19.7	8	23.5	11	3	7	1
Gemiddeld	18	17.1	12	16.9	6	17.6	10	2	2	4
Vrij weinig aandacht	23	21.9	12	16.9	11	32.4	7	5	8	3
Weinig aandacht	14	13.3	13	18.3	1	2.9	8	5	-	1
Minste aandacht	9	8.6	3	4.2	6	17.6	2	1	3	3
Totaal	105	100	71	100	34	100	50	21	22	12

### Lesactiviteiten

Voor vijf vaardigheidsdomeinen is gevraagd welke activiteiten leerkrachten in de les Engels uitvoeren. Leerkrachten konden per voorbeeld aankruisen hoe vaak de leerlingen de activiteit uitvoeren. In tegenstelling tot eerdere peilingen Engels is gedefinieerd wat onder de drie antwoordcategorieën verstaan wordt:

(Bijna) nooit = Dit gebeurt niet of bijna niet;

Af en toe = Dit komt een of twee keer per maand voor;

Vaak = Dit is minimaal één keer per week aan de orde.

Deze definiëring gaf leerkrachten een duidelijke richtlijn voor de betekenis van de categorieën, maar bemoeilijkt de vergelijking met eerdere peilingen. In eerdere peilingen is gevraagd of leerkrachten de activiteit nooit, soms of regelmatig uitvoeren. Er is gerapporteerd over het percentage leerkrachten dat regelmatig een bepaalde activiteit uitvoert, maar onduidelijk is wat de leerkracht hier precies onder verstaat – door het gebrek aan specifieke explicitering van de categorieën. Ook al is de vergelijking dus niet goed te maken, toch zijn in de tabel volledigheidshalve per vaardigheid de percentages leerkrachten die in 2012 ‘regelmatig’ aankruisten, weergegeven. Daarnaast bevat Tabel 6.20 voor deze peiling een onderscheid tussen EIBO- en VVTO-leerkrachten.

Tabel 6.20: Percentages leerkrachten dat genoemde activiteit inzet

	2018						2012
	(bijna) nooit		Af en toe		Vaak		regelmatig
	EIBO	VVTO	EIBO	VVTO	EIBO	VVTO	2012
<b>Spoken (N=109)</b>							
Naspreken van nieuwe woorden	5	-	31	26	64	74	86
Naspreken van standaard gesprekken in het Engels	15	18	50	38	35	44	66
In het Engels vertellen wat er op plaatjes te zien is	19	14	50	46	31	40	41
In het Engels vertellen over een gebeurtenis (b.v. verjaardag, sportactiviteit, weekend, vakantie)	35	31	47	49	18	20	21
Een tekst, boek of film(pje) navertellen in het Engels	64	54	32	34	4	11	19
het zingen van liedjes in het Engels	6	9	49	43	45	49	-



Tabel 6.20: percentage leerkrachten dat genoemde activiteit inzet (vervolg)

Gespreksactiviteiten (N=109)	2018						2012
	(bijna) nooit		Af en toe		Vaak		regelmatig
	EIBO	VVTO	EIBO	VVTO	EIBO	VVTO	2012
Een rollenspel spelen in het Engels	10	20	68	46	23	34	38
Voeren van een gesprek in het Engels met één of meer medeleerlingen	8	9	53	51	39	40	52
Voeren van een gesprek in het Engels met de leerkracht	29	11	58	49	13	40	25
Voeren van een gesprek in het Engels met een ( <i>near</i> ) <i>native speaker</i> (iemand anders dan de leerkracht)	87	77	11	17	3	6	12
Voeren van een gesprek in het Engels met leeftijdgenoten via internet (b.v. Skype)	93	97	7	3	-	-	-
Een spelletje spelen of activiteit doen in het Engels	26	26	54	51	20	23	29
<b>Luister- en kijkactiviteiten (N=108-109)</b>	<b>EIBO</b>	<b>VVTO</b>	<b>EIBO</b>	<b>VVTO</b>	<b>EIBO</b>	<b>VVTO</b>	<b>2012</b>
Luisteren naar een tekst of gesprek in het Engels uit de methode	10	6	22	12	69	82	89
Luisteren naar een tekst of gesprek in het Engels van tv, internet of uit een film ( <u>zonder ondertiteling</u> )	55	40	38	40	7	20	25
Luisteren naar een tekst of gesprek in het Engels van tv, internet of uit een film ( <u>met ondertiteling</u> )	24	34	57	54	19	11	19
Luisteren naar de leerkracht die alleen maar Engels spreekt	26	12	34	38	41	50	57
<u>Nederlandse</u> vragen beantwoorden bij een Engelse luistertekst	26	43	47	34	27	23	32
<u>Engelse</u> vragen beantwoorden bij een Engelse luistertekst	27	29	39	37	34	34	47
<b>Leesactiviteiten</b>	<b>EIBO</b>	<b>VVTO</b>	<b>EIBO</b>	<b>VVTO</b>	<b>EIBO</b>	<b>VVTO</b>	<b>2012</b>
Korte Engelse teksten uit de methode lezen	8	3	29	33	63	64	-
Korte Engelse teksten uit een tijdschrift of op de tablet/computer lezen	56	49	40	40	4	11	9
Langere Engelse teksten uit een boek of tijdschrift of op de tablet/computer lezen	76	69	22	29	1	3	14
<u>Nederlandse</u> vragen beantwoorden bij een Engelse leestekst	25	37	42	46	33	17	25

<b>Leesactiviteiten</b>	EIBO	VVTO	EIBO	VVTO	EIBO	VVTO	2012
Engelse vragen beantwoorden bij een Engelse leestekst	25	26	47	37	28	37	34
Woorden invullen die in een Engelse leestekst zijn weggelaten	32	31	44	40	24	29	21
Een Nederlandse samenvatting geven van een Engelse leestekst	57	51	31	40	13	9	14
<b>Woordenschatactiviteiten</b>	EIBO	VVTO	EIBO	VVTO	EIBO	VVTO	2012
De betekenis van woorden afleiden uit een Engelse context	17	6	33	46	50	49	54
Engelse woorden opzoeken in een woordenboek / -lijst	17	29	44	49	39	23	37
Engelse woorden proberen te omschrijven	19	9	54	71	26	20	-
Een tegenovergesteld woord zoeken (Engels-Engels)	64	54	31	40	6	6	15
Engelse woorden bij afbeeldingen (tekeningen / foto's) zoeken	13	9	36	31	51	60	46
De Nederlandse vertaling geven van een Engels woord	4	9	15	11	81	80	77
Engelse woorden classificeren (in categorieën plaatsen)	38	44	47	50	15	6	20
Engelse woorden uit het hoofd leren	10	11	28	37	63	51	56
Engelse woorden oefenen met de computer	50	71	29	15	21	15	38
<b>Schrijfactiviteiten</b>	EIBO	VVTO	EIBO	VVTO	EIBO	VVTO	2012
Engelse woorden bij afbeeldingen (tekeningen/ foto's) schrijven	12	9	36	26	52	66	55
Engelse zinnen bij afbeeldingen schrijven	23	17	51	51	26	31	33
De Engelse vertaling van Nederlandse woorden opschrijven	14	17	40	49	46	34	54
Engelse woorden en/of zinnen overschrijven	22	23	36	54	42	23	42
Engelse woorden en/of zinnen opschrijven in een dictee	63	49	26	40	11	11	25
Antwoorden op vragen bij een tekst opschrijven in het Engels	32	26	36	51	32	23	32
Een brief, e-mail, WhatsApp, Instagrambericht of SMS schrijven in het Engels	86	71	14	26	-	3	5
Een verhaaltje, gedichtje of liedje schrijven in het Engels	71	65	29	32	-	3	5
Een werkstuk schrijven in het Engels	90	97	8	3	1	-	2

### Doeltaal en werkvormen

Het belang van het gebruik van de doeltaal in de lessen in een vreemde taal is veelvuldig aangetoond (zie bijvoorbeeld, Verspoor, Schuitemaker-King, van Rein, de Bot, & Edelenbos, 2010). Daarnaast gaan zowel de Schijf van Vijf als de uitwerking daarvan voor de lespraktijk, het Vierfasenmodel, er beide vanuit dat leerlingen voldoende in de gelegenheid worden gesteld om de taal zelf actief te gebruiken (Westhoff, 2008). Gevraagd is naar het ingeschatte percentage tijd van de les waarin leerkrachten en/of leerlingen Engels met elkaar spreken. Er waren vijf antwoordcategorieën. De drie vragen naar bestede tijd sluiten elkaar niet uit: de leerkracht kan bijvoorbeeld vragen stellen en de hele klas uitnodigen te antwoorden voor een groot deel van de les. Een voorbeelditem is:

*Welk deel van de lessen Engels wordt gevoerd in de Engelse taal?*

	0 tot 5%	5 tot 30%	30 tot 60%	60 tot 85%	85 tot 100%
Ik spreek Engels met de leerlingen voor:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figuur 6.3 voorbeeld van een item over doeltaal

Tabel 6.21 laat de frequenties en percentages zien. 27.8 procent van de leerkrachten rapporteert voor 30-60% van de tijd Engels met de leerlingen te spreken, terwijl precies de helft (50.0%) van de leerkrachten zegt dit meer dan 60% van de tijd te doen. 7.4 procent van de leerkrachten zegt maar 0-5% van de lestijd Engels te spreken. Dat leerkrachten rapporteren een groot deel van de tijd in het Engels te spreken is in lijn met bevindingen uit de vorige peiling en uit ander onderzoek (Thijs, 2011; Geurts & Hemker, 2012).

De mediaan voor het uitnodigen van de hele klas tot spreken ligt op 30-60% van de tijd, maar alsnog geeft 45% van de leerkrachten aan dit voor 60% of meer van de tijd te doen. 12% van de leerkrachten doet dit voor 0-5% van de tijd. Relatief minder vaak spreken leerlingen met elkaar: de mediaan ligt op 5-30% van de tijd en 13% van de leerkrachten geeft aan dat dit voor 60% of meer van de tijd is. Dertien leerkrachten (17%) geven aan dat leerlingen maar 0-5% van de les Engels in het Engels met elkaar spreken.

Verschillen tussen EIBO en VVTO scholen bij de drie doeltaalitems zijn niet erg groot. Een iets groter percentage leerkrachten op VVTO-scholen spreekt voor meer dan 30% in de doeltaal (81% versus 78%) en nodigt de hele klas uit hetzelfde te doen (72% versus 63%). Het percentage kinderen dat daadwerkelijk voor meer dan 30 procent van de tijd Engels met elkaar spreekt verschilt wel: 37% (EIBO) versus 50% (VVTO). Daarbij geldt overigens dat de categorie 60-85 % van de lestijd in beide typen scholen door een evengroot deel van de klassen gesproken wordt (namelijk 12%)

Tabel 6.21: percentage doeltaal en werkvormen

	Percentage Van de les	Totaal N-108		EIBO n=73		VVTO n=35	
		n	%	n	%	n	%
Ik spreek Engels met de leerlingen voor:	0-5 %	8	7	5	7	3	9
	5-30 %	16	15	12	16	4	11
	30-60%	30	28	21	29	9	26
	60-85 %	30	28	20	27	10	29
	85-100 %	24	22	15	21	9	26
Ik nodig de algehele klas uit om met mij Engels te spreken voor:	0-5 %	13	12	9	12	4	11
	5-30 %	24	22	18	25	6	17
	30-60%	23	21	17	23	6	17
	60-85 %	29	27	19	26	10	29
	85-100 %	19	18	10	14	9	26
De leerlingen spreken Engels met elkaar voor:	0-5 %	18	17	14	19	4	12
	5-30 %	45	42	32	44	13	38
	30-60%	30	28	18	25	12	35
	60-85 %	13	12	9	12	4	12
	85-100 %	1	1	0	-	1	3

### *Klassenorganisatie Engelse les*

Op soortgelijke wijze als bij doeltaal is gevraagd naar klassenorganisatie. Ruim een derde van de leerkrachten rapporteert voor 30-60% van de tijd klassikaal te werken, terwijl 53% van de leerkrachten zegt dit meer dan 60% van de tijd te doen. Zeventien procent van de leerkrachten geeft een bijna volledig klassikale les (85% of meer klassikaal). Het samenwerken van leerlingen in groepjes of tweetallen vindt gemiddeld in 30-60% van de tijd plaats. 72% van de leerkrachten zegt dit in 5-60% van de tijd te doen. In 2% van de lessen wordt er 0-5% samengewerkt door leerlingen. Leerlingen werken relatief weinig individueel: 30% van de leerkrachten geeft aan dat zij dit 0-5% van de les doen. Leerkrachten op VVTO-scholen geven relatief vaker aan dat zij voor meer dan 60% van de les individueel werken dan leerkrachten op EIBO-scholen (17% versus 9%). In EIBO-lessen wordt een relatief iets groter deel van de les besteed aan klassikaal onderwijs.

Tabel 6.22: klassenorganisatie Engelse les

	Percentage Van de les	Totaal (N=108)		EIBO (N=73)		VVTO (N=35)	
		n	%	N	%	n	%
Engels wordt klassikaal gegeven	0-5 %	5	4.6	3	4.1	2	5.7
	5-30 %	9	8.3	6	8.2	3	8.6
	30-60%	37	34.3	24	32.9	13	37.1
	60-85 %	39	36.1	25	34.2	14	35.9
	85-100 %	18	16.7	15	20.5	3	8.6
Leerlingen werken in tweetalen of groepjes	0-5 %	2	1.9	2	2.7	-	-
	5-30 %	37	34.3	24	32.9	13	37.1
	30-60%	40	37.0	28	38.4	12	34.3
	60-85 %	23	21.3	15	20.5	8	22.9
	85-100 %	6	5.6	4	5.5	2	5.7
Leerlingen werken individueel werken aan Engels	0-5 %	32	29.6	26	35.6	6	17.1
	5-30 %	49	45.4	30	41.1	19	54.3
	30-60%	17	15.7	13	17.8	4	11.4
	60-85 %	5	4.6	1	1.4	4	11.4
	85-100 %	5	4.6	3	4.1	2	5.7

### Toetsing

Het is van groot belang te evalueren of leerlingen de gestelde taaldoelen behalen. Idealiter doen leerkrachten dit tijdens en/of aan het eind van iedere les, maar ook over de langere termijn is evaluatie van de beheersing van het geleerde relevant. Leerkrachten is gevraagd naar het soort toetsen dat zij gebruiken. Van de 101 leerkrachten die de vraag hebben beantwoord, zeggen 15 (14.9%) geheel geen toetsen af te nemen (zie Tabel 6.23). Dat is vooral bijzonder omdat het in meer dan de helft van die gevallen gaat om VVTO leerkrachten (specifiek: acht leerkrachten uit het stratum VVTO<10% gewichtsleerlingen). Zestig leerkrachten (59%) rapporteren één toets te gebruiken, terwijl 25 leerkrachten (25%) twee manieren van toetsen noemen. De meeste genoemde manieren zijn de methodetoetsen en de zelfgemaakte toetsen. EIBO leerkrachten maken relatief vaker gebruik van methodetoetsen (52.8% versus 36.1%) dan VVTO leerkrachten, als ook van zelfgemaakte toetsen (40.6% versus 27.8%). De toetsen van Engelstalige instituten en de Cito- en IEP-toets worden enkel in combinatie met toetsen uit de methode of zelfgemaakte toetsen gebruikt. Drie leerkrachten gebruiken alleen andersoortige manieren van toetsen. Drie andere leerkrachten vullen methode- en eigen toetsen aan, samen met andersoortige manieren van toetsen. Andersoortige toetsen die genoemd worden zijn dictees, woordenschattoetsen passend bij aanbod in de les, Duolingo en door een collega samengestelde toetsen.

Tabel 6.23: Gebruikte toetsen<sup>68</sup>

Mate	Totaal		EIBO		VVTO		EIBO	EIBO	VVTO	VVTO
	n	%	n	%	n	%	< 10	> 10	< 10	> 10
Toetsen uit de methode	60	59.4	47	52.8	13	36.1	33	14	6	7
Ik stel zelf toetsen samen	41	40.6	31	34.8	10	27.8	22	9	7	3
Ik zet toetsen in van Engelstalige instituten	1	1.0	-	-	1	2.8	-	-	1	-
Ik zet Me2! Engels (Cito) in	1	1.0	-	-	1	2.8	-	-	1	-
Ik zet de IEP toets Engels in	1	1.0	1	1.1	-	-	1	-	-	-
Engels wordt op school niet getoetst	15	14.9	7	7.9	8	22.2	4	3	8	-
Anders...	6	5.9	3	3.4	3	8.3	2	1	2	1
Totaal antwoorden	125	100	89	100	36	100	62	27	25	11

Leerkrachten is tevens gevraagd of (en zo ja, welke) acties zij ondernemen op basis van de toetsresultaten. Dit geeft een indicatie van het gebruik van toetsdata voor formatieve doeleinden, meer specifiek voor de acties ten behoeve van de achterblijvende leerlingen. Van de 108 leerkrachten die de vraag hebben beantwoord, geeft 69% van de leerkrachten aan (bijna) nooit op basis van achterblijvende prestaties extra ondersteuning aan te bieden. 26% (n=28) van de leerkrachten doet dit incidenteel, terwijl 4% (n=4) structureel extra ondersteuning aanbiedt op basis van zwakke prestaties. Gevraagd naar de aard van deze extra ondersteuning hebben deze 32 leerkrachten één (n=28) of twee (n=17) antwoordopties aangekruist. Meest genoemde vervolgacties waren het geven van verlengde instructie (n=24, 75% van de leerkrachten) en het aanpassen van taken en opdrachten (n=20, 63% van de leerkrachten). Het bieden van extra oefenmateriaal, Remedial teaching en extra oefening in groepjes worden respectievelijk 9 (28%), 2 (6%), en 5 (16%) keer genoemd. Twee leerkrachten hebben aangegeven iets anders te doen: één leerkracht gebruikt internetsites voor extra ondersteuning, een ander laat de achterblijvende leerling met een andere leerling samenwerken, zodat deze leerling van de betere leerling kan leren.

### Getoetste vakgebieden

Net als de mate van aandacht in de lessen voor de verschillende (deel)vaardigheden is ook gevraagd naar welke vaardigheid het *meest* getoetst wordt. In Tabel 6.24 is te zien dat veruit de meeste leerkrachten (68%) woordenschat het vaakst toetsen, gevolgd door lezen (8%), grammatica (7%) en schrijven (6%). Slechts zes leerkrachten geven aan dat zij mondelinge vaardigheden (spreken gespreksvaardigheid) het meeste toetsen.

<sup>68</sup> De gerapporteerde percentages zijn per subgroep, respondenten konden meerdere antwoorden geven.

Tabel 6.24: Vaardigheid die volgens leerkrachten het meest getoetst wordt

(Deel)vaardigheid	Totaal		EIBO		VVTO		EIBO	EIBO	VVTO	VVTO
	N	%	N	%	n	%	<10	>10	<10	>10
Lezen	7	8.0	5		2	8.0	2	3	1	1
Luisteren	4	4.5	3	4.8	1	4.0	3	-	1	-
Spreken (monoloog)	4	4.5	3	4.8	1	4.0	2	1	1	-
Schrijven	5	5.7	3	4.8	2	8.0	2	1	2	-
Woordenschat	60	68.2	46	73.0	14	56.0	33	13	5	9
Grammatica	6	6.8	1	1.6	5	20.0	1	-	3	2
Gespreksvaardigheid (interactie)	2	2.3	2	3.2	-		-	2	-	-
Totaal	88	100	63	100	25	100	43	20	13	12

### Toetsing gespreksvaardigheid

Hiervoor is beschreven dat gespreksvaardigheid niet een van de vaardigheden is waaraan leerkrachten veel aandacht besteden. Het blijkt dat leerkrachten die in enige mate aandacht besteden aan gespreksvaardigheid, N=84, deze vaardigheid weinig of nauwelijks toetsen (bijn 80% geeft aan niet tot weinig te toetsen). Twee leerkrachten hebben zelfs een code 'niet' toegevoegd. In de VVTO strata heeft geen enkele leerkracht aangegeven gespreksvaardigheid (vrij) veel of het meest te toetsen (Tabel 6.25).

Tabel 6.25: mate van toetsing van gespreksvaardigheid

mate	Totaal		EIBO		VVTO		EIBO	EIBO	VVTO	VVTO
	n	%	n	%	n	%	<10	>10	<10	>10
Meest getoetst	2	2.4	2	3.4	-	-	0	2	-	-
Veel getoetst	2	2.4	2	3.4	-	-	2	-	-	-
Vrij veel getoetst	6	7.1	6	10.2	-	-	3	3	-	-
Gemiddeld	7	8.3	4	6.8	3	12.0	3	1	2	1
Vrij weinig getoetst	6	7.1	2	3.4	4	16.0	1	1	2	2
Weinig getoetst	27	32.1	23	39.0	4	16.0	18	5	2	2
Minst getoetst	32	38.1	19	32.2	13	52.0	12	7	6	7
Niet getoetst	2	2.4	1	1.7	1	4.0	1	-	1	-
Totaal	84	100	59	100	25	100	40	19	13	12

### Huiswerk

Voor het vak Engels kunnen leerlingen in groep 8, net als bij taal en rekenen, huiswerk meekrijgen. Zeven op de tien van de leerkrachten geeft aan dat af en toe (46.8%), of zelfs vaak (24.8%) te doen. Bijna dertig procent van de leerkrachten geeft (bijna) nooit huiswerk mee.

Tabel 6.26: Mate van het geven van huiswerk voor Engels

Frequentie huiswerk	Totaal		EIBO		VVTO		EIBO <10	EIBO >10	VVTO <10	VVTO >10
	n	%	n	%	n	%	%	%	%	%
(bijna) nooit	31	28.4	20	27.0	11	31.4	25.5	30.4	36.4	23.1
af en toe	51	46.8	37	50.0	14	40.0	52.9	43.5	31.8	53.8
vaak	27	24.8	17	23.0	10	28.6	21.6	26.1	31.8	23.1
Totaal	109	100	74	100	35	100	100	100	100	100

Van de 78 leerkrachten dat aangeeft huiswerk te geven, hebben 50 ook ingevuld welke vaardigheden dat huiswerk betreft. Leerkrachten hebben soms meerdere vaardigheden aangekruist. Het vaakst genoemd is woordenschat (48 keer), gevolgd door grammatica (15 keer), schrijven (7 keer) en lezen (4 keer). Drie leerkrachten geven aan gespreksvaardigheidsoefeningen mee te geven als huiswerk (alle drie EIBO leerkrachten).

#### *Integratie andere vakken*

In het kader van doeltaal-voertaal kiezen leerkrachten er soms voor om ook tijdens andere vakken/lessen de Engelse taal te gebruiken, bijvoorbeeld in de les wereldoriëntatie, burgerschap, of lichamelijke opvoeding. Ongeveer drie op de tien leerkrachten (27.8%) zegt dit toe te passen. Verdeeld over de vier gecombineerde strata gaat het om respectievelijk 31% (EIBO 0-10), 13% (EIBO>10), 33% (VVTO 0-10) en 31% (VVTO>10)

#### *Overdracht groep 7 naar groep 8*

Aanbod op het gebied van Engels is verplicht in groep 7 en 8. Nagegaan is in hoeverre en hoe de overdracht van groep 7 naar groep 8 verloopt op het gebied van Engels (zie Tabel 6.27). Relatief veel leerkrachten zeggen de ontwikkeling op het gebied van Engels over te dragen door middel van rapportcijfers (38 %) en voor 41% van de leerkrachten was overdracht niet nodig, omdat ze zelf de groepsleerkracht in groep 7 waren. Bijna een kwart van de leerkrachten stemt de vorderingen op Engels mondeling af met de leerkracht van het voorgaande schooljaar. Vier leerkrachten geven aan dat er helemaal geen overdracht plaatsvindt m.b.t. Engels.



Tabel 6.27: Overdracht informatie over Engels van groep 7 naar 8 (*meerdere antwoorden mogelijk*)

Overdracht ....	Totaal		EIBO		VVTO	
	n	%	n	%	n	%
niet nodig, ik was ook leerkracht groep 7	41	41	28	42	13	41
vorderingen in vorm rapportcijfers	38	38	25	37	13	41
vorderingen via overdrachtdossier	9	9	3	4	6	19
mondelinge afstemming met leerkracht groep 7	23	23	14	21	9	28
er is geen overdracht	4	4	3	4	1	3
Totaal	99	100	67	100	32	100

### *Stellingen over Engels in het PO*

Aan de leerkrachten zijn 12 stellingen voorgelegd die gaan over de plus- en minpunten het vak Engels. Tabel 6.28 bevat de gemiddelde scores per stelling voor deze peiling, voor de vorige peiling Engels, en naar stratum. Leerkrachten konden kiezen uit een antwoordregime dat liep van 'helemaal eens' tot 'helemaal oneens' (5). Enkele (negatief geformuleerde) stellingen zijn gehercodeerd zodat een hogere score aangeeft dat een groter deel van de leerkrachten het eens is met de genoemde stelling.

Voor veel van de cellen in de tabel geldt dat de leerkrachten in de verschillende strata het redelijk met elkaar eens zijn. Zo vinden bijna alle leerkrachten dat de niveaoverschillen Engels tussen leerlingen erg groot zijn. De gemiddelde scores voor motivatie bij leerlingen zijn hoger bij de VVTO scholen. Er is verder een vrij hoge consensus bij EIBO en VVTO leerkrachten als het gaat om het startmoment. Leerkrachten kiezen in het algemeen eerder voor het eerste leerjaar als gewenst startmoment dan voor jaar vijf als startmoment. VVTO leerkrachten zijn, naar verwachting, meer voorstander van leerjaar één dan leerjaar vijf. Het vak Engels gaat dan ook niet echt ten koste van de centrale vakken taal en rekenen: in alle vier de groepen ligt het gemiddelde rond de 2,0; de score van leerkrachten op scholen met relatief meer gewichtenleerlingen is echter hoger. Op de vraag of Engels in de eindtoets zou moeten worden opgenomen zijn de meeste leerkrachten niet uitgesproken positief. VVTO leerkrachten uit de groep VVTO>10% zijn hier het meest een voorstander van. In alle groepen zijn leerkrachten het meer eens met de stelling dat de taalvaardigheid van leerkrachten zou moeten worden verbeterd dan met de stelling dat de diactische vaardigheden kunnen worden geoptimaliseerd.

Tabel 6.28: Gemiddelde score stellingen Engels, naar stratum

Stelling	Totaal	EIBO totaal	VVTO totaal	EIBO <10	EIBO >10	VVTO <10	VVTO >10
De meeste leerlingen zijn gemotiveerd voor Engels	3.64	3.57	3.80	3.55	3.61	3.77	3.85
De meeste leerlingen volgen de lessen Engels met plezier	3.63	3.62	3.63	3.61	3.65	3.50	3.85
Engels moet worden aangeboden vanaf groep 1	3.91	3.73	4.26	3.69	3.83	4.27	4.23
Engels moet worden aangeboden vanaf groep 5	2.64	2.86	2.13	2.94	2.70	2.11	2.17
Engels moet worden opgenomen in de eindtoets	2.68	2.56	2.94	2.67	2.30	2.82	3.15
Leerkrachten zijn gemotiveerd voor Engels	3.81	3.74	3.97	3.76	3.70	4.00	3.92
De doelen voor Engels in het primair onderwijs zijn duidelijk over wat leerlingen moeten kunnen	3.25	3.21	3.34	3.27	3.09	3.36	3.31
De meeste leerlingen kennen al aardig wat Engels voordat het op school gegeven wordt	3.08	3.19	2.85	3.33	2.91	2.73	3.08
De aandacht voor Engels gaat ten koste van het niveau voor taal en rekenen	2.09	2.15	1.97	2.08	2.30	1.78	2.31
De niveaoverschillen tussen leerlingen zijn voor Engels groot	4.44	4.44	4.43	4.41	4.52	4.45	4.38
Om Engels te kunnen geven moet de taalvaardigheid van leerkrachten verbeterd worden	3.81	3.74	3.97	3.76	3.70	4.00	3.92
Om Engels te kunnen geven moet de didactische vaardigheid van leerkrachten verbeterd worden	3.19	3.14	3.29	3.16	3.09	3.29	3.31

## 6.4 Resultaten vragenlijst leerlingen

De leerlingvragenlijst is in principe bij alle leerlingen afgenomen. Bruikbare vragenlijstinformatie is aanwezig voor 2075 leerlingen. Evenals de school- en de leerkrachtvragenlijst is de leerlingvragenlijst voor een groot deel gebaseerd op de vragenlijst uit de vorige peiling. Een groot deel van de vragen komt dan ook overeen. In de vorige peiling is echter niet nagegaan of en zo ja, welke onderliggende schalen de vragenlijst bevatte. Schaling heeft in de huidige peiling wel plaatsgevonden. Voor meer informatie over de leerlingvragenlijst en de schaling wordt verwezen naar paragraaf 5.3.2. Op basis van inhoudelijke overwegingen zijn vier schalen onderscheiden: *Attitude*, *Linguistic Self Confidence*, *Buitenschools gebruik van Engels* en *Ingeschatte Vaardigheid*. Binnen deze laatste schaal waren weer vier schalen te onderscheiden voor de vaardigheden lezen, luisteren, spreken en schrijven.

In deze paragraaf bespreken we eerst de gebruikte schalen en de inhoudelijke overwegingen die daaraan ten grondslag lagen. Gepresenteerd worden resultaten die naar de populatie zijn teruggewogen op de 8 combinatiestrata (zie § 5.4). Per schaal wordt naast de landelijke totalen tevens de verdeling in scores gegeven over de vier in dit hoofdstuk onderscheiden groepen scholen. Daarna wordt een beschrijving van de taalachtergrond van de leerlingen gegeven. De beschrijving van de algemene kenmerken van de onderzoeksgroep staat weergegeven in paragraaf 2.5).

### 6.5.1 Attitude

De attitude geeft het systeem weer van positieve of negatieve gevoelens ten opzichte van activiteiten, objecten, personen of ideeën. Wanneer het gaat om het aanleren van een vreemde taal beïnvloedt de attitude de bereidheid om de taal te leren en de openheid die de leerling heeft ten opzichte van de tweede taal en de bijbehorende (socio)culturele kenmerken. Een hogere attitude bevordert dan ook het aanleren van die tweede taal (Masgoret & Gardner, 2003). Ook de kerndoelen voor Engels in het basisonderwijs zijn gericht op het bevorderen van een positieve attitude: kerndoel 14 gaat in op het ontwikkelen van een attitude waarbij leerlingen zich durven te uiten in het Engels.

Evenals in de vorige peiling is in de huidige peiling het construct "attitude" geoperationaliseerd door vragen waarin leerlingen kunnen aangeven of ze Engels leuk, makkelijk en belangrijk vinden. De vragen naar attitude waren grotendeels gelijk aan die in de vorige peiling. Drie vragen toegevoegd. Leerlingen is gevraagd of ze Nederlands een moeilijk vak vinden en of ze Engels praten in de les. Leerlingen moesten verder noteren wat ze leuk vinden in de Engelse les (vraag 10). Naast de toevoeging van deze drie vragen is de analyse van de resultaten gewijzigd ten opzichte van de vorige peiling. Zoals in hoofdstuk 5 is beschreven, zijn voor de huidige peiling schalen geconstrueerd op inhoudelijke gronden. Inhoudelijke analyse leidde ertoe de vragen die in de vorige peiling onder de term "attitude" werden gerapporteerd, op te splitsen in vragen naar twee onderliggende constructen: attitude en linguistic self-confidence (zie 6.5.2). In Tabel 6.29 is een overzicht gegeven van de items uit leerlingvragenlijst die bij de schaal Attitude horen. Daarna worden twee voorbeelditems gegeven.

Tabel 6.29: schaal attitude

Schaalscore	Items
Attitude	1, 3, 4, 10.1, 10.2, 10.3, 10.6 <sup>69</sup>

*Voorbeelditems attitude:*

Vind je Engels op school een leuk vak?
<input type="checkbox"/> Nee, ik vind Engels niet echt leuk.
<input type="checkbox"/> Ik vind het niet zo'n heel leuk, maar ook geen vervelend vak.
<input type="checkbox"/> Ja, ik vind Engels een leuk vak.

Vind je het belangrijk voor jouw toekomst om goed Engels te leren?
<input type="checkbox"/> Nee, ik vind Engels niet zo belangrijk
<input type="checkbox"/> Ik heb er geen mening over.
<input type="checkbox"/> Ja, ik vind Engels belangrijk.

In Tabel 6.30 is de verdeling van gewogen scores weergegeven voor alle leerlingen en voor de leerlingen in de vier typen scholen. De schaal attitude voor Engels loopt van 0 – 10. Gemiddeld hebben leerlingen een (vrij) positieve attitude ten opzichte van Engels ( $M=6.63$ ). Een Anova toont dat de attitudes van de leerlingen op de vier onderscheiden groepen scholen significant verschillen ( $F(3)=13.01, p<0.01$ ). Post hoc analyses laten zien dat de attitudes voor Engels van leerlingen in EIBO-scholen met weinig gewichtenleerlingen ( $M=6.46, S.D.=1.94$ ) significant lager zijn dan die van leerlingen in EIBO-scholen met veel gewichtenleerlingen ( $M=7.05, S.D.=1.90$ ). Ditzelfde geldt voor de attitude van leerlingen in VVTO-scholen met relatief weinig gewichtenleerlingen ( $M=6.51, S.D.=1.99$ ), wier attitude ook lager is dan die van leerlingen in EIBO-scholen met relatief veel gewichtenleerlingen ( $M=7.14, S.D.=1.90$ ). Hoewel de gemiddelde attitude voor Engels van leerlingen in de VVTO-scholen met relatief veel gewichtenleerlingen hoger is, is er geen sprake van significante verschillen met de andere twee groepen (de twee EIBO-groepen). Hier speelt het geringe aantal leerlingen in cluster 4 (VVTO >10) een rol. Toetsing van het verschil tussen alle leerlingen op VVTO-scholen en alle leerlingen op EIBO-scholen, met behulp van een onafhankelijke t-toets op basis van gewogen scores, laat zien dat er geen significant verschil tussen leerlingen op beide typen scholen bestaat ( $M_{VVTO}=6.65, M_{EIBO}=6.63, t(df=2063)=0.11, p=0.87$ ). Een vergelijking van beide typen scholen (EIBO versus VVTO) laat geen significant verschil zien ( $t(df=2057)=-.19, p=0.91$ ).

<sup>69</sup> Items 10.1/2,3,6 zijn gehercodeerd.

Tabel 6.30: Verdeling gewogen scores per cluster

Attitude	N	Missing	Gem.	S.D.	Min	P25	P50	P75	Max
Totaal	2059	29	6.63	1.95	0	5	7	8	10
EIBO totaal	1836	28	6.62	1.95	0	5	7	8	10
EIBO (0-10)	1322	24	6.46	1.94	0	5	7	8	10
EIBO (>10)	514	4	7.05	1.90	0	6	7	8	10
VVTO totaal	222	1	6.64	1.98	0	5	7	8	10
Vvto (0-10)	177	1	6.51	1.99	0	5	7	8	10
Vvto (>10)	45	0	7.14	1.90	1	6	7.7	9	10

### 6.5.2 Linguistic self confidence

Naast een positieve attitude ten opzichte van de taal is het voor de ontwikkeling van een vreemde taal belangrijk dat leerlingen de taal durven te gebruiken, de 'L2 self confidence van leerlingen'. MacIntyre, Clément, Dörnyei en Noels (1998) omschrijven dit construct als de overtuiging van de tweedetaal leerder om effectief en adaptief te kunnen communiceren in de doeltaal. Het construct wordt geacht zowel een cognitieve – gericht op zelfevaluatie van de eigen vaardigheid in de tweede taal - als ook een affectieve component te bevatten – gericht op het wel of geen angst hebben om de taal te gebruiken (Clément & Kruidenier, 1985, Clément, Dörnyei, & Noels, 1994). Leerders met een hogere L2 self confidence komen meer in aanraking met zowel input en output in de tweede taal, hetgeen de verwerving van de tweede taal positief beïnvloedt. In het navolgende spreken wij van Linguistic Self Confidence.

In Tabel 6.31 is een overzicht gegeven van de items uit de leerlingvragenlijst die bij de schaal Linguistic Self Confidence horen. Daaronder worden twee voorbeelditems getoond.

Tabel 6.31: Schaal Linguistic Self Confidence

Schaalscore	Items
Linguistic Self Confidence	2, 5, 6, 7, 11

#### Voorbeelditems Linguistic Self Confidence

Stel je voor: Je bent op vakantie. Kinderen vragen in het Engels of je mee wilt spelen. Wat doe je?
<input type="checkbox"/> Je schudt nee, want je denkt dat je Engels niet goed genoeg is.
<input type="checkbox"/> Je knikt ja, maar maakt in je beste Engels duidelijk dat je ze niet zo goed verstaat.
<input type="checkbox"/> Je zegt "yes" en speelt in het Engels mee.

Stel je voor: er is op tv een Engelse tekenfilm die je graag wilt zien, maar de film heeft geen ondertiteling. Wat doe je?
<input type="checkbox"/> Ik zet de film uit.
<input type="checkbox"/> Ik probeer of ik de film kan volgen.
<input type="checkbox"/> Ik kijk de film gewoon zonder ondertiteling.

In Tabel 6.32 is de verdeling van gewogen scores weergegeven voor alle leerlingen en voor de leerlingen in de vier typen scholen. Leerlingen in Nederland scoren gemiddeld 8.13 op de schaal Linguistic Self Confidence, die van 0 tot 12 loopt. Dit betekent dat ook de Linguistic Self Confidence van leerlingen vrij positief is. Een Anova laat zien dat er een significant verschil is in attitude tussen de leerlingen op de vier onderscheiden groepen scholen ( $F(3)=3.35, p<0.02$ ). Post hoc analyses laten zien dat de Linguistic Self Confidence van leerlingen in de scholen met EIBO met 0-10% gewichtenleerlingen ( $M=8.00, S.D.=2.52$ ) significant lager ligt dan die van leerlingen in EIBO-scholen met meer dan 10% gewichtenleerlingen ( $M=8.37, S.D.=2.45$ ). De gemiddelde LSC van de overige groepen leerlingen verschilt niet significant. Wanneer gekeken wordt naar de vraag of de Linguistic Self Confidence van leerlingen ook verschilt op de hoofdcategorieën VVTO of EIBO scholen, dan wordt geen significant verschil gevonden ( $t(df=2032) = 1.40, p=0.17$ ).

Tabel 6.32: Verdeling gewogen scores per cluster

LSC	N	Missing	Gem.	S.D.	Min	P25	P50	P75	Max
NL	2034	54	8,13	2,50	0	6	8	10	12
EIBO totaal	1817	48	8.10	2.51	0	6	8	10	12
EIBO (0-10)	1314	32	8,00	2,52	0	6	8	10	12
EIBO (>10)	503	16	8,37	2,45	1	7	9	10	12
VVTO totaal	217	7	8.35	2.46	1	7	9	10	12
Vvto (0-10)	172	6	8,31	2,49	1	6.7	9	10	12
Vvto (>10)	45	1	8,49	2,35	2	7	8.6	10	12

### 6.5.3 Buitenschools gebruik van het Engels

Zoals in hoofdstuk 1 reeds is beschreven, is de mate van contact met de vreemde taal gerelateerd aan de prestaties van leerlingen. Zowel op de prestaties op lezen, luisteren als grammaticale vaardigheid zijn positieve effecten van extramuraal contact met de vreemde taal gevonden (Ellie project, Munoz, et al., 2018).

Het buitenschools gebruik van leerlingen is op een aantal manieren bevraagd. In acht vragen konden leerlingen noteren of ze een bepaalde activiteit in het Engels a) nooit, b) 1 à 2 keer of c) wel vaker uitvoeren. Bij de andere vragen wordt de leerlingen gevraagd naar de frequentie van activiteiten waarin zij gebruikmaken van het Engels.

#### *Voorbeelden van vragen over buitenschools gebruik*

Heb je wel eens in het Engels gechat, bijvoorbeeld via een chatprogramma of via een online game?
<input type="checkbox"/> Ja, dat doe ik wel vaker in het Engels.
<input type="checkbox"/> Ja, één of twee keer.
<input type="checkbox"/> Nee, dat heb ik nog nooit gedaan in het Engels.

Hoeveel tijd besteed je per dag ongeveer aan het lezen, luisteren of schrijven in het Engels? Denk aan Youtube, gamen, social media, Spotify, Netflix, series op tv of Engels praten met bijvoorbeeld je Engelstalige buurjongen of klasgenootje.
<input type="checkbox"/> 0 minuten
<input type="checkbox"/> Tussen 1 en 10 minuten
<input type="checkbox"/> Tussen 10 en 30 minuten
<input type="checkbox"/> Tussen 30 en 60 minuten
<input type="checkbox"/> Meer dan 1 uur

Zoals in paragraaf 5.3.2. staat beschreven zijn er twee schalen voor buitenschools gebruik gemaakt, die allebei betrouwbaar zijn. In dit rapport wordt over de schaal VL\_BsG15 gerapporteerd. In Tabel 6.33 is een overzicht gegeven van de items uit de leerlingvragenlijst die bij de schaal Buitenschools gebruik van het Engels horen.

Tabel 6.33: Schaal buitenschools gebruik Engels

Schaalscore	Items
Buitenschools gebruik	12, 13, 14.1/2/3/4/5/6, 15, 16, 18, 19, 20, 21, 22

In Tabel 6.34 wordt de verdeling van de gewogen scores voor buitenschools gebruik gepresenteerd. De schaal loopt van 0 tot en met 38. Gemiddeld scoren leerlingen in Nederland 18.96, hetgeen betekent dat zij buitenschools weliswaar in contact komen met het Engels via diverse activiteiten, maar dat de meeste leerlingen verre van de maximale score zitten ( $p75_{\text{landelijk}}=24$ ). Een maximale score zou echter betekenen dat leerlingen dagelijks bezig zijn met zowel Engelstalige films, muziek, internet etc, en bijvoorbeeld ook vaak met mensen op vakantie Engels praat en met onbekenden. Het is zeer de vraag of het dagelijks uitvoeren van al deze activiteiten een realistisch scenario zou zijn. Een onafhankelijke t-toets laat geen verschil in buitenschools gebruik van het Engels tussen leerlingen op EIBO of VVTO-scholen zien ( $t(df=1953) = -0.34$ ,  $p=0.74$ ). Wanneer echter gekeken wordt naar verschillen in de subgroepen met behulp van een Anova, dan zien we dat er wel een significant verschil is in buitenschools gebruik tussen de leerlingen op de vier onderscheiden groepen scholen ( $F(3)=5.38$ ,  $p=0.01$ ). Ook hier laten post hoc analyses zien dat er een significant verschil is in het buitenschoolse gebruik van leerlingen in de EIBO-scholen met 0-10% gewichtenleerlingen ( $M=18.53$ ,  $S.D.=6.98$ ) en dat van leerlingen in EIBO-scholen met meer dan 10% gewichtenleerlingen ( $M=20.01$ ,  $S.D.=6.81$ ). Leerlingen op EIBO-scholen met 0-10% gewichtenleerlingen rapporteren buitenschools minder vaak Engels te gebruiken dan leerlingen op EIBO-scholen met meer dan 10% gewichtenleerlingen. Het gemiddelde buitenschoolse gebruik van de overige groepen leerlingen verschilt niet significant van elkaar.

Tabel 6.34: Verdeling gewogen scores per cluster

Buitenschools gebruik	N	missing	Gem.	S.D.	Min	P25	P50	P75	Max
NL	1955	133	18.96	6.92	0	14	19	24	38
EIBO totaal	1754	111	18.94	6.96	0	14	19	24	38
EIBO (0-10)	1272	74	18.53	6.98	0	14	18	23	38
EIBO (>10)	482	37	20.01	6.81	4	15	20	25	38
VVTO totaal	201	22	19.11	6.61	2	14	19	24	37
VVTO (0-10)	158	20	19.10	6.58	2	14.5	19	24	37
VVTO (>10)	43	2	19.16	6.78	5	14	19	24.3	37

### 6.5.4 Zelfingeschatte vaardigheden Engels

Er zijn vier schalen geconstrueerd op basis van de zelf ingeschatte vaardigheden van leerlingen. In de vragenlijst wordt leerlingen aan de hand van stellingen gevraagd hun eigen vaardigheid voor lezen, luisteren, schrijven en spreken in te schatten. De stellingen die de leerlingen voorgelegd kregen waren vereenvoudigde versies van de *can do*-descriptoren binnen het ERK, variërend van A1, A2 en B1-niveau.

Het voorbeeld hieronder toont de stellingen waarvoor de leerlingen moesten aankruisen of ze de activiteit al wel of nog niet beheersten. De ERK-niveaus waren niet zichtbaar voor de leerlingen, maar staan hier weergegeven ter verduidelijking. Opgemerkt moet worden dat item 18 (schrijven) geen onderdeel uitmaakt van de schaal voor schrijven. Dit betreft namelijk het invullen van naam, adres en nationaliteit (zie paragraaf 5.3.2 voor meer informatie over de schaling van de ingeschatte vaardigheid). De schalen voor lezen en schrijven bevatten derhalve ieder 4 items (mogelijke scorerange 1-4) en de schalen voor luisteren en spreken bevatten 5 items (mogelijke score-range 1-5).

Wat kun jij al in het Engels? Kruis steeds 'ja' of 'nog niet' aan

Lezen van Engels		
1. Ik kan informatie vinden in menu's en in folders (A2)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
2. Ik kan borden en posters met eenvoudige woorden en zinnen begrijpen (A1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
3. Ik kan beschrijvingen van gebeurtenissen of gevoelens in brieven, mails of chats begrijpen (B1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
4. Ik kan korte, eenvoudige e-mails, sms-berichten en WhatsApp berichten lezen (A2)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
Luisteren naar Engels		
5. Ik kan simpele zinnen begrijpen in een gesprek over familie en het dagelijks leven (A1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
6. Ik kan de belangrijkste punten in omroepberichten volgen (A2)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
7. Ik kan woorden en zinnen verstaan in een gesprek over school en hobby's (A2)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet



<b>Luisteren naar Engels</b>		
8. Ik kan de hoofdzaken begrijpen van Engelstalige radio- en tv-programma's (B1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
9. Ik kan de hoofdzaken begrijpen van Engelstalige filmpjes op internet (B1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
<b>Engels spreken</b>		
10. Ik kan mijn mening in het Engels geven en daarbij een uitleg geven (B1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
11. Ik kan in eenvoudige woorden en zinnen vertellen over mijn familie of omgeving (A2)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
12. Als mensen langzaam spreken en dingen herhalen, kan ik deelnemen aan een gesprek in het Engels (A1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
13. Ik kan eenvoudige vragen stellen en korte antwoorden geven (A1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
14. Ik kan de inhoud van een boek of film navertellen in het Engels (B1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
<b>Schrijven in het Engels</b>		
15. Ik kan een vakantiekaart, e-mail, WhatsApp- of chatbericht schrijven in het Engels (A1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
16. Ik kan een tekst schrijven over bijvoorbeeld mijn huis, sport of hobby's (B1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
17. Ik kan aantekeningen maken in het Engels (A2)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet
18. Ik kan mijn ervaringen beschrijven in het Engels (B1)	<input type="checkbox"/> ja	<input type="checkbox"/> nog niet

In de Tabel 6.35 staan de gewogen gemiddeldes voor de zelf ingeschatte vaardigheden voor lezen, luisteren, spreken en schrijven weergegeven. Uit Tabel 6.35 kan worden afgeleid dat leerlingen zichzelf met name op het gebied van schrijven relatief laag inschatten. Dit kan verklaard worden door de beperkte aandacht aan schrijfvaardigheid in de methodes.

Tabel 6.35: Ingeschatte vaardigheden

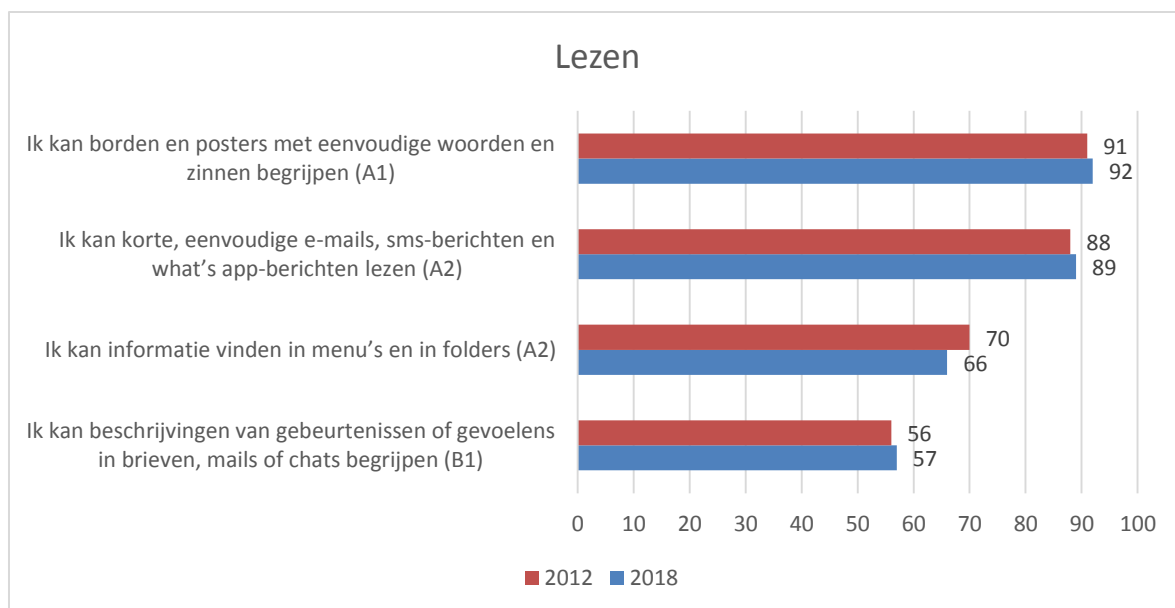
	<b>Inschatting Lezen (M, SD)</b>	<b>Inschatting Luisteren (M, SD)</b>	<b>Inschatting Spreken (M, SD)</b>	<b>Inschatting Schrijven (M, SD)</b>
<b>NL</b>	3.05 (1.03)	3.70 (1.25)	3.30 (1.19)	2.18 (1.47)
<b>EIBO totaal</b>	3.04 (1.04)	3.68 (1.26)	3.29 (1.19)	2.17 (1.47)
EIBO (0-10)	3.01 (1.05)	3.66 (1.28)	3.26 (1.19)	2.11 (1.46)
EIBO (>10)	3.13 (1.01)	3.74 (1.20)	3.36 (1.18)	2.31 (1.49)
<b>VVTO</b>	3.11 (0.97)	3.87 (1.14)	3.42 (1.17)	2.32 (1.50)
VVTO (0-10)	3.11 (0.96)	3.89 (1.13)	3.42 (1.17)	2.26 (1.52)
VVTO (>10)	3.14 (0.99)	3.81 (1.21)	3.41 (1.19)	2.58 (1.41)

Anova's laten zien dat er geen significante verschillen zijn tussen de leerlingen op de vier onderscheiden groepen scholen wat betreft hun zelfinschatting voor lezen ( $F(3)=1.89$ ,  $p=0.13$ ), luisteren ( $F(3,177.66)=1.93$ ,  $p=0.09$ ) en spreken ( $F(3)=1.55$ ,  $p=0.20$ ). Wel blijken de groepen

leerlingen significant van elkaar te verschillen wat betreft hun zelfinschatting voor schrijfvaardigheid ( $F(3)=3.33$ ,  $p=0.02$ ). Nadere analyse van de verschillen tussen de specifieke groepen, met behulp van de Tukey post hoc tests, lieten echter geen significante verschillen zien. Een vergelijking van de twee hoofdcategorieën (VVTO en EIBO) laat een ietwat ander beeld zien. Leerlingen verschillen niet in hun inschatting op de vaardigheden lezen, spreken en schrijven, maar wel op luisteren. Bij luisteren schatten de leerlingen op VVTO-scholen hun vaardigheid significant ( $p=0.04$ ) hoger in dan de totale groep EIBO-leerlingen.

De schaalgemiddelden geven niet aan welke activiteiten de leerlingen denken te beheersen en welke niet. In de Figuren 6.8 – 6.11 is voor de verschillende items weergegeven welk percentage van de leerlingen inschat deze activiteit te beheersen. Deze percentages worden vergeleken met de uitkomsten uit de vorige peiling.

Bij lezen schat een groot deel van de leerlingen in dat zij het A1-niveau beheersen (92%). De inschatting over de beheersing van de twee A2-items is meer divers: Bijna 90% van de leerlingen verwacht korte digitale berichten te kunnen lezen, terwijl 66% van de leerlingen denkt informatie in menu's en folders te kunnen lezen. Deze percentages zijn vergelijkbaar met die van de vorige peiling<sup>70</sup>.

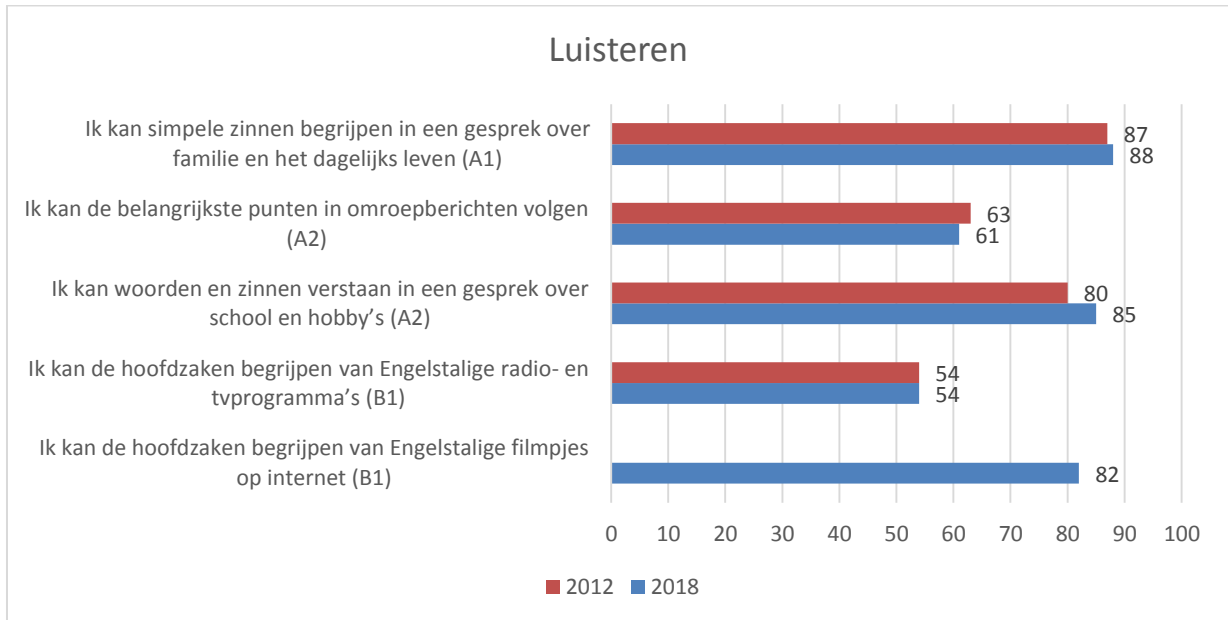


Figuur 6.8: Percentages leerlingen dat de beschreven leesactiviteiten zegt te beheersen (2012 en 2018)

Wat betreft de eigen inschatting van de luistervaardigheid blijkt uit Figuur 6.9 dat leerlingen vooral denken dat zij gesproken informatie over familie en dagelijks leven en over school en hobby's goed kunnen begrijpen (A1 en A2). Over het kunnen volgen van omroepberichten en met name van Engelstalige programma's wordt door minder leerlingen als beheerste activiteiten gerapporteerd (A2 en B1-beschrijvingen) zijn leerlingen minder positief. Ten opzichte van de vorige peiling zijn ook hier de percentages vergelijkbaar, ware het niet dat in 2012 80% van de leerlingen aangaf woorden en zinnen in een gesprek over school en hobby's te kunnen verstaan en dat 63% de belangrijkste punten in een omroepbericht kon volgen. Het kunnen begrijpen van de hoofzaken van Engelstalige filmpjes

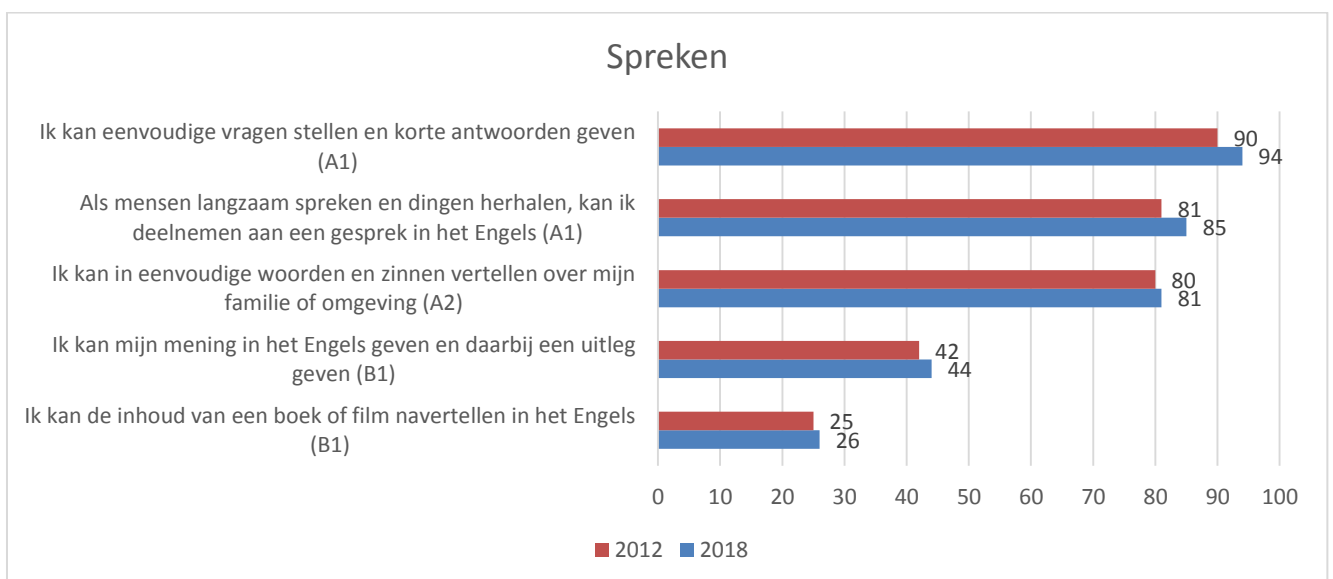
<sup>70</sup> De vergelijking is echter enkel beschrijvend van aard. Toetsing op significantie was door gebrek aan data over 2012 niet mogelijk.

op internet is ingeschaald als B1-taak, maar 82% van de leerlingen verwacht dit te kunnen<sup>71</sup>. Een mogelijke verklaring kan zijn dat leerlingen veelvuldig internetfilmpjes bekijken en hierbij de indruk hebben dat zij ze grotendeels begrijpen. De moeilijkheidsgraad van dit soort filmpjes is daarbij wellicht relatief laag.



Figuur 6.9: Percentages leerlingen dat de beschreven luisteractiviteiten zegt te beheersen (2018 en 2012)

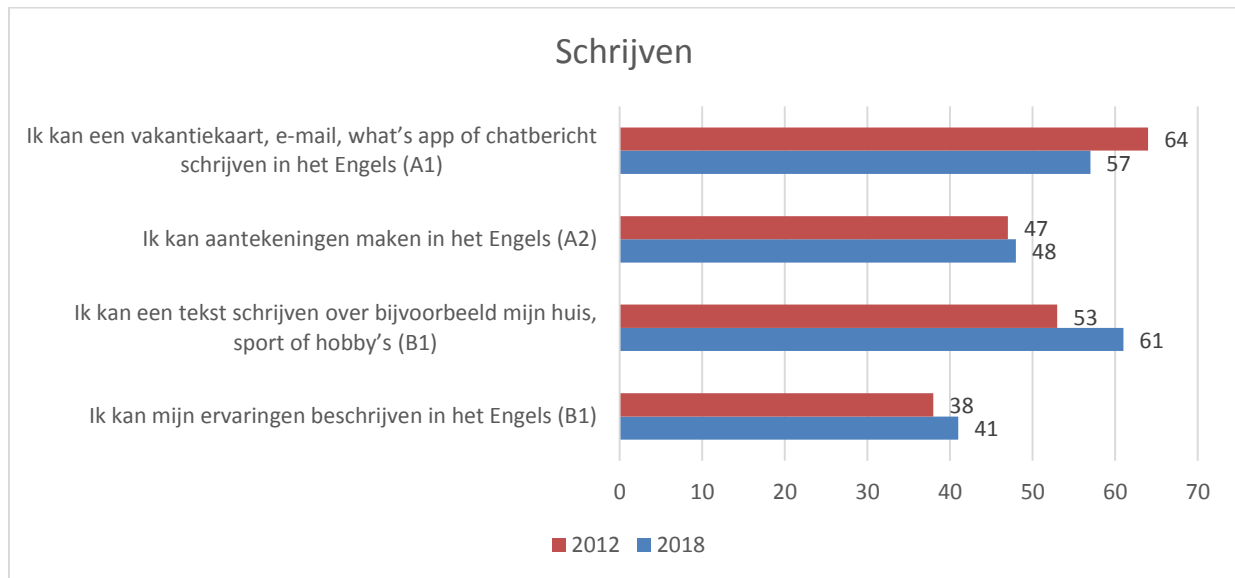
Ook over de beheersing van A1 A2-spreek/gespreksactiviteiten zijn de leerlingen positief: meer dan 80% denkt zowel de A1 A2-activiteiten te kunnen uitvoeren. Over de B1-activiteiten als het navertellen van een boek of film of een mening onderbouwen zijn zij minder zeker. Respectievelijk een kwart en 44% van de leerlingen denkt dit te kunnen. Ook de uitkomsten voor de inschatting van spreken zijn vergelijkbaar met de vorige peiling.



Figuur 6.10: percentages leerlingen dat de beschreven spreekactiviteiten zegt te beheersen (2018 en 2012)

<sup>71</sup> Deze vraag is in 2012 niet gesteld.

De laatste vaardigheid die leerlingen moesten inschatten was de schrijfvaardigheid. Ook hier zijn de uitkomsten vergelijkbaar met de vorige peiling. Een iets hoger percentage dacht in 2012 het formulier en te kunnen invullen en een tekst over hobby's te kunnen schrijven (83% en 64%), terwijl een iets lager percentage de ansichtkaart, de aantekeningen en de ervaringen in 2012 dachten te kunnen beschrijven (53%, 47% en 38%).



Figuur 6.11: percentages leerlingen dat de beschreven schrijfactiviteiten zegt te beheersen (2018 en 2012)

### 6.5.5 Taalachtergrond van de leerlingen

De taalachtergrond van leerlingen is in kaart gebracht aan de hand van feitelijke vragen naar de thuistaal ('welke talen spreek je thuis?') en naar de taal die de leerlingen met hun vader en moeder (of verzorgers) spreken. In tegenstelling tot de vorige peiling is niet gevraagd naar de moedertaal van de leerlingen. Hiervoor waren twee redenen. Allereerst is het construct 'moedertaal' complex, omdat de vraag wat precies als moedertaal moet worden beschouwd niet evident is. Ten tweede bleek dat het vragen naar de moedertaal lastig was voor leerlingen. In de pilot maakte de vraag naar moedertaal nog wel deel uit van de vragenlijst, maar leidde zij relatief vaak tot vragen bij leerlingen.

De leerlingen is gevraagd naar de taal die zij thuis spreken. Van de 2044 leerlingen die de vraag hebben beantwoord (97,9% van de 2088 deelnemende leerlingen) spreken 1985 leerlingen (97%) thuis Nederlands. Voor 59 leerlingen was dit niet het geval (3%). Tweeëntwintig procent (22%) van de leerlingen, te weten 457 leerlingen, antwoordde dat zij thuis (ook) Engels spraken. Dit is lager dan in de vorige peiling, toen dit percentage op 27% van de leerlingen lag. Ongeveer 24% van de leerlingen antwoordde thuis (ook) een andere taal te spreken dan Nederlands of Engels. Dit betrof 486 leerlingen. Een precieze vergelijking met 2012 is hier niet mogelijk vanwege de gewijzigde vraagstelling. Het percentage leerlingen dat een andere thuistaal dan Engels had lag op 11%, terwijl in 2018 20% a 21 % van de leerlingen een andere taal met respectievelijk vader of moeder zegt te spreken. Dit lijkt een hoger percentage, maar voorzichtigheid is gepast: in 2018 is niet specifiek naar thuistaal gevraagd en volledige vergelijkbaarheid is niet op zijn plaats.

Gevraagd naar de taal die de leerlingen met hun moeder en vader spreken, antwoordde 95% dat ze Nederlands met hun moeder spreken en dat 94% dat met hun vader doet. Twaalf procent (12%) van de leerlingen spreekt (ook) Engels met de moeder en 12% spreekt (ook) Engels met de vader. Een andere taal wordt door 21% van de leerlingen met de moeder gesproken, terwijl 20% van de leerlingen met de vader (ook) een andere taal spreekt. Van de 2027 leerlingen die de vraag naar het geboorteland van hun moeder hebben ingevuld, antwoordden 1672 leerlingen dat hun moeder in Nederland geboren was (83%). Het geboorteland van de vader is door 2036 leerlingen benoemd: van 1673 leerlingen is het geboorteland van hun vader Nederland (82%).

## Hoofdstuk 7. Standaardbepaling

In dit hoofdstuk staat de standaardbepaling centraal. Gestart wordt met een beknopt overzicht van doel en werkwijze bij de standaardbepaling. Daarna volgt in paragraaf 7.1 een beschrijving van de samenstelling van het expertpanel dat de standaarden heeft bepaald. Vervolgens wordt in paragraaf 7.2 de gekozen methodieken voor de standaardbepaling beschreven, alsmede de overwegingen om deze methodieken toe te passen. Paragraaf 7.3 is gewijd aan de procedure van de standaardbepaling. In deze paragraaf wordt ingegaan op de voorbereiding, planning en het uiteindelijke verloop van de standaardbepaling. De resultaten van de standaardbepaling worden in paragraaf 7.4 besproken. In de paragrafen 7.4.1 en 7.4.2. staan de bepaalde grenswaarden voor de A1-, A2- en B1-niveaus centraal, waarna in paragraaf 7.4.3 voor lezen en luisteren voorbeelden worden getoond van opgaven die leerlingen met deze niveaus *nét* of goed beheersen. Voor woordenschat konden geen standaarden worden bepaald. Interpretatie van de vaardigheid van leerlingen voor woordenschat wordt echter op een andere manier verkregen (paragraaf 7.5).

Voorafgaand dient een voorbehoud te worden gemaakt. Leerlingen die de gesprekstaken hebben uitgevoerd, hebben vanwege de begrensde beschikbare tijd een zeer beperkt aantal taken uitgevoerd waardoor de standaarden niet representatief zijn voor de beheersing van de eisen die het ERK stelt aan gespreksvaardigheid in het algemeen. Hier moet rekening mee gehouden worden bij de interpretatie van de vastgestelde standaarden. Een ander punt van aandacht is dat de beschikbare tijd voor de standaardbepaling beperkt was. Die randvoorwaarde – er kon slechts uitgegaan worden van een beperkte beschikbaarheid van de experts – heeft ook een rol gespeeld bij de keuze van de methode van standaardbepaling, met name bij de gespreksvaardigheden.

### *Standaardbepaling*

Om de resultaten op de vaardigheden van de leerlingen te kunnen koppelen aan ERK-niveaus<sup>72</sup> is op 29 en 30 augustus 2018 een standaardbepalingsonderzoek uitgevoerd. Het betreft een standaardbepaling voor de ERK-niveaus A1, A2 en B1. Deze standaarden maken het mogelijk om een inschatting te maken van het aantal leerlingen in groep 8 dat deze ERK-niveaus haalt. Daarnaast maakt de koppeling van de gebruikte vaardigheidsschaal aan de vastgestelde standaarden het mogelijk om deze vaardigheidsschaal beter te interpreteren.

Er zijn voor drie vaardigheden standaarden bepaald, te weten lezen, luisteren en gespreksvaardigheid. Per vaardigheid heeft een groep experts de hierboven genoemde ERK-niveaus (A1, A2 en B1) vastgesteld. Vocabulairekennis is niet rechtstreeks aan ERK-niveaus gekoppeld. Voor die vaardigheid zijn er dan ook geen standaarden bepaald, maar is gekozen voor een alternatieve wijze om de waarden op de vaardigheidsschaal te kunnen interpreteren (zie §7.5).

Bij het bepalen van de standaarden is aan de deelnemers gevraagd om voor ieder van de drie vaardigheden per ERK-niveau aan te geven wat de leerling die dat niveau *nét* beheerst (de zogenaamde ‘grensleerling’) kan en wat nog niet. Op deze manier zijn zogenaamde “cut-scores” (cesuren) bepaald. Dat is de discrete schaalscore die net voldoende is om indicatief te zijn voor het betreffende ERK-niveau, of met andere woorden, de scores die de grens tussen wel of niet het beheersen van het ERK-niveau weergeven. Wanneer de leerling precies die cut-score haalt, beheerst

---

<sup>72</sup> ERK is de Nederlandse afkorting voor Europees Referentiekader - in het Engels Common European Framework of Reference (kortweg CEFR of CEF). Voor meer informatie zie: [www.erk.nl](http://www.erk.nl)

de leerling het betreffende niveau nét, en wanneer de leerling net een punt lager heeft, beheerst zij/zij het nog net niet. Deze "cut-score" wordt verkregen doordat de experts bij ieder van de vaardigheden de opgaven evalueren.

Bij de schriftelijke toetsen lezen en luisteren heeft een itemgerichte standaardsetting plaatsgevonden. Bij luisteren bijvoorbeeld wordt gevraagd een inschatting te geven bij verzamelingen van luisteritems (dat wil zeggen, de combinatie van geluidsfragment en vraag) en aan te geven hoeveel een 'A1-grensleerling' goed maakt. De standaardbepaling voor productieve vaardigheden, in ons geval de gespreksvaardigheid, had een ander karakter: hier is gebruik gemaakt van een leerlinggerichte benadering, omdat de uitingen van de leerling in een gesprekstaak een eenheid vormen. Spreekvaardigheid is gemeten met een ankertaak die in de vorige peiling ook is gebruikt. Deze taak had primair tot doel een trend met het vorige peilingsonderzoek te rapporteren. Aangezien deze taak in de vorige peiling niet was geconstrueerd aan de hand van ERK-beschrijvingen konden er geen ERK-standaarden voor spreekvaardigheid worden bepaald.

Gedurende twee dagen zijn voor ieder van de drie vaardigheden drie standaarden bepaald, namelijk een A1, A2 en B1-niveau. Alle standaarden zijn bepaald met behulp van twee aparte rondes, waarin de deelnemers individueel via de laptop de leesitems en (luister/film)fragmenten respectievelijk lezen en beluisterden, en beoordeelden. Na de eerste ronde vond voor iedere vaardigheid een plenaire discussieronde plaats.

## 7.1 Het expertpanel

Het panel dat de standaarden heeft bepaald bestond uit in totaal uit 15 personen. Er is gekozen voor een brede samenstelling van het panel, waarbij zowel taalexperts, leerkrachten en toetsdeskundigen Engels zijn geselecteerd, omdat de ervaring leert dat inhoudelijke experts hun standaarden vaker aan de (te) hoge kant vaststellen. De vijf taalexperts, afkomstig van Nuffic, SLO en HvA/UvA, zijn geselecteerd vanwege hun expertise en ervaring op het gebied van vreemde taalontwikkeling, ERK en (in de meeste gevallen) onderwijs in Engels. Twee ervan waren reeds eerder bij (inter)nationale standaardsettingsprocedures betrokken. Voorts waren er zes docenten die Engels gaven. Zij hebben veel ervaring met leerlingen en hebben in ieder geval een goede indruk van wat leerlingen aan het einde van groep 8 (zouden moeten) kunnen. Vier van deze leerkrachten gaven les in het PO, waarvan twee op EIBO-scholen. De overige twee docenten waren een docent VO en een schoolleider. Drie van de docenten en de schoolleider maakten onderdeel uit van het ontwikkelteam Engels van Curriculum.nu. De drie toetsdeskundigen voegden expertise op het gebied van toetsconstructie toe, alsmede kennis over curriculum en wat leerlingen aan het eind van groep 8 kunnen. Tenslotte heeft een master student Applied Linguistics deelgenomen. Vanwege haar grondige ERK-expertise heeft zij een deel van de gespreksbeoordelingen uitgevoerd, waardoor zij ook bekend was met de doelgroep. In de voorbereiding is rekening gehouden met de mogelijkheid dat de leerkrachten minder bekend zouden zijn met de ERK-niveaus. Aangezien de experts niet gepubliceerde items en video-opnamen van leerlingen moesten beoordelen, hebben alle deelnemers van tevoren een geheimhoudingsverklaring getekend.

## 7.2 Methode

### 7.2.1 Lees- en luistervaardigheid

Zoals in de paragrafen 1.7 en 1.8 is beschreven, is voor de leesvaardigheid en luistervaardigheid bij de constructie van de opgaven een koppeling gemaakt met de ERK niveaus A1, A2 en B1. Het uitgangspunt was dat een opgave voor het A1-niveau gedrag uitlokt bij de leerling dat hoort bij het beheersen van het A1-niveau. Dat geldt ook voor de opgaven voor de niveaus A2 en B1. De opgaven zijn voorafgaand aan de standaardbepaling geëvalueerd door de experts uit het expertpanel, die van tevoren voorbereidende opdrachten hebben gemaakt (zie paragraaf 7.3.1).

#### *Gebruikte methodiek: Data Driven Direct Consensus (3DC) methode*

Voor de standaardbepaling is gebruik gemaakt van de Data Driven Direct Consensus (3DC) methode (Keuning et al., 2017)<sup>73</sup>. In de voorbereidende fase van de 3DC methode zijn clusters van opgaven samengesteld. De clusters kunnen worden gezien als minitoetsen waarbij de experts tijdens de standaardbepalingsprocedure moeten aangeven hoeveel punten (een geheel getal) zij verwachten dat de 'grens-leerling' gaat halen op dat cluster, oftewel op deze minitoets. Een grensleerling is een (denkbeeldige) leerling die nét het niveau heeft waarop de standaard bepaald wordt. Bij de hier beschreven standaardbepaling, waarbij alleen met dichotome opgaven wordt gewerkt, komt dat overeen met de vraag: *"Hoeveel opgaven zou een leerling op dit cluster goed moeten maken als zijn/haar vaardigheid zich precies op de grens A1, A2 of B1 van het referentieniveau bevindt?"*

Het voordeel van deze methode is dat deze data-gedreven is. Informatie van de verschillende clusters kunnen met behulp van IRT met elkaar in verband gebracht worden, omdat de items aan dezelfde vaardigheidsschaal te relateren zijn (zie paragraaf 5.4 voor een beschrijving van de schaling van de toetsresultaten). Ook de *scoreschaal* speelt een rol in deze methodiek. Tijdens de standaardbepaling is gebruik gemaakt van afbeeldingen, die gerelateerd waren aan de scoreschaal.

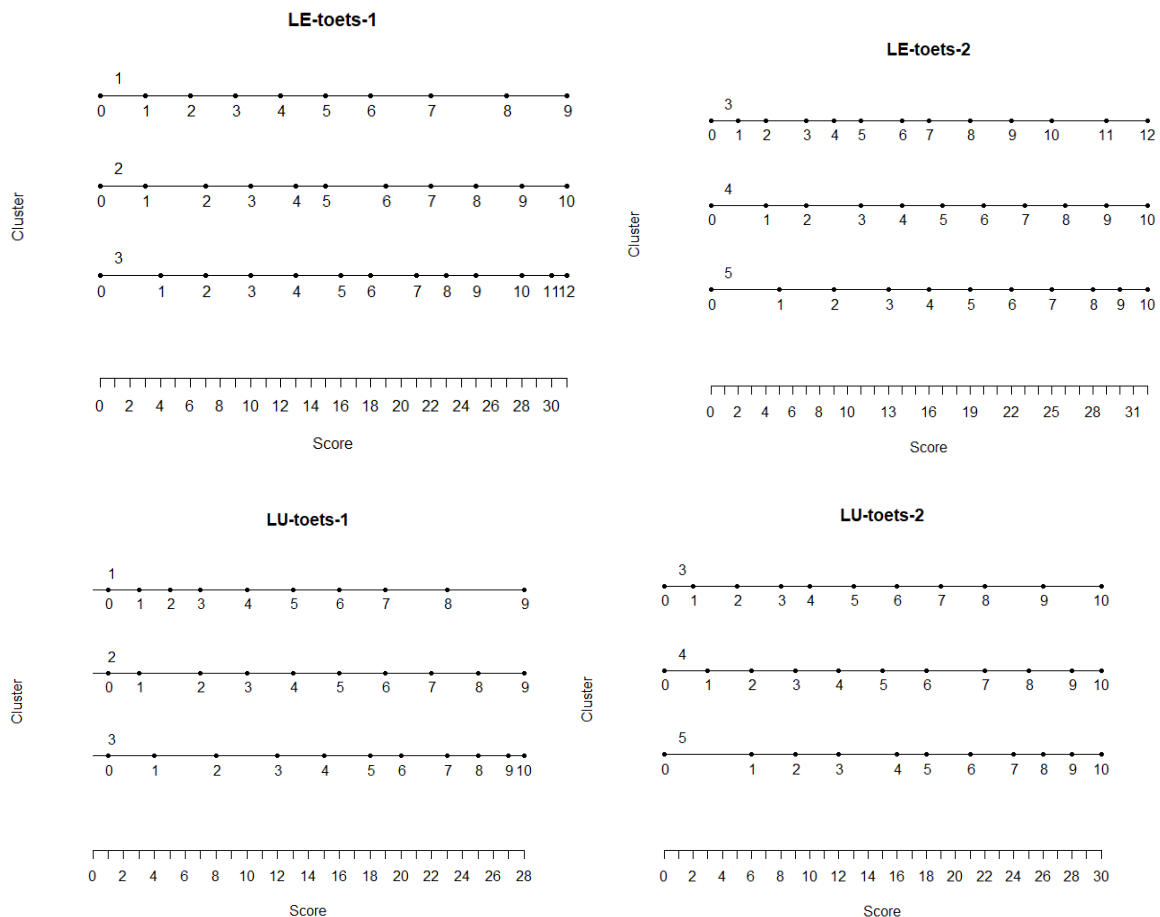
Voor de standaardbepaling voor zowel lezen als luisteren zijn per vaardigheid de opgaven opgedeeld in vijf verschillende clusters, de eerdergenoemde minitoetsen. Per vaardigheid hadden deze de labels 1 tot en met 5. De eerste twee clusters, 1 en 2, betroffen gemakkelijke opgaven die gedrag op het niveau van A1 en A2 zouden moeten ontlocken. Cluster 3 bevatte een gemengde verzameling opgaven en clusters 4 en 5 bevatten vooral wat moeilijker opgaven, die gedrag op het niveau van A2 en B1 zouden moeten uitlocken. Voor het vaststellen van zowel de A1 als A2 standaard zijn de clusters 1 tot en met 3 gebruikt en voor het vaststellen van B1 als standaard zijn de clusters 3 tot en met 5 gehanteerd. In Figuur 7.1 is de clustering weergegeven. Figuur 7.1 laat zien dat de set LE-toets-1 de clusters 1, 2 en 3 bevatte en de set LE-toets-2 de clusters 3, 4 en 5. Ditzelfde gold voor luisteren (respectievelijk *LU-toets-1* en *LU-toets-2*). Verder staan er per set (*LE-toets-1* en *LE-toets-2/ LU-toets-1* en *LU-toets-2*) vier reeksen getallen weergegeven: drie clusters behorend bij de toetsset, met daaronder de gezamenlijke scoreschaal van alle drie de clusters per set bij elkaar. Dankzij IRT zijn al deze reeksen aan elkaar te relateren, zodat per vaardigheidsscore op de vaardigheidsschaal de grenzen van de niveaus A1, A2 en B1 kunnen worden aangegeven.

In Figuur 7.1 is te zien dat de clusters niet allen van gelijke lengte zijn. Zo varieerden de clusters bij lezen tussen negen en 12 items, terwijl de clusters bij luisteren varieerden tussen negen en 10 items. Dit is ook te zien in Tabel 7.1 waarin de aantallen opgaven voor de clusters, het totaal en de sets weergegeven zijn.

<sup>73</sup> Zie ook: <https://www.cito.nl/kennis-en-innovatie/tools-voor-toetsontwikkelaars/tools-voor-toetsconstructie/3dc>



De vaardigheidsschaal kan op twee manieren informatie geven. Behalve dat alle opgaven via een vaardigheidsschaal aan elkaar te relateren zijn, kan ook de verdeling van de kandidaten op de vaardigheidsschaal worden afgebeeld. Van deze laatste mogelijkheid wordt gebruik gemaakt tijdens de 3DC methode. De consequenties van de keuzes van het expertpanel worden direct inzichtelijk gemaakt: wanneer de standaard is bepaald, kan ook voor de (geschatte) populatie worden ingeschat hoeveel procent van de leerlingen een bepaald niveau heeft. Het expertpanel ziet daarmee direct de gevolgen van de (voorlopig) vastgestelde standaard. Meer over de werkwijze van de 3DC methode is beschreven in paragraaf 7.3 (procedure).



Figuur 7.1: de afbeelding van de vijf sets van de clusters die gebruikt zijn voor standaard bepaling bij lezen van de niveaus A1 en A2 (LE-toets-1) en B1 (LE-toets-2) en luisteren op de niveaus A1 en A2 (LU-toets-1) en B1 (LU-toets-2).

Tabel 7.1: aantallen opgaven per vaardigheid gebruikt bij de standaardbepaling

aantal opgaven	Cluster					Totaal
	1	2	3	4	5	
Per cluster lezen	9	10	12	10	10	51
LE-toets 1	9	10	12			31
LE-toets 2			12	10	10	32
Per cluster luisteren	9	9	10	10	10	48
LU-toets 1	9	9	10			28
LU-toets 2			10	10	10	30

### 7.2.2 Gesprekstaken

De drie gesprekstaken zijn geconstrueerd aan de hand van kenmerken van het ERK, waardoor ook voor deze taken standaarden konden worden bepaald. De drie taken betroffen ieder een ander niveau. Taak 1 was (grotendeels gebaseerd en volledig gescoord) op A1-niveau, Taak 2 op A2-niveau en Taak 3 op B1-niveau. De drie taken bleken op basis van de schaling samen één vaardigheidsschaal te vormen. Bij elk van de drie taken zijn de beoordelingen nader verdeeld in drie onderdelen: inhoud, kwaliteit en interactie & fluency. Ook deze onderdelen bleken samen op een schaal te passen, waardoor schaling over de drie taken heen mogelijk was<sup>74</sup>. Voor iedere taak afzonderlijk is een standaard bepaald. Hierbij is per taak gebruikgemaakt van de totaalscore over alle items.

#### *Gebruikte methodiek: leerlinggerichte methode*

Voor de standaardbepalingen voor de gesprekstaken is gebruikgemaakt van een leerlinggerichte methode in plaats van de itemgerichte methode(n) die bij luister- en leesvaardigheid is gebruikt. Dat is gedaan omdat bij deze gesprekstak en het gebruikte beoordelingsmodel de leerling-respons een eenheid vormt. De standaard is daarom bepaald op basis van de uitingen van de leerlingen. Aan de beoordelaars werd nu de vraag gesteld: *“Welke vaardigheid is indicatief voor een leerling waarvan het niveau van de uitingen zich precies op de grens A1, A2 of B1 van het referentieniveau bevindt?”*

Hoewel de leerlingprestatie een geheel vormde, zijn er voor de beoordeling losse momenten (segmenten van de taak) onderscheiden waarbij per segment de prestatie op de drie genoemde onderdelen (inhoud, kwaliteit en fluency) beoordeeld werd. Per segment kon een leerling per gescoord onderdeel 0, 1 of 2 punten verdienen. In de praktijk liep de totaalscore per taak van 0 tot en met zes keer het aantal segmenten per taak. Taak 1 bestaat uit 18 segmenten (dus de scorering liep van 0 – 108), Taak 2 uit vijf segmenten (dus de scorering liep van 0 – 30), en Taak 3 uit zes segmenten (dus de scorering liep van 0 – 36).

Per gesprekstak kreeg de leerling op basis van de beoordeelde segmenten een score toegekend, binnen de genoemde scorering voor taak 1, 2 of 3. Deze scores, die leerlingen op de taken hadden gehaald, vormden de basis voor de standaardbepaling. Voor iedere taak zijn voorbeelden geselecteerd behorend bij een behaalde score. Deze voorbeelden bestonden uit

<sup>74</sup> Voor meer informatie over de schaling van de gesprekstaken wordt verwezen naar paragraaf 5.4.2.

filmpjes van de gesprekstaak van leerlingen, die de geselecteerde scores op deze gesprekstaak hadden behaald. Bij sommige scores was er een keuzemogelijkheid voor de selectie van voorbeelden. In die gevallen is met een aantal factoren rekening gehouden bij de keuze van de te tonen voorbeelden. De eerste factor was de bruikbaarheid van het filmpje, de tweede factor de beoordelaar, de derde factor was de school en laatste factor de kenmerken van de leerling. Hoe met deze factoren rekening is gehouden, is verder beschreven in Bijlage 10. De selectie van voorbeelden leverde een illustratie op van de scoreschaal (met een relatie tot de vaardigheidsschaal) waarop standaarden moesten worden bepaald. Op basis van die voorbeelden kon achterhaald worden welke vaardigheid net voldoende zou zijn om indicatief te zijn voor het ERK-niveau A1, A2 of B1.

### *Werkwijze binnen de leerlinggerichte methode*

Deze leerlinggerichte methode werkt met een aantal ronden. Het doel van de eerste ronde is om snel inzicht te krijgen in de bandbreedte waarin de gezochte standaard zich zou bevinden. Er wordt daarbij niet direct ingezoomd op een specifieke bandbreedte: door de ruime bandbreedte in deze globale ronde krijgen de beoordelaars veel vrijheid om te oordelen, omdat ze niet te veel een richting op geduwd worden.

De tweede ronde gaat meer gedetailleerd in op de exacte positie van de standaard. Welke voorbeelden (i.e., de video-opnamen van de gesprekstaken van leerlingen behorende bij specifieke scores) in de tweede ronde getoond worden is voor een groot deel afhankelijk van de resultaten van de eerste ronde. De voorbeelden van de tweede ronde liggen namelijk om en na bij de scores die in de eerste ronde als score-range aangegeven zijn, als zijnde aannemelijke scores voor de standaard. Doordat de selectie van voorbeelden in ronde 2 afhankelijk is van de oordelen in ronde 1, heeft ronde 2 een adaptief karakter. Door dit adaptieve karakter worden in ronde 2 enkel voorbeelden getoond rond het niveau van de op dat moment ingeschatte standaard.

Deze methode kent de mogelijkheid een derde ronde toe te voegen. Dat wordt gedaan als de experts het onderling nog niet met elkaar eens zijn en de cesuurgrenzen van de experts ver uit elkaar liggen. In dat geval is een gedetailleerder oordeel noodzakelijk of zijn er meer voorbeelden nodig om tot overeenstemming te komen. De in de derde ronde aangeboden voorbeelden worden weer geselecteerd op basis van de resultaten in ronde 2. De combinatie van ronden levert een zo goed mogelijke verdeling van benodigde tijd en nauwkeurigheid zonder dat de beoordelaars van te voren worden gestuurd richting het niveau dat ze verwacht worden aan te geven.

Ook bij de peiling voor mondelinge taalvaardigheid Nederlands is deze wijze van standaardbepaling gehanteerd. De ervaringen die daarbij opgedaan waren, waren positief: de experts kregen een goede indruk van wat de verschillende prestaties van de leerlingen zijn en na drie ronden lagen de oordelen dicht bij elkaar.

Een ander voordeel is dat de methode efficiënt is: het evalueren van de leerlingwerken is tijdrovend, en het totaal aantal te beoordelen werken is daardoor beperkt. Met deze methode krijgen de experts zo snel mogelijk de meest relevante leerlingwerken te zien. Zoals hierboven vermeld is het voordeel van deze methode dat direct daarna ook aangegeven kan worden hoeveel procent van de leerlingen een bepaald niveau gehaald heeft. Hoe de werkwijze specifiek bij de gesprekstaken Engels plaatsvond en hoe deze ervaren is, wordt beschreven bij de procedure.

Aangezien bij deze standaardbepaling de beschikbare tijd zeer beperkt was, is er voor de eerste ronde gebruik gemaakt van inschattingen door vier experts die op basis van kennis van de opdrachten en de scoring ervan de eerste ronde al relatief gericht konden maken zonder de experts

in het expertpanel in te perken. Hierdoor werd het mogelijk de standaarden in twee rondes te bepalen.

### ***Rationale voor de keuze van leerlinggerichte methode***

Er is voor deze methode gekozen omdat de factor tijd een belangrijke rol speelde bij de standaardbepaling. De tijd was op voorhand beperkt, doordat binnen twee dagen negen standaarden (namelijk drie per vaardigheid) moesten worden bepaald. Meer specifiek kwam daarbij dat de gesprekstaken voldoende tijd vereisten om de voorbeelden te kunnen evalueren. Het bekijken en beoordelen van een voorbeeld zou namelijk maximaal 20 minuten kunnen kosten. Hierdoor moest het aantal aan te bieden voorbeelden relatief beperkt zijn, zeker als er voor meerdere gesprekstaken standaarden moesten worden vastgesteld. Het was daarom belangrijk snel bij het relevante deel van de schaal te geraken en zo min mogelijk irrelevante voorbeelden te laten zien. Om dit te bevorderen is de standaardbepaling voorafgegaan door de hierboven genoemde inhoudelijke inschatting door een viertal experts die een ruime, aannemelijke scorering aangaven waarbinnen de standaard zeker zou zitten<sup>75</sup>. Hierdoor hoefde in de eerste algemene ronde niet de gehele scoreschaal gedekt te worden. Dit was vooral bij Taak 1 met een scoreschaal van 0 tot en met 108 relevant.

## **7.3 Procedure**

### **7.3.1 Voorbereidende activiteiten**

De beperkte beschikbaarheid van het expertpanel leidde tot de keuze voor een tweedaagse standaardsetting. Voor een goed verloop van de standaardbepaling was het noodzakelijk dat de deelnemers met voldoende kennis van het ERK aan de standaardbepaling zouden beginnen. Tijd voor een grondige opleiding tijdens de twee geplande dagen was er niet. Om die reden was bij de selectie van de deelnemers rekening gehouden met hun ERK-kennis. Alle deelnemers waren goed bekend met het ERK en een aantal ook met de procedures van standaardbepalingen. Om ons te verzekeren van deze expertise, was afgesproken dat alle deelnemers voorafgaand aan de standaardbepaling voorbereidende opdrachten zouden ontvangen. Begin juni zijn aan de deelnemers 51 leesitems en 48 luisteritems toegestuurd (toets-boekjes en usb-stick met bronmateriaal voor de luistertaken), aangevuld met ERK naslag-materiaal. Voor ieder item is de deelnemers gevraagd het ERK-niveau (A1, A2 of B1) in te schatten. Deze inschatting vulden de deelnemers in op een Excel-formulier dat ze per email ontvingen en vóór 1 juli retourneerden. Het grote bijkomende voordeel van deze huiswerkopdracht was dat a) het materiaal al bekend was bij de experts waardoor de standaarden voor lezen en luisteren vlot zouden kunnen worden bepaald en b) vooraf een indicatie werd verkregen van hoe de deelnemers beoordeelden, zodat daar tijdens de discussierondes op kon worden ingespeeld. Voor de gesprekstaken was een dergelijke procedure, waarbij de beoordelaars het te beoordelen materiaal van tevoren inzagen, niet mogelijk omdat de video-opnamen privacygevoelige informatie bevatten.

De deelnemers ontvingen feedback op hun inschattingen in de vorm van een indicatie hoe hun beoordelingen zich verhielden tot die van de overige deelnemers. Op een enkele beoordelaar na,

---

<sup>75</sup> De geconsulteerde experts waren Margreet van Aken (Cito), Wander Lowie (RuG), Alessandra Corda (HvA) en Patricia Rose (SLO)

die bij luisteren de opgaven te hoog inschatte, waren er geen opmerkelijke resultaten en werden de meeste opgaven op de juiste waarde ingeschat (mediaan 70%, met 15% hoger en 15% lager).

### 7.3.2 Planning en verloop

De eerste dag van de standaardbepaling was gewijd aan het vaststellen van de standaarden van leesvaardigheid (ochtend) en eerste globale ronde voor gespreksvaardigheid (middag). De tweede dag was de ochtend gewijd aan de tweede ronde gespreksvaardigheid en de middag aan het vaststellen van de standaarden voor luistervaardigheid. Voor de drie vaardigheden zijn drie ERK-niveaus vastgesteld, A1, A2 en B1. Het bekijken en beluisteren van zowel de luister- en leesitems als ook van de gespreksfragmenten gebeurde individueel op een computer. Na het bekijken/beluisteren stelden de deelnemers individueel hun oordeel vast. De deelnemers konden op deze manier per ronde op hun eigen tempo werken. De eerste dag werd gestart met een plenaire terugkoppeling op het gemaakte huiswerk.

#### *Lezen en luisteren*

De procedure bij lezen en luisteren was gelijk. Op de eerste van de twee dagen werden de standaarden voor lezen bepaald en op de tweede dag die van luisteren. De bepaling van de standaard A1, A2 en de standaard B1 heeft bij luisteren en lezen in twee rondes plaatsgevonden. In de eerste ronde markeerde iedere beoordelaar per niveau de score van de zogenaamde grensleerling per cluster in een figuur. Met behulp van deze markering gaf iedere beoordelaar voor bijvoorbeeld het cluster 1 bij lezen aan hoeveel van de items binnen dat cluster de grensleerling op A1-niveau goed zou hebben. Hierop volgde een plenaire discussie over de uitkomsten.

Deze discussie startte door de uitkomsten van de eerste beoordelingsronde te projecteren en aan een aantal beoordelaars te vragen om hun beoordeling bij één of twee clusters toe te lichten. In de tweede beoordelingsronde is de beoordelaars opnieuw gevraagd om in de figuur per cluster de grensscore te markeren. Op basis van de gevoerde discussie kregen de beoordelaars hierdoor de gelegenheid om hun grensscore bij te stellen indien zij dat wensten.

#### *Gespreksvaardigheid*

Bij de standaardbepaling voor gespreksvaardigheid werd per gesprekstaak iedere ronde uitgevoerd. In een eerste globale ('grove') ronde van de standaardbepaling kregen de experts een aantal filmpjes te zien van leerlingen die de gesprekstaak uitgevoerd hebben. De range van voorbeelden was relatief breed, maar door de inhoudelijke betekenis van de opgaven niet onbepaald. Zo was het beoordelingsmodel voor de gesprekstaking dusdanig opgesteld dat een score '1' min of meer overeen hoorde te komen met gedragingen die horen bij het niveau van de gesprekstaak. Bij Taak 1 betekende een score 1 bijvoorbeeld dat de leerling het item op A1-niveau uitvoerde. Om te weten in welk gebied de grens voor de score van een grensleerling precies zou liggen zijn vijf voorbeelden getoond die een brede range behelsden.

In Tabel 7.2 zijn de scores van de voorbeelden voor de eerste ronde standaardbepaling gesprekken gegeven. De keuze van de scores van de voorbeelden is gebaseerd op de range van aannemelijke waarden die voorafgaand aan de standaardbepaling door de experts waren aangegeven, alsmede de beschikbaarheid van goede voorbeelden voor die scores.

Tabel 7.2: Scores van de voorbeelden van de eerste ronde standaardbepaling gesprekken

Score per voorbeeld per taak	Taak 1	Taak 2	Taak 3
Ronde1.v1	47	12	15
Ronde1.v2	55	15	18
Ronde1.v3	65	18	21
Ronde1.v4	74	21	24
Ronde1.v5	85	24	28

De standaarden gespreksvaardigheid zijn per taak vastgesteld, startend met de eerste taak waarop het A1-niveau is vastgesteld, tot en met de derde taak waarop het B1-niveau is vastgesteld. Zoals Tabel 7.2 laat zien werden per taak de vijf voorbeelden in oplopende volgorde van beheersingsniveau (d.w.z. de plaats op de vaardigheidsschaal gespreksvaardigheid) getoond. De experts kregen dus eerst het filmpje van de leerling met het laagste vaardigheidsniveau van de schaal te zien en tot slot die van de leerling met het hoogste vaardigheidsniveau. De deelnemers bekeken de video-opnamen individueel en op eigen tempo via een eigen laptop. Nadien gaven zij individueel op het scherm aan bij welk filmpje volgens hen de grens lag voor beheersing van het bijbehorende ERK-niveau.

Het resultaat van de eerste ronde werd aan het begin van de tweede dag besproken. Daarnaast werd dit resultaat vergeleken met de verwachte percentages leerlingen die op basis van de eerste ronde het betreffende niveau zouden behalen. Hierna startte de meer gedetailleerde tweede ronde waarin werd ingezoomd op de range van de gemiddelde grensscore uit de eerste ronde. Deelnemers kregen in deze tweede ronde filmpjes te zien van leerlingen die presteren vlak onder, op en vlak boven het gemiddelde cesuurpunt van de eerste ronde. Daarna stelden zij opnieuw individueel de cesuur vast voor deze serie filmpjes. Aangezien de beoordelaars in de eerste ronde al een (vrij) hoge mate van overeenstemming hadden bereikt, volstond een tweede ronde.

## 7.4 Resultaten

### 7.4.1 Lezen en Luisteren

Voor leesvaardigheid werden eerst drie clusters beoordeeld met A1- en A2-opgaven. Hierbij werd eerst het niveau van A1 bepaald. In de eerste ronde kwam het gemiddelde uit op 13,5. De meest extreme grens die door een beoordelaar bepaald was lag bij 18 punten op de schaal met een maximum score van 31. De een-na-hoogste grenswaarde was 16 punten. De laagste grensscore was gevonden bij 11. Twee personen kwamen op die score uit. Na de tweede ronde kwam het gemiddelde uit op **12,9**. De spreiding in grensscores was lager met een range van 11 tot en met 15, en een standaardafwijking van grensscores over beoordelaars dalend van 2,0 naar 1,5. Op de discrete scoreschaal van set 1 bij lezen met maximumscore 31 heeft de leerling met de laagste score die (minstens) het niveau A1 heeft, 13 punten behaald.

Doordat telkens op de discrete schaal gewerkt is, is logischerwijs de vraag aan de experts gelijk aan de vraag wat de hoogste score op die schaal is die net geen A1 oplevert. Dat is exact 1 punt

minder dan de gestelde grens per expert. Gemiddeld is de hoogste score op de discrete schaal die volgens de expert (net) geen A1 oplevert, de score 11,9. Als we nu de grens tussen wel en geen A1 naar de continue schaal vertalen ligt die precies op het midden tussen 11,9 en 12,9 en dus bij de vaardigheid (theta) die precies de score **12,4** oplevert<sup>76</sup>. De theta-waarde die deze grenscore oplevert geeft ons de cesuur op de continue vaardigheidsschaal. De score kan door middel van IRT weer vertaald worden naar dezelfde vaardigheidsschaal waarnaar de resultaten van alle kandidaten vertaald zijn. Van daaruit is in het scorebestand per leerling aan te geven op welk niveau de prestatie ligt: <A1 of minstens op A1. In dit geval was dat bij de theta-waarde **-0,1984**. Als de leerling een vaardigheid heeft hoger dan die waarde heeft, dan heeft de leerling **(minstens) A1-niveau**.

De grens voor A2 bij lezen is op dezelfde wijze bij dezelfde clusters (cluster 1, 2 en 3) vastgesteld. Het gemiddelde van de eerste en de tweede ronde verschilde weinig (24,0 en **24,1**). De spreiding van grensscores bij de verschillende beoordelaars was in de tweede ronde wel lager, met een scorering van 21-29 in ronde 1 naar 22-28 in ronde twee en de standaardafwijking dalend van 2,3 naar 1,5. De vertaling naar de continue schaal, op dezelfde wijze als bij de A1-standaard leverde een grenswaarde voor theta op van **0,2928**. Als de leerling een vaardigheid heeft hoger dan die waarde, dan heeft de leerling **(minstens) A2-niveau**.

De grens voor B1 bij leesvaardigheid werd bepaald op de andere set van clusters (cluster 3, 4 en 5). Hierbij was een van de drie clusters in deze set ook gebruikt voor de A1 en de A2-grens, te weten cluster 3. Bij deze set was de gemiddelde grenswaarde gestegen van de eerste naar de tweede ronde: van 25,4 naar **26,1**. De spreiding in scores was wel gedaald van de eerste (2,9) naar de tweede ronde (2,0). De waarde van theta die de score 25,6 opleverde op de gebruikte set van 32 opgaven is de grenswaarde op de continue vaardigheidsschaal. Dat was hier de theta-waarde van **0,6445**, waarbij gesteld wordt dat leerlingen met vaardigheden hoger dan die theta **het niveau B1 of hoger behalen**.

De grenzen bij Luisteren zijn op een vergelijkbare manier berekend bij twee sets van clusters voor luisteren. Per niveau is in twee rondes de grenswaarde bepaald. Het belangrijkste verschil in de twee rondes was dat de spreiding duidelijk afnam, zowel wat betreft de standaardafwijking van de gegeven grenzen als wel de range (verschil tussen maximum en minimum gegeven grenswaarde). De hoogte van de grens verschilt amper van de eerste naar de tweede ronde bij luisteren. De grensscores zijn vertaald naar de continue vaardigheidsschaal op basis van de gemiddelde grenswaarden voor A1-, A2- en B1-niveau.

Op basis van de gestelde grenzen zijn de percentages leerlingen op ieder niveau te bepalen. In Tabel 7.3 zijn deze gegeven voor de geobserveerde verdelingen zoals die beschikbaar waren tijdens de standaardbepaling. In Hoofdstuk 8 waar de resultaten verder besproken worden zijn deze gegeven na weging naar de percentages in de populatie. De geobserveerde percentages zijn bepaald op de steekproef waarin zowel VVTO-scholen als de scholen met relatief meer “gewichtleerlingen” oververtegenwoordigd zijn.

---

<sup>76</sup> Nadere informatie over de theta-waarden is op te vragen bij de onderzoekers.

Tabel 7.3: Ongewogen resultaten standaard bepaling lezen en luisteren

		Lezen				Luisteren			
		<A1	A1	A2	B1	<A1	A1	A2	B1
score schaal (discreet)	maximum score		31	31	32		28	28	30
	grens (gemiddeld)		12,93	24,13	26,07		7,93	19,36	22,91
	grens na correctie		12,43	23,63	25,57		7,43	18,86	22,41
vaardigheids schaal (continu)	vaardigheidsgrens		-0,1984	0,2928	0,6445		-0,4907	0,0766	0,5880
	Cumulatief %*		99,0	68,0	20,8		93,1	54,6	26,5
	Percentage \$	0,96	31,03	47,16	20,85	6,86	38,57	28,12	26,46

#### 7.4.2 Gespreksvaardigheid

De experts gaven over twee dagen de beoordelingen van de gesprekstakingen. De eerste dag betrof de eerste globale ronde voor alle drie de taken. Per taak werd slechts één ERK-niveau bepaald, aangezien de taken specifiek voor de drie verschillende niveaus ontwikkeld waren. Hierdoor waren de taken minder geschikt zijn voor het meten van de andere ERK-niveaus. Bij alle drie de taken oordeelden de experts dat de twee voorbeelden met de laagste scores niet het beoogde niveau vertoonden. Vanaf het derde voorbeeld oordeelde minstens een van de experts dat het niveau gehaald werd. De mediaan lag in alle gevallen bij het vierde voorbeeld. Dat was ook de modus, zij het dat bij Taak 2 er sprake was van een gedeelde modus met voorbeeld 5. De resultaten van de eerste ronde zijn gegeven in Tabel 7.4.

Tabel 7.4: Oordelen van de experts in de eerste ronde

Taak 1: niveau A1				Taak 2: niveau A2				Taak 3: niveau B1			
Voorbeeld		Experts		Voorbeeld		Experts		Voorbeeld		Experts	
Nr	Score	aantal	%	Nr	Score	aantal	%	Nr	Score	aantal	%
11.1	47	0	0	21.1	12	0	0	31.1	15	0	0
11.2	55	0	0	21.2	15	0	0	31.2	18	0	0
11.3	65	2	13	21.3	18	1	7	31.3	21	3	20
11.4	74	8	53	21.4	21	7	47	31.4	24	11	73
11.5	85	5	33	21.5	24	7	47	31.5	28	1	7

Op basis van deze resultaten zijn voorbeelden gekozen voor de tweede ronde met de scores zoals gegeven in Tabel 7.5. Deze keuze is gemaakt op basis van de verdeling van de expert-oordelen in de eerste ronde: de voorbeelden waren zo gekozen dat deze in ieder geval de range van een standaardafwijking rond de gemiddelde score bevatten. Daarnaast speelde de beschikbaarheid van bruikbare voorbeelden een rol bij de keuze van de specifieke voorbeelden. Het overzicht van de verdeling van de scores van de voorbeelden van de tweede ronde standaardbepaling gesprekken is gegeven in Tabel 7.5.



Tabel 7.5: Scores van de voorbeelden van de tweede ronde standaardbepaling gesprekken

Score per voorbeeld per taak	Taak 1	Taak 2	Taak 3
Ronde2.v1	70	20	21
Ronde2.v2	72	22	23
Ronde2.v3	76	23	24
Ronde2.v4	82	24	25
Ronde2.v5	86	28	27

Op de tweede dag gaven de experts de oordelen van de tweede ronde op alle drie de taken. De verdeling van de expertoordelen van die sessie zijn gegeven in Tabel 7.6.

Tabel 7.6: Oordelen van de experts in de eerste ronde

Taak 1: niveau A1				Taak 2: niveau A2				Taak 3: niveau B1			
Voorbeeld		Experts		Voorbeeld		Experts		Voorbeeld		Experts	
Nr	Score	aantal	%	Nr	Score	aantal	%	Nr	Score	aantal	%
12.1	70	8	53	22.1	20	12	80	32.1	21	2	13
12.2	72	6	40	22.2	22	1	7	32.2	23	8	53
12.3	76	1	7	22.3	23	2	13	32.3	24	4	27
12.4	82	0	0	22.4	24	0	0	32.4	25	1	7
12.5	86	0	0	22.5	28	0	0	32.5	27	0	0

In alle drie de gevallen zijn de spreidingen van de scores lager geworden. Bij alle drie de taken zijn de beoordelaars in de tweede ronde ook milder geworden. Bij Taak 1 en Taak 2 is die daling een daling van rond de 5% van de scoreschaal. De samenvattende resultaten van ronde 2 zijn gegeven in Tabel 7.7. Daarin is ook gegeven welke vaardigheidsscore hoort bij deze niveaus. De berekening van de grensscore op de vaardigheidsschaal heeft op een vergelijkbare manier plaats gevonden als bij lezen en luisteren.

Tabel 7.7: Resultaten standaardbepaling Gespreksvaardigheid

	Taak 1: A1	Taak 2: A2	Taak 3: B1
Modus	70	20	23
Mediaan	70	20	23
Gemiddelde	71,20	20,53	23,13
Gemiddelde na continuïteitscorrectie	70,70	20,03	22,63
Eerste score met beoogd niveau	71	21	23
Theta (ENGSP_3F-schaal; WML in data)	-0,0849	0,6427	0,7598
Populatie (ongewogen) % lager of gelijk theta	33 %	84 %	89 %
Populatie (ongewogen) % groter dan theta	67 %	16 %	11 %

De percentages in Tabel 7.7 zijn de geschatte percentages gebaseerd op de ongewogen verdeling van de vaardigheid. De geobserveerde verdeling is bepaald door de gevonden vaardigheidsgrenzen toe te passen bij de 561 leerlingen die minstens een van de spreektaken heeft gemaakt. Zo is te bepalen welk niveau ieder van deze leerlingen heeft behaald. Deze aantallen zijn gegeven in Tabel 7.8. De afwijking van de gevonden verdeling met de op de ongewogen populatie geschatte percentages zijn klein te noemen.

Tabel 7.8: geobserveerde verdeling niveaus gespreksvaardigheid

	Niveau			
	<A1	A1	A2	B1(+)
Aantal leerlingen	171	290	49	54
Percentage leerlingen	30 %	51 %	9 %	10 %
Cumulatief:		A1(+)	A2(+)	B1(+)
Percentage met niveau of hoger		70 %	18 %	10 %
Cumulatief:	<A1	<A2	<B1	
Percentage met niveau of lager	30 %	82 %	90 %	

#### 7.4.3 Voorbeelditems behorende bij de standaarden

In voorgaande paragraaf is voor de drie vaardigheden beschreven welke vaardigheidsscores horen bij de bepaalde drie standaarden. Omdat aan elk item een vaardigheidsscore gekoppeld is, kan een inhoudelijke vertaling van deze standaarden gemaakt worden voor de vaardigheden lezen en luisteren: de standaarden kunnen worden geïllustreerd met behulp van opgaven die weergeven wat

leerlingen op de drie niveaus kunnen. Voor gespreksvaardigheid is deze koppeling niet goed mogelijk, omdat er sprake is van een leerlinggerichte benadering.

Hieronder worden voor lees- en luistervaardigheid voorbeelditems gepresenteerd die de bijbehorende ERK-niveaus illustreren. Voor zover mogelijk zullen per ERK-niveau twee voorbeelditems worden getoond: een opgave waarbij de leerling 50% kans heeft om de opgave goed te maken en een opgave waarbij de leerling 80% kans heeft om de opgave goed te maken. Het eerste type opgave kan als uitdagend worden beschouwd: de leerling beheerst de opgave nog niet echt, want heeft nog steeds 50% kans om de opgave niet goed te maken. Bij het tweede type opgaven kan gesproken worden van daadwerkelijke beheersing van de opgave: vier van de vijf leerlingen op dit niveau heeft de opgave goed. Dit niveau kan daarom als het meest illustratief voor het beschreven ERK-niveau worden beschouwd.

### Lezen – A1

Voor het A1-niveau voor lezen kan alleen een B50-opgave (opgave waarop leerlingen met een A1-niveau 50% kans hebben dat ze hem goed maken) worden gepresenteerd, omdat B80-opgaven niet beschikbaar waren. Dit houdt in dat er binnen de leestoetsen eigenlijk geen opgaven waren die de

A1-grenslersleerlingen echt beheersen, de opgaven waren allemaal moeilijk tot (redelijk) uitdagend voor deze leerlingen. Een voorbeeld van een B50-opgave op A1-niveaus is gegeven in Figuur 7.2.



Wat wil de gemeente met deze advertentie bereiken?  
Dat iedereen

- de poep van z'n eigen hond opruimt.
- z'n hond netjes aan de lijn laat lopen.
- z'n hond uitlaat op een honden-lostloop-plaats.

**be a responsible dog owner**

Please help to keep our town clean and tidy so that everyone can be happy with it. There are dog waste containers everywhere, but every dog owner has to tidy up after their pet.

**REMEMBER**

- Fouling by dogs is an unnecessary public problem
- Dog owners should always carry a bag with the which they can put fouling when they are taking dogs out for a walk.

For further information please contact the Environmental Health and Enforcement Unit at Berwick-upon-Tweed Borough Council  
Telephone: 01289 301821

Figuur 7.2: B50-opgave voor A1-niveau

### Lezen: A2-niveau

De toets lezen bevatte voor de A2-grenslersleerlingen (de leerlingen die nét het A2-niveau beheersen) zowel uitdagende als makkelijke opgaven, die zij beheersen. Figuur 7.3 laat een B50- en B80-opgave zien. Deze B80-opgave voor A2-niveau blijkt bijna een B50-opgave (precies: B48) voor A1-grenslersleerlingen te zijn. Waar A2-grenslersleerlingen deze opgave beheersen, is de opgave voor de A1-leerling uitdagend.

Hieronder zie je een nieuwsberichtje.

**Panda has operation on his teeth**



Ming is a 17-year old panda who lives in Berlin Zoo. One day keepers saw that it was difficult for Ming to chew his favourite food: bamboo shoots. Dentists repaired his teeth. The old panda had a two-hour operation. He got fillings in three teeth. The operation was so successful that two other pandas will get an operation too.

Waarom werd Ming geopereerd?

- omdat hij al heel oud was
- omdat hij haast niet kon kauwen
- omdat hij veel kiespijn had

"If I could give my fans a message, it would be to stay in school, stay positive. Whatever you want to be in life, just try and make your dreams come true. Don't let anybody tell you that you can't make it."

Bron: *Just Fourteen*

Wat vertelt de popster in dit interview aan zijn fans?

- hoe hij denkt over het beroemd zijn
- wat ze moeten doen om in het leven te slagen
- wat zijn toekomstplannen zijn

Figuur 7.3: B50-opgave (links) en B80-opgave (rechts) op A2-niveau

### Lezen B1-niveau

Het hoogste niveau waarvoor ERK-niveaus zijn bepaald is het B1-niveau. De toets lezen bevatte voor de B1-grenslers geen B50-opgaven, oftewel de toets was niet bijzonder uitdagend voor leerlingen op dit niveau. Er kan daarom slechts een B80-opgave worden getoond (Figuur 7.4).

**Police move Eurostar strikers from tracks**

**FRANCE** : Eurostar trains to and from Paris are running again after police cleared striking French workers from the railway tracks at the Gare du Nord. The strikers were railway workers who wanted better working conditions and the same pay as their British colleagues.

Bron: *The Independent*

Wie staakten er?

- Brits spoorwegpersoneel
- Frans spoorwegpersoneel
- de Britse politie
- de Franse politie

Figuur 7.4: B80 opgave op B1-niveau

### Luisteren – A1 niveau

Ook de luistertoets kende voor de grensleerlingen op A1-niveau geen opgaven die zij goed beheersen. Een B80-item kan daarom niet worden getoond. Figuur 7.5 bevat een voorbeeld van een B50-opgave.

#### *Fishing in school - nieuwsbericht*

<p>It's the summer holidays at the moment and we are on an activity day but we'll also be fishing in term time as well. My school in Bolton has its own lake next to the school. They even have their own fishing team.</p> <p>We are fishing for perch, roach and rudd. We have to use maggots as bait. They're pretty disgusting. An angling coach teaches us the skills. How difficult is it to start fishing?</p> <p>It's not difficult at all. There are plenty of clubs that have coaching. And other organisations like scouts, the cubs, the guides all do fishing as well.</p> <p>Is it expensive?</p> <p>No it's not expensive at all. The gear we're using today, probably cost you less than 20 pounds.</p> <p>I didn't catch any fish. But some of my classmates did, and there are always plenty more fish in the sea. Or lake.</p>	<p>Wat vertelt de visleraar?</p> <p>A Dat het heel makkelijk is om te beginnen met vissen</p> <p>B Dat je beter kunt vissen als je een dure hengel hebt</p> <p>C Dat je het beste lid kunt worden van een visvereniging</p>
---	---

Figuur 7.5: B50-opgave op A1-niveau

### Luisteren: A2-niveau

Figuur 7.6 en 7.7 geven respectievelijk een B80- en B50-opgave weer. De toets luisteren bevatte voor de grens-leerlingen op A2-niveau zowel uitdagende als gemakkelijke opgaven.

## Musical

<p>Sweeney Todd. Visit Sweeney Todd! A musical thriller. It's the West End smash hit musical production. Now at the Theatre Royal, Brighton. Best musical of the year, says The New Yorker. A bloody, brilliant musical. Kill for a ticket! Prices range from 16 to 27 pounds. Call us, at the Theatre Royal, Box Office: 08700 606 650</p>	<p>Waarvoor wordt reclame gemaakt?</p> <p>A Een film op dvd B Een muziekopleiding C Een theatervoorstelling</p>
---	---

Figuur 7.6: B80-opgave op A2-niveau

## Bern

<p>Bern - A man was locked in a phone box by people who saw him robbing a woman in the Swiss capital, Bern. The man stole a wallet out of the woman's handbag while she was making a phone call. The people helped the woman out of the phone box and locked the man in until the police arrived.</p>	<p>Wat gebeurde er bij een telefooncel in Bern?</p> <p>A De politie arresteerde er een man die er niet uit wilde B Een man werd betrapt toen hij de apparatuur kapot sloeg C Mensen hielden er tijdelijk een overvaller in gevangen</p>
---	---

Figuur 7.7: B50-opgave op A2-niveau

## Luisteren: B1-niveau

De luistertoets bevatte ook op het B1-niveau zowel uitdagende als gemakkelijke opgaven. Figuur 7.8 laat een gemakkelijk items zien voor de B1-grensleerlingen, terwijl de opgave in Figuur 7.9 uitdagend is voor deze leerlingen.

### Orinoco-rivier

<p>As the Orinoco flows down onto flat land, the trees begin to thin out and the land becomes easier for people to inhabit. Here the Orinoco is joined by a smaller river, or tributary, and it becomes bigger. It could become even bigger if there's a lot of rain. This ferryboat is passing between rocks that would be covered by the river during the rainy season.</p> <p>But this is the dry season and the river is safe for children to play in.</p>	<p>Wanneer kunnen de kinderen veilig spelen in de <i>Orinoco</i>-rivier?</p> <p><b>A</b> Als de veerboten een andere route nemen</p> <p><b>B</b> Als het lang niet geregend heeft</p> <p><b>C</b> Als ze goed kunnen zwemmen</p>
--	--

Figuur 7.8 B80-opgave op B1-niveau

### New York

<p>Last year I went on a holiday to New York. I absolutely loved the Statue of Liberty. I went all the way up to the crown. It was amazing. I recommend visiting on a clear day, so you get a great view of the New York sky line. Everyone needs to come here at least once in their lifetime.</p>	<p>Wanneer kun je het vrijheidsbeeld in New York het beste bezoeken?</p> <p><b>A</b> Als het onbewolkt is</p> <p><b>B</b> Net voor sluitingstijd</p> <p><b>C</b> Tijdens de vakantie</p>
---	--

Figuur 7.9: B50-opgave op B1-niveau

## 7.5 Woordenschat: geen standaarden, wel een referentiescore

Zoals in de inleiding is aangegeven, is het voor woordenschat niet mogelijk om standaarden te geven. De ERK niveaus zijn namelijk niet gedefinieerd in termen van *Can do*-statements voor deze vaardigheid, in tegenstelling tot bij lees-, luister- en gespreksvaardigheid.

Om de schaalwaarden van woordenschat beter te kunnen interpreteren is de wens uitgesproken een relatie te leggen tussen de vaardigheid en het percentage woorden dat een leerling kent. De uitdaging hierbij is dat de definitie van "het kennen van een woord" niet vast ligt. Dit heeft onder andere te maken met de vraag hoe "diep" of "genuanceerd" men een woord kent. De vraag is of men, als men een woord "kent", synoniemen dient te kennen, en zo ja hoeveel? En of men ook antoniemen moet kennen? Daarnaast moet bepaald worden in hoeverre de context een rol mag, dan wel moet spelen, in het kennen. En wanneer wordt woordenschat algemene kennis? Deze vragen lijken wellicht overbodig, maar als men het woord "cow" bevroegt en men laat een leerling uit de

plaatjes van een koe, een hond en een kat kiezen of uit de plaatjes van een koe, een stier en een os, dan zijn dat fundamenteel andere vragen die beiden het woord "cow" betreffen.

Om de definitie van het "kennen van de woorden" werkbaar te maken is de vraag naar het kennen van woorden vertaald naar het "kunnen maken van (vergelijkbare) opgaven over woordenschat". Die vraag is eenvoudiger te beantwoorden. Als we aannemen dat de geselecteerde woorden representatief zijn voor alle mogelijke woorden kan vanuit de geschatte vaardigheid ook een geschat aantal goed gemaakte opgaven geschat worden. Dat kan doordat vanuit de itemresponse-functies (IRFs) voor iedere vaardigheid bij alle opgaven geschat kan worden wat de kans is dat deze opgave goed gemaakt wordt. De som van deze kansen over alle 110 opgaven in dit onderzoek (peilingsafname 2018) levert dan het te verwachten aantal goede opgaven op. Het percentage goede antwoorden is dan het verwachte aantal goede antwoorden gedeeld door 110 maal 100%.

Het is ook mogelijk te stellen dat niet ieder woord even representatief is en dat sommige woorden meer over de vaardigheid van woordenschat zeggen dan andere woorden. "Belangrijke woorden" krijgen dan een hoger gewicht (staan voor meer punten of meer woorden) dan minder belangrijke woorden. Een manier om te bepalen wat belangrijke woorden zijn is door het programma OPLM te gebruiken. De a-parameters van woorden die meer indicatief zijn voor de vaardigheid, en dus om die reden belangrijker worden geacht, zijn hoger dan de a-parameters van woorden die niet zo indicatief zijn. Door nu de kansen per vaardigheidsniveau per opgave te wegen met de a-parameter en op te tellen krijgt men een "gewogen te verwachten aantal goede opgaven".

Het gewogen percentage opgaven zou wellicht een betere weergave kunnen zijn op het moment dat de a-parameters daadwerkelijk het inhoudelijk belang van de woorden weergeven. Echter, het wegen van de woorden met de a-parameters is vooral een technische aangelegenheid die niet noodzakelijkerwijs een inhoudelijke dekking heeft. Opgaven kunnen meer met elkaar samenhangen doordat ze bijvoorbeeld meer op elkaar lijken. Deze opgaven met een grotere samenhang krijgen vaak hogere a-parameters. Een woord dat representatief is voor een andere reeks opgaven van woorden die onder-gerepresenteerd zijn in de verzameling opgaven zou dan een lagere a-parameter kunnen krijgen, terwijl er inhoudelijke gronden zouden kunnen zijn op basis waarvan dit woord juist een hoger gewicht zou moeten krijgen. Ook kunnen er andere mogelijke redenen worden opgevoerd om opgaven meer te wegen, bijvoorbeeld op basis van frequentielijsten. Welke weging er ook gekozen wordt, het gewogen percentage kan bepaald worden door het gewogen aantal goed te delen door het maximaal aantal gewogen goed en dat te vermenigvuldigen met 100%. Er is een Excel-bestand beschikbaar gemaakt dat deze waarde per vaardigheidsniveau berekent nadat de gewichten per opgave zijn gekozen. Door middel van de vaardigheden als link zijn deze dan aan de data te koppelen<sup>77</sup>. In hoofdstuk 8 zullen de resultaten voor woordenschat beschreven worden in termen van de scores met "gelijke gewichten".

---

<sup>77</sup> Dit bestand kan opgevraagd worden bij de onderzoekers.





## Hoofdstuk 8. Resultaten - beantwoording van de onderzoeksvragen

Belangrijk doel van het peilingsonderzoek is om de Engelse vaardigheden en competenties van de leerlingen in groep 8 in kaart te brengen. In dit hoofdstuk worden de resultaten van de leerlingen beschreven. Met behulp van deze resultaten kan antwoord worden gegeven op de onderzoeksvragen die in dit onderzoek centraal stonden. Deze onderzoeksvragen luiden als volgt<sup>78</sup>:

1. Hoeveel procent van de leerlingen beheerst in leerjaar 8 de ERK-niveaus voor de domeinen gespreksvaardigheid, luistervaardigheid en leesvaardigheid?
2. Hoe kunnen de prestaties van de leerlingen die functioneren op <A1, A1, A2, A2+ of B1-niveau worden getypeerd?<sup>79</sup>
3. Wat is de omvang van de Engelse woordenschat van de leerlingen in leerjaar 8?
4. Wat is de onderlinge samenhang in de uitkomsten op de domeinen gespreksvaardigheid, luisteren, lezen en woordenschat en is er sprake van samenhang met de leerlingprestaties op het gebied van begrijpend lezen?
5. In hoeverre hangen kenmerken van leerlingen samen met de Engelse leerprestaties?
6. In hoeverre hangen kenmerken van het onderwijsleerproces samen met de Engelse leerprestaties?
7. In hoeverre hangen kenmerken van de leerkracht samen met de Engelse leerprestaties?
8. In hoeverre hangen achtergrondkenmerken van de school samen met de Engelse leerprestaties?

Alvorens de resultaten te beschrijven wordt in paragraaf 8.1 ingegaan op de manier waarop weging is toegepast op de resultaten en de leerlingen die zijn meegenomen in de beschreven resultaten. Daarna worden in paragraaf 8.2 de leerlingresultaten weergegeven in termen van vaardigheidsschattingen (WML-scores), waarna deze resultaten in paragraaf 8.3 worden beschreven in termen van de ERK-standaarden die ten behoeve van het peilingsonderzoek zijn bepaald (zie hoofdstuk 7). Voor de vaardigheden luisteren, lezen en gespreksvaardigheid worden de percentages leerlingen gegeven, die het A1-, A2- en B1-niveau laten zien. Voor de competentie woordenschat wordt een inschatting van de relatieve omvang van de woordenschat van de leerlingen gemaakt. Ook wordt in deze paragraaf ingegaan op de samenhang tussen de schalen. In tegenstelling tot de individuele resultaten die in de paragrafen 8.2 en 8.3 centraal staan, worden in paragraaf 8.4 de interactieresultaten op paarniveau beschreven. Ook wordt de samenhang tussen de parenscores en de gespreksresultaten op leerlingniveau in kaart gebracht. De paragrafen 8.2 en 8.3 geven daarmee dus antwoord op de onderzoeksvragen 1 en 3, terwijl informatie over de prestaties van paren (mede in relatie tot individuele scores) in 8.4 wordt toegevoegd. Paragraaf 8.5 richt zich op de samenhang tussen de vaardigheden (gebaseerd op de individuele scores), waarmee onderzoeksvraag 4 wordt beantwoord. Om na te gaan of groepen scholen en leerlingen wat betreft hun resultaten van elkaar verschillen volgt in paragraaf 8.6 een beschrijving van de relatie tussen de prestaties en de school- en individuele kenmerken op leerlingniveau. Hierbij worden de gewogen gemiddelden beschreven en is geen rekening gehouden met de hiërarchische structuur van de data. In de paragrafen 8.7 en 8.8 staan de strategie en de resultaten van de meerniveau-analyses centraal, waarin die hiërarchische

---

<sup>78</sup> De vraag naar een trendvergelijking mist in deze opsomming, maar staat in hoofdstuk 9 centraal.

<sup>79</sup> Onderzoeksvraag 2 is beantwoord in hoofdstuk 7, meer specifiek in paragraaf 7.4.3.

structuur wel is verdisconteerd. Deze analyses zijn uitgevoerd om verbanden vast stellen tussen de prestaties van de leerlingen en leerling-, klas-, leerkracht- en schoolkenmerken. Hierbij is gebruik gemaakt van ongewogen analyses, waarbij telkens is gecontroleerd voor de interactie tussen startmoment en percentage gewichtenleerlingen. In de paragrafen 8.6 en 8.8 wordt daarmee antwoord gegeven op de onderzoeksvragen 5 tot en met 8.

## 8.1 Weging en gebruikte leerlingresultaten

De weging van de leerlingen vindt plaats op basis van de gegevens zoals die bekend zijn in de populatie. Dit betreft een weging naar de combinatie van twee variabelen die in de steekproef stratificatiecriteria waren, te weten type school met twee niveaus (startmoment: EIBO – VVTO) en formatiegewichtsstratum (vier niveaus). In paragraaf 2.1 zijn deze criteria nader omschreven.

Zoals in paragraaf 2.4 is beschreven, bleek in de steekproef dat de best mogelijke gegevens over welke scholen VVTO-scholen zouden zijn niet geheel overeenkwamen met de werkelijkheid. Zo bleek dat in de steekproef vier scholen als EIBO bekend stonden maar eigenlijk sinds 2013 VVTO waren, en dat drie scholen bekend stonden als sinds 2013 VVTO, maar eigenlijk EIBO waren. Een van de oorzaken lijkt te zijn dat het startjaar niet geheel correct in de aangeleverde bestanden stond. De verwisselingen laten niet zien dat er sprake is van een systematische verschuiving. Het is niet zo dat alle VVTO-scholen eigenlijk EIBO-scholen waren, of omgekeerd. Aangezien de verwisselingen van EIBO naar VVTO en vice versa vrijwel gelijk waren, wordt aangenomen dat de gegeven verdeling in de populatie van VVTO en EIBO wel correct is. Er is weliswaar niet met 100% zekerheid vast te stellen welke school VVTO is, maar de verhouding EIBO-VVTO lijkt correct en is niet aangepast op basis van deze resultaten.

Wat betreft de strata die waren samengesteld op basis van formatiegewicht is gebruik gemaakt van de opdeling van de strata in schooljaar 2016-2017. Ten tijde van de steekproeftrekking was dat de enige beschikbare data. De afname van de peiling vond echter plaats in het daaropvolgende jaar. Dat betekende dat zowel de scholen van stratum konden verwisselen als dat de verdeling in de populatie kon veranderen. Aangezien de gegevensverzameling in het schooljaar 2017-2018 plaatsvond, is de weging gebaseerd op de informatie van het schooljaar 2017-2018. Merk daarbij op dat dit niet bij Type School speelde, aangezien deze indeling gebaseerd was op het startjaar 2013. Dit verandert niet van 2016-2017 op 2017-2018.

Bij de weging is de combinatie onderzocht van schooltype en formatiegewichtsstratum. Dat leverde een opdeling in acht combinatiecategorieën op (CC8). Per combinatiecategorie is een gewicht bepaald. Naast de vierdeling van formatiegewichtsstratum is er ook een tweedeling in formatiegewichtsstratum mogelijk: tot 10% gewichtenleerlingen (samenstelling van de eerste twee strata van de vierdeling) of meer dan 10% gewichtenleerlingen (samenstelling van de laatste twee strata van de vierdeling). In combinatie met schooltype levert dit een opdeling in vier combinatiecategorieën op (CC4). Voor CC4 zijn ook gewichten bepaald.

De wegingsfactor wordt per combinatiecategorie berekend. Die kan zowel voor de CC8 als de CC4. Hiertoe moeten voor alle combinatie-categorieën zowel de aantallen in de populatie als de aantallen in de steekproef bekend zijn. Op basis van die aantallen kunnen de percentages in de steekproef en in de populatie berekend worden. Het gewicht van een combinatiecategorie is de verhouding van die twee binnen die combinatie-categorie. Voor combinatiecategorie X (CC X) is dat

$$\text{Gewicht CC X} = (\% \text{ CC X in de Populatie}) / (\% \text{ CC X in de Steekproef}).$$

Dit is ongeacht welke combinatiecategorie of welke opdeling van de combinatiecategorieën wordt gehanteerd (CC4 of CC8). Als de percentages in de steekproef en de populatie gelijk zijn, dan is het gewicht 1. Is het percentage hoger in de steekproef dan in de populatie in die combinatiecategorie, dan is het gewicht kleiner dan 1. Is het percentage hoger in de populatie dan in de steekproef in die combinatiecategorie, dan is het gewicht groter dan 1.

In de bepaling van de gewichten was er nog een aantal zaken te bepalen. Ten eerste de vraag wat als de populatie beschouwd wordt, ten tweede welke aantallen gebruikt worden om de percentages mee te berekenen, en ten derde wat precies als de steekproef wordt beschouwd. De keuzen worden hieronder beschreven.

De weging kon plaatsvinden zowel op basis van de aantallen scholen in de populatie en de steekproef als op basis van het aantal leerlingen. Beiden zijn uitgevoerd, maar de analyses (in dit hoofdstuk) zijn gedaan op de verdelingen van de leerlingen. Dat is gedaan omdat de gegevens ook leerlinggegevens (resultaten) betroffen. Hierbij is er ook de keuze om per schaal gewichten te bepalen. Het aantal leerlingen waarbij een vaardigheid gemeten wordt verschilt enigszins bij de toetsen voor leesvaardigheid, luistervaardigheid en woordenschat. Bij het onderdeel gesprekken scheelt het duidelijk meer. Daarnaast zijn de aantallen van de vragenlijst ook te gebruiken, als wel de aantallen in de gehele dataset (2088 leerlingen). Leerlingen die 100% ontbrekende waarden hadden zijn buiten alle analyses voor de bepaling van de gewichten gelaten. Als de verhoudingen van de CC8 en CC4 in de steekproef bij de vier vaardigheden verschillen per vaardigheid of voor de vragenlijst of op het totaal, dan levert dit van al deze verschillende verdelingen ook verschillende gewichten. Om die reden zijn er per vaardigheid, voor de vragenlijst en het totaal gewichten berekend. Op het schoolniveau was overigens de variatie aanzienlijk kleiner, aangezien weliswaar niet alle leerlingen aan alles meededen, maar wel alle scholen.

Door deze werkwijze is een groot aantal gewichten bepaald. Ten eerste per type combinatiecategorie (CC8 en CC4), voor leerlingen en scholen (LLN en SCH), en voor zes varianten van de steekproef (de gehele dataset, de vier vaardigheden, en de vragenlijst). Daarnaast zijn er ook gewichten berekend met de gegevens van de 2016-2017 populatie. Al deze gewichten zijn ook vergeleken. Het bleek dat de variatie niet zeer groot was en de impact van de variatie al helemaal zeer klein was. Aangezien het de data van schooljaar 2017-2018 betreft, het over leerlingresultaten gaat in dit hoofdstuk, en doordat de gewichten voor de verschillende vaardigheden en vragenlijst niet veel van de gewichten op de gehele dataset verschilden, zijn de gewichten van de gehele dataset (ALL) als wegingsfactor gebruikt.

De resultaten van de leerlingen worden in dit hoofdstuk weergegeven als vaardigheidsschattingen van de leerlingen, evenals de standaardfout rond die schattingen. Dat zijn de schattingen zoals verkregen met het programma OPLAT; het zijn WML-schattingen. In deze rapportage worden deze weergegeven als de vaardigheden. De gerapporteerde waarden in de tabellen zijn gewogen om de populatiewaarden te verkrijgen. De weging is hier naar de combinatie van het type school (VVTO versus EIBO) en het gewichtenstratum (S4o1: geen gewichtenleerlingen op school; S4o2: meer dan 0 tot en met 10% gewichtenleerlingen; S4o3: meer dan 10 tot en met 25% gewichtenleerlingen; en S4o4: meer dan 25% gewichtenleerlingen op school). Dit levert acht verschillende gewichten op, afhankelijk van de categorisering van de school.

Bij het weergeven van de resultaten is er naast de weging ook de vraag wanneer een leerling meetelt voor het bepalen van de resultaten. De leerling bij wie 100% van de opgaven ontbreekt bij een toets of taak zal in dit onderzoek voor deze toets/taak nooit meetellen. Er kan wat betreft het evalueren van de ontbrekende waarden ook een restrictievere keuze worden gemaakt. Doel van het peilingsonderzoek is het meten van de vaardigheid Engels. Diverse andere factoren die daarbij een rol kunnen spelen, zoals een zeer lage motivatie, het missen van een pagina, het halverwege de afname de klas moeten verlaten en dergelijke, vertroebelen dat beeld. Om die reden is er een onderscheid gemaakt tussen leerlingen die wel of niet vijf opgaven op een rij niet gemaakt hebben (respectievelijk de y-serie en de x-serie<sup>80</sup>).

Omdat de y-serie ook de leerlingen bevat met meer dan 5 missende waarden op rij, bevat de y-serie altijd gegevens van meer leerlingen dan de x-serie. Het verschil in aantallen leerlingen tussen de twee groepen is echter beperkt. Het percentage leerlingen dat in de x-serie niet wordt meegenomen en dat wel in de y-serie zit is gemiddeld iets meer dan 1%: bij lezen is dat 1,3%, bij luisteren 0,4% en bij woordenschat 2,3% (ongewogen resultaten)<sup>81</sup>. Bij alle vaardigheden die met de toetsen gemeten worden, zijn er ruim 1800 leerlingen beschikbaar om de resultaten te bepalen. Kijken we naar de samenhang tussen de uitval en achtergrondvariabelen, dan zien we dat deze percentages bij EIBO en VVTO-scholen zeer vergelijkbaar zijn. De uitval van leerlingen is iets hoger bij scholen met meer gewichtenleerlingen (>10%) dan met minder gewichtenleerlingen: 1,6% versus 1,2% (ongewogen resultaten; gewogen is het verschil iets kleiner). De impact op de resultaten lijkt beperkt. De resultaten van de x-serie zijn leidend en worden in dit hoofdstuk altijd gepresenteerd. In een aantal gevallen worden ook de resultaten voor de y-serie gegeven.

Alle gerapporteerde resultaten in de paragrafen 8.2 tot en met 8.6 zijn gewogen resultaten waarbij gewogen wordt om de steekproefverdeling van 2088 leerlingen overeenkomstig de populatieverdeling van de leerlingen in schooljaar 2017-2018 gelijk te krijgen voor wat betreft de verdeling van leerlingen op VVTO- en EIBO-scholen, en voor wat betreft de verdeling van de leerlingen in de gewichtenstrata<sup>82</sup>.

## 8.2 Leerlingresultaten

Alvorens de onderzoeksvragen te beantwoorden worden eerst de gemiddelden gegeven voor de categorieën waar de weging op plaatsvindt. In Tabel 8.1 zijn de WML-scores van de x-serie en de y-serie opgenomen. Naast de gemiddelden zijn ook de standaardafwijkingen gegeven om de verschillen op de verschillende schalen te kunnen interpreteren.

---

<sup>80</sup> In paragraaf 5.1.2 worden de x- en y-serie ook genoemd, bij de beschrijving van het onderscheid dat is gemaakt in de KTT-analyses.

<sup>81</sup> Bij de gesprekstaken is er geen verschil tussen de x- en de y-serie.

<sup>82</sup> Wegingsvariabele heet "Wegingsfactor" (CC8\_LLN\_POP18\_ALL: Gewicht op basis van Leerlingen Populatie 2017-2018 Aantallen Alle 2088 lln) in het totale databestand.

Tabel 8.1a: de gemiddelde scores (WML, x- en y-serie)

serie	Gemiddelden WML x-serie											
	LE			LU			WS			GESPREK		
vaardigheid	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal
S4o1	0.36	0.58	0.40	0.26	0.44	0.30	0.32	0.60	0.38	0.11	0.25	0.13
S4o2	0.38	0.42	0.39	0.32	0.33	0.32	0.33	0.37	0.33	0.12	0.29	0.14
S4o3	0.40	0.39	0.40	0.27	0.25	0.26	0.34	0.37	0.34	0.13	-0.03	0.12
S4o4	0.33	0.38	0.33	0.18	0.25	0.19	0.31	0.38	0.31	0.25	0.07	0.23
Totaal	0.38	0.43	0.39	0.30	0.33	0.30	0.33	0.39	0.33	0.13	0.21	0.14

serie	Gemiddelden WML y-serie											
	LE			LU			WS			GESPREK		
vaardigheid	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal
S4o1	0.36	0.57	0.40	0.26	0.44	0.30	0.31	0.59	0.37	0.11	0.25	0.13
S4o2	0.37	0.40	0.37	0.32	0.33	0.32	0.32	0.36	0.32	0.12	0.29	0.14
S4o3	0.39	0.37	0.39	0.26	0.25	0.25	0.33	0.35	0.33	0.13	-0.03	0.12
S4o4	0.33	0.35	0.33	0.18	0.24	0.18	0.29	0.35	0.29	0.25	0.07	0.23
Totaal	0.37	0.41	0.37	0.29	0.33	0.30	0.32	0.38	0.32	0.13	0.21	0.14

Tabel 8.1b: de standaardafwijking (WML, x- en y-serie)

serie	Standaardafwijking WML x-serie											
	LE			LU			WS			GESPREK		
vaardigheid	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal
S4o1	0.50	0.45	0.50	0.35	0.41	0.37	0.39	0.50	0.43	0.44	0.42	0.43
S4o2	0.44	0.44	0.44	0.43	0.39	0.42	0.42	0.43	0.42	0.52	0.45	0.51
S4o3	0.43	0.44	0.43	0.41	0.35	0.40	0.41	0.39	0.41	0.50	0.68	0.52
S4o4	0.43	0.43	0.43	0.45	0.51	0.45	0.45	0.45	0.45	0.53	0.64	0.54
Totaal	0.44	0.44	0.44	0.42	0.39	0.42	0.42	0.43	0.42	0.51	0.51	0.51

serie	Standaardafwijking WML y-serie											
	LE			LU			WS			GESPREK		
vaardigheid	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal	EIBO	VVTO	Totaal
S4o1	0.50	0.47	0.50	0.35	0.42	0.37	0.39	0.50	0.43	0.44	0.42	0.43
S4o2	0.45	0.45	0.45	0.43	0.39	0.42	0.43	0.42	0.43	0.52	0.45	0.51
S4o3	0.45	0.46	0.45	0.42	0.35	0.42	0.42	0.40	0.42	0.50	0.68	0.52
S4o4	0.43	0.48	0.43	0.45	0.51	0.45	0.46	0.47	0.46	0.53	0.64	0.54
Totaal	0.45	0.46	0.45	0.43	0.39	0.42	0.43	0.43	0.43	0.51	0.51	0.51

Bij de vaardigheid lezen is het verschil tussen EIBO en VVTO ongeveer een 0,10 standaardafwijking (effectgrootte ongeveer 0,10) bij de x-serie. Bij luisteren is het effect iets kleiner (0,08). De verschillen met de y-serie zijn minimaal. Bij woordenschat en gesprekken was het effect iets groter met een effectgrootte van ongeveer 0,15. Bij de y-serie was deze weer vergelijkbaar.

Er zijn wel duidelijk verschillen tussen de verschillende gewichtenstrata. In combinatie met het startmoment valt allereerst op dat alle VVTO-strata bij de vaardigheden lezen, luisteren en woordenschat hoger scoren dan de vergelijkbare EIBO-strata, afgezien van lezen en luisteren bij stratum 3. Het grootste verschil tussen EIBO en VVTO treedt op bij S4o1 (geen gewichtenleerlingen). De groep leerlingen op S4o1 EIBO-scholen is in de populatie relatief klein, maar bij onze steekproef bestaat de groep uit ruim 200 leerlingen van negen verschillende scholen en ook de groep leerlingen uit S4o1 VVTO-scholen is met rond de 150 leerlingen van de acht scholen relatief groot. De gevonden verschillen zijn dan ook significant en met effectgroottes van 0,46 (lezen), 0,48 (luisteren), 0,63 (woordenschat) en 0,34 (gesprekken) betekenisvol<sup>83</sup>. Van de drie toetsen zien we bij het stratum S4o4 (meer dan 25% gewichten leerlingen) dat de effectgrootte bij lezen iets groter is dan gemiddeld in de populatie. Daarentegen zijn de effectgroottes bij de middelste twee strata juist iets (S4o2) tot duidelijk (S4o3) kleiner. Bij gesprekken, de praktische taak, wordt een ander resultaat gevonden: daar is juist het verschil bij S4o2 vergelijkbaar met dat van S4o1 (effectgrootte rond 0,35), terwijl de VVTO-scholen bij de strata met veel gewichtenleerlingen (meer dan 10% gewichtenleerlingen) duidelijk slechter presteren dan EIBO-scholen. Zowel bij S4o3 als S4o4 ligt het negatieve effect hier rond 0,30.

### 8.3 ERK-prestaties gemeten vaardigheden en relatieve woordenschatomvang

De in de vorige paragraaf beschreven vaardigheidsschattingen, de WML-scores, zijn vertaald naar ERK-niveaus tijdens een standaardbepaling. In paragraaf 8.3.1 worden de resultaten voor lezen, luisteren en gespreksvaardigheid beschreven in termen van ERK niveaus, waarna in paragraaf 8.3.2 de relatieve omvang voor woordenschat van leerlingen centraal staat.

#### 8.3.1 ERK-niveaus voor lees-, luister- en gespreksvaardigheid

De eerste te beantwoorden onderzoeksvraag behelst het percentage leerlingen in groep 8 dat de ERK niveaus A1, A2 en B1 behaalt. Hiervoor zijn de tijdens de standaardbepaling verkregen grenzen vertaald naar de vaardigheidsschaal. Dat is ten eerste gedaan naar de vaardigheidsschaal waarop de opgaven ook geschat zijn (met CML, behalve bij gesprekken waarbij MML gebruikt – zie hoofdstuk 5). Deze grenzen worden gegeven in Tabel 8.2. Voor leesvaardigheid betekent dit bijvoorbeeld dat de grens tussen het <A1/A1 niveau ligt op vaardigheidsscore -0.1984. Dit betekent dat leerlingen met een vaardigheidsscore lager dan -0.1984 de *Can do* beschrijvingen behorend bij het A1-niveau nog niet beheersen. Zo zijn deze leerlingen nog niet (volledig) in staat vertrouwde namen, woorden en zeer eenvoudige zinnen te begrijpen, bijvoorbeeld in mededelingen, op posters en in catalogi. Leerlingen met een score tussen -0.1984 en 0.2928 zijn hiertoe wel in staat, maar begrijpen bijvoorbeeld nog geen specifieke voorspelbare informatie in eenvoudige en alledaagse teksten. Uit Tabel 8.3 kan worden afgeleid dat minder dan 6% van de leerlingen in groep 8 dit ERK niveau <A1 heeft.

Voor een inhoudelijke invulling van de per ERK-niveau gemeten *Can do* beschrijvingen wordt verwezen naar de paragrafen 1.5, 1.7 en 1.8. Hier staan de met behulp van de toetsen/taken gemeten descriptoren per vaardigheid weergegeven. Daarnaast worden in paragraaf 7.4.3 voor lezen

---

<sup>83</sup> N.B. Bij deze analyses is geen rekening gehouden met intra-klasse correlatie. De geneste structuur wordt wel in acht genomen in de meerniveau analyses die in de paragrafen 8.7 en 8.8 staan beschreven.

en luisteren voorbeelditems getoond van de opgaven die de leerling met een grensscore op A1-, A2 en B1-niveau nét en in ruime mate beheerst.

Tabel 8.2: ERK-Grenzen op de vaardigheidsschaal voor Lezen, Luisteren en Gesprekken

grens tussen	Lezen (LE)	Luisteren (LU)	Gesprekken (GESPR)
<A1 / A1	-0.1984	-0.4907	-0.0849
A1 / A2	0.2928	0.0766	0.6427
A2 / B1	0.6445	0.5880	0.7598

De (gewogen) resultaten – de percentages leerlingen per ERK-niveau - zijn gegeven voor zowel de x- als de y-serie (cursief) gegeven in Tabel 8.3. De resultaten zijn zowel per niveau als cumulatief (+) gegeven. Merk op dat de B1-categorie eigenlijk een cumulatieve categorie is, aangezien we geen onderscheid kunnen maken tussen B1 en de hogere niveaus. De waarden voor B1 en B1(+) zijn daarom gelijk.

Tabel 8.3: Percentages leerlingen per ERK-niveau

	Lezen		Luisteren		Gesprekken
	<i>y-serie</i>	<i>x-serie</i>	<i>y-serie</i>	<i>x-serie</i>	
<A1	6.5%	5.6%	1.0%	0.7%	31.9%
A1	40.8%	41.0%	31.0%	31.1%	50.2%
A2	27.6%	28.0%	47.3%	47.5%	8.5%
B1	25.1%	25.4%	20.6%	20.7%	9.4%
A1(+)	93.5%	94.4%	99.0%	99.3%	68.1%
A2(+)	52.7%	53.4%	68.0%	68.2%	17.9%
B1(+)	25.1%	25.4%	20.6%	20.7%	9.4%

Deze resultaten kunnen ook per combinatiestratum bekeken worden. Wanneer we naar de aantallen leerlingen en daarna naar de percentages per combinatiestratum kijken, heeft weging geen zin. De wegingscategorieën zijn immers dezelfde als die waarover we hier rapporteren.

### 8.3.2 Relatieve omvang woordenschat

Woordenschat kan niet geïnterpreteerd worden in termen van het ERK. In paragraaf 7.5 staat beschreven op welke manier de relatieve omvang van de woordenschat bepaald kan worden. Hiertoe is de vraag naar het "kennen van woorden" vertaald naar het "kunnen maken van (vergelijkbare) opgaven over woordenschat". Met deze operationalisering kan aan de hand van IRT het te verwachten aantal goede opgaven woordenschat ten opzichte van alle woordenschatopgaven worden berekend. Hierbij kan er zowel een ongewogen (ieder woord heeft eenzelfde impact) als een gewogen (sommige woorden kunnen een hoger "gewicht" meekrijgen) woordenschatomvang bepaald worden. In paragraaf 7.5 wordt voor deze procedure reeds een voorbehoud gemaakt: de gegeven percentages geven niet echt de "woordenschatomvang" weer, aangezien het problematisch is het kennen van een woord goed te definiëren.

In tabel 8.4 worden de resultaten op twee manier gepresenteerd: per combinatiestratum gegeven en als gewogen resultaten voor het totaal van de relatieve woordenschatomvang (zowel



voor de gelijke gewichten en gewogen gewichten). In de tabel staan achtereenvolgens het gemiddelde percentage woorden (gem.), de standaardafwijking (st.afw.), het minimum (min.) maximum (max.).

Tabel 8.4: Omvang relatieve woordenschat

%woorden		Iedere opgave gelijk gewogen				Woorden gewogen (a-parameter)			
Type school		Gem.	St.Afw.	Min.	Max	Gem.	St.Afw.	Min.	Max
EIBO	S4o1	65.6	18.3	26.1	95.0	66.4	19.4	24.2	96.6
	S4o2	65.3	18.5	5.2	98.2	66.1	19.5	3.8	99.1
	S4o3	65.9	18.3	15.8	97.1	66.7	19.3	13.7	98.3
	S4o4	63.3	20.1	18.6	98.3	64.0	21.2	16.5	99.2
VVTO	S4o1	75.6	17.2	24.9	98.6	76.9	17.9	23.0	99.3
	S4o2	67.4	18.4	18.2	98.3	68.3	19.4	16.1	99.2
	S4o3	67.4	18.3	18.6	95.6	68.4	19.4	16.5	97.1
	S4o4	67.2	21.4	11.4	94.0	68.1	22.7	9.4	95.7
Totaal		65,6	18.6	5.2	98.6	66.4	19.6	3.8	99.3

#### 8.4 Resultaten interactie (paarscores)

In paragraaf 8.2 zijn de individuele leerlingprestaties op de toetsen en gesprekstaken besproken. Naast een individuele gespreksbeoordeling is ook een interactiescore per leerlingpaar berekend. Met behulp van de interactiescores wordt antwoord gegeven op twee vragen: 1) Is er sprake van interactie per tweetal? en 2) In hoeverre is er sprake van Engelstalige interactie in leerlingparen uit groep 8? Bij ieder van de taken worden hiertoe een drietal scores gegeven, te weten a) de Interactiescore, b) de Strikte score voor 'Interactie in het Engels' en c) de Milde score voor 'Interactie in het Engels', zie paragraaf 5.2.3 voor meer informatie. In Tabel 8.5 worden de maximumscores voor de drie verschillende scores gegeven. De scorering verschilt per taak en per type score. Bij taak 1 en 3 worden de paren op vier subtaken beoordeeld, terwijl dat bij Taak 2 op vijf subtaken is. Bij de Milde beoordeling heeft de score GT(1), dat wil zeggen dat de leerlingen volledig Engels spraken tijdens de subtaak, een gewicht 2 waardoor ook het maximum met twee vermenigvuldigd wordt. De minimale score was in alle gevallen 0.

Tabel 8.5: Maximum scores bij de verschillende Taken en type scores

Maximum score	Taak 1	Taak 2	Taak 3
Interactiescore:	4	5	4
Int_Engels_strikt:	4	5	4
Int_Engels_mild:	8	10	8

De eigenschappen van de scoreverdelingen die gevonden zijn bij de verschillende taken worden gegeven in Tabel 8.6. Daarin worden van de P'-waarde<sup>84</sup> en de score de mediaan, het gemiddelde en de standaardafwijking gegeven.

Tabel 8.6: Verdelingseigenschappen van de scores bij de drie taken

Type score Taak		Interactie			Mild			Strikt		
		Tk 1	Tk 2	Tk 3	Tk 1	Tk 2	Tk 3	Tk 1	Tk 2	Tk 3
Maximumscore		4	5	4	8	10	8	4	5	4
aantal observaties		286	228	85	286	228	85	286	228	85
P'-waarde	Mediaan	100	100	100	63	55	63	25	10	25
	Gemiddelde	100	100	100	62	61	64	24	22	29
Score	Mediaan	4	5	4	5	5.5	5	1	0.5	1
	Gemiddelde	4.00	4.99	4.00	4.94	6.07	5.15	0.95	1.08	1.15
	Standaardafwijking	0.06	0.15	0.00	0.94	1.38	1.21	0.93	1.36	1.21

Wanneer we kijken naar de scores is duidelijk dat de 'algemene interactiescore' weinig variatie oplevert. Bijna alle paren interacteren bij alle onderdelen met elkaar. Er was een drietal uitzonderingen, te weten een paar dat bij Taak 1 bij één van de vier subtaken geen interactie had. Bij Taak 2 was dit bij één paar bij 1 van de 5 subtaken het geval en bij één paar bij 2 van de 5 subtaken. De scores 0, 1 en 2 worden bij Taak 2 nooit geobserveerd. Dit is een positieve bevinding. Alle leerlingen zijn blijkbaar in staat om met elkaar in een gesprekstaak tot enige vorm van interactie te komen, al dan niet in het Engels en/of met gebruik van gebaren. Wanneer we echter beogen onderscheid te maken tussen paren, voegt deze maat niet veel toe.

De variatie bij de andere twee scores was groter en zeer vergelijkbaar over de taken heen. Het is duidelijk dat zuiver Engelstalige interactie (gemeten met de strikte score) een stuk moeilijker is: de gemiddelde P'-waarde is 24 over paren en taken heen. Dit is ook te zien aan de relatief lage medianen en gemiddelden behorend bij de strikte score. Bij de Milde score was de gemiddelde P'-waarde 64. Bij de Strikte score wordt de zo minimaal mogelijke score van 0 geobserveerd bij alle taken. Bij de Milde score is de laagst geobserveerde score '3' bij Taak 2 (30% van de maximumscore). Beide scores lijken bruikbaar om paren mee te omschrijven.

#### 8.4.1 Relatie tussen de beoordeling van de paren en de individuele gespreksscores

De paarsgewijze scores – te weten de Interactiescore (Sc 1), de "milde" Engels interactiescore (Sc 2) en de "strengere" Engels interactiescore (Sc 3) - zijn gegeven per taak (taak 1, taak 2 en taak 3). Naast de scores per taak is ook de som van scores op taak 1 en taak 2 en op taak 1, 2 en 3 berekend. Die combinaties van paren zijn de te verwachten combinaties gezien de opdracht. De relatie tussen de beoordeling van de paren, die de gesprekken voerden, en de individuele gespreksscores kunnen op verschillende manieren worden weergegeven. Eerst zal deze gegeven worden als correlatie (Tabel 8.7).

Een aantal zaken in Tabel 8.7 valt op. Ten eerste dat de Interactiescore (Sc 1) laag samenhangt met de gespreksvaardigheidsscore (T1, T2, T3, paarT1T2 en paarALLT). Dit wordt vooral

<sup>84</sup> Gemiddelde score gedeeld door het verschil tussen de maximum en de minimumscore.

veroorzaakt doordat deze interactiescore weinig variatie vertoont. Zoals eerder vermeld heeft in bijna alle gevallen iedereen de maximale score. Dat is te verwachten omdat men alleen niet de maximale score verkrijgt als een paar ergens de score “geen interactie” heeft verkregen en dat gebeurt zelden. In een aantal gevallen (zie: **Geen variatie in scores**) heeft iedereen de maximumscore.

Kijken we naar de score 2 en 3, de milde en een strenge score voor interactie in het Engels, dan zien we dat de correlaties sterk op elkaar lijken. Dat is ook te verwachten aangezien de scores in de meeste gevallen slechts een constante waarde van elkaar verschillen (zie de formules voor de berekening van de paarscore in paragraaf 5.2.3). Dit is alleen niet het geval als bij een paar ergens “geen interactie” geobserveerd is. In theorie zijn verschillen anders dan met een constante waarde ook het geval als bij dat paar de interactie beoordeeld is als “Deels Nederlandstalig, deels non-verbaal”, “Volledig non-verbaal” of “Volledig Nederlandstalig”, maar dat was in dit onderzoek bij geen van de paren het geval. De gevonden correlaties met de vaardigheid liggen bij score 2 en 3 om en nabij de 0,45. De correlatie bij taak 3 is lager (0,3). Doordat er relatief minder variatie was in de vaardigheid van de leerlingen die taak 3 maakten is de correlatie met de paarscore relatief lager.

De vaardigheidsscores kunnen ook per paarsgewijze score worden beschouwd. In dat geval bleek een hogere paarsgewijze score niet altijd gemiddeld een hogere vaardigheid op te leveren. Als paarsgewijze scores echter samengenomen worden (taak 1 en 2; taak 1, 2 en 3) is er een monotoon stijgend verband tussen deze paarsgewijze score en vaardigheidsscore zichtbaar.

Tabel 8.7: correlatie tussen verschillende scores van de beoordelingen van de paren (paarscore en individuele score)

Taak 1	Correlatie	Sign	N
T1 Sc 1	0.152	0	536
T1 Sc 2	0.443	0	536
T1 Sc 3	0.436	0	536
Taak 2	Correlatie	Sign	N
<b>T2 Sc 1</b>	<b>0.018</b>	<b>0.705</b>	<b>428</b>
T2 Sc 2	0.488	0	428
T2 Sc 3	0.491	0	428
Taak 3	Correlatie	Sign	N
<b>T3 Sc 1</b>	<b>Geen variatie in scores</b>		
T3 Sc 2	0.307	0	158
T3 Sc 3	0.307	0	158
Taak 1&2	Correlatie	Sign	N
<b>PaarT1T2Sc1</b>	<b>0.018</b>	<b>0.711</b>	<b>424</b>
PaarT1T2Sc2	0.516	0	424
PaarT1T2Sc3	0.518	0	424
ALL tkn (1.2&3)	Correlatie	Sign	N
<b>PaarALLTSc1</b>	<b>Geen variatie in scores</b>		
PaarALLTSc2	0.474	0	152
PaarALLTSc3	0.474	0	152

## 8.5 Samenhang tussen de vaardigheden

Onderzoeksvraag 3 is gericht op de onderlinge samenhang tussen de vaardigheden van de leerlingen<sup>85</sup>. Hiertoe is de correlatie berekend tussen de verschillende (individuele) vaardigheidsschattingen van de leerlingen. Deze correlaties staan weergegeven in Tabel 8.8, die eerder in paragraaf 5.4.4 is weergegeven (Tabel 5.12). De correlaties zijn redelijk hoog, hetgeen betekent dat er sprake is van een matig tot sterk verband tussen de vier gemeten vaardigheden. Zo presteren leerlingen met hoge scores op lezen over het algemeen ook goed op luisteren, gespreksvaardigheid en met name woordenschat. Met name het verband tussen woordenschat en lees- en gespreksvaardigheid is vrij sterk te noemen. De correlatie tussen luisteren en gespreksvaardigheid is met 0.56 relatief lager.

8.8: correlaties tussen de vaardigheidsscores<sup>86</sup>

	lezen	luisteren	woordenschat	gespreksvaardigheid
LE		0.64	0.76	0.65
LU	0.64		0.64	0.56
WS	0.76	0.64		0.75
GESPRESK	0.65	0.56	0.75	

## 8.6 Koppeling prestaties leerlingen aan school- en individuele kenmerken

In deze paragraaf worden de resultaten gegeven op schoolniveau. Hierbij is geen rekening gehouden met het feit dat leerlingen gegroepeerd zijn binnen scholen. De resultaten zijn dus enkel op leerlingniveau. De gemiddelden zijn de *gewogen* gemiddelden, met weging naar gewichtenstratum en startmoment (CC8). In paragraaf 8.8 worden de resultaten van leerlingen nogmaals in relatie gebracht met leerling-, klas-, leerkracht- en schoolkenmerken in meerniveau modellen waarbij wel rekening is gehouden met de clustering van leerlingen in scholen. Deze meerniveau analyses zijn ongewogen.

### 8.6.1 Schoolvariabelen

Het belangrijkste onderscheid op schoolniveau wordt op twee kenmerken gegeven, te weten a) is de school een EIBO-school of een VVTO-school? en b) wat is het percentage gewichtenleerlingen op de school (vier gewichtenstrata)? De resultaten in Tabel 8.9 laten de resultaten per variabele zien (startmoment en gewichtenstratum), maar ook de combinatie van de twee (startmoment X gewichtenstratum). Deze laatste opdeling is ook de basis voor de weging, dus hier verschillen de gewogen en de ongewogen resultaten niet van elkaar. Tot slot worden de resultaten ook gegeven voor de combinatie van startmoment en een indikking van de gewichtenstrata, te weten de scholen tot en met 10% gewichtenleerlingen en scholen met meer dan 10% gewichtenleerlingen. De

<sup>85</sup> Onderzoeksvraag 3 is tevens gericht op de samenhang tussen de leerlingresultaten voor Engels en hun resultaten voor begrijpend lezen Nederlands. Deze resultaten worden in paragraaf 8.6.2 beschreven.

<sup>86</sup> De getoonde correlaties hebben betrekking op de individuele vaardigheidsscores van leerlingen. Dit geldt ook voor gespreksvaardigheid.

resultaten worden hier in Tabel 8.9 alleen gepresenteerd voor de x-serie<sup>87</sup>. Kijken we naar de y-serie dan zijn de resultaten weinig anders.

De vergelijking van de resultaten tussen EIBO- en VVTO-scholen laat zien dat leerlingen op VVTO-scholen iets beter presteren, met name bij lezen, woordenschat en bij gesprekken. Kijken we naar de gewichtenstrata dan worden iets opvallender resultaten gevonden. De verschillen tussen S1, S2 en S3 zijn relatief klein. Kijken we echter naar S4-scholen (>25% gewichtenleerlingen) zien we dat deze leerlingen op de toetsen lezen, luisteren en woordenschat minder goed presteren dan leerlingen uit de andere drie strata, maar bij de gesprekken beter.

Bij de combinatiegroepen zien we dat de best presterende groep de VVTO-scholen zonder gewichtenleerlingen is. Ongeacht het gewichtenstratum hebben de leerlingen op de VVTO-scholen gemiddeld bij de toetsen een hogere vaardigheid, behalve bij S3 voor lezen en luisteren. Bij gesprekken zien we dat de leerlingen op de S3 en S4 scholen binnen het VVTO gemiddeld duidelijk slechter presteren dan die op de S1 en S2 scholen. Bij de reguliere scholen vinden we bij spreken juist een omgekeerd effect. Wat hiervoor de verklaring is, is niet bekend.

Tabel 8.9: leerling-resultaten op schoolniveau (gewogen resultaten)

	Gemiddelde (WML-schaal)				Standaardafwijking (WML)			
	x-serie		x&y-serie		x-serie		x&y-serie	
	LE	LU	WS	GESPR	LE	LU	WS	GESPR
<b>Totaal</b>	<b>0.39</b>	<b>0.30</b>	<b>0.33</b>	<b>0.14</b>	<b>0.44</b>	<b>0.42</b>	<b>0.42</b>	<b>0.51</b>
Regulier	0.38	0.30	0.33	0.13	0.44	0.42	0.42	0.51
VVTO	0.43	0.33	0.39	0.21	0.44	0.39	0.43	0.51
S1: 0% gew lln	0.40	0.30	0.38	0.13	0.50	0.37	0.43	0.43
S2: >0 - 10% gew lln	0.39	0.32	0.33	0.14	0.44	0.42	0.42	0.51
S3: >10 - 25% gew lln	0.40	0.26	0.34	0.12	0.43	0.40	0.41	0.52
S4: > 25% gew lln	0.33	0.19	0.31	0.23	0.43	0.45	0.45	0.54
Regulier - S1(0)	0.36	0.26	0.32	0.11	0.50	0.35	0.39	0.44
Regulier - S2(>0-10)	0.38	0.32	0.33	0.12	0.44	0.43	0.42	0.52
Regulier - S3(>10-25)	0.40	0.27	0.34	0.13	0.43	0.41	0.41	0.50
Regulier - S4(>25)	0.33	0.18	0.31	0.25	0.43	0.45	0.45	0.53
VVTO - S1(0)	0.58	0.44	0.60	0.25	0.45	0.41	0.50	0.42
VVTO - S2(>0-10)	0.42	0.33	0.37	0.29	0.44	0.39	0.43	0.45
VVTO - S3(>10-25)	0.39	0.25	0.37	-0.03	0.44	0.35	0.39	0.68
VVTO - S4(>25)	0.38	0.25	0.38	0.07	0.43	0.51	0.45	0.64
Regulier - S1&2(0-10)	0.38	0.32	0.33	0.12	0.44	0.42	0.42	0.51
Regulier - S3&4(>10)	0.38	0.24	0.33	0.16	0.43	0.42	0.42	0.51
VVTO - S1(0-10)	0.44	0.35	0.40	0.29	0.44	0.39	0.44	0.44
VVTO - S3(>10)	0.39	0.25	0.37	0.00	0.43	0.38	0.39	0.65

<sup>87</sup> De meeste gegevens in Tabel 8.1a en 8.1b zijn identiek aan de gegevens in Tabel 8.9, zij het dat ze op andere wijze worden gepresenteerd. Tabel 8.9 bevat echter additionele informatie over de opdeling naar startmoment en de indikking van de gewichtenstrata.

### 8.6.2 Individuele leerling-variabelen

De individuele variabelen die meegenomen zijn, zijn geslacht, formatiegewicht en leeftijd. Deze laatste variabele is meegenomen als "leertijd". Het door de docent ingeschatte doorstroomniveau naar het voortgezet onderwijs is ook meegenomen. Daarnaast is gevraagd naar het ingeschatte niveau Engels en de prestaties voor begrijpend lezen Nederlands. Bij veruit de meeste leerlingen waren waarden opgegeven voor deze variabelen. Bij de variabele geslacht is deze van 99,5% van de leerlingen bekend. Bij formatiegewicht was dat het geval bij bijna 95% van de leerlingen. Bij leeftijd is dat bij 93% het geval en bij het doorstroom advies is dat 92%. De resultaten in Tabel 8.10 zijn de gewogen resultaten en worden gegeven na verwijdering van de leerlingen met vijf ontbrekende waarden op een rij. Het betreft hier dus de x-serie.

Tabel 8.10: leerling-resultaten op schoolniveau (gewogen resultaten)

	Gemiddelde (WML-schaal)				Standaardafwijking (WML)			
	x-serie		x/y		x-serie		x/y	
	LE	LU	WS	GESPR	LE	LU	WS	GESPR
<b>Totaal</b>	0.39	0.30	0.33	0.14	0.44	0.42	0.42	0.51
<b>Geslacht</b>								
Man/Jongen	0.40	0.34	0.35	0.13	0.44	0.45	0.41	0.53
Vrouw/Meisje	0.37	0.26	0.32	0.15	0.44	0.39	0.43	0.49
<b>Formatiegewicht</b>								
Geen (0,0)	0.40	0.31	0.34	0.15	0.44	0.42	0.42	0.50
Laag (0,3)	0.35	0.23	0.31	0.18	0.47	0.45	0.42	0.47
Hoog (1,2)	0.27	0.21	0.20	0.04	0.39	0.46	0.41	0.67
<b>Leertijd/Leeftijd</b>								
Regulier (11 jr)	0.40	0.32	0.35	0.14	0.44	0.42	0.43	0.50
Vertraagd (12-13 jr)	0.25	0.19	0.17	0.01	0.45	0.47	0.40	0.56
Vervroegd (10 jr)	0.44	0.30	0.40	0.21	0.41	0.36	0.41	0.50
<b>Doorstroom advies</b>								
BB	0.04	0.06	0.04	-0.37	0.35	0.36	0.35	0.75
KB	0.24	0.16	0.18	-0.11	0.37	0.33	0.35	0.45
GT	0.28	0.21	0.23	0.10	0.40	0.38	0.37	0.44
Ha	0.47	0.39	0.42	0.29	0.39	0.43	0.39	0.39
VW	0.70	0.54	0.60	0.35	0.45	0.43	0.39	0.43

Bij de toetsen zien we dat de jongens iets beter presteren dan de meisjes. Bij Gesprekken is dat omgekeerd, zij het dat het verschil iets kleiner is. Van de gevonden resultaten is alleen het resultaat bij luisteren significant<sup>88</sup>. Bij Luisteren is de effectgrootte 0,20; bij de drie andere vaardigheden is deze in absolute waarde kleiner dan 0,10. Bij formatiegewicht zien we dat de leerlingen zonder formatiegewicht beter presteren dan de leerlingen met een formatiegewicht. Alleen de verschillen tussen geen gewicht en een hoog gewicht zijn significant bij woordenschat, met een effectgrootte van 0,34. De verschillen bij Lezen en Luisteren zijn hierbij alleen significant bij  $\alpha=0.05$  (effectgroottes 0,29 en 0,22). De overige verschillen zijn geen van alle significant.

Bij de variabele leeftijd hebben we gekeken naar de leeftijd op 1 oktober 2017. Dat was van oorsprong de grensdatum om in groep 2 te bepalen of een leerling door mocht naar groep 3. Hoewel de grens niet officieel meer geldt, is deze voor de leerlingen die nu in groep 8 zitten nog wel relevant, aangezien die over het algemeen nog wel gehanteerd lijkt bij die groep: de verdeling van de 11-jarige

<sup>88</sup> Wanneer er over significantie gesproken wordt betreft dit significant bij  $\alpha=0.01$ , tenzij anders vermeld.

leerlingen is door de jaren redelijk constant rond de 75%. Door de jaren heen<sup>89</sup> is het percentage vervroegde leerlingen in groep 8 (jonger dan 11 op 1 oktober; in dit geval alleen 10-jarigen) toegenomen en het percentage vertraagde leerlingen in groep 8 (ouder dan 11 op 1 oktober; vooral 12-jarigen en een enkele 13-jarige) afgenomen. Nu is het percentage vertraagde en de vervroegde leerlingen ongeveer gelijk (beiden tegen de 12%). De vertraagde leerlingen presteren op alle onderdelen slechter dan de reguliere en vervroegde leerlingen. Bij de toetsen zijn deze verschillen significant; bij gesprekken is het verschil tussen vertraagd en vervroegd alleen significant bij  $\alpha=0,05$ . De effectgrootten bij de toetsen zijn het kleinst bij Luisteren (rond 0,30) en het grootst bij Woordenschat (rond 0,50). De verschillen tussen de reguliere en vervroegde leerlingen zijn niet significant.

Bij het doorstroomadvies voortgezet onderwijs zien we dat hoe hoger het advies, hoe hoger de score. Dit geldt voor alle vaardigheden. Bij de toetsen is het kleinste verschil tussen vmbo-KB en vmbo-GT. Dit verschil is ook niet significant. De overige verschillen zijn bij de toetsen wel significant. De verschillen zijn daarbij het kleinst bij luisteren, en het grootst bij lezen. Bij gesprekken is het verschil het kleinst tussen Havo en VWO. Dat verschil is ook niet significant. Er zijn ook resultaten berekend voor de gemengde adviezen (bijvoorbeeld havo-vwo). Over het algemeen zien we dan ook een stijging in vaardigheden bij een hoger advies, al wordt daarbij in een enkel geval ook gevonden dat het gemengde advies een iets ander resultaat laat zien. Zo is het resultaat bij KB/GT lager dan bij vmbo-KB bij Lezen, bij havo/vwo lager dan bij havo bij Luisteren en bij BB/KB hoger dan bij vmbo-KB (en vmbo-KB/GT) bij Gesprekken. Deze verschillen zijn echter niet significant, aangezien het kleine verschillen zijn bij relatief kleine groepen leerlingen.

Twee variabelen waar ook gegevens voor verzameld zijn, zijn het ingeschatte niveau Engels en de prestaties voor begrijpend lezen Nederlands. De percentages bruikbare gegevens lagen hier een stuk lager. Bij het ingeschatte niveau Engels was dat lager omdat dit alleen aan het begin van het onderzoek is bevraagd. Reden hiervoor was dat toen de optie nog openstond om de paren van de gesprekstaken samen te stellen op basis van het ingeschatte niveau Engels van de leerlingen. Om praktische en methodologische redenen is daar van afgezien en is deze variabele niet meer opgevraagd. De resultaten hier betreffen zodoende een specifiek deel van de steekproef, te weten de eerste scholen die zich aangemeld hebben. Bij iets minder dan de helft van de leerlingen is het ingeschatte niveau bekend. Bij de prestaties voor begrijpend lezen Nederlands was deze informatie bij ruim 90% van de leerlingen ingevuld, maar de ingevulde gegevens waren slechts bij ongeveer 35% bruikbaar, omdat deze op dezelfde schaal gegeven waren. Bij de bruikbare data was de vaardigheid gegeven als een LOVS-niveau I (hoogst) tot en met V (laagst). In de overige 55% werden diverse andere niet vergelijkbare waarden opgegeven die vanwege onvergelykbaarheid niet bruikbaar waren om analyses mee uit te voeren.

Zoals te verwachten hangt het ingeschatte niveau Engels sterk samen met het vertoonde niveau. De gemiddelde effectgrootte van laag naar midden ligt bij Lezen, Woordenschat en Gespreksvaardigheid rond de 0,90. Bij Luisteren is dat 0,70. De gemiddelde effectgrootte van midden naar hoog ligt rond de 0,60 bij Lezen en Luisteren en rond 0,75 bij Woordenschat en Gesprekken. Dit betreft slechts (minder dan) de helft van de leerlingen. Vergelijken we de verdelingen van de groep waarover niets bekend is met die waar een waarde opgegeven is, dan verschillen na weging de gemiddelden van de twee groepen niet significant van elkaar, ook niet bij een  $\alpha=0,05$ . Alleen bij Woordenschat is de standaardafwijking van de groep waar gegevens van bekend waren iets groter

---

<sup>89</sup> Zie <https://www.cito.nl/kennis-en-innovatie/kennisbank/p048-jaarlijkse-meting-taal-en-rekenen-2016>

dan bij de groep waar geen gegevens bekend waren. Bij de leerlingen waarvan het niveau begrijpend lezen bekend is, zien we resultaten die te verwachten waren. Hoe hoger het niveau begrijpend lezen Nederlands, hoe groter ook het taalniveau Engels. De effectgroottes zijn wat groter bij de toetsen dan bij gesprekken. In Tabel 8.11 worden alleen de samengestelde niveaus I tot en met V gegeven<sup>90</sup>.

Tabel 8.11: leerling-resultaten op schoolniveau (gewogen resultaten)

	Gemiddelde (WML-schaal)				Standaardafwijking (WML)			
	x-serie		x/y		x-serie		x/y	
	LE	LU	WS	GESPR	LE	LU	WS	GESPR
<b>Totaal</b>	<b>0.39</b>	<b>0.30</b>	<b>0.33</b>	<b>0.14</b>	<b>0.44</b>	<b>0.42</b>	<b>0.42</b>	<b>0.51</b>
<b>Ingeschat niveau Engels</b>								
Laag	0.03	0.01	-0.03	-0.30	0.31	0.32	0.31	0.55
Midden	0.36	0.29	0.33	0.11	0.43	0.40	0.40	0.41
Hoog	0.59	0.53	0.63	0.44	0.43	0.46	0.45	0.35
<b>Begrijpend Lezen NLS</b>								
Score niveau I	0.65	0.53	0.57	0.39	0.40	0.39	0.40	0.39
Score niveau II	0.46	0.40	0.43	0.32	0.43	0.42	0.43	0.50
Score niveau III	0.42	0.31	0.42	0.09	0.43	0.41	0.43	0.56
Score niveau IV	0.26	0.22	0.29	0.02	0.37	0.32	0.39	0.49
Score niveau V	0.08	0.05	0.06	0.01	0.36	0.28	0.35	0.34

### 8.6.3 Individuele leerling-variabelen op basis van de vragenlijst

Vervolgens zijn de gewogen resultaten van de x-serie gerelateerd aan de achtergrondvariabelen van de leerlingen, die met behulp van de vragenlijst zijn verzameld. Gerapporteerd wordt over de relatie tussen leerlingprestaties en gesproken thuistaal met vader en moeder, buitenschools taalcontact, Linguistic Self Confidence en de attitude ten opzichte van Engels.

Te zien is dat de leerlingen die thuis geen Nederlands spreken het beter doen dan leerlingen die dat wel doen, met name als er thuis Engels gesproken wordt. In alle gevallen levert het thuis met moeder of met vader Engels spreken een significant hogere vaardigheid op. De grootste effecten worden daarbij voor Woordenschat gevonden (tussen 0,41 en 0,54). Ook is duidelijk dat het contrast tussen een wel of niet Engels sprekende moeder groter is dan een wel of niet Engels sprekende vader. De gemiddelde effecten bij het Engels spreken met moeder ligt op 0,44 en bij vader op 0,32. Het gemiddelde effect over alle vaardigheden heen van thuis Engels spreken ligt hier precies tussenin. Opmerkelijk is dat het thuis met moeder of vader Nederlands spreken eerder een negatief dan een positief effect heeft op de vaardigheid bij Engels. Deze effecten zijn niet significant (bij  $\alpha=0,01$ ). Thuis een andere taal spreken heeft over het algemeen eerder een positief dan een negatief effect, maar ook hier zijn de verschillen klein en niet significant. Als de vader of moeder niet in Nederland geboren is, heeft dat regelmatig een positief effect op het Engels, waarbij dit in een aantal gevallen ook significant is. Dit is bijvoorbeeld het geval bij Gesprekken en Lezen, wanneer de moeder niet in Nederland is geboren. De gewogen resultaten staan weergegeven in Tabel 8.12.

<sup>90</sup> Bij de waarden in Tabel 8.7 zijn de typefouten en alternatieve invulvormen van I tot en met V (zoals i en v) bij de juiste categorie geplaatst.



Tabel 8.12: Resultaten van Thuis taal

Je spreekt thuis...	Gemiddelde				Standaardafwijking			
	LE	LU	WS	Gesprekken	LE	LU	WS	Gesprekken
Geen Nederlands	0.41	0.35	0.45	0.39	0.48	0.48	0.44	0.45
Nederlands	0.39	0.30	0.33	0.14	0.44	0.42	0.42	0.51
Geen Engels	0.35	0.28	0.29	0.08	0.43	0.41	0.40	0.50
Engels	0.51	0.39	0.49	0.30	0.46	0.46	0.45	0.49
Geen andere taal	0.38	0.31	0.32	0.11	0.44	0.42	0.41	0.51
Een andere taal	0.43	0.27	0.38	0.25	0.44	0.43	0.46	0.50
Je moeder spreekt met jou...	LE	LU	WS	Gesprekken	LE	LU	WS	Gesprekken
Geen Nederlands	0.36	0.32	0.38	0.32	0.41	0.45	0.47	0.51
Nederlands	0.39	0.30	0.34	0.13	0.44	0.42	0.42	0.51
Geen Engels	0.36	0.29	0.31	0.11	0.43	0.41	0.41	0.50
Engels	0.58	0.42	0.54	0.32	0.44	0.46	0.44	0.55
Geen andere taal	0.38	0.31	0.33	0.14	0.44	0.42	0.41	0.50
Een andere taal	0.42	0.29	0.39	0.18	0.42	0.42	0.46	0.54
Je vader spreekt met jou...	LE	LU	WS	Gesprekken	LE	LU	WS	Gesprekken
Geen Nederlands	0.44	0.33	0.43	0.31	0.43	0.46	0.49	0.56
Nederlands	0.38	0.30	0.33	0.13	0.44	0.42	0.42	0.51
Geen Engels	0.37	0.29	0.31	0.12	0.44	0.41	0.42	0.50
Engels	0.49	0.40	0.49	0.30	0.43	0.48	0.45	0.53
Geen andere taal	0.38	0.31	0.33	0.13	0.44	0.42	0.41	0.50
Een andere taal	0.41	0.27	0.37	0.18	0.41	0.42	0.46	0.56
Geboorteland ouders	LE	LU	WS	Gesprekken	LE	LU	WS	Gesprekken
Moeder, niet Nederland	0.45	0.30	0.40	0.33	0.43	0.44	0.47	0.56
Moeder, Nederland	0.38	0.30	0.33	0.11	0.44	0.42	0.41	0.50
Vader, niet Nederland	0.43	0.28	0.39	0.27	0.41	0.42	0.46	0.56
Vader, Nederland	0.38	0.31	0.32	0.12	0.44	0.42	0.42	0.50

Kijken we naar de overige onderdelen uit de vragenlijst dan betreft dat de beschreven schalen uit de vragenlijst. In Tabel 8.13 zijn de correlaties tussen de schaalscores op de leerlingvragenlijsten, de toetsen en de gesprekstaken opgenomen. Alle gevonden correlaties zijn positief en significant groter dan 0. De correlaties met de ingeschatte vaardigheid zijn het grootst. Daarnaast blijkt per vaardigheid de bijbehorende ingeschatte vaardigheid hoger te zijn dan de vaardigheidsinschatting voor de andere domeinen. De ingeschatte vaardigheid voor luisteren correleert bijvoorbeeld hoger met de luistertoets dan de ingeschatte vaardigheid voor lezen, spreken en schrijven. Voor woordenschat was er geen ingeschatte vaardigheid: daar is de hoogste correlatie gevonden met de ingeschatte vaardigheid bij spreken.

Tabel 8.13: Correlaties vragenlijstschalen en vaardigheden

<b>Vragenlijst</b>	<b>LE</b>	<b>LU</b>	<b>WS</b>	<b>Gesprekken</b>
Attitude	0.43	0.31	0.46	0.41
Linguistic Self Confidence	0.52	0.45	0.57	0.50
Buitenschools gebruik (15 items)	0.46	0.41	0.50	0.47

<b>Ingeschatte vaardigheid</b>	<b>LE</b>	<b>LU</b>	<b>WS</b>	<b>Gesprekken</b>
Alle vaardigheden	0.58	0.48	0.62	0.55
Lezen	0.51	0.41	0.52	0.44
Luisteren	0.49	0.42	0.49	0.42
Spreken	0.51	0.41	0.55	0.52
Schrijven	0.40	0.33	0.45	0.44

## 8.7 Meerniveau analyses

In paragraaf 8.6 is de relatie tussen verschillende school- en leerlingkenmerken en de leerlingprestaties aan de orde gekomen (onderzoeksvragen 5 en 8). Hierbij zijn gewogen analyses gepresenteerd, waarin rekening is gehouden met de overschatting van VVTO-scholen en bepaalde strata. In de in paragraaf 8.6 gepresenteerde analyses wordt de hiërarchische structuur van de data niet in aanmerking genomen. Echter, binnen de peiling is de data niet onafhankelijk van elkaar: leerlingen zijn geclusterd binnen klassen/scholen. Om verbanden vast stellen met leerling-, klas-, leerkracht- en school-kenmerken zijn meerniveau-analyses uitgevoerd, waarin wel wordt uitgegaan van de aanwezige afhankelijkheid in de data. Deze resultaten worden gepresenteerd in de paragrafen 8.7 en 8.8, waarin de onderzoeksvragen 5-8 worden beantwoord met data afkomstig uit de leerling-, leerkracht en schoolleidersvragenlijst (intake).

### 8.7.1 Strategie meerniveau analyses

Voor de beantwoording van de vragen wordt gebruik gemaakt van een serie van statistische modellen. De modellen zijn vergelijkbaar in de zin dat ze dezelfde afhankelijke variabelen, een vergelijkbare datastructuur en vergelijkbare toetsing hebben. Ze verschillen alleen in de set van voorspellende kenmerken (predictoren) die in de analyses zijn opgenomen. De afhankelijke variabelen in alle modellen zijn de prestaties van de leerlingen op de vier getoetste domeinen van Engels; de vaardigheid voor lezen, luisteren, woordenschat en gesprekken. De gespreksvaardigheid is de individuele score voor gespreksvaardigheid.

De getrapte steekproeftrekking van Peil Engels zorgt voor een hiërarchische structuur van de data. Het ontwerp van de steekproeftrekking impliceert bovendien dat bij ongeveer driekwart van de leerlingen een geldige score voor gespreksvaardigheid ontbreekt (zie hiervoor hoofdstuk 3). Verder is het ontbreken van informatie over kenmerken van het onderwijsleerproces (afkomstig is uit de leerkrachtvragenlijst) van zorgelijke omvang: hoewel het percentage leerkrachten met meer dan 10% missende waarden ten opzichte van de totale vragenlijst aanvaardbaar is (9%), hebben relatief veel leerkrachten één of meer missende waarden op de deelvragen over het onderwijsleerproces. Als hier geen rekening mee zou worden gehouden en alleen met volledige data per leerling geanalyseerd zou worden (*listwise deletion*), blijft een erg beperkte dataset over en wordt waardevolle informatie

over kenmerken die wél beschikbaar is niet gebruikt. Om te komen tot complete data voor iedere leerling is meerniveau multiple imputatie toegepast (Van Buren, 2012).

### 8.7.2 Multiple imputatie

Bij de meerniveau multiple imputatie zijn er twee niveaus in de dataset onderscheiden: leerlingen (*Level 1*) en klassen/scholen (*Level 2*). Het was niet mogelijk om onderscheid te maken tussen het klas- en schoolniveau. Dit kan verklaard worden door het gering aantal scholen dat met meer dan één klas aan de peiling heeft meegedaan.

De meerniveau multiple imputatie is uitgevoerd in twee stappen. **De eerste stap** betreft het imputeren van alleen de vier afhankelijke variabelen, te weten de vaardigheid voor lezen, luisteren, woordenschat en gesprekken. Deze variabelen hebben een continu meetniveau en zijn feitelijk schattingen van de vaardigheid van een leerling voor het betreffende domein van Engels, de Weighted Maximum Likelihood (WML) schattingen, waarvan de uitkomsten in paragraaf 8.2 staan beschreven. De individuele WML-schatting van de vaardigheid met bijbehorende standaardfout en variantie-covariantie matrix dienen als basis voor de multiple imputatie. Bij een ontbrekende score voor bijvoorbeeld gespreksvaardigheid van een leerling wordt een voorspelling van deze score gedaan middels meerniveau lineaire regressiemodellen die gebruikmaken van informatie over de WML-vaardigheidsscores voor lezen, luisteren en/of woordenschat van de betreffende leerling en de verdeling van deze scores voor lezen, luisteren, woordenschat en gesprekken in de gehele steekproef. Er zijn vijf imputatieronden uitgevoerd die hebben geresulteerd in vijf geïmputeerde datasets. Iedere dataset bevat steeds complete records van iedere leerling; de score van een leerling is zijn geldige WML schatting voor de betreffende vaardigheid of – indien sprake was van een ontbrekende waarde – een voorspelde vaardigheidsscore verkregen uit multiple imputatie. Een overzicht van de binnen en tussen-datasetvariantie voor de vier afhankelijke variabelen is weergegeven in Tabel 8.14.

Tabel 8.14: *Variantie binnen en tussen de vijf geïmputeerde datasets per domein van Engels vaardigheid*

	Leesvaardigheid	Luistervaardigheid	Woordenschat	Gespreksvaardigheid
<i>Binnen-datasetvariantie</i>				
Dataset 1	0.1930	0.1724	0.1955	0.2826
Dataset 2	0.1957	0.1748	0.1904	0.2672
Dataset 3	0.1923	0.1736	0.1903	0.2641
Dataset 4	0.1958	0.1720	0.1905	0.2676
Dataset 5	0.1938	0.1748	0.1896	0.2856
Overall	0.1941	0.1735	0.1912	0.2733
<i>Tussen-datasetvariantie</i>	3.2005E-06	2.3013E-06	6.8017E-06	1.0229E-04

Tabel 8.14 toont dat de binnen-dataset variantie voor gespreksvaardigheid groter is dan voor lees- en luistervaardigheid en woordenschat. Hieraan ligt het beperkt aantal leerlingen met een gespreksvaardigheid-meting in de klas ten grondslag, waardoor de schattingen gevoeliger worden

voor ruis. Voor de vier domeinen Engels is de tussen-datasetvariantie vrijwel gelijk aan 0. Er is geen sprake van vertekening.

**De tweede imputatiestap** betreft het multiple imputeren van predictoren afkomstig uit de leerling-, leerkracht- of schoolleidersvragenlijst (intake) volgens de zogenaamde *chained equation* methode (Snijders & Bosker, 2012, p 145). Deze methode – ook wel *predictive mean matching* genoemd – komt overeen met de *2l.pmm*-methode in de R-package *Mice*. Er is rekening gehouden met de geneste structuur van de data (L1: leerling; L2: klas/school). Ontbrekende scores bij discrete ordinale predictoren - zoals formatiegewicht van de leerling - zijn geïmputeerd met logistische regressie modellen en ontbrekende scores bij continue predictoren - zoals aantal jaren ervaring in het geven van Engels – met lineaire regressie modellen. Ontbrekende waarden bij nominale predictoren - zoals gebruikte didactische methode voor Engels - zijn *niet* geïmputeerd. Bij dit type variabelen is 'onbekend' als extra antwoordcategorie toegevoegd. Zo wordt voorkomen dat door het ontbreken van een score op een of meerdere nominale variabelen een leerling alsnog geheel uit de analyses wordt verwijderd (*listwise deletion*). Door multiple imputatie kan zo veel mogelijk informatie over relevante predictoren uit de drie vragenlijsten toch worden meegenomen in de analyses. Er zijn vijf imputatieronden toegepast in deze tweede imputatiestap die ook geleid hebben tot vijf geïmputeerde datasets. De vijf datasets van de tweede imputatiestap zijn samengevoegd met de vijf datasets uit de eerste imputatiestap en vormen de basisbestanden voor de meerniveau analyses.

### 8.7.3 Meerniveau analyses

Vanwege de hiërarchische structuur van de geïmputeerde datasets zijn meerniveau analyses toegepast om de onderzoeksvragen te beantwoorden. In de meerniveau analyses zijn dezelfde niveaus onderscheiden als bij de multiple imputatie; leerlingen (L1) zijn genest binnen klassen/scholen (L2). De variantie op klas- en schoolniveau is weer samengenomen in niveau 2 en het gaat hier steeds om *random intercept* modellen (Snijders & Bosker, 2012, p 41). Er is geen weging naar de populatie toegepast. Wel is in alle modellen gecorrigeerd voor de interactie tussen startmoment en stratum. Alle meerniveau modellen zijn vijf keer geschat, voor elke geïmputeerde dataset afzonderlijk (Van Buren, 2012, p 17). Vervolgens zijn per analysemodel de vijf schattingen voor de regressiecoëfficiënten en varianties op leerling- en klas/schoolniveau gepoold om te komen tot correcte schattingen van de regressiecoëfficiënten, varianties en hun standaardfouten (Snijders & Bosker, 2012, p 135). Voor de toetsing van de samenhang tussen een predictor en een afhankelijke variabele zijn 95% betrouwbaarheidsinterval rondom de gemiddelde regressiecoëfficiënt (over de vijf datasets) berekend en is de gepoolde standaardfout berekend (gemiddelde regressiecoëfficiënt  $\pm 1.96 * SE_{\text{gepooled}}$ ). Bij de bepaling van significante resultaten is dus uitgegaan van tweezijdige toetsing met een  $\alpha$  van 5%.

De onderzoeksvragen zijn beantwoord met zeven verschillende meerniveau modellen. De opzet ervan wordt hieronder besproken, een toelichting op de resultaten volgt in paragraaf 8.8.

- Model 1 betreft een leeg meerniveau model. Het belangrijkste doel is nagaan in hoeverre verschillen in Engelse vaardigheidsscores gerelateerd zijn aan verschillen tussen leerlingen (leerlingniveau) en verschillen tussen klassen (klas/schoolniveau). Indirect gaat het hier om de vraag of het zinvol is om zowel een leerling- als klas/schoolniveau in de data-analyses te onderscheiden.

- Model 2 bevat slechts een predictor, te weten de interactieterm startmoment\*stratum. Deze variabele is een combinatie van EIBO/VVTO schooltype en het percentage leerlingen met een formatiegewicht op een school. Het model geeft antwoord op de vraag in hoeverre EIBO en VVTO scholen van elkaar verschillen in Engelse vaardigheidsscores als rekening wordt gehouden met het percentage leerlingen met een formatie gewicht van .3 of 1.2. In de overige modellen is deze interactieterm startmoment\*stratum meegenomen als covariaat. In de gepresenteerde modellen wordt deze covariaat weergegeven onder de noemer 'startmoment'.
- Model 3 geeft inzicht in de samenhang tussen leerlingkenmerken en de vier Engelse vaardigheidsscores, na correctie voor startmoment\*stratum. Er is onderscheid gemaakt in algemene (Model 3a) en domeinspecifieke leerlingkenmerken (Model 3b). Bij Model 3a gaat het om de predictoren sekse, leerlinggewicht, leeftijd en schooladvies voor voortgezet onderwijs. Bij Model 3b om taalgebruik thuis, buitenschools gebruik van Engels, linguistic self confidence, attitude en eigen inschatting van de betreffende vaardigheid. Model 3 is bedoeld om antwoord te geven op de eerste onderzoeksvraag: *In hoeverre hangen kenmerken van leerlingen samen met de Engelse leerprestaties?*
- Bij Model 4 worden diverse kenmerken van het onderwijsleerproces als predictor toegevoegd aan Model 2. Het betreft een lijst van 24 predictoren. Voorbeelden zijn onderwijstijd, didactische methode, bij- of nascholing voor Engels, samenwerken in groepen door leerlingen en coördinator/werkgroep Engels hebben. Startmoment\*stratum dient weer als covariaat. De tweede onderzoeksvraag is hier aan de orde: *In hoeverre hangen kenmerken van het onderwijsleerproces samen met de Engelse leerprestaties?*. Het model geeft inzicht in welke kenmerken van het onderwijsleerproces relevant zijn voor de vaardigheidsscores Engels en welke kenmerken deels de eventuele verschillen tussen EIBO/VVTO scholen met correctie voor leerlinggewicht kunnen verklaren.
- Model 5 heeft betrekking op de invloed van achtergrondkenmerken van leerkrachten. In dit model zijn, naast startmoment\*stratum, zes achtergrondkenmerken van de leerkrachten meegenomen: opleidingsniveau, ervaring in het geven van Engels, moedertaal, buitenschools gebruik Engels en inschatting van eigen niveau van de betreffende vaardigheid. De predictoren geven een beeld van de Engelse taalvaardigheid en de vaardigheid om lessen Engels te geven van de leerkracht. Met Model 5 wordt de derde onderzoeksvraag beantwoord: *In hoeverre hangen kenmerken van de leerkracht samen met de Engelse leerprestaties?*
- Model 6 is bevat predictoren die betrekking hebben op de algemene kenmerken van scholen. Het gaat om een percentage leerlingen met een niet-Nederlandse thuistaal, urbanisatiegraad van de schoollocatie, denominatie, percentage gewichtenleerlingen en gemiddelde schooladvies van groep acht. Bij dit model gaat het om de vierde onderzoeksvraag: *In hoeverre hangen achtergrondkenmerken van de school samen met de Engelse leerprestaties?*
- Model 7 is het compacte totaalmodel. Het model is als volgt opgebouwd. Model 4 met de gehele set van onderwijsleerprocespredictoren dient als basis. Deze set is in eerste instantie uitgebreid met alleen de significante domeinspecifieke leerkracht- en leerlingkenmerken (Model 5 en 3b) en vervolgens met alle algemene leerling- en schoolkenmerken (Model 3a en Model 6). Het resulteert een uitgebreid totaal model waarin ook predictoren zitten die niet langer significant zijn. Middels een *backward selection* procedure zijn één voor één de predictoren verwijderd die

de geringste samenhang hebben met de afhankelijke variabelen. Het eindresultaat is een compact totaalmodel met predictoren die een significante samenhang hebben met tenminste één van de vier vaardigheidsscores voor Engels.

Ter ondersteuning van de interpretatie van de resultaten van de meerniveau modellen zijn de effectgroottes van de significante predictoren berekend. De effectgrootte (ES) voor een niet-continue predictor X1 (categorisch of discrete ordinaal) is berekend door de geschatte regressiecoëfficiënt  $b_{X1}$  van deze predictor te delen door de wortel van de variantie op leerlingniveau, oftewel de tussen-leerlingen variantie ( $\text{Var}_{\text{tussen leerlingen}}$ ) (Hedges, 2007). In formule:

$$ES_{\text{niet-continue predictor X1}} = b_{X1} / \sqrt{\text{Var}_{\text{tussen leerlingen}}}$$

De effectgrootte voor een continue predictor X2 is berekend middels de partiële correlatie R van deze predictor X2 met de afhankelijke variabele Y (Hedges, 2007). Deze partiële correlatie is dan gelijk aan de gestandaardiseerde regressiecoëfficiënt:

$$R_{\text{continue predictor}} = b_{X2} * SD_{X2} / \sqrt{\text{Var}_{\text{tussen leerlingen}}}$$

In deze formule is  $b_{X2}$  is de ongestandaardiseerde regressiecoëfficiënt van X2 en  $SD_{X2}$  de spreiding in de geobserveerde scores van X2, waarbij rekening is gehouden met het hiërarchisch niveau waarop de predictor van toepassing is. Bij een predictor op schoolniveau gaat het om de spreiding van de scores op X2 tussen scholen.

Dan is de effectgrootte voor een continue predictor:

$$ES_{\text{continue predictor X2}} = 2 * R_{\text{continue predictor}} / \sqrt{(1 - R_{\text{continue predictor}}^2)}$$

De interpretatie van de effectgrootte komt overeen met de standaard voor effectgroottes van Cohen (1977).

## 8.8 Resultaten meerniveau analyses

### 8.8.1 Model 1, leeg model

In Tabel 8.15 zijn de gepoolde resultaten van de meerniveau analyses voor de leerlingprestaties op de domeinen Engels lezen, luisteren, woordenschat en gespreksvaardigheid gepresenteerd. Het gaat hier om Model 1, het model zonder predictoren.

Tabel 8.15: Meerniveau modellen zonder predictoren (Model 1)

	Leesvaardigheid			Luistervaardigheid			Woordenschat		Gespreksvaardigheid			
<i>Fixed Part</i>	<i>b</i>	<i>SE(b)</i>		<i>b</i>	<i>SE(b)</i>		<i>b</i>	<i>SE(b)</i>		<i>b</i>	<i>SE(b)</i>	
Intercept	0.416	0.018	*	0.317	0.018	*	0.374	0.020	*	0.101	0.020	*
<i>Random Part:</i>												
Klas/schoolniveau	0.020	0.008	*	0.019	0.007	*	0.024	0.007	*	0.021	0.008	*
Leerlingniveau	0.175	0.035	*	0.155	0.032	*	0.167	0.032	*	0.252	0.033	*
ICC	0.10			0.11			0.13			0.08		
Aantal scholen	113			113			113			113		
Aantal leerlingen	2059			2059			2059			2059		

Het intercept van Model 1 geeft het algemeen gemiddelde in de steekproef weer. De gemiddelde Engelse leesvaardigheid van de leerlingen is 0,416 ( $SEb = ,018$ ). Mogelijk wijkt dit enigszins af van het 'gewone' rekenkundige gemiddelde (Tabel 8.1a), omdat rekening is gehouden met de geneste structuur van dataset en de gegevens deels zijn geïmputeerd. In de meerniveau modellen is de totale variantie in Engelse vaardigheidsscores is tweeën gesplitst. Model 1 geeft inzicht in welk deel van de totale variantie is toe te schrijven aan verschillen tussen leerlingen (variantie op leerlingniveau) en welke deel aan verschillen tussen scholen (variantie op schoolniveau). Voor Engelse leesvaardigheid is de variantie op leerlingniveau gelijk aan ,175 ( $SEb = ,035$ ) en op klas/schoolniveau gelijk aan ,020 ( $SEb = ,008$ ). Beide delen zijn statistisch significant afwijkend van nul (aangegeven met een \* in deze en navolgende tabellen). Dit betekent dat leerlingen onderling van elkaar verschillen in leesvaardigheid en scholen ook, maar de verschillen tussen leerlingen (variantie = ,175) is aanzienlijk groter dan de verschillen tussen scholen (variantie = ,020). Dit volgt ook uit de intraklasse correlatie (ICC). De ICC geeft aan welk deel van de totale variantie op schoolniveau ligt. Voor Engelse leesvaardigheid is de ICC gelijk aan ,10. Dit betekent dat 10% van de totale variantie in leesvaardigheid is toe te schrijven aan verschillen tussen scholen. Voor luistervaardigheid is dit 11%, voor woordenschat 13% en voor gespreksvaardigheid 8%. De analyses van Model 1 zijn steeds gebaseerd op de geïmputeerde datasets met vaardigheidsscores van 2059 groep-8-leerlingen. Deze leerlingen zitten in 113 verschillende klassen (onderste twee regels van Tabel 8.15).

### 8.8.2 Model 2, startmoment

De gepoolde resultaten van de meerniveau analyses voor Model 2 zijn gepresenteerd in Tabel 8.16. Het model geeft inzicht in welke mate de Engelse vaardigheidsscores tussen EIBO- en VVTO-basisscholen verschillen als ook rekening wordt gehouden met het percentage leerlingen met een formatiegewicht. EIBO-basisscholen met minder dan 10% leerlingen met een formatiegewicht (0.3 of 1.2) is de referentiecategorie.

Tabel 8.16: Meerniveau modellen voor startmoment\*stratum (Model 2)

	Leesvaardigheid			Luistervaardigheid			Woordenschat			Gespreksvaardigheid		
<i>Fixed Part</i>	<i>b</i>	<i>SE(b)</i>		<i>b</i>	<i>SE(b)</i>		<i>b</i>	<i>SE(b)</i>		<i>b</i>	<i>SE(b)</i>	
Intercept	0.387	0.026	*	0.309	0.025	*	0.333	0.027	*	0.062	0.029	*
EIBO - S3&4(>10%)	0.007	0.045		-0.040	0.043		0.010	0.047		0.024	0.051	
VVTO - S1(0-10%)	0.098	0.043	*	0.075	0.041		0.129	0.044		0.130	0.047	*
VVTO - S3(>10%)	0.017	0.058		-0.026	0.055		0.068	0.060		0.017	0.063	
<i>Random Part:</i>												
Klas/schoolniveau	0.020	0.008	*	0.015	0.008		0.019	0.008	*	0.016	0.008	*
Leerlingniveau	0.175	0.035	*	0.156	0.035	*	0.169	0.035	*	0.254	0.036	*
ICC	0.101			0.088			0.101			0.060		
Aantal scholen	113			113			113			113		
Aantal leerlingen	2059			2059			2059			2059		

In Tabel 8.16 staan sterretjes bij de positieve regressiecoëfficiënten (*b*) van de categorie 'VVTO-S1(0-10%)' voor de drie van de vier domeinen van de Engelse vaardigheidsscores. Leerlingen op VVTO-scholen met minder dan 10% gewichtenleerlingen presteren significant beter bij lezen, woordenschat en gespreksvaardigheid dan leerlingen op EIBO-scholen met minder dan 10% gewichtenleerlingen (referentie). Het gaat steeds om kleine effecten. De bijbehorende effectgrootten (ES) zijn 0,23 voor lezen, 0,32 voor woordenschat en 0,26 voor gespreksvaardigheid. De twee overige categorieën van de interactie tussen startmoment en stratum, te weten 'EIBO- S3&4(>10%)' en 'VVTO-S3(>10%)', hebben bij geen enkel domein een sterretje. Dit duidt erop dat zowel EIBO-scholen met veel gewichtenleerlingen als VVTO-scholen met veel gewichtenleerlingen niet statistisch significant beter of slechter presteren dan de EIBO-scholen met weinig gewichtenleerlingen.

### 8.8.3 Model 3, leerlingkenmerken

De samenhangen tussen de algemene kenmerken van leerlingen en de Engelse vaardigheidsscores zijn gepresenteerd in Tabel 8.17. Het gaat hier om de gepoolde resultaten van Model 3a. Startmoment\*stratum is meegenomen als covariaat in de modellen.



Tabel 8.17: Meerniveau modellen voor algemene leerlingkenmerken (Model 3a)

	Leesvaardigheid		Luistervaardigheid		Woordenschat		Gespreksvaardigheid	
<i>Fixed Part</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>
Intercept	0.436	0.258	0.156	0.257	0.596	0.247 *	0.220	0.314
EIBO - S3&4(>10%)	0.060	0.044	0.000	0.042	0.062	0.044 *	0.075	0.051
VVTO - S1(0-10%)	0.055	0.041	0.046	0.040	0.091	0.041	0.090	0.046 *
VVTO - S3(>10%)	0.105	0.056	0.039	0.054	0.160	0.057 *	0.106	0.062
Jongen	0.017	0.019	0.052	0.020 *	0.033	0.019	0.012	0.029
Leerlinggewicht:								
0.3	0.018	0.047	-0.049	0.048	0.033	0.047	0.027	0.062
1.2	-0.013	0.045	0.017	0.046	-0.043	0.043	-0.022	0.059
onb	0.032	0.071	0.013	0.069	0.100	0.074	0.045	0.087
Leeftijd	-0.027	0.022	-0.006	0.022	-0.046	0.021 *	-0.036	0.027
Advies VO:								
Vmbo-gl/tl	0.195	0.028 *	0.163	0.026 *	0.188	0.029 *	0.214	0.038 *
Havo-vwo	0.445	0.025 *	0.328	0.024 *	0.412	0.025 *	0.423	0.034 *
<i>Random Part:</i>								
Klas/schoolniveau	0.016	0.007 *	0.014	0.007	0.017	0.007 *	0.015	0.008 *
Leerlingniveau	0.142	0.032 *	0.137	0.031 *	0.139	0.032 *	0.223	0.032 *
ICC	0.099		0.094		0.107		0.064	
Aantal scholen	113		113		113		113	
Aantal leerlingen	2059		2059		2059		2059	

Noot. 'onb' = onbekend

Tabel 8.17 toont dat, na correctie voor startmoment\*stratum, leerlingen met een Vmbo-hoog of Havo-vwo advies significant beter presteren op alle vier domeinen van Engels dan de leerlingen met een Vmbo-laag advies (de referentiecategorie). De effectgroottes variëren tussen 0,44 en 0,52 voor het Vmbo hoog advies en tussen 0,89 en 1,18 voor Havo-vwo advies, hetgeen duidt op een respectievelijk matig en erg groot effect. Verder is gevonden dat jongens iets beter presteren bij luistervaardigheid dan meisjes ( $b = ,052$ ;  $ES = 0,14$ ) en dat oudere leerlingen in groep 8 een iets geringere woordenschat hebben dan jongere groepsgenoten ( $b = -,046$ ,  $ES = -0,12$ ). Het gaat hier om onafhankelijke associaties; de verbanden gelden na correctie voor de overige predictoren in het model.

De associaties tussen de domeinspecifieke leerlingkenmerken en de Engelse vaardigheidsscores zijn gepresenteerd in Tabel 8.18. Het betreft hier de gepoolde resultaten van Model 3b.

Tabel 8.18: Meerniveau modellen voor domeinspecifieke leerlingkenmerken (Model 3b)

	Leesvaardigheid		Luistervaardigheid		Woordenschat		Gespreks- vaardigheid	
	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>
<i>Fixed Part</i>								
Intercept	-0.505	0.040 *	-0.373	0.042 *	-0.584	0.039 *	-0.833	0.049 *
EIBO - S3&4(>10%)	-0.043	0.035	-0.063	0.035	-0.044	0.036	-0.034	0.042
VVTO - S1(0-10%)	0.062	0.033	0.038	0.034	0.093	0.033 *	0.084	0.038 *
VVTO - S3(>10%)	-0.032	0.045	-0.055	0.045	0.013	0.045	-0.048	0.051
Taalgebruik thuis	-0.016	0.008	-0.024	0.009 *	-0.011	0.008	-0.002	0.011
Buitenschools gebruik								
Engels	0.003	0.002	0.006	0.002 *	0.007	0.001 *	0.006	0.002 *
Linguistic self confidence	0.040	0.006 *	0.036	0.006 *	0.045	0.005 *	0.036	0.007 *
Attitude	0.036	0.006 *	0.010	0.006	0.031	0.006 *	0.026	0.008 *
Ingeschatte vaardigheid lezen/luisteren/spreken#	0.098	0.012 *	0.070	0.009 *	0.084	0.011 *	0.101	0.014 *
<i>Random Part:</i>								
Klas/schoolniveau	0.008	0.007	0.009	0.007	0.009	0.007	0.007	0.008
Leerlingniveau	0.123	0.031	0.123	0.031	0.113	0.031	0.197	0.036
ICC	0.061		0.065		0.075		0.035	
Aantal scholen	113		113		113		113	
Aantal leerlingen	2059		2059		2059		2059	

*Noot.* # Bij woordenschat is de ingeschatte vaardigheid voor lezen als predictor in het model opgenomen.

Tabel 8.18 toont dat, na correctie voor startmoment\*stratum, diverse domeinspecifieke leerlingkenmerken een deel van de variantie op leerling- en klasniveau kunnen verklaren. Daarnaast daalt de ICC van bijna 10% (Model 2) naar ongeveer 6%. Dit is vooral te herleiden tot de ingeschatte vaardigheid van leerlingen. Van de significante predictoren zijn de effectgroottes van de ingeschatte vaardigheid voor een specifiek domein van Engels het grootst (ES = 0,51 tot ES = 0,59), hetgeen staat voor een middelgroot effect. Daarna volgt linguistic self confidence met een klein tot middelgroot effect (ES = 0,41 tot ES = 0,69), afhankelijk van het domein. De predictoren attitude (ES = 0,23 tot ES = 0,41) en buitenschools gebruik Engels (ES = 0,21 tot ES = 0,28) hebben een klein en taalgebruik thuis een zeer klein effect (ES = -0,07).

#### 8.8.4 Model 4, onderwijsleerproceskenmerken

De gepoolde resultaten van de meerniveau analyses voor Model 4 zijn gegeven in Tabel 8.19. Het model geeft inzicht in welke kenmerken van het onderwijsleerproces samenhangen met de vaardigheidsscores van leerlingen voor Engels lezen, luisteren, woordenschat en gespreksvaardigheid.

Tabel 8.19: Meerniveau modellen voor onderwijsleerproces (Model 4)

	Leesvaardigheid		Luistervaardigheid		Woordenschat		Gespreks vaardigheid	
	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>
<i>Fixed Part</i>								
Intercept	0.088	0.113	0.189	0.121	0.038	0.110	-0.514	0.169 *
EIBO - S3&4(>10%)	0.031	0.043	-0.011	0.040	0.033	0.043	0.070	0.053
VVTO - S1(0-10%)	-0.018	0.059	-0.012	0.061	-0.033	0.064	-0.025	0.072
VVTO - S3(>10)	-0.069	0.067	-0.027	0.065	-0.026	0.069	-0.097	0.071
Onderwijstijd	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000
Didactische methode #:								
Just do it	0.029	0.078	0.033	0.073	0.098	0.073	0.098	0.083
Step to	-0.083	0.063	-0.046	0.066	-0.038	0.064	-0.048	0.074
Groove me	-0.014	0.042	-0.093	0.040 *	-0.073	0.043	-0.063	0.046
Overig	0.007	0.047	-0.058	0.049	0.045	0.050	-0.006	0.052
methode onb	0.136	0.091	-0.087	0.086	0.033	0.094	0.039	0.096
Extra materiaal	0.041	0.039	0.030	0.036	0.065	0.040	0.057	0.048
Extra materiaal onb	-0.163	0.153	0.033	0.139	0.158	0.162	0.067	0.178
Differentiatie	0.070	0.041	0.015	0.034	-0.003	0.040	-0.002	0.042
Faciliteiten Engels	-0.077	0.039 *	-0.074	0.038	-0.075	0.038 *	-0.056	0.048
Faciliteiten onb	-0.066	0.180	0.033	0.177	-0.287	0.184	-0.197	0.197
Bijscholing Engels	0.064	0.055	0.029	0.054	0.053	0.062	0.068	0.069
Bijscholing onb	0.091	0.052	0.020	0.048	0.010	0.051	0.027	0.074
Inzet authentiek Engels materiaal	0.008	0.016	-0.007	0.017	-0.005	0.018	-0.005	0.020
Engels als voertaal	-0.004	0.014	0.012	0.014	-0.005	0.015	0.004	0.016
Werken in groepen	0.050	0.020 *	0.004	0.021	0.014	0.020	0.040	0.019 *
Huiswerk Engels	-0.021	0.031	-0.026	0.028	-0.037	0.030	-0.035	0.036
Integratie Engels in andere vakken	-0.006	0.038	0.040	0.037	-0.031	0.039	-0.029	0.042
Integratie Engels onb	0.151	0.183	0.223	0.168	0.094	0.178	0.119	0.208
Toetsing Engels	-0.056	0.067	-0.079	0.059	-0.067	0.068	0.004	0.080
Toetsing Engels onb	-0.141	0.123	-0.179	0.113	-0.283	0.120 *	-0.195	0.148
Overdracht VO	-0.047	0.035	-0.003	0.034	-0.047	0.036	-0.051	0.043
Overdracht VO onb	-0.106	0.112	-0.126	0.110	-0.041	0.114	-0.169	0.130
Coördinator Engels	0.030	0.046	0.095	0.040 *	0.081	0.043	0.076	0.047
Coördinator Engels onb	-0.043	0.115	0.011	0.105	0.109	0.117	0.099	0.140
Prioritering in les:								
Domein in algemeen	0.011	0.009	0.025	0.009 *	0.005	0.011		
Solo spreken							0.006	0.009
Gespreksvaardigheid							0.014	0.011
Prioritering bij toetsing:								
Domein in algemeen	0.036	0.014 *	-0.001	0.012	0.023	0.011 *		
Solo spreken							0.011	0.015
Gespreksvaardigheid							0.025	0.017

Lesactiviteiten:									
Impliciet	0.002	0.021	0.027	0.016					
Gericht	-0.010	0.013	0.006	0.013					
Receptieve WS					0.006	0.018			
Productieve WS					0.073	0.022	*		
Spreeken met medeleerling/lkr								0.047	0.022 *
<i>Random Part:</i>									
Klas/schoolniveau	0.006	0.008	0.006	0.008	0.007	0.008	0.004	0.008	
Leerlingniveau	0.176	0.035 *	0.156	0.035 *	0.169	0.035 *	0.252	0.036 *	
ICC	0.034		0.037		0.042		0.016		
Aantal scholen	113		113		113		113		
Aantal leerlingen	2059		2059		2059		2059		

Noot: # De didactische methode *Take it easy* is de referentie. 'Onb' = onbekend, 'lkr' = leerkracht, WS = woordenschat.

Negen van de in 24 kenmerken van het onderwijsleerproces hebben een significante onafhankelijke associatie met één of meer domeinen Engels (Tabel 8.19); het gaat om twee school- en zeven klaskenmerken. De ICC daalt van bijna 10% (Model 2) naar ongeveer 4% in dit model. De twee significante schoolkenmerken zijn een coördinator of werkgroep Engels hebben bij luistervaardigheid (ES = 0,24) en faciliteiten voor Engels op school hebben bij leesvaardigheid en woordenschat (ES = -0,18 voor beide domeinen). De negatieve coëfficiënt van faciliteiten voor Engels impliceert dat op een school die leerkrachten faciliteert bij het geven van Engels, de vaardigheidsscores voor lezen en woordenschat in het algemeen iets lager zijn. De significante klaskenmerken uit Model 4 zijn het oefenen met productieve woordenschat bij woordenschat (ES = 0,34), de didactische methode bij luisteren (ES = -0,24 voor 'Groove me'), het werken in tweetallen of groepjes bij lees- en gespreksvaardigheid (respectievelijk ES = 0,15 en ES = 0,23), het spreken met medeleerlingen of de leerkracht bij gespreksvaardigheid (ES = 0,18) en het prioriteren van luisteren in de les bij luisteren (ES = 0,25) en woordenschat bij toetsing bij woordenschat (ES = 0,20). Het negatieve verband voor 'Groove me' duidt erop dat de luistervaardigheid van leerlingen die 'Groove me' als methode gebruiken iets lager is dan voor 'Take it easy', de referentiecategorie en meest gebruikte didactische methode in de peiling. Alle onderzochte onderwijsleerproceseffecten zijn klein of zeer klein. Desondanks zijn het hebben van een coördinator of werkgroep Engels, de didactische methode, het oefenen met de productieve woordenschat en het prioriteren van luisteren in de les te beschouwen als de meest relevante kenmerken van het onderwijsleerproces (ES > 0,20).

#### 8.8.5 Model 5, leerkrachtkenmerken

De gepoolde resultaten van Model 5 met betrekking tot de samenhang tussen de domeinspecifieke kenmerken van leerkrachten en de vaardigheidsscores Engels worden getoond in Tabel 8.20.

Tabel 8.20: Meerniveau modellen voor de domeinspecifieke leerkrachtkenmerken (Model 5)

	Leesvaardigheid		Luistervaardigheid		Woordenschat		Gespreksvaardigheid	
	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>
<i>Fixed Part</i>								
Intercept	0.207	0.101 *	0.124	0.094	0.183	0.109	-0.113	0.117
EIBO - S3&4(>10%)	0.002	0.045	-0.024	0.041	0.005	0.048	0.019	0.053
VVTO - S1(0-10%)	0.071	0.045	0.069	0.043	0.110	0.048 *	0.102	0.052 *
VVTO - S3(>10)	0.013	0.058	-0.028	0.052	0.073	0.061	0.029	0.064
Werkervaring in PO	-0.001	0.016	-0.016	0.016	0.003	0.017	0.005	0.019
Ervaring Engels lesgeven	0.028	0.019	0.045	0.017 *	0.013	0.019	0.015	0.021
Opleiding:								
verkort pabo	0.007	0.057	-0.040	0.052	0.008	0.062	0.011	0.071
acad pabo	0.001	0.059	-0.029	0.058	-0.008	0.064	-0.012	0.075
opleiding onb	-0.033	0.071	-0.027	0.067	-0.081	0.077	-0.074	0.087
Moedertaal:								
Engels	0.039	0.178	0.109	0.162	0.116	0.194	0.120	0.206
Anders	0.092	0.096	0.025	0.088	0.033	0.105	0.009	0.112
Moedertaal onb	0.082	0.071	0.044	0.067	0.077	0.077	0.068	0.086
Buitenschools gebruik Engels	0.021	0.033	0.068	0.033 *	0.006	0.036	0.062	0.043
Ingeschatte vaardigheid lezen / luisteren#	0.017	0.027	-0.013	0.028	0.033	0.029	-0.035	0.036
Ingeschatte vaardigheid spreken			-0.022	0.017			0.016	0.020
<i>Random Part:</i>								
Klas/schoolniveau	0.014	0.008	0.011	0.008	0.018	0.008	0.014	0.008
Leerlingniveau	0.176	0.035	0.156	0.035	0.168	0.035	0.253	0.036
ICC	0.073		0.065		0.096		0.054	
Aantal scholen	113		113		113		113	
Aantal leerlingen	2059		2059		2059		2059	

*Noot.* # Bij woordenschat is de ingeschatte vaardigheid voor lezen als predictor in het model opgenomen, bij gespreksvaardigheid de ingeschatte vaardigheid voor luisteren, 'Onb' = onbekend.

Uit Tabel 8.20 volgt dat, met correctie voor startmoment\*stratum, enkele leerkrachtkenmerken een deel van de verschillen tussen klassen/scholen kunnen verklaren. De ICC daalt van bijna 10% in Model 2 naar ongeveer 6% in Model 5. Er zijn twee kenmerken die een statistisch significante samenhang hebben met (alleen) luistervaardigheid: ervaring in het geven van het vak Engels en buitenschools gebruik van Engels door de leerkracht. De samenhangen zijn positief. Naarmate de leerkracht meer ervaring heeft in het lesgeven van het vak Engels en vaker buitenschools Engels gebruikt, zijn de luistervaardigheidsscores van de leerlingen enigszins beter. De bijbehorende effectgroottes duiden op een klein effect (respectievelijk,  $ES = 0,29$  en  $ES = 0,22$ ). De overige potentieel relevante leerkrachtkenmerken hebben geen significante samenhang met de vaardigheidsscores Engels.

### 8.8.6 Model 6, schoolkenmerken

De gepoolde resultaten van de meerniveau analyses voor de samenhang tussen algemene schoolkenmerken en de vaardigheidsscores voor de vier domeinen van Engels zijn gegeven in Tabel 8.21 (Model 6). Startmoment\*stratum is weer meegenomen als covariaat.

Tabel 8.21: Meerniveau modellen voor de schoolkenmerken (Model 6)

	Leesvaardigheid		Luistervaardigheid		Woordenschat		Gespreks- vaardigheid		
	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	
<i>Fixed Part</i>									
Intercept	0.163	0.256	-0.030	0.256	0.086	0.259	-0.193	0.355	
EIBO - S3&4(>10%)	0.053	0.051	0.005	0.050	0.052	0.052	0.072	0.064	
VVTO - S1(0-10%)	0.053	0.039	0.026	0.040	0.073	0.041	0.073	0.046	
VVTO - S3(>10)	0.065	0.065	0.020	0.063	0.105	0.066	0.061	0.076	
Thuis taal:									
% leerlingen met Nederlands-dialect	0.000	0.001	0.001	0.001	0.000	0.001	0.000	0.002	
% leerlingen niet Nederlands	-0.001	0.002	-0.001	0.002	0.000	0.002	0.000	0.003	
Stedelijkheid locatie #:									
Weinig	0.006	0.050	0.047	0.051	0.011	0.051	0.003	0.061	
Matig	0.041	0.048	0.029	0.047	0.023	0.048	0.039	0.064	
Sterk	0.104	0.047	*	0.048	0.048	0.087	0.048	0.112	0.061
Zeer sterk	0.188	0.053	*	0.112	0.053	*	0.200	0.054	*
Denominatie #:									
Openbaar	0.026	0.055	-0.037	0.055	-0.027	0.056	-0.007	0.062	
Protestants christelijk	-0.035	0.055	-0.137	0.055	*	-0.059	0.057	-0.041	0.065
Rooms katholiek	-0.025	0.057	-0.063	0.056	-0.060	0.059	-0.064	0.064	
Overig	-0.074	0.112	-0.093	0.113	-0.142	0.122	-0.123	0.132	
% leerlingen met formatiegewicht	-0.002	0.002	0.000	0.001	-0.002	0.002	-0.002	0.002	
Schoolgemiddelde advies VO	0.124	0.069	0.194	0.068	*	0.134	0.071	0.106	0.098
<i>Random Part:</i>									
Klas/schoolniveau	0.008	0.008	0.008	0.008	0.009	0.008	0.007	0.008	
Leerlingniveau	0.177	0.035	*	0.156	0.035	*	0.169	0.035	*
ICC	0.042		0.051		0.050		0.028		
Aantal scholen	113		113		113		113		
Aantal leerlingen	2059		2059		2059		2059		

*Noot.* # Bij de predictor stedelijkheid van de schoollocatie is 'niet-stedelijk' de referentiecategorie, bij denominatie is dat 'algemeen bijzonder'.

Tabel 8.21 toont dat, met correctie voor startmoment\*stratum, drie van de vijf schoolkenmerken significant bijdragen aan de verklaring van de variantie in de vaardigheidsscores Engels tussen klassen/scholen. De ICC is nu ongeveer 4%, tegen 10% in Model 2. Stedelijkheid hangt samenhang

met alle vier Engelse domeinen. Scholen in een zeer sterk stedelijke omgeving zijn het meest onderscheidend. De lees-, luister, gespreksvaardigheid en woordenschat van leerlingen op dergelijke scholen is beter dan op scholen in een niet-stedelijke omgeving, de referentiegroep (ES = 0,28 tot ES = 0,49). Verder is gevonden dat leerlingen op Protestants Christelijke basisscholen in het algemeen lagere prestaties voor luisteren hebben dan leerlingen op algemeen bijzondere basisscholen (de referentiecategorie) (ES = -0,35). Stedelijkheid en denominatie hebben een klein effect. Het schoolgemiddelde advies voor het voortgezet onderwijs heeft een klein effect op de luisterprestaties van de leerlingen (ES = 0,29). Naarmate dit hoger is, scoren de leerlingen in het algemeen hoger voor luistervaardigheid. De twee overgebleven schoolkenmerken – percentage leerlingen met een formatiegewicht en thuistaal van de leerlingen op schoolniveau – hebben geen onafhankelijke samenhang met één van de vaardigheidsscores Engels.

#### **8.8.7 Model 7, compact totaalmodel**

Als laatste stap in de meerniveau analyses zijn uit de Modellen 3 tot en met 6 de predictoren geselecteerd die met tenminste één domein van Engels een significante samenhang hadden (voor een uitgebreide beschrijving van de methode zie paragraaf 8.7). Het uitgebreide totaalmodel met 35 predictoren is vervolgens teruggebracht tot een compact totaalmodel met 12 significante predictoren. De gepoolde resultaten van het compacte totaalmodel zijn gepresenteerd in Tabel 8.22.

Tabel 8.22: Meerniveau modellen met Model 7: compact totaalmodel

	Leesvaardigheid		Luistervaardigheid		Woordenschat		Gespreks- vaardigheid	
<i>Fixed Part</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>	<i>b</i>	<i>SE(b)</i>
Intercept	-0.724	0.057 *	-0.601	0.066 *	-0.728	0.058 *	-1.045	0.074 *
EIBO - S3&4(>10)	0.004	0.031	-0.034	0.031	-0.005	0.032	0.027	0.041
VVTO - S1(0-10)	-0.005	0.033	-0.036	0.036	0.023	0.036	0.025	0.044
VVTO - S3(>10)	-0.021	0.038	-0.094	0.039 *	0.023	0.040	-0.023	0.049
<b>Model 4: Onderwijsleerproceskenmerken</b>								
Didactische methode:								
Just do it	0.152	0.053 *	0.155	0.053 *	0.148	0.054 *	0.173	0.064 *
Step to	-0.014	0.043	-0.011	0.046	-0.012	0.045	-0.004	0.059
Groove me	-0.009	0.031	-0.048	0.032	-0.058	0.034	-0.040	0.039
Overig	-0.023	0.034	-0.007	0.035	-0.018	0.035	-0.010	0.043
Methode onb	0.045	0.038	0.007	0.038	-0.005	0.037	-0.018	0.044
Faciliteiten Engels	-0.062	0.027 *	-0.076	0.029 *	-0.061	0.027 *	-0.047	0.032
Faciliteiten onb	0.042	0.098	0.044	0.102	0.027	0.100	-0.093	0.129
Werken in groepen	0.025	0.012 *	0.027	0.013 *	0.009	0.013	0.024	0.016
Coördinator Engels	0.069	0.028 *	0.130	0.028 *	0.088	0.028 *	0.096	0.038 *
Coördinator onb	-0.068	0.081	-0.064	0.086	-0.043	0.087	0.046	0.112
<b>Model 5: domeinspecifieke leerkrachtkenmerken</b>								
Ervaring Engels lesgeven	0.032	0.009 *	0.032	0.010 *	0.022	0.009 *	0.024	0.010 *
<b>Model 3b: domeinspecifieke leerlingkenmerken</b>								
Buitenschools gebruik								
Engels leerling	0.003	0.001 *	0.005	0.001 *	0.006	0.001 *	0.006	0.002 *
Linguistic Self								
Confidence	0.038	0.005 *	0.033	0.005 *	0.043	0.005 *	0.032	0.007 *
Attitude	0.032	0.006 *	0.006	0.006	0.027	0.006 *	0.023	0.008 *
Ingeschatte vaardigheid lezen / luisteren / spreken								
	0.066	0.011 *	0.056	0.009 *	0.054	0.010 *	0.081	0.015 *
<b>Model 3a: algemene leerlingkenmerken</b>								
Leerlinggewicht:								
0.3	-0.015	0.041	-0.074	0.043	-0.003	0.040	-0.001	0.057
1.2	-0.051	0.039	-0.003	0.043	-0.086	0.037 *	-0.049	0.055
onb	-0.032	0.049	-0.030	0.052	0.056	0.056	0.001	0.067
Advies VO:								
vmbo-gl/tl	0.142	0.023 *	0.125	0.024 *	0.140	0.025 *	0.163	0.034 *
havo-vwo	0.329	0.022 *	0.234	0.023 *	0.299	0.021 *	0.304	0.031 *
<b>Model 6: algemene schoolkenmerken</b>								
Stedelijkheid:								
Weinig	-0.003	0.036	0.083	0.040 *	0.013	0.039	-0.004	0.046
Matig	0.005	0.033	0.014	0.036	0.005	0.035	0.008	0.047



Sterk	0.058	0.033		0.028	0.035	0.049	0.036	0.049	0.045
Zeer sterk	0.094	0.037 *		0.064	0.043	0.123	0.039 *	0.094	0.054
<i>Random Part:</i>									
Klas/schoolniveau	0.003	0.008		0.003	0.008	0.004	0.008	0.002	0.008
Leerlingniveau	0.106	0.035 *		0.114	0.035 *	0.099	0.035 *	0.181	0.036 *
ICC	0.028			0.027		0.039		0.012	
Aantal scholen	113			113		113		113	
Aantal leerlingen	2059			2059		2059		2059	

*Noot:* # De didactische methode *Take it easy* is de referentiecategorie, 'Onb' = onbekend.

Tabel 8.22 laat zien dat van de omvangrijke set van **onderwijsleerprocespredictoren** in Model 4 er vier significante resterende, namelijk didactische methode, coördinator Engels, faciliteiten Engels en werken in groepen. De didactische methode *Just do it* is enigszins beter en *Groove me* is niet langer slechter (Model 4) dan *Take it easy* (referentie) voor alle Engelse vaardigheidsscores (ES = 0,41 tot ES = 0,47 voor *Just do it*). De overige methodes onderscheiden zich niet van *Take it easy*. Verder is gevonden dat op scholen met een coördinator of werkgroep Engels alle vaardigheidsscores beter zijn dan op scholen die dit niet hebben (ES = 0,21 tot ES = 0,39). Het gaat om kleine effecten bij deze twee predictoren. De twee overige significante predictoren van het onderwijsleerproces – het bieden van faciliteiten voor Engels aan leerkrachten en het werken in groepjes door leerlingen – zijn wel statistisch significant, maar de effecten zijn als zeer klein te beschouwen (ES = -0,19 tot ES = -0,22 voor faciliteiten voor Engels en ES = 0,14 en ES = 0,15 voor werken in groepen bij respectievelijk lezen en luisteren). De regressiecoëfficiënt van het hebben van faciliteiten voor Engels is opnieuw negatief (zie Model 4). Dit impliceert dat op een school die leerkrachten faciliteert bij het geven van Engels in het algemeen iets lagere vaardigheidsscores voor lezen, luisteren en woordenschat zijn gevonden. Het werken in tweetallen of groepjes door leerlingen hangt daarentegen samen met iets hogere scores voor lezen en luisteren. Ook dit komt overeen met Model 4.

Van de set van zes domeinspecifieke **leerkrachtkenmerken** in Model 5 is alleen ervaring in het Engels lesgeven overgebleven (ES = 0,14 tot ES = 0,25) in het compacte totaalmodel (Tabel 8.22). Voor alle Engelse vaardigheidsscores geldt dat naarmate de leerkracht meer ervaring heeft in het Engels lesgeven, de scores van de leerlingen in het algemeen iets beter zijn.

Vier van de vijf **domeinspecifieke leerlingenkenmerken** hebben ook in het compacte totaalmodel een significante samenhang met de Engelse vaardigheidsscores (Tabel 8.22). De kenmerken die bevorderend zijn voor alle Engelse vaardigheidsscores zijn ingeschatte vaardigheid voor een bepaald domein door de leerling zelf (ES = 0,36 tot ES = 0,46), linguistic self confidence (ES = 0,38 tot ES = 0,72) en buitenschools gebruik van Engels door de leerling (ES = 0,14 tot ES = 0,29). Attitude (ES = 0,21 tot ES = 0,39) is positief voor Engels lezen, woordenschat en gespreksvaardigheid. Deze samenhangen zijn wel statistisch significant, maar de effectgroottes duiden op kleine effecten. Linguistic self confidence vormt hierop de uitzondering. Bij lezen, luisteren en woordenschat gaat het om een middelgroot effect; naarmate leerlingen meer zelfvertrouwen hebben in hun Engelse taalvaardigheid zijn deze vaardigheidsscores in het algemeen hoger.

Het **algemene leerlingenkenmerk** dat er het meeste toe doet in het compacte totaalmodel is het advies voor het voortgezet onderwijs (VO) (ES = 0,38 tot ES = ,43) (Tabel 8.22). Naarmate het schooladvies van leerlingen hoger is, zijn alle Engelse vaardigheidsscores hoger. In tegenstelling tot Model 3a hebben leeftijd en sekse in dit model geen significante onafhankelijke associaties meer met

de vaardigheidsscores, maar formatiegewicht van de leerling wel ( $ES = -0,27$ ). Gevonden is dat leerlingen met een 1.2 gewicht iets lager presteren bij (alleen) woordenschat dan leerlingen zonder formatiegewicht.

Van de drie significante **algemene schoolkenmerken** uit Model 6 komt alleen de stedelijkheid van de schoollocatie terug als significante predictor in het compacte totale model (Tabel 8.22). De samenhang is minder sterk geworden. Opnieuw is gevonden dat op scholen in een zeer sterk stedelijke omgeving de leesvaardigheid en woordenschat van de leerlingen iets beter is dan op scholen in een niet-stedelijke omgeving ( $ES = 0,29$  en  $ES = 0,39$ , respectievelijk). Verder is gevonden dat op scholen met een weinig stedelijke locatie de luistervaardigheid in het algemeen iets beter dan op scholen met een niet-stedelijke locatie ( $ES = 0,25$ ).

In het **compacte totaalmodel** is ook startmoment\*stratum opgenomen als covariaat. Uit de vergelijking met Model 2 volgt dat de oorspronkelijke verschillen tussen EIBO en VVTO scholen - met correctie voor het aandeel gewichtenleerlingen – niet meer significant zijn. Uit Model 2 kwam naar voren dat VVTO scholen met weinig gewichtenleerlingen iets beter presteren bij lezen, woordenschat en gespreksvaardigheid dan EIBO scholen met weinig gewichtenleerlingen ( $ES = 0,23$  tot  $ES = 0,32$ ). Dit verschil is 'weg'-verklaard door de leerling- leerkracht- en schoolkenmerken van Model 7. Nadere bestudering van de Modellen 3 tot en met 6 leert dat het vooral om de onderwijsleerproceskenmerken gaat. Deze set van predictoren brengen de grootste reductie in de schattingen van de regressiecoëfficiënten voor VVTO scholen met weinig gewichtenleerlingen teweeg. Model 7 maakt duidelijk om welke kenmerken van het onderwijsleerproces het dan gaat. Gesorteerd op effectgrootte gaat het in de eerste plaats om didactische methode die de scholen gebruiken, gevolgd door het hebben van een coördinator of werkgroep Engels, dan het bieden van faciliteiten voor Engels aan de leerkrachten en tot slot het laten werken van leerlingen in tweetallen of groepjes.



## Hoofdstuk 9. Trendvergelijking leerlingprestaties Engels

Om een trendvergelijking te kunnen maken tussen de leerlingprestaties in 2012 en 2018 worden in dit hoofdstuk de leerlingresultaten van de vorige peiling vergeleken met de resultaten van de huidige peiling. Centraal staat daarbij de beantwoording van de volgende onderzoeksvraag: *Welke verschillen zijn er tussen de leerprestaties met het peilingsonderzoek Engels van 2012?* (onderzoeksvraag 9).

In paragraaf 9.1 worden de resultaten van de vergelijking tussen de uitkomsten voor de toetsen lezen, luisteren en woordenschat beschreven. In paragraaf 9.2 staat de vergelijking tussen prestaties voor spreekvaardigheid (monologisch spreken) in 2012 en 2018 centraal. Een dergelijke vergelijking is voor de gesprekstaken niet mogelijk, omdat deze taken in 2012 niet zijn afgenomen.

### 9.1 Trendvergelijking lezen, luisteren en woordenschat

Om de resultaten van de eerdere peiling te vergelijken met de huidige peiling moeten de resultaten bij elkaar op een schaal gebracht worden. Dat wordt gedaan met behulp van ankeropgaven. Dit zijn opgaven die zowel in de peiling van 2012 gebruikt zijn als in de peiling van 2018. Voor de verschillende vaardigheden zijn de volgende aantallen ankeropgaven meegenomen: 17 voor Lezen, 21 voor Luisteren en 33 voor Woordenschat. Door middel van OPLM worden de schalen via deze opgaven aan elkaar gerelateerd.

Bij de OPLM analyses zijn er twee mogelijkheden om de schalen aan elkaar te koppelen. Dat is ten eerste door middel van een vrije kalibratie. Hierbij worden alle parameters van de opgaven opnieuw geschat met de oude data, gekoppeld aan de nieuwe data. Een tweede optie is om met gefixeerde parameters van de gehele schaal te werken. Het voordeel van de eerste methode is dat er meer gegevens zijn om de parameters goed te schatten en de analyses niet worden gedomineerd door één van de twee jaren. Het voordeel van de tweede methode is dat alle schattingen, die eerder gedaan zijn en waarmee bijvoorbeeld de resultaten uit hoofdstuk 8 zijn verkregen, bruikbaar blijven: die veranderen namelijk niet. Daarnaast zou de 'nieuwe oude data' eigenlijk alleen serieuze impact op de resultaten hebben als er sprake is van Differential Item Functioning (DIF). Wanneer dat het geval is, is er al geen goede jaarvergelijking te maken. Beide werkwijzen zijn uitgevoerd.

Tabel 9.1a en b: P-waarden in 2018 en 2012 van de ankeropgaven Lezen (LE) en Luisteren (LU)

Label	aantal leerlingen		p-waarden		Verschil
	2018	2012	2018	2012	
LE1-20	716	529	0.59	0.68	-0.10
LE1-12	716	786	0.66	0.73	-0.07
LE1-09	604	520	0.54	0.61	-0.06
LE1-03	475	512	0.52	0.58	-0.06
LE1-10	716	786	0.63	0.67	-0.04
LE1-11	716	786	0.64	0.68	-0.04
LE1-02	621	512	0.70	0.74	-0.03
LE1-07	716	520	0.73	0.76	-0.03
LE1-16	475	529	0.74	0.76	-0.03
LE1-17	674	529	0.77	0.80	-0.02
LE1-01	708	512	0.67	0.68	-0.01
LE1-04	674	512	0.57	0.58	-0.01
LE1-08	475	520	0.77	0.76	0.01
LE1-14	708	535	0.52	0.51	0.01
LE1-15	475	529	0.56	0.53	0.04
LE1-05	708	512	0.84	0.80	0.04
LE1-06	604	520	0.78	0.73	0.04

Label	aantal leerlingen		p-waarden		verschil
	2018	2012	2018	2012	
LU1-19	713	659	0.72	0.85	-0.14
LU1-01	713	663	0.57	0.69	-0.12
LU1-03	713	663	0.71	0.81	-0.10
LU1-05	693	663	0.76	0.86	-0.10
LU1-21	453	659	0.66	0.75	-0.09
LU1-15	693	656	0.80	0.88	-0.08
LU1-14	453	660	0.78	0.85	-0.08
LU1-08	599	660	0.70	0.76	-0.06
LU1-06	713	663	0.63	0.68	-0.06
LU1-07	599	660	0.66	0.71	-0.05
LU1-16	713	656	0.86	0.91	-0.05
LU1-12	599	660	0.55	0.59	-0.04
LU1-13	693	660	0.80	0.84	-0.04
LU1-09	599	660	0.62	0.65	-0.02
LU1-17	453	659	0.91	0.92	-0.01
LU1-11	599	660	0.42	0.42	0.00
LU1-18	693	659	0.64	0.64	0.00
LU1-20	453	659	0.92	0.91	0.01
LU1-04	693	663	0.80	0.79	0.01
LU1-22	713	659	0.86	0.85	0.02
LU1-02	453	663	0.81	0.74	0.07

Bij de kalibraties waarbij de oude data gekoppeld is aan de nieuwe data is dus enkel gekeken naar de ankeropgaven. Als we in de Tabellen 9.1a en 9.1b alleen al naar p-waarden kijken valt een aantal zaken op. Bij Lezen en Luisteren is te zien dat het verschil tussen de p-waarden loopt van -0,10 tot 0,04 bij Lezen (Tabel 9.1a-boven; ordening naar verschil in p-waarde) en van -0,14 tot en met 0,07 bij Luisteren (Tabel 9.1b-onder). Voor beide vaardigheden zijn drie vergelijkbare conclusies te trekken: a) er is bij de meeste opgaven sprake van een daling van de p-waarde, b) het grootste verschil tussen

2012 en 2018 is een daling en c) er is gemiddeld sprake van een daling ten opzichte van 2012: -0,04 bij Lezen en -0,02 bij Luisteren. De p-waarden lagen in 2012 over het algemeen wat hoger dan in 2018.

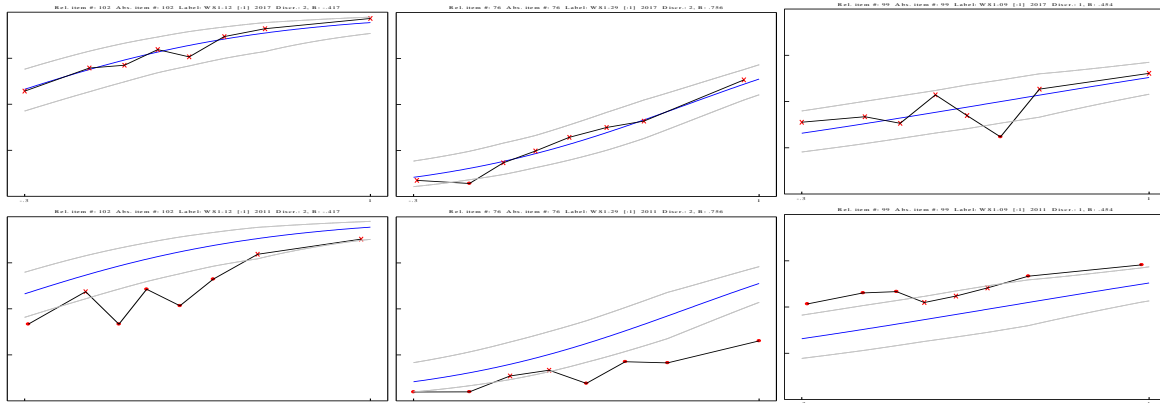
Bij Woordenschat zien we een ander beeld. Met één opgave meer is er a) vaker een stijging dan een daling (17 om 16), b) het grootste verschil is een stijging en c) gemiddeld over alle opgaven is er juist sprake van een stijging (+0,01). Het overzicht hiervan staat in Tabel 9.1c.

Tabel 9.1c: P-waarden in 2018 en 2012 van de ankeropgaven Woordenschat

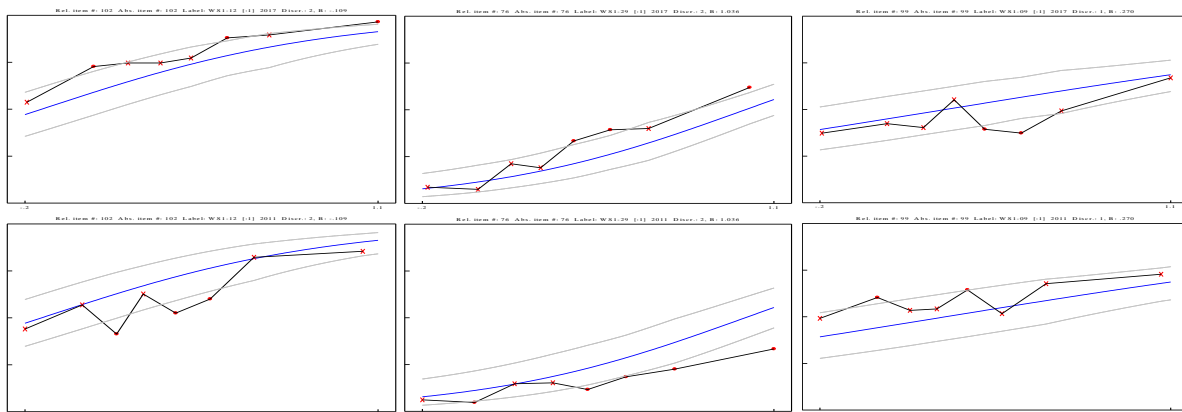
Item	P/R	Label	Aantal leerlingen		P-waarden		verschil
			2018	2012	2018	2012	
WS1-09	R	poverty	596	521	0.48	0.61	-0.125
WS1-31	P	seventy	588	517	0.29	0.41	-0.117
WS1-03	R	autumn	715	521	0.60	0.69	-0.089
WS1-25	R	perhaps	596	521	0.39	0.47	-0.078
WS1-33	R	hide	596	495	0.64	0.69	-0.053
WS1-11	R	fortune	596	521	0.78	0.82	-0.041
WS1-21	P	morning	659	521	0.56	0.60	-0.038
WS1-06	R	stomach	676	521	0.40	0.43	-0.033
WS1-24	R	cloudy	588	517	0.72	0.75	-0.029
WS1-04	R	tall	659	521	0.54	0.57	-0.029
WS1-15	P	knife	715	521	0.39	0.42	-0.028
WS1-20	P	tree	659	495	0.51	0.54	-0.027
WS1-22	P	spring	588	521	0.69	0.72	-0.024
WS1-26	R	sure	452	521	0.49	0.51	-0.022
WS1-14	P	eye	383	521	0.69	0.70	-0.010
WS1-02	R	behind	659	521	0.64	0.65	-0.008
WS1-16	P	chair	588	495	0.46	0.46	0.002
WS1-27	R	hard	676	521	0.75	0.73	0.022
WS1-01	R	shout	588	521	0.56	0.54	0.024
WS1-32	R	farm	715	495	0.71	0.68	0.027
WS1-23	R	early	676	517	0.79	0.75	0.037
WS1-30	P	today	596	517	0.64	0.61	0.038
WS1-17	P	pencil	676	495	0.46	0.42	0.048
WS1-08	R	church	676	521	0.56	0.50	0.057
WS1-19	P	ear	588	495	0.75	0.69	0.063
WS1-05	R	thumb	676	521	0.62	0.55	0.065
WS1-18	P	duck	452	495	0.82	0.75	0.070
WS1-07	R	theatre	676	521	0.82	0.75	0.072
WS1-10	R	peak	596	521	0.61	0.52	0.087
WS1-13	R	bottom	596	521	0.73	0.61	0.118
WS1-28	R	fortunate	715	521	0.57	0.45	0.119
WS1-29	P	Tuesday	715	517	0.32	0.17	0.149
WS1-12	R	rush	596	521	0.80	0.63	0.171

Noot: P/R staat voor 'productief'/'receptief'

De spreiding in de verschillen bij woordenschat is groter dan bij lezen en luisteren. Dit zou op DIF kunnen duiden. De drie opgaven met de grootste verschillen (WS1-12, WS1-29 en WS1-09) staan gegeven in de figuren 9.1a (gebaseerd op de gefixeerde schaal) en 9.1b (gebaseerd op de vrij geschatte schaal).



Figuur 9.1a: Item WS1-12 (links), WS1-29 (midden) en WS1-09 (rechts): boven 2018, onder 2012 - weergegeven op de schaal met gefixeerde parameters.



Figuur 9.1b: Item WS1-12 (links), WS1-29 (midden) en WS1-09 (rechts): boven 2018, onder 2012 - weergegeven op de schaal met vrije parameters.

De weergegeven opgaven behoorden wat betreft de passing van het model ook tot de minst goed passende opgaven. De opgaven zijn echter niet verwijderd, aangezien dit waarschijnlijk werkelijke verschillen zijn bij deze woorden en deze verschillen dus relevant zijn. Dat verschilt met de kunstmatige verschillen zoals gevonden door een andere scoringsregel van de opgaven (zoals bijvoorbeeld met of zonder spelling). De DIF bij woordenschat was echter dusdanig gering dat de parameters opnieuw konden worden geschat.

Bij lezen en luisteren zijn geen vergelijkbare opgaven met DIF gevonden, zoals de p-waarden al lijken te suggereren. Als we kijken naar de correlaties tussen de p-waarden in de verschillende jaren dan zien we dat deze bij lezen en luisteren hoog zijn (0,92 bij lezen en 0,91 bij Luisteren). De schalen voor lezen en luisteren leveren daarom verder geen vragen meer op. Bij woordenschat is die correlatie iets lager (0,88; met een correlatie van 0,92 bij de productieve taken en 0,81 bij de receptieve taken).

Om nu tot een optimale schaal te komen zijn in OPLM bij de jaarvergelijking de parameters van opgaven opnieuw geschat. De informatie van de peiling in 2012 heeft hierdoor ook invloed op de schattingen van de a- en b-parameters. Het heeft tot gevolg dat de waarden van parameters op de theta-schaal veranderen. Om de latente gemiddelden en standaardafwijkingen te schatten worden deze weer opnieuw gefixeerd voor beide jaren. Doordat deze op een nieuwe schaal gezet zijn, zijn schattingen van het gemiddelde en de standaardafwijking in de populatie niet meer te vergelijken met de eerder gegeven WML-waarden in hoofdstuk 8. Aangezien het hier met MML-geschatte latente verdelingen betreft zijn de standaardafwijkingen ook kleiner dan wanneer er met de WML-waarden gerekend wordt: bij de MML-schattingen is er voor meetfout gecorrigeerd.

De geschatte populatieverdelingen voor 2012 en 2018 staan in Tabel 9.2, evenals de effectschattingen. Bij controles is gevonden dat resultaten met gefixeerde parameters in vergelijking met de vrij geschatte parameters alleen een lineaire verschuiving laten zien bij de verdelingen. De effectgroottes zijn zeer vergelijkbaar over de twee methoden.

Tabel 9.2: MML-schattingen van de latente verdelingen van 2012 en 2018 op de nieuwe parameterschaal waarbij data 2012 meegenomen is.

		<b>Gemiddelde</b>	<b>Standaardafwijking</b>	<b>Effectgrootte</b>	<b>p</b>
Lezen	2018	0.43	0.37	-0.09	0.002
	2012	0.46	0.35		
Luisteren	2018	0.31	0.30	-0.28	0.00
	2012	0.40	0.30		
Woordenschat	2018	0.44	0.39	0.05	0.057
	2012	0.42	0.38		

De resultaten laten zien dat er bij lezen en luisteren een daling gevonden wordt en bij Woordenschat een stijging. De p-waarden van de ankeropgaven preludeerden hier al op. De verschillen tussen de twee jaren bij lezen en woordenschat zijn beperkt, maar de daling bij luisteren is fors te noemen. Als de vuistregels van Cohen gehanteerd worden is alleen het effect bij luisteren betekenisvol (klein effect). Wat de verklaring is voor de daling bij luisteren is niet bekend. Op basis van de nu beschikbare gegevens lijkt het niet te gaan om een kunstmatig verschil dat door methodologische keuzes naar voren komt. De correlaties tussen opgaven laten zien dat de ankers goed gekoppeld lijken en er is ook geen sprake van DIF. Als er iets aan de hand is bij deze opgaven geldt dat voor alle opgaven.

Bij de hierboven gerapporteerde resultaten is nog wel een aantal kanttekeningen te maken. De eerste kanttekening is dat dit ongewogen gegevens zijn. Bij de gegevens van 2018 weten we dat de gemiddelden van de gewogen gegevens lager uitkomen dan in het ongewogen geval. Dat zou hier dus betekenen dat de werkelijke gemiddelden van 2018 nog iets lager liggen. Ook bij 2012 zouden echter de gemiddelden lager liggen, aangezien er ook in 2012 relatief veel VVTO-leerlingen in de steekproef zaten.

Om nu de resultaten van de jaarvergelijking beter te kunnen vergelijken met die resultaten die alleen voor 2018 zijn gerapporteerd zijn ook voor 2012 WML-schattingen van de vaardigheden gemaakt voor ieder van de leerlingen. Deze 2012 WML-schattingen zijn gebaseerd op de schaalwaarden waarbij de parameters van 2018 gefixeerd zijn, om zo de waarden van 2018 in de eerdere tabellen te kunnen vergelijken met die van 2012. Ook wordt er bij deze werkwijze gewogen. Bij 2018 is dit op de combinatie van formatiegewicht-stratum en het onderscheid tussen EIBO en



VVTO. Bij de data van 2012 was alleen een weging naar het onderscheid tussen EIBO en VVTO noodzakelijk.

De weging in 2012 heeft invloed op de resultaten voor 2012. In 2012 waren er 7 VVTO-scholen, goed voor 13,4% van de leerlingen. Deze scholen waren sinds 2004 of eerder VVTO-school. Als we de gegevens gebruiken zoals opgenomen in het beschikbare bestand (PEN.VVTOScholen2011.sav), dan is te zien dat in 2012 0,5% van de scholen een VVTO school was (sinds 2004 of eerder). Alle 7 scholen waren ook scholen die in stratum 1 zaten. De verdeling van de strata binnen EIBO was wel vergelijkbaar met de populatieverdeling<sup>91</sup>. Met in de populatie een half procent VVTO-scholen was de impact op de stratumverdeling na weging van de VVTO-scholen nihil. Als we nu met deze informatie de scores van het peilingsonderzoek uit 2012 gaan wegen, kunnen de twee gewogen gemiddelde van 2012 en 2018 met elkaar vergelijken.

Tabel 9.3: WML-schattingen voor 2012 en 2018, x-serie

		Gemiddelde	Standaardafwijking	Effectgrootte	p
Lezen	2018	0.39	0.44	-0.08	0.022
	2012	0.42	0.44		
Luisteren	2018	0.30	0.42	-0.13	0.000
	2012	0.35	0.39		
Woordenschat	2018	0.33	0.42	0.03	0.406
	2012	0.32	0.43		

De schaalwaarden van tabel 9.2 en 9.3 betreffen twee verschillende typen schalen. De effectgrootten zijn echter wel vergelijkbaar, aangezien die onafhankelijk van de schaal gemaakt zijn. Daar zien we dat de resultaten voor lezen en woordenschat in Tabel 9.3 niet veel verschillen van die van Tabel 9.2, maar dat het effect bij luisteren wel duidelijk verschilt. Dit is deels te verklaren uit het feit dat bij de WML de standaardafwijking wat groter is. Als die factor meegenomen wordt, dan zou de daling ES = -0,17 worden.

## 9.2 Trendvergelijking spreekvaardigheid

### 9.2.1 Spreektaak 'de Brug'

De mondelinge taalvaardigheden zijn lastiger over de jaren heen te vergelijken. Het type gesprekstaken dat in 2018 gebruikt wordt om mondelinge vaardigheid te evalueren werd niet toegepast in 2012. In 2012 zijn wel zes spreektaken gebruikt (monologisch spreken) op basis waarvan een jaarvergelijking mogelijk zou zijn. Deze taken waren voor de huidige afname echter niet zo geschikt meer. Om toch enige indruk te krijgen van de ontwikkeling van de spreekvaardigheid is ervoor gekozen om één enkele taak van de zes ook in 2018 af te nemen. Dit was de taak 'De Brug'. Bij deze taak moeten de leerlingen een verhaal vertellen naar aanleiding van 6 plaatjes die hen getoond worden. Deze uitvoering wordt later door een beoordelaar beoordeeld.

<sup>91</sup> De definitie van stratum in 2011 was wel fundamenteel anders dan in 2018. Deze was gebaseerd op het schoolgewicht, zoals dat in 2011 gedefinieerd was. De strata zijn over de jaren niet vergelijkbaar. Wel was binnen een jaar te evalueren of de verdeling overeenkomstig de populatie was of niet.

Tijdens de beoordeling is het de eerste taak voor de beoordelaar om te vast te stellen of de taak te beoordelen is. Na de evaluatie van de beoordeelbaarheid van het geheel is er een beoordeling per plaatje. Het eerste punt waarop een plaatje wordt beoordeeld is of de leerling in het Engels inhoudelijk op het plaatje ingaat. Als dat niet het geval is volgt er verder geen beoordeling van de prestatie bij dat plaatje. Is de uiting wel inhoudelijk en in het Engels, dan wordt beoordeeld of bepaalde woorden of een verzameling van woorden correct gebruikt worden. Het aantal van dergelijke beoordelingen verschilt per plaatje van 4 tot en met 9 beoordelingen. Het totaal aantal van dergelijke beoordelingen is 37. Bij een vijftiental van deze opgaven is het ook mogelijk het antwoord deels correct te geven, al worden niet bij al deze opgaven de bijna-goede antwoorden geobserveerd. Tot slot wordt er gevraagd waarom er geen beoordeling is (response afwezig, onverstaanbaar of in het Nederlands). Deze vraag hoeft alleen ingevuld te worden als het plaatje niet beoordeeld is. Aan het eind van de te beoordelen taak worden nog twee algemene indrukken bevraagd. Dat betreft de globale beoordeling van de uitspraak en de algemene indruk van de invloed van de toetsleider.

Van deze taak zijn er 210 observaties uit 2012 en 297 observaties uit 2018. Van de observaties uit 2018 zijn er 14 dubbel beoordeeld wat een totaal van 311 beoordelingen in 2018 oplevert. Van de dubbel beoordeelde observaties is er willekeurig één beoordeling gekozen om verder te analyseren. De 14 dubbele beoordelingen zijn geëvalueerd om overeenstemming te evalueren.

De werken zijn in 2012 door 6 verschillende beoordelaars beoordeeld: een die 21 observaties beoordeeld had en vijf die 37 tot 40 observaties beoordeelden. In 2018 waren er drie beoordelaars: een met 15 beoordelingen, een met 107 en een met 189 beoordelingen. Meer informatie over de overeenstemming tussen de beoordelaars in 2018 wordt gegeven in paragraaf 5.2.2.

### 9.2.2 Resultaten

De resultaten worden eerst gegeven voor wat betreft de beoordeelbaarheid. Dan zien we dat er in 2012 3% (6 observaties) niet beoordeelbaar was en 1% (2 observaties) in 2018. Dit kan aan de opnamen liggen en hoeft zodoende niet iets te zeggen over de kwaliteit van de prestaties. Deze oordelen worden verder buiten beschouwing gelaten. Hierdoor blijven er voor 2012 nog 204 observaties over en voor 2018 nog 295.

Tabel 9.4: beoordeelbaarheid van de prestaties bij de afbeeldingen

JAAR	gemiddelde	
	2018	2012
Afb 1 beoordeelbaar	0.94	0.96
Afb 2 beoordeelbaar	0.94	0.89
Afb 3 beoordeelbaar	0.85	0.79
Afb 4 beoordeelbaar	0.89	0.86
Afb 5 beoordeelbaar	0.93	0.85
Afb 6 beoordeelbaar	0.87	0.84

Wat wel met de prestatie van de leerlingen te maken heeft is de beoordeelbaarheid per plaatje. Als de leerling een inhoudelijke respons in het Engels geeft dan levert dit een score 1 op, anders een score 0. De gemiddelde gevonden resultaten zijn gegeven in Tabel 9.4. Uit deze tabel is af te leiden dat bij 5 van de 6 plaatjes de prestaties in 2018 beter waren. Alleen bij het eerste plaatje is de

response in 2012 vaker adequaat. Als deze zes opgaven samen een schaal zouden vormen zou deze een betrouwbaarheid van 0,75 hebben (zowel Lambda-2 als alpha). Deze somscore betreft het aantal plaatjes dat in het Engels wordt besproken. Dat zijn er in totaal meer in 2018 dan in 2012 (zie Tabel 9.5a). Dit verschil is echter niet significant. De frequentie van de totaal aantallen is gegeven in Tabel 9.5b.

Tabel 9.5a: Verdelingseigenschappen totaal beoordeelbaarheid van de prestaties bij de afbeeldingen

JAAR	aantal	gemiddelde	standaardafwijking	SE(gemiddelde)
2018	295	5,42	1,09	0,06
2012	204	5,20	1,44	0,10

Tabel 9.5b: Frequentie totaal beoordeelbaarheid van de prestaties bij de afbeeldingen

Aantal afbeeldingen	Aantal observaties		Percentage		Cumulatief %		Cumulatief % (omgekeerd)	
	2018	2012	2018	2012	2018	2012	2018	2012
0	2	4	0.7	2.0	0.7	2.0	100.0	100.0
1	4	7	1.4	3.4	2.0	5.4	99.3	98.0
2	4	3	1.4	1.5	3.4	6.9	98.0	94.6
3	9	11	3.1	5.4	6.4	12.3	96.6	93.1
4	19	13	6.4	6.4	12.9	18.6	93.6	87.7
5	59	33	20.0	16.2	32.9	34.8	87.1	81.4
6	198	133	67.1	65.2	100.0	100.0	67.1	65.2

Een andere globale maat is de kwalificatie van de Engelse uitspraak door de leerlingen. Hier zijn vier categorieën te bepalen. Bij 3% van de leerlingen, zowel in 2012 als 2018, was het oordeel niet gegeven of niet interpreteerbaar. De verdeling van de overige leerlingen is te vinden in Tabel 9.6. Daar is te zien dat het percentage leerlingen dat “heel natuurlijk Engels” spreekt in 2012 groter was dan in 2018. Het betreft hier wel weinig leerlingen. Het percentage leerlingen dat “erg Nederlands” sprak was in 2012 lager.

Tabel 9.6: resultaten bij Uitspraak

Uitspraak	Aantal observaties		Percentage		Cumulatief %		Cumulatief % (omgekeerd)	
	2018	2012	2018	2012	2018	2012	2018	2012
Onverstaanbaar	9	8	3.1	4.1	3.1	4.1	100.0	100.0
Erg Nederlands	191	122	66.8	61.9	69.9	66.0	96.9	95.9
Best behoorlijk Engels	82	56	28.7	28.4	98.6	94.4	30.1	34.0
Heel natuurlijk Engels	4	11	1.4	5.6	100.0	100.0	1.4	5.6

Als de ordinale uitspraak-oordelen lineair gescoord zouden worden met “onverstaanbaar” score 0 en “heel natuurlijk Engels” score 3 dan zien we dat in 2012 hoger gescoord wordt dan in 2018 (1,36 om 1,28), maar dat dit verschil niet significant is. Ook als we de categorie “onverstaanbaar” niet meenemen, is het verschil tussen de twee jaren niet significant (score 1,41 om 1,32). Wat bij deze globale beoordeling wel lastig is, is dat het moeilijk te achterhalen is of in 2012 de verschillende uitspraken wel op dezelfde manier beoordeeld werden als in 2018. Richtlijnen voor wanneer welke

categorie in 2012 werd toegekend waren niet te achterhalen, maar zijn voor 2018 wel besproken. Ook het beoordelaarseffect kan hier een rol spelen. Als de 13 dubbel beoordeelde uitspraken in 2018 als richtlijn moeten nemen dan is dat niet hoopgevend<sup>92</sup>. Bij de random toewijzing van eerste en tweede beoordelaar is te zien dat de correlatie tussen die twee met een waarde 0,08 niet van 0 verschilt. Kijken we verder naar de twee beoordelaars dan is te zien dat ze het bij 5 leerlingen eens zijn, maar bij 8 leerlingen geeft één van de twee systematisch een oordeel in categorie een plek hoger (7 keer “Best behoorlijk Engels” tegenover “Heel erg Nederlands”; 1 keer “Heel natuurlijk Engels” tegenover “Best behoorlijk Engels”). Als er met scores gewerkt wordt is bij 13 dubbel beoordeelde leerlingen dit verschil zelfs significant<sup>93</sup>. Wat er in 2012 gebeurd is en of de beoordelaars van 2012 onderling en met 2018 vergelijkbaar zijn is echter lastig te beoordelen. De vergelijkbaarheid over de jaren is daarmee ook twijfelachtig voor dit punt.

De opgaven waarbij het gebruik van woorden onderling vergeleken kan worden zijn ook bekeken over de jaren heen. Hierbij zijn een aantal keuzes te maken. Ten eerste of deze voor alle leerlingen vergeleken worden of alleen per afbeelding voor de leerlingen die op die afbeelding beoordeeld zijn. Het eerste geeft het beste het oordeel over de vaardigheid van de leerlingen weer: als leerlingen uitvallen die slecht presteren omdat ze geen Engels praten, geeft dat een overschatting van de prestatie. De tweede optie geeft het beste de indruk van de specifieke opgaven, omdat een opgave niet echt te evalueren is, als de afbeelding niet beoordeeld is. Een andere keuze betreft wat te doen met de 9 trichotome opgaven. Het onderscheid tussen “goed” en “bijna goed” lijkt soms kunstmatig en willekeurig gekozen bij opgaven. Om die reden is er ook gewerkt met het dichotome oordeel waarbij het onderscheid tussen “goed” en “bijna goed” niet meer gemaakt wordt. Dit zijn de dichotome varianten van de opgaven. Bij de originele varianten wordt de score gedeeld door de maximale score om deze opgaven niet dubbel mee te laten tellen. Dit levert in totaal een viertal mogelijkheden op. Welke manier ook gekozen wordt: het verschil tussen de jaren is zeer klein.

Als we de itemscores als schaal nemen dan is de betrouwbaarheid in de dichotome variant 0,78 (Lambda-2). Er zijn een paar opgaven die de schaal wat minder betrouwbaar maken omdat ze met de helft of meer van de opgaven negatief samenhangen. Dat zijn de opgaven 1.3, 2.2, 3.2, 3.5, 4.2, 4.3, 4.4, 5.2 en 5.7. De betrouwbaarheid wordt iets hoger zonder deze opgaven (0,79). Als we de 6 algemene opgaven “drukt de leerling zich in het Engels uit” per plaatje meenemen bij deze schaal wordt de betrouwbaarheid 0,83, maar daarbij moet opgemerkt dat hier dus wel een afhankelijkheid tussen deze vraag en de vervolgvragen per afbeelding zit. De gemiddelde scores bij deze verschillende scores zijn gegeven in Tabel 9.7. Van de verschillen zijn alleen de somscore over alle items en de zes oordelen over de gehele plaatjes omdat het Engels en inhoudelijk was, significant<sup>94</sup>. De scores waren in 2018 iets beter dan in 2012. De effectgrootte is daarbij 0,18.

---

<sup>92</sup> Het totaal aantal dubbel beoordeelde leerlingen was 14, maar bij een van deze 14 had een van twee beoordelaar geen oordeel over de uitspraak gedaan.

<sup>93</sup> Als alle leerlingen bekeken worden dan is het verschil niet meer significant, maar daar zitten dus vooral verschillende leerlingen tussen. Dit zou kunnen betekenen dat de milde beoordelaar de minder goede leerlingen beoordeelde.

<sup>94</sup> De standaardafwijkingen waren in geen van de vier scores significant verschillend van elkaar.

Tabel 9.7: verschillen tussen 2012 en 2018 bij de verschillende mogelijk scores (overall scores) op de spreektaak

	JAAR	N	Mean	Std. Deviation	Std. Error Mean
Som alle dichotoom gescoorde trichotome items	2018	295	3.66	1.84	0.11
	2012	204	3.37	1.79	0.13
Som alle items die van origine al dichotoom waren	2018	295	5.07	2.82	0.16
	2012	204	4.63	3.10	0.22
Som alle dichotome en gedichotomiseerde items	2018	295	8.73	4.13	0.24
	2012	204	8.00	4.47	0.31
Som alle dichotome en gedichotomiseerde items én oordeel over Engels gesproken	2018	295	14.15	4.82	0.28
	2012	204	13.20	5.42	0.38

## Hoofdstuk 10. Verdiepend onderzoek

Na afloop van het hoofdonderzoek is een verdiepend onderzoek uitgevoerd bij een deel van de deelnemende scholen. Daarbij stonden de volgende twee onderzoeksvragen centraal: *Wat zijn, na correctie voor de stratum en startmoment, de best presterende scholen?* en *Hoe geven deze scholen vorm aan de relevante factoren uit het onderwijsleerproces?*

In dit hoofdstuk staat de beantwoording van bovenstaande vragen centraal. Van de 95 deelnemende scholen heeft op 21 scholen (22 klassen) een verdiepend onderzoek plaatsgevonden. Doel van dit onderzoek was het opleveren van inspirerende lespraktijken en best-practices, zodat deze door andere scholen in het onderwijs benut kunnen worden. Daarnaast diende met het onderzoek inzicht te worden verkregen in het onderwijsleerproces op de best-presterende scholen uit het hoofdonderzoek. Hiertoe zijn lesobservaties en interviews uitgevoerd. De data zijn verzameld door zeven onderwijsinspecteurs. In dit hoofdstuk worden allereerst de randvoorwaarden van dit onderzoek (paragraaf 10.1), de selectie en werving van scholen (respectievelijk de paragrafen 10.2 en 10.3) en de uiteindelijke onderzoeksgroep (paragraaf 10.4) beschreven. Daarna wordt in paragraaf 10.5 het instrumentarium en de training van de onderzoekers besproken. Paragraaf 10.6 gaat in op de dataverwerking. De uitkomsten van het observatie-onderzoek en de relevante facetten uit de interviews worden in de paragrafen 10.7 t/m 10.10 besproken. De resultaten worden, waar relevant, aan de ongewogen resultaten van het hoofdonderzoek verbonden. Dit om de aansluiting met hoofdstuk 6 te waarborgen. Het hoofdstuk wordt afgerond met een conclusie en discussie in paragraaf 10.11.

### 10.1 Randvoorwaarden voor het onderzoek

Voorafgaand aan de uitvoering waren aan het verdiepende onderzoek enkele randvoorwaarden vastgesteld:

- Het onderzoek vindt enkel plaats op goed presterende scholen.
- Het onderzoek op de scholen wordt uitgevoerd door onderwijsinspecteurs die met behulp van door onderzoekers aangeleverd instrumentarium informatie over de lespraktijk komen verzamelen.
- Voor zover mogelijk zijn inspecteurs geselecteerd die affiniteit hebben met taalonderwijs en, zo mogelijk, kennis en ervaring hebben op het gebied van vreemde talen onderwijs.
- Afsproken is dat het regulier inspectietoezicht op de scholen en het verdiepende onderzoek volledig van elkaar zijn gescheiden. De betrokken inspecteurs zijn niet verbonden aan het reguliere toezicht van de betreffende school én er vindt geen communicatie plaats met de 'reguliere' inspecteur over de bevindingen van dit schoolbezoek.
- De inspecteur komt in de rol van onderzoeker: hij/zij registreert feitelijk het objectief waarneembare gedrag van de leerkracht (en leerlingen). Er wordt zowel tijdens/na de lesobservatie als ook tijdens het interview geen oordeel uitgesproken.
- Deelnemende inspecteurs hebben allen ervaring in het observeren en beoordelen van de generieke onderwijskwaliteit. Kennis van vakdidactiek kan op voorhand niet verondersteld worden.
- Er wordt gewerkt met een observatie-instrument bestaande uit zowel een didactisch én vakdidactisch gedeelte. Ieder deelinstrument kent een eigen focus.
- Voor de training van inspecteurs is één dag beschikbaar, waarin enkel het vakdidactische gedeelte van het instrument aan bod komt. Reden hiervoor is met name dat het pedagogisch-didactische deel (ICALT) al bekend is bij inspecteurs en zij hier reeds ruime ervaring mee hebben.

- Het onderzoek wordt uitgevoerd tussen half september en eind oktober 2018.
- Er wordt zo mogelijk aangesloten bij de planning van de lessen Engels op de scholen/bij de leerkrachten in bovengenoemde periode.
- De Inspectie van het Onderwijs heeft zeven inspecteurs voor het onderzoek beschikbaar om de haalbaarheid van het onderzoek te garanderen, opdat alle schoolbezoeken in de beoogde periode kunnen worden uitgevoerd.

## 10.2 De selectie van scholen

De bedoeling was dat het verdiepende onderzoek op de 32 best presterende scholen zou plaatsvinden. Conform het uitgangspunt dat goede voorbeelden moesten worden verzameld die relevant zijn voor verschillende typen scholen is er bij de selectie van scholen rekening gehouden met vier strata/clusters. Deze strata waren gebaseerd op twee variabelen: het startmoment Engels op de school (vvto- of eibo-scholen) en het percentage gewichtenleerlingen (meer of minder dan 10%). Met deze selectiewijze werd getracht recht te doen aan de herkenbaarheid van de uitkomsten voor verschillende typen scholen.

Besloten is om twee verschillende uitkomstmaten te hanteren. Vanwege de benodigde tijd voor de individuele gespreksbeoordelingen kon er voor de selectie van scholen niet tijdig een totaalscore worden berekend over de vier gepeilde domeinen. Op een deel van de data is in het voorjaar 2018 gekeken naar de samenhang tussen de klassikale taken en de gesprekstaken. Uit de geringe correlatie kwam naar voren dat de selectie op de twee uitkomstmaten deels verschillende scholen zou opleveren. Uitgaande van de twee uitkomstmaten, zou er gestreefd worden naar deelname door de zestien hoogst-presterende scholen op 'totaalscore Engels' - samengesteld op basis van de klassikale taken (lezen, luisteren, woordenschat) - en naar de zestien hoogst-presterende scholen op 'score voor gespreksvaardigheid'. Op deze manier probeerden we recht te doen aan het belang dat in het werkveld wordt gehecht aan gespreksvaardigheid Engels in het PO (Rose, 2016).

Voor de selectie van scholen is een meerniveau model met een leerling- en schoolniveau gebruikt om de Engelse taalvaardigheidsscores te analyseren. In dit model is, afhankelijk van de uitkomstmaat, een aantal relevante (en significante) covariaten opgenomen. Voor de selectie op klassikale taken zijn de volgende leerlingkenmerken als covariaat meegenomen: formatiegewicht, geslacht, leeftijd, thuistaal en frequentie van gebruik Engels (lezen, luisteren, spreken en schrijven) per dag. De frequenties van het gebruik van Engels door de leerlingen zijn opgedeeld in vijf categorieën. Voor de selectie op gespreksvaardigheid zijn weer formatiegewicht, geslacht en leeftijd als covariaat meegenomen, alsmede de score voor klassikale taken. Doordat hierbij reeds gecorrigeerd werd voor klassikale taken zijn thuistaal en buitenschools gebruik van Engels in deze analyses niet meer meegenomen. Gekeken is naar de schoolniveau-residuen van de modellen. Deze zijn gerangordend van laag naar hoog en de scholen met de hoogste rangnummers per stratum zijn geselecteerd. Deze rangordening heeft afzonderlijk plaatsgevonden. Omdat de 'totaalscore Engels' voor alle scholen in de peiling al eerder was vastgesteld, heeft de rangordening voor de klassikale taken plaatsgevonden vóór de rangordening op basis van de (score op) de gesprekstaken. Ondanks het gebruik van de score voor klassikale taken als covariaat in de selectie van de scholen op basis van gesprekstaken, zijn 11 benaderde scholen voor beide rangordeningen geselecteerd (zie Tabel 10.1). Op het moment dat een school reeds voor de klassikale taken was geselecteerd is bij de selectie op basis van de gesprekstaken de school met het daaropvolgende rangnummer geselecteerd.

Aangezien per school/klas maximaal zes leerlingen de gesprekstaken hebben uitgevoerd, is mogelijke non-representativiteit van de leerlingen die zijn geselecteerd voor de gesprekstak in een andere

analyse (via weging) verdisconteerd. Dit leverde vergelijkbare resultaten op in termen van rangordening. De selectie-resultaten worden op basis van de ongewogen analyses beschreven.

De 16 best presterende scholen per uitkomstmaat zijn benaderd voor deelname. Afgesproken is dat in geval van non-respons de eerstvolgende best presterende school in het betreffende cluster werd geselecteerd, op voorwaarde dat deze school bovengemiddeld scoorde op de betreffende uitkomstmaat (klassikale totaalscore of score gespreksvaardigheid). Daarom wordt in het vervolg gesproken over de bovengemiddeld presterende scholen.

### 10.3 Werving van scholen

Alle aan de peiling deelnemende scholen zijn tijdens het hoofdonderzoek reeds geïnformeerd over de mogelijkheid van selectie voor het verdiepende onderzoek. Na feitelijke selectie van de scholen zijn eerst de schooldirecteuren en vervolgens de leerkrachten benaderd via een e-mailbericht. Voorwaarde voor deelname was dat de leerkrachten in 2017-2018 Engels in groep 8 verzorgden en ook in 2018-2019 Engels in de bovenbouw zouden geven (dan wel bereid waren een les in groep te 8 te verzorgen). Na deze e-mail zijn leerkrachten telefonisch benaderd voor nadere informatie over de selectie en de inhoud van het verdiepende onderzoek. Daarbij lag de nadruk op wat deelname de school en de leerkracht zou kosten (tijdsinvestering, eventuele noodzaak tot vervanging) en wat het onderzoek de leerkracht en het onderwijsveld zou opleveren. Vanzelfsprekend is besproken hoe met de data werd omgegaan en dat er anoniem gerapporteerd zou worden. Leerkrachten ontvingen een bevestigingse-mail van hun deelname aan de verdiepende studie. De werving van de scholen vond plaats in de periode eind juni tot en met begin september 2018. Vanwege verwachte logistieke problemen (om het verdiepende onderzoek binnen de tijd uit voeren), is de werving begin september gestaakt.

Ongeveer een week voor het lesbezoek is aan de leerkracht (en schoolleider) nogmaals een herinneringse-mail gestuurd met daarin datum, tijdstip, afspraak (lesobservatie en interview van ongeveer een uur), naam van de bezoekende inspecteur en een algemeen telefoonnummer van de inspectie bij urgente wijzigingen.

### 10.4 De onderzoeksgroep

In juni is gestart met het selecteren van zestien scholen op basis van de behaalde scores op de klassikale taken. De eerste zestien scholen (vier per cluster) zijn benaderd en bij non respons werd de eerstvolgende school van dat cluster in de rangordening benaderd. Vanaf half juli is eveneens gestart met de werving voor de gesprekstaken. Bij (reserve)scholen die reeds voor de klassikale taken waren geselecteerd werd ook weer de eerstvolgende school binnen dat cluster in de rangordening geselecteerd. Enkel de scholen die bovengemiddeld presteerden op de uitkomstmaat zijn geselecteerd.

In totaal zijn 49 scholen/leerkrachten benaderd in de periode eind juni-begin september en hadden op 10 september 24 leerkrachten van 23 scholen toegezegd deel te nemen. Na 10 september zijn twee leerkrachten uitgevallen als gevolg van langdurige ziekte binnen de school en een overlijdensgeval, waardoor uiteindelijk 21 scholen (22 klassen/leerkrachten) bezocht konden worden. Tabel 10.1 laat zien hoe de scholen verdeeld zijn over de vier clusters en over de twee uitkomstmaten. In totaal namen zes scholen (29%) deel die geselecteerd zijn op basis van de klassikale taken, tien scholen (48%) op basis van de



gespreksvaardigheid en voor vijf scholen (24%) gold dat zij zowel bij de klassikale als gesprekstaken bovengemiddeld presteerden. Bijna de helft van de deelnemende scholen (48%) behoorde per cluster tot de vier best-presterende scholen voor gespreksvaardigheid. Voor de klassikale taken behoorde een derde van de scholen tot de beste presteerders per cluster. In de groep benaderde scholen, die niet wilden deelnemen, liggen deze percentages wat lager: 32% van de benaderde, niet-deelnemende scholen behoorde tot de beste presteerders per cluster voor klassikale taken en 21% voor de gesprekstaken. Bij de cluster 4-scholen (vvto-scholen, met meer dan 10% leerlingen met formatiegewicht) waren er bij gespreksvaardigheid buiten de vier best-presterende scholen geen reservescholen die bovengemiddeld presteerden; bij de klassikale taken waren er slechts twee bovengemiddeld presterende reservescholen die konden worden benaderd – de andere scholen presteerden ondergemiddeld. De werving in cluster 4 is daarom op een bepaald moment gestaakt. Bij cluster 2 (eibo-scholen, met meer dan 10% leerlingen met formatiegewicht) presteerden slechts vier reservescholen bovengemiddeld. Ook hier moest de werving op een bepaald moment worden gestopt bij gebrek aan bovengemiddelde scholen.

De verdeling over de clusters is redelijk evenwichtig. Vijf scholen zijn afkomstig uit cluster 1 (eibo, met weinig leerlingen met formatiegewicht), vijf scholen uit cluster 2 (eibo, met meer dan 10% leerlingen met formatiegewicht), zeven scholen uit cluster 3 (vvto-scholen, met weinig leerlingen met formatiegewicht) en vier scholen uit cluster 4 (vvto-scholen, met meer dan 10% leerlingen met formatiegewicht).

Tabel 10.1: Overzicht van de deelnemende scholen

Benaderde scholen	Stratum	Selectie: klassikale taken	Selectie: gespreksvaardigheid	Selectie: beide uitkomstmaten	Totaal	Behorend tot beste 4 scholen in cluster – klassikale taken	Behorend tot beste 4 scholen in cluster - gesprekken
Deelnemende scholen	1	1	2	2	5	2	2
	2	1	2	2	5	1	3
	3	2	4	1	7	2	3
	4	2	2	0	4	2	2
Totaal		6	10	5	21	7	10
Niet deelnemende scholen	1	3	4	3	10	2	2
	2	4	3	1	8	3	1
	3	2	1	1	4	2	1
	4	3	2	1	6	2	2
totaal		12	10	6	28	9	6

Alle benaderde scholen en leerkrachten is gevraagd naar de reden om af te zien van deelname. Hoewel de school de eenheid in de steekproef was is zowel met de schoolleider als de groepsleerkracht gecorrespondeerd over mogelijke deelname. In de praktijk bleek deelname echter in alle gevallen afhankelijk van de leerkracht. Schoolleiders lieten de keuze aan hen over. Van 25 leerkrachten die niet deelnamen, is bekend waarom zij niet deel wilden nemen aan het verdiepende onderzoek (zie Tabel 10.2). Doordat het verdiepende onderzoek na de zomervakantie plaatsvond, bleken wijzigingen op de school een duidelijke rol te spelen. Leerkrachten werkten bijvoorbeeld in het nieuwe schooljaar niet meer op school, hadden andere

taken gekregen (b.v. directietaken), waardoor ze niet/nauwelijks meer voor de klas stonden, of bleken niet meer in groep 8 te staan en daarom van deelname af te zien<sup>95</sup>.

Tabel 10.2: door leerkrachten aangedragen redenen om niet deel te nemen

Reden	Aantal leerkrachten
Werkdruk/mate van vernieuwingen op school	4
Geen interesse, ongemak over observaties	5
Leerkracht is van school	6
Leerkracht heeft andere taken	5
Leerkracht heeft andere klas	2
Zwangerschapsverlof	3

## 10.5 Instrumentarium en training inspecteurs

Binnen het verdiepende onderzoek is bij elke leerkracht een les Engels geobserveerd en naderhand een interview afgenomen. Leerkrachten is gevraagd een reguliere les Engels te verzorgen, zoals deze op het programma stond. Tevens was in de werving vermeld dat het interview zich vooral zou richten op hun beweegredenen en keuzes met betrekking tot lessen Engels. Ten behoeve van een gestandaardiseerde gegevensverzameling is een observatie-instrument (10.5.2) en een interviewleidraad (10.5.3) ontwikkeld aan de hand van enkele theoretische uitgangspunten (10.5.1). Het gebruik van de instrumenten is getraind met inspecteurs in een intensieve bijeenkomst (10.5.4).

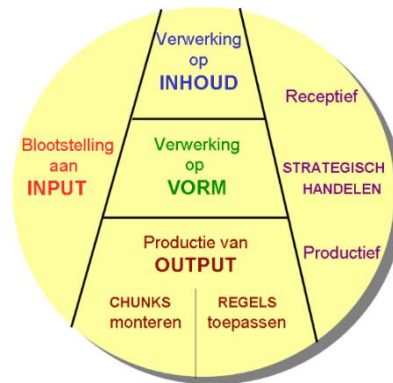
### 10.5.1 Theoretische onderbouwing instrumentarium

Voor het in kaart brengen van de generieke kwaliteit van de les is het ICALT-PO observatie-instrument gebruikt (Van de Grift, Van der Wal & Torenbeek, 2011). Het ICALT-PO instrument is een gevalideerd instrument, waarmee pedagogisch-didactische vaardigheden van leerkrachten in het basisonderwijs kunnen worden gemeten die bewezen effectief zijn. De ICALT-PO is nauw verwant aan het voormalige beoordelingskader dat door de Inspectie van het Onderwijs werd gehanteerd. De verwachting voorafgaande aan de schoolbezoeken was dan ook dat inspecteurs veel ervaring hadden met deze beoordelingswijze en het beoordelingsformulier betrouwbaar zouden kunnen invullen. De ICALT-PO kende binnen dit onderzoek echter een belangrijk beperking. Het meet namelijk slechts de generieke leerkrachtvaardigheid en het biedt geen informatie over relevante elementen in lessen Engels. Derhalve is ervoor gekozen een tweede, vakdidactisch gedeelte toe te voegen aan het instrument.

Het vakdidactische observatie-instrument en de vragen voor het interview zijn gebaseerd op theoretische uitgangspunten over succesvol onderwijs in de moderne vreemde talen en praktische bevindingen over de vakdidactiek Engels in het Primair Onderwijs (PO; Thijs & Tuin, 2015). Het onderwijs van het Engels in het Nederlandse PO heeft overigens wel een bijzondere status. Doordat de kerndoelen van het PO geen specifieke richtlijnen bieden en de kennis van de PO leerkrachten niet specialistisch is, verschilt het Engels in het PO van andere onderwijscontexten, zoals het VO. Hierdoor kunnen algemene kwaliteitscriteria over het onderwijs in de moderne vreemde talen niet zonder meer worden toegepast op talenonderwijs in het PO.

<sup>95</sup> Aan het begin van de wervingsperiode is leerkrachten die in de onderbouw lesgeven niet gevraagd alsnog een les in groep 8 te geven. Om deelname te bevorderen is deze vraag in een later stadium wel gevraagd – met wisselend succes.

Een belangrijk theoretisch uitgangspunt voor het vreemdetalenonderwijs in Nederland is de Schijf van Vijf voor het vreemdetalenonderwijs van Westhoff (2008) die in eerste instantie is gericht op het voortgezet onderwijs. Deze uitgangspunten worden ook veel toegepast op het PO (Corda, De Kraay, & Feuerstake, 2012; Corda, Philipsen, & De Graaff, 2014; Bodde-Alderlieste & Schokkenbroek, 2011). Om een taal te leren, dienen volgens de Schijf van Vijf 5 componenten aan bod te komen in taallessen (zie figuur 1; Westhoff, 2008), waaraan Westhoff ook bepaalde eisen stelt. Zo moet input bijvoorbeeld omvangrijk, gevarieerd, functioneel en levensecht zijn, en moet de input van een net iets hoger niveau zijn dan het huidige taalniveau van de leerling. Met strategisch handelen worden strategieën bedoeld die een taalleerder kan inzetten om te compenseren voor gaten in zijn taalkennis.



*Figuur 1 (Westhoff, 2008, p. 24)*

Het Vierfasenmodel van Barneveld & Van der Sanden (1987) biedt de leerkracht PO handvatten om de taalwetenschappelijke principes van de Schijf van Vijf te implementeren (Corda e.a., 2012; Meijerink e.a., 2012). Het is in feite een lessencyclus bestaande uit vier fasen in een vaste volgorde (zie figuur 2). Het Vierfasenmodel is een gangbaar model in vakdidactische handboeken geschreven voor leerkrachten PO (Oskam, 2008; Bodde-Alderlieste & Schokkenbroek, 2011). Om deze redenen zijn de uitgangspunten van Schijf van Vijf en het Vierfasenmodel vertegenwoordigd in het vakdidactisch-observatie instrument.

Fase:	Inhoud:
<b>1. Introductiefase (fase 1): pré receptief</b> Voorbereiden op thema en de te leren taal. Leren voorspellende strategieën gebruiken.	<ul style="list-style-type: none"> <li>• het <b>activeren</b> van buitenschoolse voorkennis over het nieuwe thema m.b.v. actieve werkvormen; inventariseren van woorden, zinnen en ervaringen</li> <li>• <b>motiveren</b> van de leerlingen voor het onderwerp en de eindopdracht; opwarmen en voorbereiden input</li> <li>• <b>herhalen</b> uit eerdere lessen</li> </ul>
<b>2. Inputfase (fase 2): receptief</b> Leren verstaan en begrijpen van de nieuwe taal. Lees-, luister-, kijkstrategieën leren toepassen.	<ul style="list-style-type: none"> <li>• <b>presenteren</b> van de nieuwe stof/taal; inputdialoog, kerndialoog en nieuwe woorden</li> <li>• <b>input verwerken</b> op inhoud en vorm</li> <li>• <b>controleren</b> of de input duidelijk is voor alle leerlingen</li> </ul>
<b>3. Oefenfase (fase 3): productief</b> Oefenen van de nieuwe taal. Spreek- en schrijfstrategieën leren gebruiken.	<ul style="list-style-type: none"> <li>• <b>oefeningen aanbieden</b> over de nieuwe stof in gesloten oefeningen, zowel mondeling als schriftelijk, van receptief naar (re)productief, van gesloten naar minder gesloten</li> </ul>
<b>4. Transferfase (fase 4): productief</b> Toepassen van de nieuw geleerde taal, samen met de voorkennis. Spreekstrategieën toepassen.	<ul style="list-style-type: none"> <li>• leerlingen begeleiden om de nieuwe stof samen met de voorkennis in vrijere opdrachten toe te passen, als tussenstap naar het gebruik van Engels buiten school</li> <li>• differentiatiemogelijkheden aanbieden</li> </ul>

*Figuur 2 (Corda, De Kraay & Feuerstake, 2012, p. 36)*

Een kanttekening dient geplaatst te worden bij de component 'vormgerichte verwerking' van de Schijf van Vijf. Deze is minder vertegenwoordigd in het observatie-instrument, omdat bij beginnende taalverwerving met name rijke en veelvuldige taalinput cruciaal is, met daarin slechts incidentele aandacht voor vormaspecten (Andrews, 2006; Ellis, 1998; Westhoff, 2006; Meijerink e.a., 2012). De rol van grammatica is echter niet geheel buiten beschouwing gelaten, aangezien "grammatica populair is onder basisschoolleerkrachten" (Corda e.a., 2014). Vanuit taalwetenschappelijk en curriculaire perspectief verwachten we niet dat grammatica, opgevat als aandacht voor vorm gecombineerd met metalinguïstische kennis, veelvuldig onderdeel zal zijn van de les Engels in het PO. De verwachting is wel dat er aandacht besteed wordt aan vormaspecten zonder die metalinguïstische informatie (uitspraak). Vanwege mogelijke overtuigingen en verwachtingen van de leerkracht over hoe een taal geleerd wordt en de rol die vormgerichtheid en grammatica daarin spelen wordt in het interview gevraagd naar de visie op moderne vreemde talen in het PO van de leerkracht (en van de school).

Binnen de context van vormgericht onderwijs omvat het observatie-instrument ook items over de manier waarop de leerkracht correctieve feedback geeft, in navolging van Westhoff (2008), Lyster & Ranta (1997) en Oskam (2008). Hierbij kan gedacht worden aan de vorm van correctieve feedback, bijvoorbeeld de leerling meteen verbeteren, of de leerling stapsgewijs zelf de fout laten verbeteren.

Daarnaast is het van belang om te kijken welke taal de leerkracht op welk moment hanteert en stimuleert (Westhoff, 2008; Oskam, 2017). Het belang van bewust doeltaal-voertaalgebruik is al door tal van onderzoek gestaafd (zie bijvoorbeeld, Verspoor, Schuitemaker-King, van Rein, de Bot, & Edelenbos, 2010). De rol van de moedertaal in de vreemdetaalles is meer omstreden (Lightbown, 2014) met het oog op het effect op vreemdetaalontwikkeling bij de verschillende vaardigheden. Echter, de leerkracht kan het Nederlands ook inzetten voor andere doeleinden die indirect de vreemdetaalontwikkeling ten goede komt (Oskam, 2017). Hierbij kan bijvoorbeeld gedacht worden aan het creëren van een veilig leerklimaat wat tezamen met een uitnodigend leerklimaat door leerkrachten als belangrijkste competentie gezien wordt (Tuin & Thijs, 2015). Ook Dönszelmann (2016; 2019) nuanceert het gebruik van de doeltaal. Het kan zeer nuttig zijn om de doeltaal te gebruiken, maar vooral als leermiddel en niet alleen als communicatiemiddel. De doeltaal is "een instrument om leerlingen taal te laten leren, afgestemd op de doelgroep en met

herkenbare elementen ('ankers') die leerlingen kans geven om te begrijpen en leren" (Dönszelmann, 2016, p. 38). Dit houdt in dat de inzet van de doeltaal gebonden is aan didactische eisen. Dönszelmann (2019) spreekt in dit verband dan ook liever van 'doeltaal-leertaal' dan van 'doeltaal-voertaal'.

### 10.5.2 Het observatie instrument

Hoofddoel van de observaties was om te komen tot waardevolle praktijkvoorbeelden die met het werkveld gedeeld zouden kunnen worden. Tevens werd getracht enig zicht te krijgen op de daadwerkelijke lespraktijk op scholen waar leerlingen goed presteren. Voor de observaties van de Engelse lessen is een instrument ontwikkeld dat uit twee onderdelen bestond: een op de generieke onderwijskwaliteit gericht deel en een vakspecifiek deel (zie Bijlage 12). Dit instrument is tijdens de les, ter plekke op papier ingevuld. Er was ook een digitale variant van het observatieformulier beschikbaar, waarin de inspecteurs naderhand alle aspecten konden aanvinken.

#### *Opbouw observatie-instrument*

Het observatie-instrument had, conform de theoretische onderbouwing (zie boven), een generiek en een vakdidactisch inhoudelijk deel. Daarnaast had het een voorpagina waarop algemene gegevens worden genoteerd, zoals naam van de school, datum en lesduur. Ook was er ruimte voor het noteren van klaskenmerken, zoals het aantal leerlingen in de klas, het tijdstip van observatie en het type klas (combinatieklas of niet), en van een indicatie van de mate van vloeiendheid en correctheid van het Engels van de leerkracht.

Het generieke gedeelte (ICALT-PO) bevatte 24 items. Deze meten zes onderliggende constructen, te weten: (1) een veilig en stimulerend klasklimaat, (2) efficiënte lesorganisatie, (3) duidelijke en gestructureerde instructie, (4) intensieve en activerende les, (5) afstemmen op verschillen en (6) leerstrategieën aanleren. De 24 items werden door de observatoren gescoord op een 4-punts Likertschaal (1= overwegend zwak, 2= meer zwak dan sterk, 3= meer sterk dan zwak en 4= overwegend sterk). De items zijn in het ICALT hiërarchisch geordend, van meer basale naar meer complexe vaardigheden. Een hogere somscore veronderstelt een hogere pedagogisch-didactische vaardigheid weer te geven. Naast de door de leerkracht uitgevoerde activiteiten wordt tevens de betrokkenheid van de leerlingen bij de les m.b.v. het observatieschema gemeten. Voor het leerproces is het immers van groot belang dat leerlingen actief meedoen met de les.

Het vakdidactische gedeelte bevatte 22 items. De les werd gescoord op zes dimensies van de les: (1) gebruik van taaldoelen, (2) uitleg en instructie, (3) interactie, (4) (vak)didactisch repertoire, (5) evaluatie van de les en (6) algemene indruk. Met de items bij *algemene indruk* wilden we meten in hoeverre de leerkracht de les vooral richt op het aanleren van het Engels of dat de leerkracht de les ook wil gebruiken als middel om leerlingen te enthousiasmeren voor het Engels/vreemde talen. Ook het aspect 'aansluiting bij de belevingswereld' werd in deze dimensie meegenomen. Per item kon worden aangekruist of het item werd gezien (ja) of niet (nee). Een aantal items is daarbij opgesplitst in subitems waarvoor ook moest worden ingevuld of het gezien werd of niet. Wanneer de leerkracht het gedrag behorend bij een (sub)item liet zien, én wanneer dit door de observator als  *mooi voorbeeld* werd geclassificeerd, kon de observator met behulp van steekwoorden kort omschrijven wat de leerkracht (precies) deed en hoe hij/zij dat uitvoerde. *Uitzonderingen voor deze werkwijze waren* items omtrent de taaldoelen en de geobserveerde (deel)vaardigheden. Bij deze items diende altijd te worden genoteerd welke doelen aan bod kwamen, respectievelijk welke activiteiten werden uitgevoerd.

Het vakdidactische deel van het observatieformulier kende een invulversie en een uitgebreidere versie. De korte versie was opgenomen in het gebruikte instrument. De uitgebreide versie (Bijlage 14) is een nadere uitwerking van de invulversie, waarin voorbeelden of beschrijvingen zijn gegeven van leerkrachtgedragingen waaraan het te scoren gedrag kan worden afgelezen. Deze uitgebreide versie is opgesteld ten behoeve van de bespreking van de items in de training van de inspecteurs en fungeerde tevens als naslagdocument tijdens het observeren.

### 10.5.3 Het interview

Het interview is na afloop van de observatie van de les Engels gehouden en had als hoofddoel "good practices" te verzamelen. Daarnaast wilden we inzicht krijgen in de beweegredenen en keuzes van leerkrachten bij het ontwerpen en geven van lessen Engels. De geplande duur van het interview was ongeveer een uur. Van het interview zijn audio-opnamen gemaakt op basis waarvan de interviews nader zijn uitgewerkt.

Het interview kende 3 centrale thema's: 1) visie op Engels in het PO, 2) informatie over een standaardles Engels, in relatie tot de geobserveerde les en 3) opleiding en professionele ontwikkeling van de leerkracht. Door bij thema 2 (de onderwijspraktijk) naar een standaardles Engels en de geobserveerde les te vragen werd leerkrachten ruimte geboden vrij over hun lessen te vertellen. Hierdoor hoopten we zoveel mogelijk voorbeelden voor de praktijk te verzamelen. Thema 2 was derhalve de kern van het interview en de instructie naar de inspecteurs toe was dat de focus van het interview op dit thema zou moeten liggen, ook qua tijdsplanning.

Onderzoek van Thijs, Trimbos en Tuin (2011) laat zien dat PO-leerkrachten twijfelen of zij over het vereiste B1-niveau beschikken. Dit werd al eens bevestigd in de gemeten zelfinschattingen van productieve vaardigheden in peiling Engels IV (Geurts & Hemker, 2013). Met de toevoeging van thema 3 is geprobeerd inzicht te verkrijgen in de ervaringen aangaande opleiding en professionalisering Engels en manieren waarop leerkrachten met eventuele lacunes in de opleiding omgaan.

#### De interviewleidraad

Voor het interview is een open 'leidraad' ontwikkeld die in totaal 12 vragen kende. Voor een vlotte verwerking is de leidraad omgezet naar een invulformulier (Bijlage 13). Het interview had name bij thema 2 een open karakter. Om er zeker van te zijn dat belangrijke aspecten van de lespraktijk Engels aan bod kwamen, waren de aandachtspunten in steekwoorden geformuleerd, die ook in het invulformulier waren weergegeven. De interviewer diende ervoor te zorgen dat alle genoemde aspecten besproken werden.

Er is gekozen voor deze open opzet, omdat ervaring leert dat dergelijke gesprekken zich niet houden aan strakke interviewleidraden. Stringente en sturende leidraden vragen veel van de interviewer (grote flexibiliteit en grondige kennis van de leidraad). Daarnaast verwachtten we dat sturende vragen meer sociaal-wenselijke antwoorden zouden opleveren. De bedoeling was derhalve dat de inspecteur uitnodigend en open in gesprek zou gaan met de leerkracht en zich zou focussen op het doorvragen *waarom* de leerkrachten bepaalde dingen wel of niet doet én of ze concrete voorbeelden konden geven of een collega zouden kunnen adviseren. Door daarbij een koppeling te maken met de geobserveerde les kon de leerkracht de geobserveerde les toelichten en nader duiden.

### 10.5.4 Training inspecteurs

Op 4 september 2018 zijn de zeven betrokken inspecteurs getraind. Ter voorbereiding hadden de inspecteurs 'De Schijf van Vijf van het vreemde talenonderwijs' (Westhoff, 2008) bestudeerd. Aan deze vijf

uitgangspunten (en de praktische vertaling in het Vierfasenmodel) werd in de training aandacht besteed. Tijdens de training zijn tevens de (sub)items in het observatie-instrument, hun specifieke betekenis en relevantie besproken. Ook is aan de hand van video-opnamen uitvoerig met de inspecteurs gesproken over hun scores en de beoogde scoringswijze. Daarbij is gebruik gemaakt van video-opnamen (afkomstig van Taalklavertje Vier) en de opname van een volledige les Engels. De volledige les is met behulp van het volledige observatie-instrument (ICALT PO en vakdidactisch gedeelte) door de inspecteurs gescoord en naderhand nabesproken. Na afronding van de scoring kregen de beoordelaars de richtlijn geen wijzigingen meer aan te brengen op het formulier tijdens de bespreking van de scores en overwegingen op de vakdidactische gedeelte. Het percentage overeenstemming dat bereikt werd voor ICALT PO was 81.5% en voor het vakdidactische deel 78.2%.

In de training is ook de interviewleidraad besproken, zijn de items daaruit toegelicht, en is de inspecteurs verteld dat het gesprek vooral een open karakter moest hebben en dat de focus van het gesprek vooral moest liggen op de items uit thema 2 (standaardles en geobserveerde les). Na afloop van de training is het protocol verder uitgebreid, aangescherpt en aan inspecteurs opgestuurd. Tevens ontvingen zij een voorbeeld-uitwerking van de les die tijdens de training door hen was geobserveerd, alsmede de powerpoint presentatie. Ook zijn de belangrijkste uitgangspunten en aandachtspunten nog een keer door de onderzoekers met de inspecteurs doorgenomen, zowel via e-mail als via telefonisch contact.

## 10.6 Dataverwerking

Tijdens de observaties hebben de inspecteurs in de klas op papier twee observatieformulieren ingevuld. Tijdens het interview noteerden de inspecteurs per vraag de belangrijkste bevindingen in steekwoorden/korte zinnen. Voor de begrijpelijkheid zijn – indien nodig - deze aantekeningen na afloop door de inspecteur nader uitgewerkt. Alle ingevulde interviewformulieren zijn in word-versie aan de onderzoekers opgestuurd. Daarnaast zijn audio-opnamen van de gesprekken met de leerkrachten gemaakt, waarvoor de leerkrachten voorafgaand om toestemming is gevraagd. Om een breder beeld van de algemene bevindingen te krijgen heeft op 7 november 2019 een focusinterview met de betrokken inspecteurs plaatsgevonden<sup>96</sup>. Gedurende deze bijeenkomst is de inspecteurs gevraagd naar een overall beeld van de geobserveerde lessen en interviews – mede aan de hand van opvallende uitkomsten in de data, en konden de inspecteurs een oordeel geven over datgene wat zij hadden gezien. Ook zijn onduidelijkheden in de notatie aan hen voorgelegd.

### Observaties

Na invoering van de observatiedata in SPSS zijn de 24 scores voor de ICALT gedichotomiseerd. Hiertoe zijn de scores 1 & 2 en 3 & 4 samengenomen tot een score van respectievelijk '0' en '1'. De gedichotomiseerde scores zijn samengevoegd tot een somscore per dimensie en een algehele somscore (zie Van de Grift, et al., 2011).

Voorts is er een totaalscore gemaakt voor het vakdidactische handelen van de leerkracht in de les. De 22 items uit de oorspronkelijke observatielijst zijn daarbij omgezet naar 15 overkoepelende vakdidactische leerkracht-activiteiten (zie Bijlage 12) – dit zijn gedichotomiseerde scores, die aangeven of de activiteiten al dan niet geobserveerd zijn in de les. De 15 vakdidactische activiteiten betroffen: 1) Het gebruik

---

<sup>96</sup> Twee inspecteurs waren helaas verhinderd.



van (taal)doelen, 2) het activeren van kennis bij leerlingen, 3) het controleren of leerlingen het aanbod begrepen hebben, 4) het gebruik van de doeltaal in de les, 5) het begrijpelijk maken c.q. aanpassen van eigen doeltaalgebruik, 6) het gebruik van werkvormen voor communicatieve doeleinden, 7) het aandacht besteden aan taalfouten, 8) het geven van positieve feedback, 9) het gebruik van differentiatie, 10) het stimuleren van gebruik van compensatiestrategieën, 11) het laten oefenen met een aangeleerde regel, 12) het gebruik van samenwerking, 13) de aansluiting van werkvormen bij de taaldoelen, 14) het afsluiten van de les door terug te grijpen op de taaldoelen en 15) het motiveren van leerlingen voor Engels<sup>97</sup>. Deze activiteiten geven leerkrachtgedrag weer dat als bevorderlijk voor het leren van een vreemde taal wordt beschouwd.

De dichotome scores voor de leerkrachtactiviteiten zijn als volgt berekend. Allereerst zijn dichotome scores berekend voor de overkoepelende leerkrachtactiviteiten 'taaldoelen', 'voorkennis', 'doeltaal in interactie' en 'differentiatie'<sup>98</sup>: een score '1' (aanwezig in de les) werd toegekend aan het item als één van de subitems zichtbaar was in de les. De subitems behorend bij deze 4 items staan vermeld in de kolom 'Dit kun je zien aan....' in bijlage 14. Ten tweede is het oorspronkelijke item 19 ('werkvormen' – zie bijlage 14) hernoemd tot 'Samenwerking', waarbij de subitems 'leerlingen werken in tweetallen' en 'leerlingen werken in groepjes' zijn meegenomen in de berekening van de score 'Samenwerking'. Ook hier betekent de aanwezigheid van één van beide subitems in de les dat het item als 'aanwezig' is gecodeerd. Voor de overige vier leerkrachtactiviteiten met subitems (controle op begrip, begrijpelijk maken eigen spreken, aandacht voor taalfouten en compensatiestrategieën) zijn ook dichotome scores gemaakt (als één van de subitems positief was gescoord, dan is het item gecodeerd als 'aanwezig'). Uiteindelijk is van de vijftien dichotome scores een somscore gemaakt, die een globale indicatie geeft van de vakinhoudelijke kwaliteit van de les.

## De interviews

De beschrijvingen op het invulformulier van het interview zijn nader uitgewerkt door de interviews te beluisteren en hier uitgebreide gespreksverslagen van te maken. Aan deze gespreksverslagen zijn letterlijke quotes toegevoegd, die voorbeelden, onderbouwingen of redeneringen bevatten die bruikbaar konden zijn voor de beschrijving van de visie op Engels in het PO en van 'good practices' van de leerkrachten.

De uitgewerkte verslagen zijn ingevoerd in het verwerkingsprogramma *Atlas.ti*. Bij de codering is gebruik gemaakt van op voorhand vastgestelde thema's behorend bij de inhoud van de (sub)vragen (1 en 3) en de besproken aspecten bij vraag 2. Deze codes zijn gecategoriseerd en aangevuld met thema's en concepten die uit de interviews naar voren kwamen. Ook aan relevante, woordelijke uitspraken van leerkrachten zijn codes toebedeeld. Bij de verwerking van de interviewdata is nagegaan in hoeverre het startmoment (vvto, eibo) en het gewichtenpercentage van de leerlingen op de school een rol speelde bij het denken of handelen van de leerkrachten. Waar relevant is het onderscheid beschreven.

---

<sup>97</sup> Drie (sub)items uit Bijlage 14 zijn in de analyses buiten beschouwing gelaten vanwege a) een gebrek aan eenduidige relatie tussen aanwezigheid van het subitem en het taalleerproces (te weten de subitems 'zelfstandig' en 'klassikaal' bij werkvormen -, item 19) en b) onvoldoende aandacht voor het construct in de training en derhalve interpretatieproblemen bij de observatoren ('de leerkracht past zijn doeltaalgebruik doelbewust aan'). Oorspronkelijk bevatte de lijst ook een item (item 17) waarmee werd geturfd aan welke vaardigheden of competenties de leerkracht aandacht besteedde. Deze telling is niet meegenomen in de vakdidactische score (en het overzicht in tabel 10.5), maar wordt apart beschreven in paragraaf 10.8.4.

<sup>98</sup> Deze leerkrachtactiviteiten waren oorspronkelijk als losse items in het formulier opgenomen.



## 10.7 Resultaten verdiepend onderzoek: interviews en observaties

In totaal zijn er 21 scholen bezocht. Op deze scholen zijn 22 lessen geobserveerd, waaronder één ‘content and language integrated learning’ (CLIL)-les - een les waarin andere vakken dan Engels in het Engels worden gegeven, zoals in het tweetalig onderwijs (tto) gebruikelijk is. Daarnaast zijn er 21 interviews met leerkrachten afgenomen<sup>99</sup>. Op drie scholen is het onderzoek bij een vakleerkracht afgenomen. Tijdens deze interviews kwamen drie hoofdthema’s aan bod: 1) de visie op Engels in het PO, b) een karakterisering van (standaard)lessen Engels en 3) de initiële scholing en verdere professionalisering van leerkrachten op het gebied van Engels. In de volgende paragrafen worden de resultaten van de observaties en interviews beschreven. Hierbij wordt waar relevant en mogelijk een koppeling gemaakt met bevindingen uit het hoofdonderzoek. De observaties en interviews bieden daarbij een verdieping op wat er op basis van de antwoorden in de leerkrachtvragenlijst uit het hoofdonderzoek is gevonden. Deze vragenlijst is door 109 deelnemende leerkrachten ingevuld.

### 10.7.1 Visie van leerkracht op het vak Engels in het PO

Bij de start van het interview is leerkrachten gevraagd naar hun eigen visie en die van de school op Engels in het basisonderwijs en naar de mate waarin die twee overeenkomen. De vraagstelling was gericht op wat de leerkrachten belangrijk vonden in de lessen Engels en welke doelen zij wilden bereiken. De voornaamste aspecten die uit de vraag naar de visie voortkwamen zijn: 1) leren communiceren (spreken en begrijpen), 2) het zich veilig/vertrouwd voelen om de vreemde taal te gebruiken, 3) het bevorderen van plezier in de vreemde taal, 4) het leggen van een basis en 5) het belang van een vroege start.

#### *Communiceren in het Engels*

Ruim de helft van de leerkrachten (4 vvto en 8 eibo) benoemt expliciet dat de lessen Engels er vooral op gericht moeten zijn om (op basaal niveau) te kunnen communiceren. Met name het kunnen gebruiken van de taal in authentieke situaties staat daarbij centraal: “Ik vind het belangrijk dat ze zich verstaanbaar kunnen maken. Op het moment dat ze op een camping vriendjes of vriendinnetjes tegenkomen, dan wil ik dat ze zich, al is het met handen en voeten, verstaanbaar kunnen maken.” Opvallend is dat de helft van de leerkrachten die de focus met name leggen op communicatie in het Engels, deze direct koppelen aan het werken aan voldoende zelfvertrouwen en durf om de taal te spreken. Dezelfde leerkracht zegt: “Fouten maken mag. Ik heb liever dat je het heel fonetisch uitspreekt en dat het niet klopt maar dat je het wel geprobeerd hebt.”

De observaties bevestigen het belang dat gehecht wordt aan het kunnen communiceren in het Engels, aangezien er in 17 (77%) van de 22 lessen sprake is van communicatieve werkvormen, zoals het gebruik van spelletjes in het Engels als “Wie is het?”, van samenwerkingsvormen en van dialogen. Sprekend over het communicatieve hoofddoel gaan leerkrachten regelmatig in op de relatie tussen het communicatief gebruiken van de taal (met name mondeling) en het aanleren van andere, meer vormgerichte vaardigheden, zoals grammatica en spelling. Vier leerkrachten geven stellig aan dat grammatica ondergeschikt zou moeten zijn aan dit communicatieve doel, zoals door de volgende leerkracht geformuleerd: “We hebben een paar jaar een vakdocent gehad en die heeft ons ook een keer als team een studiedag gegeven, waarin eigenlijk centraal stond: ‘Je moet gewoon heel veel Engels praten en je moet ook de kinderen zoveel mogelijk Engels

---

<sup>99</sup> Eén interview vond plaats met twee leerkrachten, maar hiervoor was de inbreng van beide leerkrachten niet van elkaar te onderscheiden. Dit interview wordt als één interview met één leerkracht geïnterpreteerd.

laten praten.' [...]Veel laten spreken, veel laten luisteren, dan maak je het je eigen. (...) Als je dat gewoon regelmatig doet, dan zijn ze straks 14-15 jaar en dan kunnen ze gewoon Engels. En dan kunnen ze op gevoel de zinsbouw en zo [grammatica] ook voor elkaar krijgen."

### *Het durven gebruiken van de taal/bevorderen van zelfvertrouwen*

Zestien leerkrachten, waarvan 10 vvtto-leerkrachten, noemen als belangrijkste doel van hun Engelse lessen dat leerlingen de taal durven te gebruiken en zich voldoende op hun gemak voelen wanneer zij dit doen. Enkele leerkrachten spreken daarbij over het gebruik van een breder scala aan vaardigheden Engels (spreken, luisteren, lezen), terwijl de meeste leerkrachten bij dit punt specifiek de spreek- en gespreksvaardigheid noemen. Zoals een eibo-leerkracht zei: "We willen bereiken dat kinderen vrijuit durven te praten, ook in situaties waarin ze misschien voor het eerst met iemand in gesprek komen. Dat is het voornaamste." Leerkrachten geven in dit verband aan dat zij willen dat leerlingen zich vrij voelen om Engels te gebruiken en dat er geen sprake moet zijn van angst. Om dit te bereiken proberen de leerkrachten het zelfvertrouwen van de leerlingen te bevorderen en het Engels 'niet te zwaar te maken'. Een leerkracht zegt daarom vaak tegen de leerlingen dat ze al veel meer kunnen dan ze denken. Vier leerkrachten benadrukken expliciet bij de leerlingen dat fouten maken niet erg is. Zij corrigeren de leerlingen dan ook niet of nauwelijks en afhankelijk van het niveau van de leerling. In de geobserveerde lessen is teruggezien dat de leerkrachten inzetten op het welbevinden van de leerlingen in de Engelse les. In 17 lessen (77%) is er sprake positieve feedback van de docent, met name via prijzende opmerkingen.

Twee leerkrachten stellen in dit kader specifiek het gebruik van toetsen (en dit peilingsonderzoek) ter discussie. De druk die toetsing van het Engels bij leerlingen veroorzaakt zou ten koste kunnen gaan van het durven: "Je legt er wel weer gelijk een hoop druk op. Engels moet ook wel leuk blijven, vind ik, dat ze niet het gevoel hebben dat ze daar ook nog in moeten presteren." Hoewel enkele leerkrachten er moeite mee hebben om leerlingen tot spreken te motiveren, noemt een vvtto-leerkracht de kracht van de gewenning. Op deze school wordt, vanaf groep 1-2, thematisch gewerkt binnen de lessen en ieder thema wordt afgesloten met presentaties in het Engels door de leerlingen. Het gemak waarmee dat gebeurt verbaast en motiveert de leerkracht. Sprekend over de deelnemende groep 8-leerlingen: "Die hielden al hele verhalen in groep 6 en natuurlijk zitten er wel allemaal fouten in, maar dat maakt niet uit. Het gaat meer om het feit dat ze dat durven. En dan kun je al zoveel bereiken eigenlijk..."

### *Bevorderen van plezier in de vreemde taal*

Negen leerkrachten benoemen bij de gevraagde visie op Engels het bevorderen van plezier in Engels. De drie eibo-leerkrachten koppelen dit plezier aan durf en ditzelfde geldt voor vijf van de zes vvtto-leerkrachten. Het enthousiasmeren van leerlingen voor Engels is volgens deze leerkrachten goed voor hun leerproces en dit gebeurt door speelse en afwisselende werkvormen, waarin ruimte is voor liedjes, filmpjes en activerende werkvormen (rollenspelen, dialogen, spelletjes). Drie leerkrachten benoemen expliciet dat zij hun eigen enthousiasme en liefde voor de taal op de leerlingen willen overbrengen. Een ander aspect dat door twee leerkrachten naar wordt gebracht is dat ze de natuurlijke nieuwsgierigheid van leerlingen ook proberen te prikkelen, bijvoorbeeld door de leerlingen uit te dagen op zoek te gaan naar overeenkomsten tussen het Nederlands, het dialect en het Engels. Dat spreekt de leerlingen aan. Een leerkracht noemt verder dat leerlingen vaak ook onbewust met de taal bezig zijn en bijvoorbeeld willen weten waar liedjes over gaat. De leerlingen zijn dan sterk gemotiveerd om het liedje te vertalen. Deze bewuste omgang met taal van de leerlingen motiveert de leerkracht ook weer.

### *Leggen van een basis/kennismaking met de taal*

Acht leerkrachten benadrukken dat het Engels in het PO met name gericht moet zijn op het leggen van een algemene basis in het Engels bij de leerlingen. Leerlingen moeten volgens de leerkrachten vooral veel met Engels (in allerlei vormen, als films, internet, boeken, liedjes) in aanraking worden gebracht. Ook wordt gesproken over een kennismaking met de taal en de cultuur. Dit laatste is met name relevant voor scholen met een duidelijke internationale focus. Om de blik van de leerlingen te vergroten wordt op een internationaal georiënteerde vvto-school bijvoorbeeld ook een 'pen-pal'-project – een schrijfmaatjesproject - met een school in Schotland georganiseerd, waarbij de leerlingen het hele jaar schrijven met een maatje in Schotland en op een bepaald moment op bezoek gaan. Hierbij leren ze veel over de stad en regio, de taal, de geschiedenis en het landschap. Later in het jaar komen de Schotse leerlingen dan naar Nederland. Ook zij leren veel over ons land en doen bijvoorbeeld mee met projecten, zoals een project over de Tweede Wereldoorlog (uitstapje naar bunkers).

### *Belang van een vroege start*

Zeven leerkrachten (4 vvto en 3 eibo) noemen het vroeg beginnen met Engels, wanneer hen gevraagd wordt naar hun visie op Engels. Met name vanwege het belang van Engels in onze (toekomstige) samenleving vinden zij het belangrijk om in de kleuterklas of in groep 4 met Engels te beginnen. Dit zou dan op een speelse manier moeten gebeuren, met veel aandacht voor het spreken, voor liedjes en voor spelletjes. Juist wanneer al op jonge leeftijd wordt begonnen ervaren leerlingen geen schroom om Engels te spreken, waar ze op latere leeftijd van kunnen profiteren.

### **10.7.3 Belang Engels in het PO**

Naast de visie op Engels is leerkrachten gevraagd naar het belang dat zij hechten aan Engels en de wenselijkheid van het opnemen van Engels in de eindtoets in groep 8. Achttien leerkrachten onderschrijven het belang van Engels volmondig, ook in het basisonderwijs. Eén van de leerkrachten maakt daarbij de kanttekening dat een andere vreemde taal, zoals Duits of Frans, ook relevant zou kunnen zijn, afhankelijk van de regio van de school. Niet alle leerkrachten beargumenteren waarom ze Engels belangrijk vinden. Van de leerkrachten die dit wel doen, benoemen negen leerkrachten dat het Engels vooral ook belangrijk is voor de toekomst van de leerlingen. Engels wordt steeds belangrijker in de maatschappij; de internationalisering en globalisering nemen toe en ook in de latere schoolloopbaan neemt Engels een steeds belangrijker plek in. Deze leerkrachten noemen het daarom relevant dat leerlingen al vroeg (in het PO) in contact komen met Engels en de taal leren gebruiken. Een enkele leerkracht wijst er op dat Engels ook in groep 8 al relevant is in hun dagelijks leven (internet, televisie): "In de bovenbouw vragen kinderen ook wel eens: 'Wat heb je hier nou aan? Waarom leer je het?' [...] En bij Engels kan ik dit heel goed uitleggen. Het is heel makkelijk in het dagelijks leven, want je komt overal, op internet en TV, Engels tegen. Als je naar het buitenland gaat is het nog steeds in vrijwel elk land de meest gevoerde taal dus je kunt er gewoon goed mee terecht." Vijf van de leerkrachten koppelen het belang van Engels voor PO-leerlingen ook aan wat leerlingen nodig hebben in het voortgezet onderwijs. Met name het in de brugklas starten met voldoende bagage en het daar goed mee kunnen komen vinden zij belangrijk. Een leerkracht van een vvto-school noemt beide aspecten: "Die vroege taalperiode hebben we toen ingezet om een behoorlijk niveau te krijgen aan het eind van groep 8. En dan hebben we echt voor ogen de aansluiting bij het voortgezet onderwijs, maar ook gericht op later. De samenleving verengelt."

Drie leerkrachten zijn minder uitgesproken: zij vinden dat de focus in het basisonderwijs op de Nederlandse taal (en rekenen) zou moeten liggen. Dit zou prioriteit moeten krijgen, zeker ook gezien hun wat zwakkere leerlingpopulatie. Dit zijn leerkrachten op scholen met meer dan 10% gewichtenleerlingen.

### *Engels als onderdeel van de eindtoets?*

Over de toevoeging van Engels aan de eindtoets bestaan grotere verschillen van mening. Negen leerkrachten zien de meerwaarde van een dergelijke toevoeging niet in. Volgens vier van deze leerkrachten sluit het gewenst aanbod binnen het PO, dat met name gericht is op communicatie en plezier, niet goed aan bij schriftelijke toetsing. Andere genoemde redenen zijn: de druk die toetsing oplegt bij de leerlingen (alsook hun ouders) en het mogelijke (negatieve) effect op het plezier in Engels. Ook het gebrek aan aansluiting bij de algemene onderwijsvisie van de school en de leerkracht en de prioriteit die leerkrachten geven aan Engels wordt genoemd: Nederlands wordt belangrijker geacht en/of ook andere vakken zouden dan in aanmerking moeten komen voor de eindtoets.

Twaalf leerkrachten zijn juist wel van mening dat Engels in de eindtoets zou moeten worden opgenomen. Het belang van Engels - in het kader van eerder genoemde globalisering en overgang/doorstroom VO ('het niveau van de leerlingen aan het VO laten zien') - en nieuwsgierigheid naar het uiteindelijke niveau van de groep 8-leerlingen worden hiervoor aangevoerd. Wel worden wenselijke randvoorwaarden genoemd, zoals het verplicht stellen van Engels vanaf groep 3 of 5, het verplicht stellen van een vakleerkracht en het relatieve gewicht dat aan Engels zou moeten worden gehecht. Deze leerkracht geeft dergelijke overwegingen weer: "Als je er een verplicht vak van zou maken vanaf groep 3 (...) dan zou ik het zeker willen. Je moet je afvragen hoe zwaar je dat dan laat meetellen, maar Engels is zo'n essentieel onderdeel van onze maatschappij. Je ontkomt er niet aan. Misschien moet het aan de zijlijn meetellen, de hoofdmoot blijft rekenen en begrijpend lezen." Een andere (vak)leerkracht ziet in opname van Engels in de eindtoets de mogelijkheid om Engels nadrukkelijker in het PO-aanbod uit te werken en om verschillen in het niveau van de vaardigheden van de leerlingen in groep 8 te verkleinen: "... we [po-scholen] zijn nu allemaal een beetje aan het pionieren nog steeds, en de ene school vindt het wel belangrijk en de andere pff, nou ja,...en er zijn ook zoveel andere dingen die ook aandacht vragen, maar ik denk dat als je dit ook echt goed wilt en ook het idee hebt dat, hier hebben kinderen echt op de langere termijn wat aan, dan moet het ook aangepakt worden binnen het middelbaar onderwijs, dan moet het gewoon gelijk getrokken worden voor alle scholen." Deze leerkracht pleit voor een groter belang van Engels in het PO, maar met name ook voor het maken van duidelijke keuzes voor/door de scholen. Dit punt van de grote diversiteit in het aanbod van Engels tussen scholen en de onwenselijkheid daarvan gezien de eisen van Engels in het VO, wordt vaker genoemd.

#### **10.7.4 Doorgaande leerlijnen po-vo**

Zoals hierboven reeds aan bod kwam is de aansluiting tussen PO-VO op het gebied van Engels als thema in de interviews niet expliciet aan bod gekomen. Ruim twee derde van de leerkrachten (72%) bracht het echter zelf ter sprake. Hiervan benadrukten negen leerkrachten dat zij zich inspannen om een behoorlijk tot goed niveau in de Engelse vaardigheden bij hun leerlingen te bereiken, zodat de leerlingen op het VO direct goed mee kunnen komen of zelfs een voorsprong op andere leerlingen hebben. Twee leerkrachten wijzen daarbij expliciet op het belang daarvan voor juist *hun* leerlingpopulatie met relatief veel gewichtenleerlingen. In twee interviews betreuren de leerkrachten dat leerlingen door het hoge PO-niveau op het VO vrij veel herhaling krijgen en dat het resultaat van het intensieve aanbod daarmee verloren gaat. Een ander aandachtspunt is het verschil dat bestaat tussen de op communicatie gerichte visie (en het daarbij

aansluitende aanbod) in het PO en de verwachtingen in het VO, waar men uitgaat van meer aandacht voor grammatica en schriftelijke vaardigheden. Zoals één leerkracht het verwoordde: “Voor basisschoolleerlingen is het belangrijk dat ze Engels op een leuke manier doen en dat het niet zo heel erg schools is, dat kan straks altijd nog op het voortgezet onderwijs. Daar zit dan ook wel meteen een groot probleem, want het voortgezet onderwijs gaat soms uit van een bepaald startniveau, dat we al wat aan grammatica gedaan hebben en aan zinsopbouw, ja dat is niet altijd van toepassing .... daar zit gewoon een verschil tussen waar het vo vanuit gaat en waar het po mee bezig is.” Tussen de groep 8-leerkrachten en het VO vindt contact plaats over het (algemene) niveau van de ‘afgeleverde’ leerlingen, al dan niet tijdens algemenere terugkomdagen. Ook zijn er leerkrachten die zelf het initiatief nemen om met docenten Engels in het VO af te stemmen hoe hun leerlingen presteren in de brugklas en wat zij verwachten. Dit kan leiden tot aanpassingen, zoals voldoende aandacht voor (specifieke) woordenschat en zinsbouw.

## 10.8 De invulling van de lessen Engels

Het tweede thema binnen het interview was de invulling van een standaardles Engels. De leerkrachten is gevraagd een reguliere les Engels te beschrijven. Dit leidde tot open beschrijvingen, waarbij de interviewers zoveel mogelijk – indien de leerkracht het zelf niet ter sprake bracht – de vooraf vastgestelde onderwerpen inbrachten. Leerkrachten was op voorhand gevraagd een reguliere les te geven. Onderdeel van het interview was de vraag of de geobserveerde les inderdaad representatief was voor de reguliere lessen Engels die door de leerkracht verzorgd worden. Slechts op één school gaf de leerkracht aan dat de les iets anders verliep, omdat, vanwege ziekte, een deel van de les reeds door een andere leerkracht was uitgevoerd. De overige leerkrachten noemden de les representatief voor hoe zij de les normaal gesproken geven. Een voorbehoud dat daarbij meermaals werd genoemd, was dat de leerkrachten de groep over het algemeen nog niet goed kenden, omdat de lessen aan het begin van het schooljaar werden geobserveerd. Gevolgen daarvan waren dat de leerkrachten het Nederlands naar eigen zeggen veelvuldiger gebruikten dan later in het schooljaar en dat zij het niveau van de leerlingen (en de klas) nog niet goed hadden kunnen vaststellen. In deze paragraaf worden de lesaspecten van thema 2 één voor één besproken, alsmede additionele onderwerpen die door leerkrachten naar voren werden gebracht (vanaf paragraaf 10.8.2). Deze beschrijvingen worden, daar waar relevant, aangevuld met informatie uit de observaties en bevindingen uit het hoofdonderzoek. Gestart wordt met een beknopte beschrijving van de observaties, meer specifiek van het algemene niveau van pedagogisch-didactisch handelen en van het vakdidactische handelen van de geobserveerde leerkrachten (paragraaf 10.8.1).

## Kader 1: Samenwerking tussen po en vo: doorgaande leerlijnen

Op één school wordt een nauwe aansluiting tussen het aanbod op het PO en VO gezocht door de inzet van een werkgroep waarin zowel po-leerkrachten als vo docenten Engels zijn vertegenwoordigd. Alle deelnemers in de werkgroep zijn werkzaam voor hetzelfde schoolbestuur. Doelstelling is om doorgaande leerlijnen voor de afzonderlijke vaardigheden Engels te ontwikkelen. Hierdoor zou het vak voor leerlingen als een opbouwend geheel worden ervaren en ervaren de leerlingen een minder grote overgang wanneer ze van po naar vo gaan. Leerlingen in de brugklassen zouden verder minder te maken krijgen met herhaling, en de kwaliteit van het Engels in zowel groep 8 als in de brugklassen zou op termijn voor alle scholen (in hetzelfde bestuur) minder variatie moeten laten zien. Doorlopende leerlijnen vergemakkelijken het voor leerkrachten (aan beide kanten van de PO-VO overgang) om te werken aan (gezamenlijke) doelstellingen en om tevens beter hun individuele leerlingen te kunnen bedienen.

In de werkgroep wordt besproken wat het VO van de leerlingen in bepaalde klassen verwacht en hoe en wat leerlingen in de eerste drie leerjaren VO (zouden moeten) leren. Het blijkt dat dit voor veel leerkrachten zeer interessant en leerzaam is: "Dan krijg je zicht op hoe snel het in het VO met het Engels gaat". Naast overleg over de inhoud van de lessen en het curriculum Engels, vinden er regelmatig uitwisselingen tussen scholen plaats. PO-leerkrachten bezoeken (brugklas)lessen van VO-docenten. VO-docenten komen op hun beurt regelmatig in de PO-scholen om daar voorbeeldlessen te geven (in alle groepen), om de lessen van leerkrachten te observeren en om de leerkrachten te adviseren.

In de lijn van het ERK worden in de werkgroep nadere *Can-do* statements geformuleerd om tot de verbeterde leerlijnen te komen. Daarbij is gestart bij de doelstellingen die in het VO worden gehanteerd. Vervolgens is terug geredeneerd naar (mogelijke) doelstellingen voor PO, tot en met groep 1-2. De doelstellingen liggen op het gebied van woordenschat bijvoorbeeld in inhoudsgebied ('over welke thema's zou en kind in groep x iets moeten weten?') en aantallen ("hoeveel woorden moet de leerling kennen?").

Op het moment van het schoolbezoek voor de peiling is de werkgroep bezig met de vraag hoe men leerlingen in groep 8 kan uitsplitsen (per deelvaardigheid) naar de ERK niveaus A1, A2, B1. Doel daarvan is dat "[je] de bovenste laag, met name havo-/vwo-leerlingen, maar ook andere leerlingen die [zich] op een andere manier het Engels eigen hebben gemaakt, kunt identificeren". De uitsplitsing naar niveau is iets wat het VO-veld in de werkgroep graag ziet: op die manier kunnen kinderen in de brugklassen beter worden bediend. Als PO en VO gezamenlijk aan dezelfde doelstellingen en leerlijnen werken is het voor de docent in het VO beter duidelijk waar hij/zij in een groep, met een specifiek niveau, mee verder kan (bijvoorbeeld een havo-klas met relatief veel kinderen met een A2 niveau uit de PO-scholen). Ook voor het beter differentiëren in de hoogste groepen van het basisonderwijs is deze kennis van belang.

De leerkrachten van de basisscholen die meedoen in de werkgroep worden enkele uren vrij geroosterd om aan de afstemming met het VO te werken. Dat betekent dat deelname niet vrijblijvend is en dat door het bestuur, op termijn, ook resultaten worden verwacht. Daarbij is voldoende draagvlak voor samenwerking onontbeerlijk. In dat kader moet opgemerkt worden dat in dit geval de aansturing van het gezamenlijke bestuur het samenwerken tussen PO en VO zeker makkelijker heeft maakt. Samenwerking levert inzichten tot verbetering op, maar implementatie van die verbeteringen moeten nog altijd op schoolniveau worden gerealiseerd. Alle afzonderlijke scholen in de werkgroep gaan met de afspraken aan de slag. De specifieke school probeert op maat (niveau) het aanbod af te stemmen, de "leerlijnen te bedienen", maar ervaart dat dit nog moeilijk is: "zeker als het gaat om differentiatie m.b.t. sterke leerlingen".

### 10.8.1 Niveau van pedagogisch-didactisch handelen

Met behulp van de ICALT-PO is het generieke niveau van leerkrachthandelen door de inspecteurs in kaart gebracht. De ICALT-PO bevat 24 items. Om te komen tot het generieke niveau is van deze 24 items een somscore berekend, lopend van minimaal 0 tot maximaal 24 punten. Gezien de opbouw in vaardigheidsniveau staat een hogere score daarbij voor een hogere niveau van pedagogisch-didactisch handelen. Gemiddeld behaalden de leerkrachten een ICALT-score van 15 punten tijdens hun lessen Engels, met een standaarddeviatie van 3.61. Gezien de oplopendheid van de schaal kan een gemiddelde score van 15 punten als volgt worden geïnterpreteerd: gemiddeld zijn de leerkrachten tijdens hun lessen Engels in staat om een gunstig klasklimaat te creëren, om hun les goed te organiseren, hun instructie helder over te brengen en leerlingen (vrij goed) te activeren. De complexe vaardigheden ‘afstemming op verschillen’ en ‘het aanleren van leerstrategieën’ zijn in de geobserveerde Engelse lessen niet of nauwelijks terug gezien. Dit is zichtbaar in Tabel 10.3, waarin, naast de totale somscore, tevens de somscores voor de onderliggende dimensies staan weergegeven. Overigens verschilden leerkrachten wel degelijk van elkaar op het gebied van hun generieke pedagogisch-didactische vaardigheid: ICALT-scores varieerden tussen 8 en 19 punten.

Het beeld dat uit de ongewogen resultaten van geobserveerde lessen naar voren komt stemt overeen met de bevinding uit het hoofdonderzoek dat leerkrachten weinig differentiëren in hun lessen Engels: gevraagd naar acties ten behoeve van achterblijvende leerlingen gaf in het hoofdonderzoek slechts 4% van de ondervraagde leerkrachten aan structureel extra ondersteuning aan te bieden, een kwart van de leerkrachten zei dit incidenteel te doen en meer dan twee derde van de leerkrachten (69%) gaf aan dit nooit te doen.

Tabel 10.3: ICALT-resultaten geobserveerde Engelse lessen.

	N	Gemiddelde	S.D.	Minimum	Maximum	Max haalbare score
ICALT-score (totaal)	22	15.05	3.61	8.00	19.00	24
Klimaat	22	3.77	0.69	1.00	4.00	4
Organisatie	22	4.05	1.33	1.00	5.00	5
Instructie	22	3.86	1.49	1.00	5.00	5
Activeren	22	2.14	0.89	0.00	3.00	3
Afstemmen	22	0.77	0.75	0.00	3.00	4
Leerstrategieën	22	0.41	0.73	0.00	2.00	3

Het vakdidactische deel van het observatie-instrument bevatte uiteindelijk 15 gedichotomiseerde vakdidactische leerkracht-activiteiten.<sup>100</sup> In Tabel 10.5 staan de frequenties weergegeven van de activiteiten die geobserveerd zijn in de lessen, eventueel uitgesplitst naar de subitems. Hiermee wordt een overzicht van de aangetroffen activiteiten gegeven. Een nadere beschrijving, waar mogelijk met voorbeelden, van de geobserveerde activiteiten in de les wordt in onderstaande paragrafen gekoppeld aan de interviewdata (vanaf paragraaf 10.8.2).

Verondersteld wordt dat de 15 (vak)didactische activiteiten van leerkrachten bevorderlijk zijn voor de vreemde taalverwerving van leerlingen. Een indicatie van de totale inzet van effectieve elementen in de

<sup>100</sup> Oorspronkelijk bevatte de lijst ook een item (item 17) waarmee werd geturfd aan welke vaardigheden of competenties de leerkracht aandacht besteedde. Deze telling is niet meegenomen in het vakdidactisch overzicht in tabel 10.5.



geobserveerde lessen kan worden weergegeven door een somscore te maken van de 15 leerkrachtactiviteiten, waarbij elke activiteit een score van 0 of 1 krijgt. Tabel 10.4 laat het gemiddelde en de spreiding van die somscore zien. In de geobserveerde lessen bleken de leerkrachten gemiddeld bijna tien effectieve elementen toe te passen, hoewel de variatie aanzienlijk was, variërend van drie tot veertien effectieve elementen (van de maximaal 15 elementen).

Tabel 10.4: Resultaten voor de vakdidactische scoring van de lessen Engels

	N	Gemiddelde	S.D.	Minimum	P25	P50	P75	Maximum
Vakdidactische score	22	9.73	2.85	3	7.75	9.5	12	14

Omdat de items effectief vakdidactisch gedrag weergeven zou de verwachting zijn dat een hogere score een meer succesvolle les weergeeft dan een lagere. Echter, een nuancering is op zijn plaats. Omdat de dichotome scores enkel weergeven of bepaalde activiteiten al dan niet zijn geobserveerd, geven zij geen indicatie van de kwaliteit waarmee een activiteit is uitgevoerd. Neem bijvoorbeeld het item 'geven van positieve feedback'. Deze leerkrachtactiviteit is positief gescoord (score = '1') op het moment dat de leerkracht positieve feedback aan de leerling gaf, ongeacht of deze feedback gericht is op bevordering van het leerproces (zoals bij feedback op het niveau van de taak, proces en zelfregulatie) of niet (feedback op het niveau van de persoon, zoals bij complimenten als 'you did it!', 'well done') (Hattie & Timperley, 2008). De impact van het type feedback op het leerproces is verschillend, maar dit onderscheid wordt in de gebruikte vragenlijst niet gemaakt. Een ander voorbeeld is het gebruik van taaldoelen. Alom wordt verondersteld dat de bespreking van taaldoelen aan relevantie wint door de leerlingen uit te leggen waarom deze doelen relevant zijn. In het huidige instrument wordt echter al een positieve score gegeven voor het benoemen van de taaldoelen, ongeacht of ze gemotiveerd worden – een kwalitatief verschil dat niet tot uiting komt in het instrument en de totaalscore. Omdat er dus geen indicatie is van de kwalitatieve invulling van de activiteiten, is de conclusie dat een hogere vakdidactische score automatisch staat voor een effectievere les niet juist. De score geeft op zijn hoogst een indicatie en de resultaten dienen zeer voorzichtig te worden geïnterpreteerd.

### Globaal beeld observatoren

Na afloop van de schoolbezoeken is een focusinterview gehouden met vijf van de zeven inspecteurs die de observaties en interviews hebben uitgevoerd. Doel van dit interview was om te komen tot een algeheel beeld van de ervaringen tijdens de schoolbezoeken. Daarnaast werd beoogd de positieve aspecten en eventuele ontwikkelpunten in kaart te brengen. Hoewel de scholen waren geselecteerd op bovengemiddelde prestaties op Engels, vonden de inspecteurs de lessen van wisselende kwaliteit.

Bij de meeste leerkrachten werd een groot enthousiasme voor en sterke betrokkenheid bij het geven van Engels opgemerkt, hetgeen in de lessen zichtbaar was en de leerlingen ook daadwerkelijk leek te motiveren. De lessen werden door de inspecteur over het algemeen als activerend en afwisselend gekenmerkt: het merendeel van de lessen bevatte verschillende activiteiten en werkvormen, waardoor de lessen verre van saai waren. Ook speelse elementen, zoals wedstrijdjes en spelletjes, droegen bij aan de motivatie en activering van de leerlingen. Wat bijdroeg aan de positieve ervaring was de beheersing van het Engels van de leerkrachten. Een deel van de leerkrachten sprak veel en zeer goed Engels in de les. Alhoewel beoordeling van de les of school in geen enkel opzicht doel was van het onderzoek, spraken de observatoren vooral in positieve zin over de scholen die het vak Engels op een meer doelgerichte en geïntegreerde manier leken aan te bieden en de scholen waar sprake is van een zekere schoolvisie op het vak. Een dergelijke schoolvisie leidde tot specifiekere keuzes omtrent zaken als professionalisering van leerkrachten, een



doorgaande leerlijn en/of de vakdidactische inhoud van de lessen Engels. En deze keuzes leken zich weer te vertalen in meer weloverwogen inhoud en structuur van de geobserveerde lessen.

Eventuele ontwikkel- en aandachtspunten die de observatoren constateerden gaan in eerste instantie over het beheersingsniveau van de taal door de leerkracht: enkele leerkrachten beheersten het Engels op een basaal niveau of spraken (bijna) geen Engels in de les. Als mogelijke belemmering om het eigen niveau goed in te kunnen schatten werd door de inspecteurs geopperd dat de leerkrachten geen duidelijke maatstaf voor het gewenste niveau hadden. De opleiding leek hen in dit opzicht niet te hebben gefaciliteerd. Ander terugkerend aandachtspunt was het opmerkelijke gebrek aan differentiatie, terwijl de verschillen tussen leerlingen in de geobserveerde klas evident waren. Ook in geobserveerde combinatieklassen maakten de leerkrachten zowel wat betreft beoogde doelen als in de uitvoering geen onderscheid tussen leerlingen van verschillende leerjaren en/of niveaus. Voor de observatoren is het dan de vraag of en hoe doorgaande leerlijnen worden geïmplementeerd. Een derde aandachtspunt was dat de nadruk op en het aanbod van Engels sterk afhankelijk leek van de leerkracht. De bezochte leerkrachten leken de meest gemotiveerde leerkrachten op de school te zijn, terwijl vervolgens in de interviews bleek dat er op de meerderheid van de scholen geen sprake was van een duidelijke doorgaande leerlijn, een focus op Engels of de implementatie van een specifieke visie op Engels over meerdere leerjaren heen.

Samenvattend stellen de observatoren in het focusinterview dat visie, doelgerichtheid en cyclisch werken op het gebied van Engels de kwaliteit van de lessen lijkt te bevorderen, maar dat dat op een beperkt deel van door hen bezochte scholen zichtbaar wordt: “weliswaar zijn bijna alle lessen leuk, energiek en afwisselend, maar ze missen vaak een kop en een staart”

Tabel 10.5: frequenties van geobserveerde leerkrachtactiviteiten in de les

Item	Aantal lessen	Subitem (in oorspronkelijke observatielijst)	Aantal lessen
1 Taaldoelen	11	Benoemt doelen	11
		Motiveert doelen	2
		SMART doelen	7
2 Voorkennis	18	Terugblik vorige les	15
		Aanwezige kennis onderwerp	9
3 Controle op begrip	22	Controlerende vragen	18
		Leerlingen geven verbaal antwoord	20
		Gebruik kaartjes	0
		Wisbordjes	2
		Digitale leermiddelen	3
		Anders	3
4 Doeltaal (in interactie met/tussen leerlingen)	13	Leerkracht-leerling interactie in Engels	10
		Leerlingen spreken onderling Engels	12
		Actie als leerlingen NL spreken	6
5 Aanpassing taalgebruik	16	Tempo verlagen	8
		Gebruik gebaren of lichaamstaal	10
		Visuele ondersteuning (plaatje, voorwerpen)	13
		Herhaling in andere bewoordingen	12
6 Werkvormen en activiteiten voor communicatieve doeleinden	17		
7 Aandacht voor taalfouten (in output leerlingen)	19	Focus op veelgemaakte fouten	5
		Correctie tijdens spreken	5
		Correctie na spreken	11
		Herhaling met correcte bewoording (niet expliciet)	12
		Uitleg grammaticaregel	5
		Leerlingen zichzelf laten herstellen (mbv vraag)	5
		Andere leerling vragen om correcte bewoording	14
8 Positieve feedback	17		
9 Differentiatie	1	Aanpassing leerstof op niveau	1
		Aanpassing leerstof op interesse/leerprofiel	0
10 Stimuleren gebruik compenserende strategieën door leerlingen	14	Omschrijven woord	8
		Synoniemen	6
		Gebaren	7
		Gebruik context	4
11 Laten toepassen gegeven regel/woorden	18		
12 Samenwerking	19	Zelfstandig werken	11
		Tweetallen	14
		Groepjes	10
		Klassikaal	21
13 Aansluiting werkvorm – taaldoelen	16		
14 Evaluatie taaldoelen	4		
15 Leerlingen worden gemotiveerd voor Engels	9		

### 10.8.2 Methode versus eigen materialen

In het hoofdonderzoek is leerkrachten gevraagd naar het gebruik van methoden voor Engels. De ongewogen resultaten laten zien dat de helft van de leerkrachten (52%) gaf aan dat zij uitsluitend de methode gebruiken. De overige leerkrachten maakten (ook) gebruik van andere leermiddelen. De methoden "Take it easy", "Groove me" en "Step to" worden het meest gebruikt, op respectievelijk 34%, 25% en 8% van de aan de peiling deelnemende scholen. Over de gebruikte methoden oordeelden de leerkrachten over het algemeen positief.

Op de bezochte scholen binnen het verdiepende onderzoek wordt, net als op de scholen in het hoofdonderzoek, op een derde van de scholen de methode *Take it Easy* gebruikt (op 4 eibo- en op 3 vvtto-scholen), terwijl *Groove me* op vier scholen (19%) wordt ingezet (waarvan 1 vvtto-school) en *Step to* op één vvtto-school. Op de overige negen scholen wordt met acht verschillende methoden gewerkt; de negende school werkt helemaal niet volgens een methode.

Tijdens het interview is de leerkrachten gevraagd naar het gebruik van de methode dan wel eigen materialen. Bij veertien leerkrachten (67%) is de methode leidend in hun lessen. Conform de verwachting zijn de leerkrachten op eibo-scholen hierbij oververtegenwoordigd (negen van de 10 geïnterviewde eibo-leerkrachten versus 5 van de 11 geïnterviewde vvtto-leerkrachten). De leerkrachten die iets over hun methode zeggen zijn over het algemeen tamelijk positief. Methodes worden motiverend genoemd, doordat zij interactie bevorderen, aansluiten bij interesses van leerlingen en leerlingen activeren, bijvoorbeeld door de inzet van liedjes. Niet alle geïnterviewde leerkrachten noemen de methode aansprekend. Daarom zetten sommige leerkrachten bewust ook andere materialen en werkwijzen in, zoals de toevoeging van coöperatieve werkvormen of spelvormen. Een leerkracht geeft aan vaak aan te haken bij toevalligheden of lastige onderwerpen: "Soms ontpopt zich een heel leuk gesprek en hebben ze ineens heel veel belangstelling voor iets, en daar haak ik dan wel op in ... En als ik merk dat ze iets heel leuk vinden, of heel lastig, ..., dan stop ik zoiets ook in de weektaak. Dan krijgen ze een extra dingetje voor Engels."

Een leerkracht die normaliter met een andere methode werkt, gebruikt ook wel eens aansprekende liedjes uit *Groove me*: "Af en toe pak ik nog even een liedje van een andere methode, *Groove Me*. Dat is puur voor de fun, voor de afwisseling." Ook wordt gebruik gemaakt van thematisch/ projectonderwijs. Een vvtto-school organiseert de eerder genoemde Schotlandreis, waar diverse voorbereidings- en kennisactiviteiten aan gekoppeld worden. Het thema Engeland wordt genoemd, waarvoor leerlingen een poster mogen maken, maar ook thema's die dichtbij de belevingswereld liggen, zoals 'school', 'sports' of 'Harry Potter'. Overigens kan de aandacht die wordt besteed aan projectonderwijs afhankelijk zijn van het niveau van de groep en niet standaard tot het programma behoren. Het merendeel van de leerkrachten gebruikt spelletjes, die leerlingen in het Engels spelen. Te denken valt aan (taal)kwartet, 'I spy with my little eye' [Ik zie, ik zie, wat jij niet ziet], '30 seconds', zelfgemaakte rollenspellen, 'wie ben ik' en Ganzenbord.

Ook wordt gebruik gemaakt van 'visual aids' (bijvoorbeeld via websites als [www.Twinkle.co.uk](http://www.Twinkle.co.uk) –, op het bord, eigen tekeningen en knutselwerkjes) om productie van het Engels bij leerlingen te stimuleren. Een leerkracht beschrijft hoe hij/zij dergelijke visual aids gebruikte bij een storyline project over gefingeerde families, waarvoor de leerlingen de karakters zelf verzonnen. Met name dat laatste aspect motiveerde de leerlingen om hun fantasie te gebruiken. Leerlingen vonden het leuk om hun familie te mogen knutselen en de opdracht zette de leerlingen volgens de leerkracht aan tot het spreken van Engels en het herhalen van de woorden die ze bij het thema in de methode hadden geleerd. De opdracht en een afsluitend filmpje voor de ouders waarin de leerling vertelde over zijn/haar gefingeerde familie, waren een groot succes. Overigens wijst deze leerkracht wel op het gevaar dat leerlingen tijdens het knutselen Nederlands spreken met elkaar. De leerkracht maakte hier zelf geen punt van, want het verwoorden van je verhaal over de gefingeerde familie in het Nederlands zou de leerlingen helpen wanneer ze het later in het Engels moesten verwoorden.

Twee leerkrachten benoemen expliciet dat ze leerlingen veel Engels laten lezen – te denken valt aan een Engelse bibliotheek met boekjes, stripboeken of simpele tekstjes. Hun ervaring is dat leerlingen dit aansprekend vinden. Het gebruik van teksten in de les is in dertien lessen geobserveerd. Eén leerkracht maakte daarbij gebruik van het lezen van een ‘de schriftelijke uitwerking’ behorend bij een filmpje dat de leerlingen zagen/hoorden, waarna leerlingen de opdracht kregen om zelf mensen te gaan beschrijven (schrijfopdracht). Dit werd gevolgd door mondelinge beschrijvingen van leerlingen in de klas met behulp van het spel ‘Wie is het?’).

### 10.8.3 Taaldoelen, monitoring en evaluatie

Voor een weloverwogen aanbod van Engels is het van belang langere termijn en korte termijn-taaldoelen vast te stellen. Op deze manier wordt de opbouw van de taaldoelen binnen het curriculum geëxpliciteerd en ontstaat er een helder beeld van de vaardigheden waar leerlingen zich in zouden moeten ontwikkelen. Evaluatie van de doelen met behulp van formele/informele toetsing geeft zicht op de vraag of en in welke mate deze ontwikkeling ook daadwerkelijk heeft plaatsgevonden. Het werken met doelen en evaluatie van deze doelen op lesniveau en op de langere termijn geeft een cyclische manier van werken weer, waarmee een school inzicht in de opbrengsten van haar aanbod en inspanningen voor Engels kan verkrijgen.

Er is gevraagd naar de lange termijn-doelen en naar het gebruik van taaldoelen per les. Op de meeste scholen zijn de methodedoelen leidend, maar leerkrachten noemen ook andere doelen. In totaal noemen 11 leerkrachten (7 eibo, 4 vvto) de methodedoelen leidend voor hun lessen, ook al werken twee leerkrachten via een eigen invulling van de methode aan deze doelen: “Ik vind de methode heel fijn. Maar ik vind het leuker en interessanter om hem eraan te leggen. Dus dat ik wel weet welke kan ik op moet gaan, maar hoe ik dat ga doen, dat vind ik wel leuk om dat zelf uit te zoeken. Dan ben je ook veel ontspanner voor de klas.” Het uitgaan van methodedoelen geldt voornamelijk voor de eibo leerkrachten en in mindere mate voor de vvto-leerkrachten. De vvto-leerkrachten benoemen daarbij de aansluiting tussen methode en kerndoelen. Vier leerkrachten, waaronder één eibo-leerkracht, zeggen te werken aan de kerndoelen, waarbij één van de vvto-leerkrachten deze kerndoelen gebruikt in combinatie met de SLO-tussendoelen. Twee andere vvto-leerkrachten noemen voor het opstellen van leerdoelen deze kerndoelen juist te vaag, waardoor zij op zoek zijn gegaan naar andere doelen voor hun Engelse lessen: de ene leerkracht is in gesprek gegaan met het VO (een school waarop Cambridge English wordt aangeboden) om zicht te krijgen op wat voor leerlingen relevant is om te leren en de andere leerkracht gaat uit van het Anglia-curriculum, waarbij in een Teacher’s guide behorend bij het eigen materiaal per les leerdoelen zijn geformuleerd. De Anglia-doelen worden in totaal door twee leerkrachten genoemd. Twee leerkrachten noemen ERK-doelen. Eén leerkracht in combinatie met kerndoelen, maar zonder uitleg. De andere leerkracht beschrijft hoe er in samenwerking tussen PO en VO wordt gewerkt aan ERK-doelen, vanuit de *Can do*-beschrijvingen (Zie kader 1). Drie leerkrachten geven aan dat het werken aan doelen beter kan: ze zijn niet altijd vooraf vastgesteld of wat losgelaten doordat niet (meer) met een methode wordt gewerkt.

Niet alle leerkrachten hebben aangegeven of ze aan het begin van de les hun lesdoelen expliciet bespreken. De lesbezoeken laten zien dat in de helft van lessen (50%) is geobserveerd dat de leerkracht het taaldoel benoemt en in zeven lessen (32%) waren deze doelen SMART geformuleerd. Motivering van lesdoelen gebeurde echter weinig: slechts in 9% (twee lessen) maakte de leerkracht duidelijk waarom de taaldoelen relevant zijn. De manier waarop de leerkracht dit deed kan op basis van de aanwezige data niet worden achterhaald. De aansluiting bij de voorkennis – bijvoorbeeld met behulp van een mindmap over het onderwerp - of het terugblikken op de vorige les wordt door achttien leerkrachten uitgevoerd.

Een cyclische manier van werken, waarin het behalen van de doelen centraal staat, vraagt om voldoende evaluatie. Van de negen leerkrachten die vertellen over of en hoe zij het lesdoel evalueren ontstaat een wisselend beeld. Drie vvto-leerkrachten zeggen dit niet te doen, bijvoorbeeld omdat het er bij in schiet: “Ik wil iets teveel in zo’n les proppen [...] Want ik gun het ze zo, ik vind altijd zoveel leuke dingen, en dan denk ik: jeetje, je hebt maar drie kwartier.” Drie andere leerkrachten evalueren dit doel door tussentijdse observaties, bijvoorbeeld door na te gaan of leerlingen in hun gesprekje de doelwoorden gebruiken. Twee eibo-leerkrachten zeggen aan het eind na te gaan of leerlingen het doel hebben gehaald en voor één leerkracht is actieve deelname aan de les de maatstaf. In het merendeel van de geobserveerde lessen werd bij de afsluiting niet nagegaan of de doelen bereikt waren: in slechts 18% van de lessen werden de taaldoelen geëvalueerd.

Op zestien (7 eibo, 9 vvto) van de 21 scholen zeggen leerkrachten het niveau van de leerlingen met behulp van toetsen vast te stellen, maar de manier van toetsing is verschillend. De helft van de leerkrachten (5 vvto en 5 eibo) werkt met methodetoetsen. Evenwel varieert de frequentie van het toetsen: van één toets per blok van 3 weken tot maximaal twee methodetoetsen per jaar. Ook eigengemaakte toetsen worden ingezet (eventueel in combinatie met de methodetoetsen). Deze zijn met name gericht op woordenschat – evenals de methodetoetsen. Dit beeld is grotendeels ook terug te vinden in de (ongewogen) resultaten in het hoofdonderzoek: van de aan de peiling deelnemende leerkrachten zegt 59% toetsen uit de methode te gebruiken en woordenschat wordt verreweg het meest getoetst (68% van de leerkrachten gaf aan woordenschat het vaakst te toetsen).

De leerkracht op een school die volledig inzet op mondelinge vaardigheid evalueert haar aanbod met name door presentaties aan het eind van het blok (zie Kader 2). Bij het thema ‘jouw school’ wil de leerkracht aan eind van het thema dat de leerlingen in tweetallen een rondleiding in de school geven aan ouders/jongere leerlingen. Door na te gaan of ze de geleerde woorden gebruiken evalueert de leerkracht dan het geleerde. Wel zegt de leerkracht nog zoekend te zijn in deze opzet. Daarnaast gebruikt deze leerkracht woordenschattoetsjes, maar vooral om de leerlingen voor te bereiden op de werkwijze binnen het VO. Ook een andere leerkracht nuanceert het belang van dergelijke toetsresultaten, omdat het elkaar verstaan, het zich kunnen uiten en het praten met elkaar belangrijker wordt geacht.

In het vvto worden (ook) andere toetsen en toetsvormen ingezet: Anglia-toetsen (op 2 scholen), Me2! van Cito (1 school) en het gebruik van portfolio’s (op 2 scholen). Ook zegt een vvto-leerkracht te evalueren aan de hand van presentaties, boekverslagen en scrapbooks, waarin de leerlingen verslag doen over een uitwisseling. Op drie scholen, waaronder twee vvto-scholen, worden geen toetsen ingezet om het Engelse taalniveau te meten vanwege de pedagogische visie van de school; dit is een keuze op schoolniveau. Wel zeggen deze leerkrachten de ontwikkeling van leerlingen, met name op het gebied van mondelinge vaardigheid, door observaties vast te stellen. Op een andere school geeft de leerkracht aan dat de methodetoetsen te moeilijk, want te talig, zijn voor hun leerlingpopulatie, maar blijft het onduidelijk of het niveau al dan niet met toetsen wordt vastgesteld. Aanpassingen in het aanbod op basis van evaluatieresultaten worden slechts door twee leerkrachten benoemd. Eén van deze leerkrachten expliciteert dit: “Een onderdeel wordt getoetst en daar pik je dan uit welke kinderen daar nog moeite mee hebben en die krijgen daar dan nog extra uitleg over.”

## Kader 2: focus op mondelinge taalvaardigheid

Op een eibo-school is op teamniveau besloten de gebruikelijke methode los te laten en het aanbod met name te richten op mondelinge taalvaardigheid, omdat men dit het belangrijkste doel vindt voor leerlingen in het basisonderwijs. Specifiek gaat het volgens de leerkracht om: “[het] zich vrij voelen om in het Engels te praten (te durven), om een basiswoordenschat van Engels [te] hebben waarmee ze zich zouden kunnen redden (zich voorstellen, simpele zinnestelsels gebruiken, een gesprekje kunnen voeren, boodschappen doen).” Woordenschat is daarbij erg belangrijk; lezen, schrijven en grammatica komen wel aan bod maar zijn minder belangrijk.

Voorheen besteedde de school aandacht aan alle deelvaardigheden volgens de methode, maar dit bleek veel te veel, voor zowel de leerlingen als de leerkrachten zelf. Enkele leerkrachten gaven aan niet alle deeltaken te overzien en uit te voeren, waardoor er soms helemaal niets gebeurde. Gezamenlijk hebben de leerkrachten toen afgewogen wat ze echt belangrijk vonden: “Willen we dat ze de grammatica [uit de methode] al helemaal perfect beheersen, terwijl ze dat ook in de eerste klas van de middelbare school weer krijgen of willen we gewoon dat ze zich goed kunnen uitdrukken en dat ze durven spreken?”

Door uit te gaan van de SLO- leerlijn ‘mondelinge taalvaardigheid’, waarin woordenschat een deelvaardigheid is, wordt sinds het huidige schooljaar de visie geïmplementeerd. De leerlijn gaat uit van thema’s voor elk leerjaar en er zijn ongeveer 5 thema’s van gemiddeld acht weken per jaar. Met behulp van de leerlijn werkt de school systematisch aan de einddoelen van de leerlijn. Om in de toekomst deze doelen te bereiken ‘transformeert’ de school geleidelijk naar een vvto-school. Juist door het systematisch uitvoeren van de leerlijn door de hele school verwachten de leerkrachten dat jongere leerlingen er sterk van zullen profiteren en later, in groep 7-8, zich mondeling beter kunnen uitdrukken dan de huidige lichten.

Alle groepen ontvangen in totaal een uur Engels per week, verspreid over drie blokken van 20 minuten per week. De school heeft de indruk dat met deze tijdsindeling meer wordt bereikt dan met één lesuur per week volgens de methode. De tijd is systematisch ingepland in de week, in alle klassen.

De thema’s zijn vastgelegd in de jaarkalender van de drie combinatiegroepen (1-2-3, 4-5-6, 7-8). Iedere leerkracht werkt vervolgens de thema’s afzonderlijk voor zijn/haar groep uit, aan de hand van 20 basiswoorden die de leerkracht zelf kiest. De basiswoorden worden via verschillende activiteiten en spelletjes aangeleerd. Een voorbeeld is het thema sport. In les één wordt er bijvoorbeeld in de klas gepraat over een sportwedstrijd. In de volgende les moeten leerlingen dan in het Engels verslag uitbrengen van de sport die ze zelf in het weekend beoefend hebben.

De leerlijn kent een teachers’ guide en voor verschillende lessen zijn losse materialen, zoals flash cards, beschikbaar. Bij de uitvoering van het lesprogramma wordt les 1 in de week altijd uitgevoerd door de vakleerkracht en les 2 en 3 door de groepsleerkracht. Elk thema wordt afgesloten met een project (in alle klassen), bijvoorbeeld door leerlingen iets te laten presenteren over ‘planeten’, of zoals het voorbeeld in Tabel 10.6: door een Engelstalige rondleiding door de school te geven. Dit is tevens het moment om te evalueren of leerlingen de leerdoelen hebben bereikt. Deze werkwijze wordt in de gehele school, vanaf de kleuters met simpele opdrachten, uitgevoerd.

In Tabel 10.6 is een samenvatting van de SLO leerlijn voor groep 7/8 opgenomen, zoals gebruikt op deze school. Hier gaat het om het subdomein woordenschat van de leerlijn mondelinge taalvaardigheid. De leerlijn kent verder doelstellingen en een beschreven didactische aanpak voor de subdomeinen ‘spreken’, ‘luisteren’ en ‘gesprekjes voeren’. Alle vier subdomeinen komen terug in de thema’s. De tabel bevat een ingekorte uitwerking van de lessen behorend bij het thema ‘My school’, als ook enkele woorden die leerlingen binnen dat thema moeten kennen en gebruiken. De SLO leerlijn is voor iedere school beschikbaar, zie: <http://tule.slo.nl/Engels/F-L14.html>

Tabel 10.6 (bij kader 2): Samenvatting van de leerlijn mondelinge taalvaardigheid voor groep 7/8.

Subdomein 1.1 woordenschat en woordgebruik		
Resultaat eind groep 8:	Leerling beschikt over voldoende woorden om zich te kunnen uiten. Het kan soms nog nodig zijn een omschrijving van een onbekend woord te geven, Ook is de leerling gefocust op correct woordgebruik en uitbreiding van woordenschat	
	Passief:	Woordenschat van gem. 8500 woorden
	Actief:	Actieve productie van gem. 1000 woorden
	Betekenisuitbreiding:	+ Begrijpt spreekwoorden, uitdrukkingen, beeldspraak + kan bevestigende en ontkennende taaluitingen geven + kan variëren in woordgebruik
Thema's:	1. Myschool	Rondleiding geven door de school
	2. Me in the future	Beroepen. Opleiding
	3. Health	Gezonde leefstijl, sporten, eten, slapen
	4. Restaurant	Restaurant opzetten, bestellen, afrekenen
	5. Travel	Reisbureau, landen, route bepalen
Lessen 'My school':	Week 1 (les 1-3)	Voorkennis activeren (mindmap) / plattegrond klas met Eng. woorden / Duolingo activeren
	Week 4 (les 10-12)	Alfabet leren (ABC-rap)/Galge met themawoorden / 'wie ben ik' met themawoorden
	Week 8 (les 18-20)	Rondleiding geven door de school
Themawoorden 'My school'	o.a. teacher, student, primary education, language, homework, pencil, eraser, desk, blackboard, calculator, notebook	

#### 10.8.4 Aanbod verschillende vaardigheden

Leerkrachten is gevraagd aan welke (deel)vaardigheid zij de meeste aandacht besteden. Bijna alle leerkrachten (n=20) geven aan de meeste aandacht te besteden aan mondelinge vaardigheden, zoals (monologisch) spreken en gesprekken. In het kader van spreken worden uitspraak oefeningen (bijvoorbeeld door het zingen van liedjes), spreekbeurtjes en presentaties genoemd en voor gesprekken worden voorgeschreven en meer vrij in te vullen gesprekjes vermeld. Ook interactieve spelletjes als 'binnen-buitenkring', 'Wie ben ik?', 'Wat zijn de dagen van de week', kwartetten, bordspelletjes of '*I spy with my little eye*' [Ik zie, ik zie, wat jij niet ziet] worden genoemd. Ongeveer een derde van de leerkrachten geeft overigens aan dat ze mondelinge taalvaardigheid een moeilijke vaardigheid vinden om te bevorderen. Deze leerkrachten wijzen dan vooral op de onzekerheid van de leerlingen en dat het lastig is om ze aan het praten te krijgen, zoals ook al naar voren kwam in de beschrijving van de leerkrachtvisie. Manieren die door leerkrachten worden ingezet om leerlingen te prikkelen en interactie te bevorderen zijn bijvoorbeeld het eerdergenoemde storyline project over familie. Afsluitende activiteiten, zoals het maken van filmpjes werken daarbij motiverend. Dit blijkt ook uit de ervaringen van de leerkracht in Kader 2. Een andere leerkracht betreft leerlingen ook bij het bepalen van doelen voor mondelinge vaardigheid door te vragen wat ze straks op de camping willen kunnen en dat in een dubbele kring te oefenen. Leerlingen vertelden elkaar in het Engels wie ze zijn, wat ze leuk vinden om te doen, etc. Een aantal leerkrachten laat leerlingen door de klas lopen om aan wisselende gesprekspartners iets te vertellen, bijvoorbeeld wie je bent. Dit zorgt voor een ongedwongen sfeer en voor actieve deelname van alle leerlingen. Naast de moeilijkheid om leerlingen aan het praten te krijgen worden de grote verschillen tussen leerlingen genoemd als reden



waarom aandacht voor mondelinge vaardigheid lastig uitvoerbaar is in de klas: “Het niveauverschil is groter dan bij welk vak dan ook. Sommigen zijn haast al vloeiend, [leerlingen] die veel gamen en gewoon ook aanleg ervoor hebben. Anderen hebben een heel beperkt [bereik]—die kunnen misschien de kleuren en tot twintig tellen, maar verder komen ze gewoon niet.”

Evenals bovenstaande leerkracht koppelt twee derde van de leerkrachten de aandacht voor mondelinge vaardigheid aan aandacht voor andere vaardigheden: vaak is er sprake van een combinatie van vaardigheden waar de nadruk op ligt. De meest voorkomende combinaties zijn mondelinge vaardigheid en luistervaardigheid (3 leerkrachten), woordenschat (5 leerkrachten) of beide (6 leerkrachten). Enkele leerkrachten noemen dat de methode (*Groove me*) luistervaardigheid en woordenschat op een aantrekkelijke manier bevordert. Door bij songteksten stil te staan bij de woorden die leerlingen al kennen en bij de inhoud van de liedjes ontwikkelen beide (sub)vaardigheden. Ook worden filmpjes, spelletjes en verhalen genoemd. Vergroting van de Woordenschat, als voorwaarde voor mondelinge vaardigheid, wordt als belangrijk aandachtspunt voor het basisonderwijs genoemd: “Als je een start kunt maken met praten, als je dat durft omdat je een basiswoordenschat hebt zeg maar, en dat je de angst hebt overwonnen om te praten, denk ik dat je een heel mooi begin hebt en dan kun je [dat] dus verder heel mooi uitbouwen. En het is de bedoeling dat je in het voortgezet onderwijs nog meer je woordenschat uitbreidt en dan kunnen ze daar weer mee verder.” Bij woordenschat worden ook het aanleren van ‘chunks’ genoemd, de aandacht voor vaste uitdrukkingen.

Aan de vaardigheden lezen, spelling en grammatica wordt volgens de leerkrachten beduidend minder aandacht besteed. Slechts twee leerkrachten noemen lezen als belangrijkste vaardigheid. Eén van beide leerkrachten richt zelfs één van de twee lessen in de week op het lezen van boeken en teksten. Over de rol die grammatica in het aanbod zou moeten spelen wordt verschillend gedacht, met name binnen de groep eibo-leerkrachten. Vier eibo-leerkrachten zijn van mening dat grammatica in het PO niet erg belangrijk is en dat dat een taak is voor het VO, terwijl drie eibo-leerkrachten juist aangeven bewust meer aandacht aan grammatica te besteden dan de methode. Genoemde reden is de voorbereiding op wat er van leerlingen gevraagd wordt, hetzij van de havo-vwo leerlingen in het VO, hetzij in de methodetoets – zoals een leerkracht het verwoordt: “In de methode wordt [grammatica] wel lichtelijk weggelaten en dan komt er aan het eind: ‘Deze zinnen moet je onthouden.’ Nou, dat gaan we dus niet aan het eind doen. [...] Die haal ik naar voren. [...] Ik maak ze even wat belangrijker met de wetenschap dat ze straks wel in een toets worden gevraagd. [...] Elke keer komt daar weer grammatica naar voren, dus ik weet dat ik hem prominenter moet maken .... Dat ze een bepaalde volgorde kunnen onthouden als het gevoelsmatig niet meer lukt.” Van de vto-leerkrachten spreken vijf leerkrachten zich expliciet uit over de aandacht voor grammatica in hun lessen. Grammatica zou volgens hen geen prominente positie in moeten nemen. Twee van hen vinden het wel belangrijk aandacht aan deze basis te besteden, maar ‘zoals het komt’, niet vanuit expliciete doelen.

### **Aandacht voor de verschillende (sub)vaardigheden**

De nadruk die leerkrachten zeggen te leggen op *mondelinge vaardigheid* is deels teruggevonden in de geobserveerde lessen. De aandacht voor monologisch spreken, gespreksvaardigheid en uitspraak is geobserveerd. In het merendeel van de lessen (77%) is aandacht besteed aan de uitspraak van het Engels, terwijl de aandacht voor spreken en gesprekken lager was, in respectievelijk 55% en 27% van de lessen werd aan deze vaardigheid aandacht besteed. Aandacht voor *spreekvaardigheid* was daarbij gedefinieerd als het laten oefenen met een korte monoloog, waarbij de leerlingen de tekst niet van te voren hadden uitgeschreven, maar hooguit een paar steekwoorden konden gebruiken en ter plekke zelf de uitingen moesten formuleren. Leerlingen voorwerpen of andere leerlingen laten beschrijven, het houden van een



korte presentatie door groepje leerlingen over één van de seizoenen en het vertellen over een poster zijn hiervan geobserveerde voorbeelden. Als voorbeelden van aandacht voor *gespreksvaardigheid* werden de volgende activiteiten geobserveerd: het leerlingen in het Engels laten associëren over een onderwerp (t.b.v. een presentatie), het in groepjes laten overleggen over een invuloefening en een vrije invulling van een rollenspel of gesprekje. Bij gespreksvaardigheid diende altijd sprake te zijn van ‘echte’ interactie tussen leerlingen onderling of tussen leerling en leerkracht, waarbij op elkaars uitingen werd gereageerd. Dit is derhalve niet hetzelfde als het voorlezen van een voorgestructureerd/geschreven gesprek. Deze vorm werd in enkele klassen teruggezien, maar is gecategoriseerd als ‘aandacht voor uitspraak’, omdat er geen eigen inbreng van de leerlingen is. Ook het voorlezen van of luisteren naar een tekst zonder in te gaan op de betekenis is gescoord als uitspraak. Ook het meezingen met karaokeliedjes (zonder aandacht voor betekenis van de tekst) viel daarom onder *uitspraak*. In ruim driekwart van de lessen is dat gezien. Andere voorbeelden waren het naspreken van woorden met aandacht voor de specifieke uitspraak (b.v. *mathematics*) of aandacht voor zowel de uitspraak als schrijfwijze van woorden (b.v. *knee, true/through, three, tree*).

De meeste aandacht is in de geobserveerde lessen besteed aan *woordenschat*: in 86% van de lessen is aandacht besteed aan aanleren, herhalen of verwerken van woorden en het aanleren van chunks (voor twee lessen specifiek genoemd). Zo kon een les om een woord (en bijbehorende woorden) draaien, werd er aandacht besteed aan het bevragen van de betekenis van woorden, b.v. aan de hand van plaatjes. Plaatjes werden ook gebruikt ter ondersteuning van de betekenis van nieuw aangeleerde woorden. Leerkrachten besteedden aandacht aan moeilijke woorden (door woorden die in een filmpje zijn benoemd bijvoorbeeld op het bord te schrijven), lieten leerlingen de woorden in hun schrift schrijven of in zinnen gebruiken. Ook werden onbekende woorden door de leerkracht in het Engels omschreven. In de geobserveerde CLIL-lessen besteedde de leerkracht aandacht aan vaktaal en in weer een andere klas liet de leerkracht groepjes leerlingen een wedstrijdje doen, waarin zij per groepje als eerste vier woorden moesten bedenken die te maken hadden met een bepaalde hobby, kleur, dieren of dagen van de week. Het competitie-element leek de leerlingen sterk te motiveren. Ook *luistervaardigheid* kwam in veel lessen aan bod (77%). Hiervoor was de vereiste dat er aandacht besteed werd aan de betekenis van het geluisterde. Dit werd gedaan aan de hand van teksten/verhalen, uitleg van de leerkracht, flimpjes, videoclips, audiofragmenten.

Waar leerkrachten in de interviews aangaven over het algemeen niet veel aandacht te besteden aan lezen en grammatica bleek dat 58% van de geobserveerde lessen *lezen* (en de betekenis) van zinnen of teksten en 45% van de lessen *grammaticale aandacht* bevatten. Leerkrachten besteedden aandacht aan werkwoordsvervoeging (van ‘to be’, ‘to have’), zinsbouw/samentrekkingen (‘let’s...; ‘we’ve...’) of woordsoorten (onderwerp, gezegde, voorzetsels, persoonlijke en bezittelijke voornaamwoorden). Aandacht voor *spelling* van woorden of zinnen was er in een kwart van de lessen (23%). Te denken valt bijvoorbeeld aan de schrijfwijze van tijden, van samentrekkingen (zoals ‘let’s’) en aandacht voor verschillen in uitspraak en schrijfwijze (zie hierboven). Aandacht voor het *schrijfvaardigheid* is maar beperkt geobserveerd, namelijk in drie lessen (14%). Dat is in lijn met de beperkte nadruk die leerkrachten er in de lessen aan zeggen te geven.

Tenslotte is nagegaan in hoeverre leerkrachten aandacht besteedden aan de Anglosaksische cultuur. Ook dit bleek slechts in drie lessen het geval te zijn, zoals wanneer eigen buitenland- ervaringen verteld werden of een filmpje over London werd getoond. In een derde van de lessen richtten de leerkrachten de aandacht van de leerlingen op verschillen tussen de Engelse en Nederlandse taal. Hierbij werden zowel verschillen tussen beide talen (bijvoorbeeld in betekenis bij tijdsaanduidingen - ‘half past 9’ -, in uitspraak of het persoonlijk voornaamwoord ‘you’) als overeenkomsten (zoals het woord ‘museum’) benadrukt. Eén leerkracht richtte zich tevens op verschillen tussen Brits en Amerikaans Engels.

Om een indicatie te geven van de hoeveelheid behandelde vaardigheden in de lessen, en daarmee van de invulling van deze lessen, is een gemiddelde berekend van het aantal in de lessen geobserveerde (sub)vaardigheden<sup>101</sup>. De aandacht voor cultuur en verschillen tussen de talen is daarin niet meegenomen. Gemiddeld besteedden leerkrachten aan bijna 5 vaardigheden aandacht (gemiddelde: 4.7, s.d.: 1.7; mediaan: 4), maar wederom lijken er vrij grote verschillen te zijn: de range loopt van 2 tot 8 vaardigheden per les. Benadrukt moet worden dat de interpretatie van dit gemiddelde onduidelijk is, omdat de hoeveelheid behandelde vaardigheden niets over de kwaliteit van die onderdelen zegt. Hiervoor is de kwalitatieve uitvoering van instructie en uitvoering relevanter. Wel indiceert het gemiddelde van 4.7 vaardigheden per les dat de lessen veel afwisseling bevatten. Dit gemiddelde en de beschrijvingen hierboven laten bovendien zien dat er in de meeste lessen aandacht is voor zowel input als output. Helaas laten de data geen conclusies toe over de vraag of dit ook leidde tot een weloverwogen opbouw (volgens het Vierfasen-model).

#### 10.8.5 Balans in werkvormen

Zoals in de theoretische onderbouwing voor het observatie-instrument is beschreven, vormen de taaltheoretische uitgangspunten van de Schijf van Vijf en de daaruit voortkomende vertaling naar de praktijk, het Vierfasen-model, relevante elementen van een les in een moderne vreemde taal. Een weloverwogen opbouw en afwisselende werkvormen kenmerken effectieve lessen. Veertien leerkrachten (6 eibo en 8 vvto-leerkrachten) hebben zich hierover uitgelaten en zeggen te streven naar voldoende afwisseling in werkvormen. Dit wordt gerelateerd aan de afwisselende werkvormen in de methode, maar er zijn ook leerkrachten die zelf extra afwisseling aanbieden, omdat a) de methode dit onvoldoende doet, b) omdat zij ook in alle andere lessen gebruik maken van coöperatieve werkvormen en dit ook aan de Engelse lessen toevoegen, bijvoorbeeld tijdens de invuloefeningen en verwerkingsopdrachten, en/of c) om de leerlingen te activeren, zoals deze leerkracht het uitdrukt: *Afwisseling van werkvormen doe ik zelf. Want anders zit je gewoon echt klassikaal en dan luister je, maar daar hou ik niet van. Dat doe ik met alle lessen. Veel afwisselende werkvormen streven we na, aansluitend op de visie van de school. We doen ook veel buiten. Dat zijn vormen die we ons hebben aangeleerd als school en die brengen we bij al die verschillende vakken [...] omdat de betrokkenheid dan beter, gevarieerder en leuker is.* Van de veertien leerkrachten benoemen slechts twee vvto-leerkrachten expliciet dat zij hebben nagedacht over de opbouw binnen en tussen lessen en de balans tussen de vaardigheden. Eén leerkracht noemt het Vierfasen-model als één van de pijlers in de lessen en houdt deze fasen in de lessen aan. Na een inputfase (waarin bijvoorbeeld aandacht is besteed aan een tekst) volgt altijd een outputfase: *“We proberen dus echt ook altijd heel erg het [geleerde] in gesprekjes te moeten toepassen of een stukje tekst te moeten schrijven, dat ze in ieder geval altijd wel iets actiefs ermee moeten doen.”*

---

<sup>101</sup> In kaart gebracht met item 17 uit de observatielijst.

### Kader 3. Onderbouwd aanbod & integratie vakdidactische modellen

Een vvtto-school zet in op didactisch weloverwogen, zelf samengestelde lesmaterialen en leerlijnen, waarin verschillende pedagogisch-didactische modellen voor vreemde talenonderwijs zijn opgenomen.

In de school zijn zowel een vakleerkracht als de groepsleerkracht verantwoordelijk voor het uitvoeren van de lessen Engels. De lesmaterialen, leerlijnen en didactische modellen zijn ontwikkeld door een vakleerkracht vanuit de gedachte dat a) lesmaterialen voor leerlingen aansprekend moeten zijn, b) er een thematisch en systematisch werkplan met einddoelen voor de leerlingen zou moeten zijn, en c) de groepsleerkrachten in de vvtto school ook voldoende kwaliteit hebben als het gaat om Engels. De (geïnterviewde) vakleerkracht begeleidt en traint de groepsleerkrachten in de school met het toepassen van de leerlijn en materialen.

Op deze school is, didactisch gezien, sprake van een gecombineerde aanpak die ook is opgenomen in de leerlijnen en het ontworpen lesmateriaal. In de onderbouw wordt vooral gebruik gemaakt van Total Physical Response (TPR), een taalmethodiek waarin het aanleren van taal gepaard gaat met het gebruik van gebaren. In de midden- en bovenbouw worden o.a. het Vierfasen-model en in enige mate ook CLIL toegepast. Daarnaast worden ook didactische activiteiten uit bestaande methodes gebruikt. In de lesmaterialen zitten bijvoorbeeld ook liedjes die op eenzelfde wijze als in de methode *Groove me* worden gebruikt. De liedjes zijn echter zorgvuldig gekozen vanwege het raakvlak met het thema van de les, of gewoon omdat het goed aanslaat bij de doelgroep (bijvoorbeeld groep 8).

Volgens de leerkracht zijn de standaard verkrijgbare methodes te eenzijdig opgebouwd en zijn deze te weinig gericht op het interactieve en communicatieve waar het Engels in het PO op gericht zou moeten zijn. Centraal in de aanpak staat daarom *communicative language teaching*: “Daarbij doe je heel veel spreekopdrachtjes waarbij leerlingen moeten samenwerken en er een *information gap* tussen leerlingen is. Dus de één heeft informatie die de ander niet heeft. Via bijvoorbeeld ‘wie is het?’, en dat kan in verschillende varianten, kan je als leerkracht het op het thema toepassen”. Andere opdrachten in deze lijn zijn tekenopdrachten en spelletjes: “*twee leerlingen hebben een blaadje met allerlei objecten voor zich dan moet de één de ander instructie geven over wat ze zien: colour, the school bag, blue and yellow.*”

De reden voor het gebruik van wisselende didactiek is vooral het voorkomen van herhaling en verveling. Omdat de werkwijze heel sterk is gericht op communicatie, is didactische variatie een goed middel om tot meer en betere communicatie tussen leerlingen te komen. Het geeft ook voor de leerlingen een meerwaarde dat ze in verschillende lessen verschillende activiteiten moeten uitvoeren. Daarbij, stelt de leerkracht, is het voordeel dat je ook minder leuke, meer formele zaken in de lesopbouw op een plezierige manier kan aanbieden. Bijvoorbeeld het toepassen van de *past simple* in een spelletje over het verleden.

De ervaringen met de werkwijze zijn voornamelijk positief: “*de ene leerkracht is enthousiaster dan de ander en vindt het relevanter dan de ander. Het uitgangspunt in de modellen en de lesmaterialen, dat leerlingen in ieder geval veel moeten spreken en dat moeten durven - het communicatieve - dat onderschrijft iedereen*”.

### 10.8.6 Doeltaal is voertaal

Het gebruik van de doeltaal als voertaal is in de interviews besproken vanuit twee perspectieven: het gebruik van Engels door de leerkracht en het gebruik ervan door de leerlingen zelf.

#### *Doeltaalgebruik door leerkrachten*

Het bewust gebruik maken van de doeltaal door de leerkracht is bevorderlijk voor het aanleren van de vreemde taal. Leerkrachten is gevraagd of en in welke mate zij het Engels in hun lessen gebruiken. Dertien leerkrachten (48% - 5 eibo en 5 vvto leerkrachten) zeggen zoveel mogelijk tot alleen maar Engels in de les te spreken, omdat ze het belangrijk vinden dat leerlingen de taal zoveel mogelijk horen: “Ik ben van mening dat als je Engelse les geeft dat je zoveel mogelijk in het Engels moet spreken, met veel aanwijzen. [De leerlingen] schrikken daar altijd verschrikkelijk van in het begin. De eerste les zitten ze echt van: ‘Huh, waar heeft ze het over?!’ Maar dit is les drie die ze kregen en het gaat eigenlijk al heel goed. Ze kunnen me in ieder geval volgen.” Drie leerkrachten noemen daarbij dat ze in de school of klas een teken hebben afgesproken, zodat het de leerlingen duidelijk is dat er van taal gewisseld wordt. In de lagere klassen worden hier een knuffel of bolhoedje voor gebruikt, in de hogere klassen wordt bijvoorbeeld afgeteld in het Engels en wordt daarna alleen Engels gesproken. Acht leerkrachten (38%) zeggen doelbewust af te wisselen en het Nederlands doelbewust in te zetten, in ieder geval bij het uitleggen en bij instructies voor taken: “Ik heb het een keer geprobeerd om een hele les Engels te praten, maar toen gaven ook kinderen aan dat ze na drie regels al afhaken voor de rest van de les, omdat ze er niets van begrijpen. Dus ik kies er heel bewust voor om het wel in het Nederlands uit te leggen en te bespreken.” Ook wordt het Engels soms doelbewust in het Nederlands herhaald en wordt er in combinatiegroepen Nederlands met de groep 7-leerlingen gesproken. Twee leerkrachten zeggen in het begin van het jaar alleen Nederlands te spreken en later in het jaar steeds meer Engels toe te passen. Eén leerkracht zegt altijd alleen maar Nederlands te spreken.

In de 22 geobserveerde lessen wordt door achttien leerkrachten Engels gebruikt, variërend van min of meer gedurende de hele les tot gedeeltelijk afgewisseld met Nederlands. Het overstappen op het Nederlands wordt met name gesignaleerd wanneer het gaat om het geven van instructie (bijvoorbeeld van grammatica) of om moeilijke woorden, die in het Nederlands worden vertaald. Vier leerkrachten spreken (bijna) de hele les in het Nederlands. De observatoren hebben daarbij ook het Engelse taalniveau gewaardeerd. Van zeventien leerkrachten (77%) werd het Engels als voldoende vloeiend en correct gescoord. Het eigen Engelse taalniveau lijkt daarmee bij ruim driekwart van de geobserveerde leerkrachten voldoende voor het geven van Engelse les in het po. Ook is geobserveerd of de leerkracht leerkracht-leerling(en)-interactie uitlokt in de doeltaal (zie Tabel 10.5). Dit blijkt in 10 van de 22 lessen het geval te zijn. In dergelijke situaties gaan leerkrachten bijvoorbeeld in gesprek met de leerling(en), bijvoorbeeld na een spontane reactie van een leerling of wordt gebruik gemaakt van rollenspelletjes. Het in het Engels bespreken van een mindmap of van een poster ter voorbereiding op een presentatie is ook geobserveerd. De leerkracht laat daarbij duidelijk zien dat hij/zij verwacht dat de leerling Engels terugpraat: “You're talking Dutch, so I don't understand what you're saying.”

Ook in het hoofdonderzoek is gevraagd het gebruik van de doeltaal bij Engels. Hier is specifiek gevraagd naar het percentage van de lestijd dat de leerkracht Engels spreekt. 28% van de respondenten gaf daarbij aan voor 30-60% van de tijd Engels te gebruiken, terwijl de helft van de leerkrachten dit 60% of meer van de tijd zegt te doen (ongewogen resultaten). Hoewel zowel in hoofd- als verdiepend onderzoek ca 80% van de leerkrachten aangeeft Engels te spreken in de les, is een goede vergelijking tussen beide uitkomsten niet mogelijk, omdat er in het verdiepende onderzoek niet is gemeten welk percentage van de les de leerkracht daadwerkelijk Engels spreekt.

### Doeltaalgebruik door leerlingen

Op de helft van de scholen (52%, bij 11 leerkrachten) wordt van de leerlingen verwacht dat zij de doeltaal, het Engels, zoveel mogelijk gebruiken. Met name op vvtto-scholen verwachten leerkrachten dat hun leerlingen Engels spreken (7 leerkrachten). Dat kan een streven van de school zijn: “En in alle klassen is het zo dat alle leerlingen Engels gesproken moeten hebben tijdens de les en anders is je les eigenlijk mislukt... Dat is een beetje de hoofdmoot. Dat dus echt álle kinderen stuk voor stuk Engels spreken.” Een andere vvtto-leerkracht benoemt dat zij leerlingen altijd motiveert tot doeltaalgebruik: “[Lessen zijn] helemaal in het Engels, zoals nu, tenzij ze er echt niet uitkomen, dan zeg ik ‘Try it in Dutch’ en dan gaan we met z’n allen bespreken wat het dan in het Engels is.” Vier eibo-leerkrachten laten hun leerlingen ook Engels spreken in de les, al dan niet bij het onderdeel ‘gesprekjes voeren’. Deze eibo-leerkrachten benoemen dat zij onderscheid maken wat betreft het niveau van leerlingen. Van leerlingen die wat zwakker zijn in Engels verwachten zij geen Engels. Een eibo-leerkracht geeft ook aan dat het er vooral om gaat dat leerlingen met elkaar meedenken in plaats van dat het taalgebruik correct is: “Zoveel mogelijk Engels en ik verwacht ook dat ze elkaar hier en daar corrigeren, van: ‘Hé, maar dit spreek je wel zo uit, maar volgens mij...’ Komen ze er niet uit, dan kunnen ze het vragen aan mij natuurlijk. Je merkt gewoon dat er wel heel veel Nederlands tussendoor komt en dat vind ik prima hoor. Maar de wil om het zoveel mogelijk in het Engels te doen staat voorop.” Zes leerkrachten (4 eibo en 2 vvtto-leerkrachten) wisselen doeltaalgebruik door leerlingen af op basis van lef, niveau of het lesonderdeel. Een vijfde van de leerkrachten, vier leerkrachten (3 eibo en 1 vvtto leerkracht), verwachten niet van hun leerlingen dat ze Engels gebruiken in de les. Eén vvtto-leerkracht heeft zich niet over het doeltaalgebruik van leerlingen uitgelaten.

In de observaties is op twee manieren in kaart gebracht of en hoe de leerkracht het gebruik van de doeltaal door de leerlingen stimuleert. In 45% van de lessen (10 lessen) motiveert de leerkracht leerlingen Engels te spreken, bijvoorbeeld door bij aanvang van het spelletje ‘30 seconds’ en ‘who is it?’ direct aan te geven dat er alleen Engels gesproken mag worden (‘You are not allowed to speak Dutch’). Een leerkracht bevordert het gebruik van Engels door het geven van complimenten en, tijdens een klassikale opdracht, een complete zin te maken van een enkel woord, die dan met de hele groep wordt nagezegd en toegepast. Daarnaast is geobserveerd of leerkrachten actie ondernemen als leerlingen Nederlands spreken in plaats van Engels of terugvallen in het Nederlands. 71% van de leerkrachten doet dit niet. Het eerder genoemde punt dat leerkrachten leerlingen met name op hun gemak willen laten voelen kan hier een rol spelen.

Net als naar het percentage doeltaalgebruik door de leerkracht is in het hoofdonderzoek ook gevraagd naar het percentage tijd dat leerlingen worden uitgenodigd om met de leerkracht Engels te spreken: de ongewogen mediaan ligt op 30-60% van de lestijd uit te nodigen om Engels met hem/haar te spreken. Gevraagd naar het percentage tijd waarin leerlingen onderling Engels met elkaar spreken ligt de ongewogen mediaan op 5-30%. Een goede vergelijking van deze ongewogen resultaten met het kwalitatieve onderzoek is niet mogelijk vanwege de verschillen in meting.

### 10.8.7 Differentiatie

Hoewel leerkrachten zich er van bewust zijn dat de verschillen tussen de leerlingen op het gebied van Engelse taalbeheersing groot zijn, is er bij vijftien leerkrachten geen sprake van planmatige differentiatie. Wel worden meer incidentele aanpassingen genoemd als het zelf houden van een gesprekje met een zwakke leerling, het stellen van moeilijke vragen aan de betere leerlingen of meer verwachten van deze betere leerlingen (b.v. doeltaalgebruik), maar dit heeft geen structureel karakter. Een vvtto-leerkracht geeft aan dat de methode geen richtlijnen voor differentiatie biedt, maar dat er ook bewust wordt gekozen voor een klassikale les: “Je zou misschien kinderen sneller aan het werk kunnen laten gaan, maar je wil ook iets van ze

horen. Dat vind ik belangrijker dan dat ze de opdrachtjes maken in het werkboek.” Tien leerkrachten differentiëren vooral voor de goede leerlingen, met name in de verwerking, maar niet structureel. Deze leerlingen krijgen extra werkbladen, opdrachten, activiteiten, mogen een liedje vertalen of kunnen in extra leesboekjes werken. Op één school heeft een leerling een musical in het Engels geschreven, die door de klas wordt uitgevoerd. Hoewel er in de Engelse les meestal geen sprake is van structurele differentiatie, noemen drie leerkrachten dat in verrijgings- of plusklassen extra aandacht aan Engels gegeven, bijvoorbeeld door het lezen van moeilijker teksten. Vier eibo-leerkrachten zeggen rekening te houden met zwakke leerlingen door de samenstelling van heterogene koppels/groepjes, zodat zij van de betere leerlingen kunnen leren. Op één school wordt geprobeerd het aanbod op maat af te stemmen (op basis van de leerlijnen – zie kader PO-VO overgang). Hoewel deze afstemming met name gericht is op de betere leerlingen, wordt er ook bij Engels gebruik gemaakt van verlengde instructie voor de zwakkere leerlingen. Opmerkelijk is dat leerkrachten van combinatieklassen het aanbod niet lijken aan te passen aan de lagere en hogere jaargroepen. Wel maken zij (soms) onderscheid tussen wat zij eisen van jongere of zwakkere leerlingen: zij hoeven niet in het Engels te praten of hoeven slechts enkele woorden te zeggen in plaats van (vol)zinnen.

Naast de vraag naar differentiatie naar niveau is ook bevraagd of leerkrachten hun lessen ook aanpassen aan de interesses (van een deel) van de leerlingen. De eerder beschreven lessituatie, waarbij leerlingen mochten meedenken over hun eigen leerdoelen, is hier een voorbeeld van: leerlingen wilden graag mondeling een korte beschrijving van zichzelf aan een andere leerling kunnen geven. Drie leerkrachten zeggen aan te haken bij interesses van de leerlingen door bijvoorbeeld een extra taak over een aansprekend onderwerp in de weektaak in te voegen, extra aandacht aan een onderwerp in de klas te besteden of door leerlingen, die erom vragen, een powerpoint in het Engels over het besproken onderwerp te laten maken. De school waar, naast de reguliere les, wekelijks tijd is ingeroosterd voor het lezen van Engelse boekjes heeft boekjes van verschillende niveaus en onderwerpen en laat leerlingen deze kiezen op basis van interesse. Aanpassing naar leerprofiel lijkt niet plaats te vinden. Dit betekent dat leerkrachten in hun lessen niet structureel rekening lijken te houden met factoren als gender, de leefomgeving van de leerlingen en (socio)culturele aspecten die het leerproces kunnen beïnvloeden.

*Tijdens het onderzoek* is slechts in één les geobserveerd dat de leerkracht de leerstof aanpast aan het niveau van de leerling. Aanpassing aan interesse of leerprofiel is niet gezien. Deze bevindingen zijn in lijn met het hoofdonderzoek, waarin leerkrachten ook aangeven weinig te differentiëren (69% van de leerkrachten<sup>102</sup> geeft aan dit niet structureel te doen). Dit kan mede veroorzaakt worden door hun oordeel dat de methode weinig richtlijnen voor differentiatie biedt.

---

<sup>102</sup> Dit is een ongewogen resultaat.

#### Kader 4: De 'methodeschool'

Bijna alle scholen in het hoofdonderzoek geven aan dat ze een methode gebruiken (94 % van 89 scholen) en meer dan de helft van de leerkrachten (52%) geeft aan dat zij geen additionele (eigen) materialen inzetten in de les. Paragraaf 10.8.2 laat zien dat in de verdiepende studie slechts één school geen methode hanteert en dat voor twee derde van de bezochte leerkrachten de methode leidend is. Het overgrote deel van de leerkrachten in de verdieping is positief en enthousiast over de gehanteerde methode; de leerkrachten die dat niet (altijd) zijn gebruiken vaker additionele (eigen) materialen of gebruiken lessen of opdrachten uit andere (online) methoden.

Uit bovenstaande wordt duidelijk dat het uitvoeren van de methode in het basisonderwijs een belangrijke centrale plaats heeft. Een groot deel van de leerkrachten in het verdiepende onderzoek geeft aan *alleen maar* de lesstof uit de methode te behandelen. Wanneer een methode een thematisch hoofdstuk (bijvoorbeeld de 'seizoenen') kent met daarin drie aansluitende en systematisch opgebouwde lessen van 45 minuten, wordt zij veelal ook op die manier uitgevoerd. Hoewel in dit onderzoek niet precies kan worden aangegeven op welke wijze de verschillende methoden een theoretische onderbouwing kennen, lijkt het er op dat zij allen zijn ontwikkeld vanuit de door de overheid gestelde kerndoelen Engels in het basisonderwijs, en dat het Vierfasen-model ten grondslag ligt aan de vakdidactische opbouw van de methoden. De leerkrachten die louter de methode gebruiken krijgen daarom richtlijnen om in meer of mindere mate de volgende fasen aan te houden: a) een introductiefase ('wat hebben we de vorige keer geleerd, en wat ga je vandaag leren', b) een deel input waarin de leerkracht de nieuwe stof en nieuwe woorden/zinnen presenteert – meestal via het digibord - en controleert of de leerlingen het hebben begrepen, c) een verwerkingsfase waarin de leerlingen via methode gebonden opdrachten de stof laat oefenen en eventueel d) een fase waarin leerlingen de nieuwe kennis in combinatie met de stof uit voorgaande lessen moeten produceren, bijvoorbeeld via gesprekjes of spelletjes.

De rest van het hoofdstuk laat zien op welke wijze de methode en de deelvaardigheden en verschillende werkvormen die daar aan gekoppeld zijn, terugkomen in de geobserveerde lessen. Daarbij wordt duidelijk dat bij een 'standaardles uit de methode' de essentieel geachte onderdelen niet altijd even goed uit de verf komen. Taaldoelen worden bijvoorbeeld niet genoemd, er ontbreekt een terugkoppeling aan het einde van de les en er wordt bijna niet gedifferentieerd. Leerkrachten geven aan dat de door hen gehanteerde methode aangaande deze aspecten ook soms tekort schiet. Met name het ontbreken van differentiatiemogelijkheden in de methode wordt gemist.

De scholen in de verdiepende studie zijn geselecteerd vanwege bovengemiddelde resultaten bij de leerlingen. Het (grotendeels) vasthouden aan de methode lijkt dan ook een vruchtbare werkwijze.

#### 10.8.8 Gebruik van ICT

De mogelijkheden van ICT voor de algemene lespraktijk worden de laatste jaren steeds groter. Bijna alle klassen in basisscholen hebben een digibord en het kan voor bijna alle vakken worden gebruikt. De vakinhoudelijke mogelijkheden van het digibord zijn bij het vak Engels meestal gekoppeld aan de methode die op de school wordt gebruikt. Het digibord kan als middel worden ingezet voor centrale (groeps)instructie, voor individuele verwerkingsopdrachten (in en buiten de methode), voor presentaties van leerlingen, als ook voor klassikale toetsing (bijvoorbeeld een luistertoets). Waar het digibord vaak een centrale functie in de les heeft kunnen Ipads, PC's, en zelfs mobiele telefoons worden ingezet om lesstof op individuele wijze te verwerken.

Uit de geobserveerde lessen blijkt inderdaad dat de meeste leerkrachten het digibord inzetten om instructie aan de hele groep te geven. Volgens enkele leerkrachten zijn alle lessen in de gehanteerde methode gekoppeld aan instructie en verwerkingsopdrachten via het digibord: "wat dat betreft is deze les demonstratief voor wat er digitaal allemaal mogelijk is". Naast de instructie gebruiken de leerkrachten het



digibord ook voor verwerkingsopdrachten: op het bord staan bijvoorbeeld foto's en via een 'wie ben ik' spelletje krijgt één leerling van de andere leerlingen aanwijzingen welke persoon op de foto hij/zij is. Meerdere leerkrachten geven aan dat leerlingen soms verwerkingsopdrachten individueel via de pc of een tablet moeten/kunnen uitvoeren.

Ook in de interviews is stil gestaan bij de rol en het gebruik van ICT in de lespraktijk van het vak Engels. Ondanks dat er verschillende methoden worden gebruikt zijn de didactische materialen uit de methode die worden ingezet vergelijkbaar: geluid- en videofragmenten, luisteropdrachten, muziekvideo's met verwerkingsopdrachten, digitale aanwijs- en invulopdrachten, interactieve games, digitaal voorlezen en digitaal klokkijken. De digitale materialen die buiten de methode worden ingezet zijn o.a. youtube filmpjes, websites met taalspelletjes, en websites waar de lesstof geoefend kan worden (en die niet bij de methode horen, zoals [www.allestoetsen.nl](http://www.allestoetsen.nl)).

Meerdere leerkrachten geven aan dat het werken met (het digitale deel van) de methode goed bevalt en dat de software zeer gebruiksvriendelijk is: "Iedereen kan deze methode aanbieden en er les in geven, je hoeft eigenlijk alleen maar een klikje op het bord te kunnen geven en dan loopt die les. De digital teacher vertelt alles". Een gevaar dat schuilt in het afdraaien van de methode, merken twee andere leerkrachten op, is dat het digibord te leidend wordt en dat leerlingen er passief van worden: "ik kan me voorstellen dat mensen zeggen 'laat dat digibord dat maar doen', lekker makkelijk". De leerkrachten die een digitale methode volgen geven aan dat ze toch vaak zelf een draai aan de les geven en dat de digitale onderdelen selectief worden ingezet. De digitale teacher kan bijvoorbeeld worden uitgezet zodat de leerkracht zelf de stof op het bord kan toelichten en af en toe onderdelen kan weglaten of overslaan. Sommige passages in de digitale les gaan bijvoorbeeld te snel of er wordt te veel stof/tekst in één keer aangeboden. Op dergelijke momenten "probeer [ik] er wel een vertaalslag in te maken, van: wat heeft ze nu daadwerkelijk gevraagd?"

De methode is volgens enkele leerkrachten te groepsgestuurd, wat betekent dat er in digitaal opzicht weinig mogelijkheden voor differentiatie zijn: er is geen software voor individuele begeleiding beschikbaar, er zijn geen individuele verwerkingsopdrachten via de computer en een leerkracht betreurt: "het zou leuk zijn als leerlingen op hun tablet ook Engelse oefeningen kunnen doen. De methode voorziet hier niet in". Een leerkracht merkt op dat de digitale methode helemaal niet ingaat op grammatica en dat hij dat zelf moet doen.

Ook de drie vakleerkrachten in de onderzoeksgroep geven aan dat zij wel met digibord/ICT werken, maar het digitale materiaal dat door hen wordt ingezet is minder gebonden aan een vaste methode. Een van hen merkt op dat "dat niet zo vreemd is, want ik heb veel meer voorbereidingstijd en kennis van de digitale materialen dan de groepsleerkrachten natuurlijk".

In de interviews richtten de leerkrachten zich met name op het gebruik van het digibord en daarmee de inzet van ICT als technisch hulpmiddel. Wat leerlingen doen met ICT buiten de les, zoals het gebruik van social media enz., wordt beperkt aangehaald. Het specifiek ingaan op de activiteiten die leerlingen buitenschools op het gebied van ICT uitvoeren, lijkt geen veelvoorkomend gebruik bij leerkrachten.

### 10.8.9 Motivatie van leerlingen

Een belangrijk aspect van effectieve lessen is dat leerlingen gemotiveerd met de lesstof bezig zijn. De meeste leerkrachten geven aan dat hun leerlingen zeer gemotiveerd zijn voor Engels. In de interviews is de leerkrachten gevraagd op welke manieren zij hun leerlingen motiveren voor het vak Engels.

Zowel impliciet als expliciet geven de leerkrachten aan dat de motivatie van leerlingen natuurlijk zeer sterk is gekoppeld aan de algemene pedagogisch-didactische aanpak in de klas. Dus ook voor het Engels



vinden de leerkrachten dat je als leerkracht vooral leerlingen zelfvertrouwen moet geven via complimenten en succeservaringen en leerlingen niet moet ontmoedigen door te veel te corrigeren en nadruk te leggen op “de formele kant van de taal.” Verder menen zij dat je te allen tijde iedereen in de les ‘erbij’ moet betrekken, en dat eigen enthousiasme voor het vak ook motiverend werkt. Motivatie van leerlingen wordt volgens de leerkrachten verder vooral bereikt door a) aan te sluiten bij de belevingswereld en interesses van de leerlingen, bijvoorbeeld in onderwerpkeuze, b) te zorgen voor veel uitdaging in de les en bij de verwerkingsopdrachten, door met name in te steken op nieuwe informatie en vaardigheden (in tegenstelling tot herhaling), c) afwisseling aan te brengen in werkvormen en opdrachten (zie paragrafen 10.8.4 en 10.8.5), d) het belang van Engels voor de toekomst en het dagelijkse leven te benadrukken, zoals reeds bij de beschrijving van de leerkrachtvisie naar voren kwam en e) aandacht te besteden aan de Engelse cultuur en geschiedenis, met behulp van verhalen, projecten of het bespreken van eigen (buitenland) ervaringen. Ook de organisatie van leuke activiteiten bevordert de motivatie voor Engels en de Engelse cultuur. Te denken valt aan een cakesale, hotdoglunch en Christmas market. Maar ook uitwisselingsprojecten (fysiek of via bijvoorbeeld brief of e-mailwisselingen met leerlingen uit het buitenland) dienen dit doel.

Uit de observaties blijkt dat de onderzoekers in twee vijfde van de lessen (9 lessen) hebben gezien dat de leerkracht expliciet leerlingen motiveert. Dat doen ze volgens de inspecteurs vooral door het aanbieden van afwisselende werkvormen, activiteiten en spelletjes: “Door de vele verschillende activiteiten blijft de groep gefocust op de Engelse taal”. Ook het enthousiasme van de leerkracht in één les is duidelijk een manier om de leerlingen voor de lesstof te motiveren. Overigens is het de vraag of de motivatieverhogende activiteiten in iedere les even goed uit de verf komen. In één observatie werd bijvoorbeeld een sterke diversiteit aan activiteiten gezien die het verloop van de les voor de leerlingen minder overzichtelijk maakte.

#### 10.8.10 Motivatie van de leerkracht

In de interviews is ook gevraagd naar de motivatie voor het vak van de leerkracht zelf. Alle leerkrachten geven aan dat ze meestal goed gemotiveerd zijn voor het geven van het vak, vooral omdat ze zelf geïnteresseerd zijn in de taal (via literatuur, films, muziek), het in privé-situaties regelmatig gebruiken en ze het leuk vinden om Engelse lessen te ontwikkelen. Zeven leerkrachten zijn echter bijzonder gemotiveerd vanwege hun voorliefde voor vreemde talen of omdat ze Engels gewoonweg “heel erg leuk en belangrijk” vinden - een enkele leerkracht noemt zichzelf een Anglofiel. Deze leerkrachten vermelden in de interviews vaker dan de overige leerkrachten dat ze ervaringen hebben met het lesgeven in het buitenland (in het Engels), als ook met coördinatietaken voor vvto.

Een drietal leerkrachten geeft in het interview aan dat het eigen enthousiasme van belang is om leerlingen ook te motiveren, om ze tot spreken aan te zetten, en om ze de stof te laten beheersen: “met name het durven spreken en plezier er in hebben vind ik belangrijk. Ik hou ontzettend van talen en ik hoop daar iets [van het enthousiasme] over te brengen.”, en: “het enthousiasme van de leerkracht werkt enthousiasmerend voor de leerlingen en daardoor gaan ze *misschien* wel beter presteren”.

Enkele leerkrachten erkennen dat de eigen motivatie niet vanzelfsprekend is. Er moet een zekere affiniteit zijn met het vak, de taal en de cultuur: “Als je collega Engels helemaal niet leuk vindt, misschien dan uitrust van talenten”. Deze leerkracht had met een collega een uitrust geregeld en werd daarmee als het ware ingezet als de vakleerkracht voor de groepen 7 en 8. Op een vvto school maakt de leerkracht ook gebruik van Engels in het periode-onderwijs (projecten), “maar dit [het toepassen] is afhankelijk van persoonlijke affiniteit”. Op een derde school (eibo) geeft de leerkracht alle lessen Engels in alle klassen maar wijt dat

vooral aan aanleg en persoonlijke interesse: “ik merkte bij collega’s dat dit niet altijd zo is, vandaar [dat ik alle lessen geef]”.

## 10.9 Opleiding en professionele ontwikkeling

De opleiding en toerusting van leerkrachten is in het interview via verschillende vragen aan bod gekomen. Er is gevraagd naar de aandacht voor Engels in de hoofdopleiding (meestal de PABO). Daarbij konden de leerkrachten aangeven welke onderdelen in de opleiding zij als relevant zien en welke onderdelen zij misten. Daarna is gevraagd of en welke professionaliseringsactiviteiten de leerkrachten hebben gevolgd. Ten slotte is de vraag gesteld of de leerkracht zichzelf voldoende toegerust vindt om goede lessen Engels in groep 8 te geven.

### 10.9.1 Initiële opleiding

Uit de presentatie van ongewogen resultaten in paragraaf 6.2 blijkt dat de leerkrachten in het hoofdonderzoek bijna allen beschikken over een PABO-opleiding. Ongeveer vier vijfde heeft een reguliere Pabo-opleiding gevolgd, terwijl de rest van de groep de verkorte pabo heeft gedaan (10%) of de academische pabo heeft afgemaakt (9%). Geconstateerd is dat niet elke leerkracht met Engels in aanraking is geweest tijdens de opleiding. Ongeveer een vijfde gaf aan dat er totaal geen aandacht voor het Engels was in de opleiding, de rest volgende een verplichte of een keuzemodule Engels. Het meest opmerkelijke uit de leerkrachtvragenlijst is dat slechts 19% tevreden zegt te zijn over het aanbod op de Pabo.

De groep van 22 geïnterviewde en geobserveerde leerkrachten is qua initiële opleiding min of meer vergelijkbaar met de gehele groep uit het hoofdonderzoek. Echter zijn er nu drie leerkrachten bij die momenteel als vakleerkracht Engels werkzaam zijn. Deze vakleerkrachten hebben zich of na de PABO verder op het gebied van Engels ontwikkeld via extra opleidingen en/of een arbeidsverleden in buitenland of op een internationale basisschool. Twee vakleerkrachten uit de geïnterviewde groep hebben (deels) een universitaire opleiding Engels gevolgd voordat ze, met lesbevoegdheid, in het basisonderwijs terecht kwamen. Een andere vakleerkracht heeft een 1e-graads lerarenopleiding Engels gevolgd waarbij ze de vakdidactische scholing op de lerarenopleiding (gericht op het vo), zelf heeft vertaald naar het po en toegepast.

Voor het overgrote deel van de leerkrachten is de opleiding al enige tijd geleden (tot zelfs 40 jaar). Bijna alle leerkrachten geven aan dat er niet of nauwelijks in de opleiding tot leerkracht aandacht werd besteed aan Engels: “geen enkele aandacht voor het vak Engels”, “nauwelijks in reguliere curriculum PABO”, “twee lessen en een toets”, “één keer een les Engels vanwege stage”. Toch zijn er ook leerkrachten die een intensieve module Engels hebben gevolgd: “8 weken met veel praktische opdrachten”. Vier leerkrachten geven aan dat er wel (een vak) Engels was in de opleiding maar dat dit zo minimaal was dat ze zich niet goed kunnen herinneren wat en of ze daar iets van hebben geleerd. De ervaring met en training in de Engelse taal komen, voor de groep die de PABO volgde, vooral uit de HAVO- en VWO-vooropleidingen en via het aanleren van de taal via eigen interesses (vakanties, films, games, muziek etc.). Daarbij geven twee leerkrachten aan een tweetalige (TTO) opleiding te hebben afgerond.

### Relevante en gemiste onderdelen

Dat er weinig aandacht voor het Engels was in de PABO-opleidingen betekent dat de PABO- leerkrachten ook niet goed de elementen uit de opleiding kunnen benoemen die relevant en nuttig zijn geweest voor het vak

Engels in het basisonderwijs. Een paar zaken die door de andere leerkrachten worden benoemd zijn: het veel lezen en praten in het Engels tijdens de opleiding “geen schroom hebben, gewoon doen”, het toepassen van het vier-fasen model in de vakdidactiek (één leerkracht), en klassenmanagement in de docentenopleiding (ook toegepast in het po).

Gemiste onderdelen zijn meer benoemd in de interviews en daarvan is aandacht voor de eigen vaardigheden het voornaamste: “de Pabo was uitsluitend gericht op didactiek van een Engelse les geven en het hield zich helemaal niet bezig met eigen vaardigheden in de Engelse taal”. Een andere leerkracht vult aan “Eigenaardig dat er toen wel werd geleerd hoe je het vak moest geven, maar dat je geacht werd met VWO voldoende woordenschat en grammatica te beheersen”. Het eigen kunnen (grammatica, uitspraak en woordenschat) was dus zeer onderbelicht tijdens de opleiding, volgens de leerkrachten.

Gemist werden ook de mogelijkheden voor stages op een buitenlandse po-school, of überhaupt stagelessen die gaan over het Engels. De leerkrachten hebben zelf nooit de mogelijkheid gehad om Engelse lessen te oefenen, maar, zoals een leerkracht opmerkt “ook van de stagiaires die nu op de scholen komen, wil of gaat er geen enkele Engels verzorgen.” Er is kortom vaak onvoldoende praktijkervaring voordat je op een school terecht komt.

De mogelijkheden tot differentiatie in de klas bij het vak Engels wordt door drie leerkrachten aangekaart als een ervaren manco in de opleiding, bijvoorbeeld richting meertalige leerlingen, maar ook in algemene zin: “hoe bepaal je het beginniveau van een leerling, en waar kan je vervolgens naar toewerken”.

### 10.9.2 Aanvullende professionalisering

Na- en bijscholing is een manier om de taal beter te beheersen en/of de didactiek van het vak Engels te verbeteren. Uit het hoofdonderzoek blijkt dat maar 14% van de leerkrachten na de initiële opleiding een na- of bijscholingsprogramma of cursus op het gebied van Engels heeft ontvangen.

Van deze groep zitten er relatief veel in de verdiepende studie: acht leerkrachten geven aan in de afgelopen jaren een of twee programma's te hebben gevolgd om a) de eigen vaardigheden te verbeteren, b) de didactiek in de klas te verbeteren of c) de coördinatie van het vak binnen de school te organiseren. Meer specifiek gaat het om:

- Early Bird scholingen, bijvoorbeeld ‘the classroom English’ (2 scholen);
- CLIL cursussen in Engeland;
- Workshops en uitwisselingen van ESPRIT en ABC bookshop;
- Cursus Teaching Proficiency by Reading and Story telling (TPRS);
- Schoolbrede deelname aan cursus Cambridge Engels na invoering vvto door bestuur;
- Het uitwisselen van ervaringen en bezoeken van lessen op VO scholen via een PO-VO werkgroep (zie kader 1);
- Een nascholing tot coördinator Engels (Marnix-academie);
- Een vvto coördinatie-opleiding.

Na- en bijscholing komt vooral bij de leerkrachten van de vvto-scholen voor. Enkele leerkrachten van deze scholen vinden dat nascholing essentieel is voor het behalen van je vvto-doelstellingen: “als je immers vvto-school bent, dan moeten leerkrachten ook voldoende niveau hebben en houden”. Op algemene studiedagen is er daarom ook aandacht voor Engels, bijvoorbeeld via ‘Classroom English’ van Early Bird dat door de leerkrachten van twee scholen is gevolgd. Een van de opbrengsten van die cursus, benoemt een leerkracht, is “dat je beter de leerlingen kan laten spreken, en met name jonge leerlingen kan stimuleren te spreken in de klas”.

Op een andere school hebben alle leerkrachten (groep 1 tot 8) de basis Cambridge English opleiding moeten volgen gedurende twee schooljaren. Ook de leerkrachten van andere scholen onder hetzelfde bestuur hebben de cursus moeten volgen. Doordat alle leerkrachten de cursus hebben gehaald, is een zekere basis binnen de school gegarandeerd. De leerkrachten van de school hebben echter de vrijheid om een eigen invulling aan het Engels te geven.

Professionalisering kan ook voortkomen vanuit het delen van onderwijspraktijken. De po-vo werkgroep zoals beschreven in paragraaf 10.7.4 is daar een mooi voorbeeld van. Ook andere leerkrachten geven aan dat een kijkje bij een collega in de les en/of het uitwisselen van lesmaterialen je eigen lespraktijk kan verbeteren. Een leerkracht heeft bijvoorbeeld op een VO-school geïnformeerd naar de inhoud van Cambridge Engels. Hierdoor had ze “een beter beeld van hoe ze de leerlingen [die naar deze school zouden uitstromen] moest toerusten”.

Vier leerkrachten geven aan dat bij- en nascholing op het gebied van Engels in principe via financiering door de school mogelijk is, wanneer zij zelf aangeven dat ze dat noodzakelijk achten. Deelname aan een cursus is echter nog niet van de grond gekomen omdat “Engels wordt nooit aangeboden”, “er nog geen concrete behoefte aan is geweest” of “bijscholing voor het niveau van groep 7/8 “[echt] niet nodig is”.

Gevraagd naar de omissies in de genoten professionalisering geeft een leerkracht aan dat er nog niet echt aandacht is voor het differentiëren bij Engels in de nascholingsprogramma's.

### 10.9.3 Voldoende toerusting lessen Engels te verzorgen

De meeste leerkrachten in dit verdiepende onderzoek achten zichzelf voldoende toegerust om kwalitatief goede lessen Engels te geven. Dit ondanks de beperkte aandacht voor Engels in de opleiding. Enthousiasme, interesse in de taal en cultuur, buitenlandervaringen, voorkennis (eigen opleiding op HAVO-VWO niveau (soms TTO)) en algemene didactische vaardigheden zijn volgens de leerkrachten de elementen die daar een rol in spelen. Voor enkele leerkrachten is de sterke inhoudelijke expertise als ook routine (universitair Engels, vto-coördinator en toch ook de genoten bij- en nascholingen) een bewijs en voorwaarde dat ze goede Engelse lessen kunnen geven.

Meerdere keren wordt de heldere en handzame methode aangedragen als de basis van de eigen toerusting om Engels te geven: “de methode geeft houvast”, de “methode is voor iedereen te beheersen”, “Ja, je pakt gewoon de methode erbij en je doet het!”. Signalen die in lijn liggen van het succes van de ‘methode-school’, zie kader 4.

In de groep van 21 zijn er echter ook enkelen die minder zeker zijn van hun eigen kunnen, bijvoorbeeld ten aanzien van spreekvaardigheid en uitspraak: “vloeiend en vlot spreken kan beter. Ik ken de woorden wel, maar het toepassen is lastig”, “[ik] werd altijd uitgelachen op middelbare school”, en “ik zou [om de leerlingen te helpen] wel een goede taalcursus Engels willen, voor de uitspraak”.

Of de leerkracht zichzelf voldoende vindt toegerust om Engels te geven lijkt overigens een kwestie van de interpretatie van de vraag. Zo geven twee leerkrachten aan dat het eigen niveau voor de basisschool vermoedelijk wel voldoende is, “maar niet verder dan dat!”, en “ik maak misschien wel wat grammaticale fouten en het is niet altijd vloeiend, maar kan me voldoende begrijpelijk maken en het is voor dit niveau voldoende.”. Een vakleerkracht geeft als antwoord dat het vakinhoudelijk en vakdidactisch wel in orde zit, maar “klassenmanagement kan wel beter. Ik kom daar maar een keer per week binnen. In je eigen groep is dat anders”.

### *Verskil vak- en groepsleerkracht*

Het verschil tussen een vak- en groepsleerkracht bij het vak Engels zit volgens de vakleerkrachten vooral in de vakinhoudelijke kennis (grammatica, woordenschat), maar ook in de vakdidactiek en het zich vrijer kunnen bewegen door gebruik maken van de taal: het “zonder schroom praten”, met wellicht iets meer zelfvertrouwen en overtuiging : “[gewone] leerkrachten denken soms dat leerlingen beter Engels spreken dan zichzelf, dat kan een belemmering zijn”. De kwaliteit van het Engels (voor de doelgroep) is volgens een vakleerkracht minder een issue: “sommige leerkrachten durven eigenlijk toch niet zo goed (ook al kunnen ze het wel), voelen zich onzeker [...] en daardoor heb je misschien ook minder plezier in het Engels geven”.

Een ander verschil met sommige groepsleerkrachten is wellicht dat vakleerkrachten vaker beschikken over visie en overzicht, over een integrale vakinhoudelijke en vakdidactische aanpak, met een bepaalde opbouw in de tijd. Of zoals een zegt: “Ik weet waarom ik dit op die manier doe, en ik kan het vertalen in eigen lessen, die uitvoerbaar zijn. [Door minder overzicht] zit het vak minder in het systeem bij groepsleerkrachten, waardoor zij soms keuzes maken die b.v. door tijd zijn ingegeven (zo van: ik heb nog zoveel tijd, dus ik laat de leerlingen even een werkblad invullen). Dat is niet vreemd, maar soms wel jammer”.

## **10.10 Inbedding van en focus op Engels in de school**

Uit bovenstaande wordt duidelijk dat leerkrachten (sterk) verschillen in hun visie over het vak Engels, de inhoudelijke invulling van de lessen en in hun didactische aanpak. Daarbij lijkt, zoals we zien in kader 4, bij de meeste leerkrachten de methode als voornaamste leidraad te fungeren: deze bepaalt de inhoud van de individuele les, de aan te leren vaardigheden als ook de werkvormen die daar bij horen.

Binnen het hoofdonderzoek is niet specifiek gevraagd naar de visie van de school, maar wel naar de manier waarop het Engels is georganiseerd. Op de onderzochte scholen bleken er op vvto-scholen, zoals verwacht, vaker een coördinator/werkgroep Engels aanwezig op school en relatief vaker sprake te zijn van faciliteiten ten behoeve van Engels (zoals budget, partner-/uitwisselingsprojecten) dan op eibo-scholen. Op eibo scholen bleken er in de helft van de gevallen geen (extra) voorzieningen voor het vak te zijn (zie paragraaf 6.1.2).

Wanneer we kijken naar wat leerkrachten in de interviews zeggen over de schoolvisie aangaande Engels blijkt dat op zeven scholen de leerkracht aangeeft dat de school geen duidelijke visie op Engels heeft. Voor de andere veertien scholen wordt wel een visie weergegeven, maar deze betreft met name een schoolbrede focus in termen van specifieke vaardigheden. De schoolvisie is volgens leerkrachten op tien scholen (4 eibo en 6 vvto) gericht op de mondelinge vaardigheid en communicatieve redzaamheid van de leerlingen – een visie die (meestal) aansluit bij die van de leerkracht. Twee leerkrachten benoemen expliciet dat er weliswaar een schoolbrede focus op specifieke vaardigheden is, maar dat de school zoekend is naar de juiste balans van vaardigheden en naar een antwoord op de vraag welke vaardigheden nu het belangrijkste zijn binnen het po.

Een andere genoemde schoolvisie is dat leerlingen vooral moeten *kennismaken* met de vreemde taal. Deze schoolvisie wordt door vijf leerkrachten genoemd, in twee gevallen in combinatie met de communicatieve focus. Een specifieke invulling van deze focus op kennismaking is de gerichtheid (op twee scholen) op dat wat de methode aanbiedt (in combinatie met de algemene visie op onderwijs) of op de kerndoelen. Dit laatste wordt door één van de leerkrachten als volgt beschreven: “Dus willen zij vooral dat je als het ware een soort douche over ze uitstroomt, zo van ‘nou dit is Engels en je pikt ervan op wat je ervan

op kunt pikken' en zo werkt het ook. We dwingen ze niet, [...] het is echt wat ze ermee doen en ik denk natuurlijk 'alles wat ik erin kan stoppen dat stop ik erin', gewoon omdat ik het zelf heel leuk vind."

Meerdere keren wordt genoemd dat het onderwijs en aanbod op het gebied van Engels met name 'aan de koffietafel' wordt besproken, als het al besproken wordt. Er zijn echter ook scholen waar Engels een vast onderdeel op de vergaderagenda is. De (extra) aandacht voor de Engelse lessen op scholen wordt overigens in een aantal gevallen (4 scholen) als belastend ervaren, met name omdat rekenen en taal op de school en door de leerkracht belangrijker geacht worden. Deze perceptie kan sterker naar voren komen als er sprake is van een zwakkere leerlingpopulatie. "Ja, want dan is het: 'We moeten al zoveel en er staat best veel tijd voor Engels.' Eerst was dat 45 minuten en in de bovenbouw is dat het dubbele geworden en we zitten in de onderbouw ook al aan een uur."

Ondanks dat er in de interviews in beperkte mate aandacht is besteed aan de uitwerking en implementatie van de schoolvisie, bleken er vier min of meer onderscheidende vormen te zijn waarop Engels op schoolniveau wordt vormgeven. Deze uitwerkingen worden in sterkere of mindere mate door de teamleden gedragen. Drie scholen lijken een duidelijke focus in hun onderwijs hebben aangebracht, namelijk op 1) doorgaande leerlijn en samenwerking PO-VO, 2) mondelinge taalvaardigheid, 3) een onderbouwd aanbod en het integreren van vakdidactische modellen. Daarnaast leken meerdere scholen getypeerd te kunnen worden als een 'methodeschool', scholen waar de methode uitgangspunt en leidraad in het handelen is. Overigens kenmerken alle scholen zich door de implementatie van elementen van deze uitwerkingen, al dan niet in combinatie.

## 10.11 Conclusie en discussie

Na afloop van het hoofdonderzoek is een verdiepend onderzoek uitgevoerd. Getrainde inspecteurs hebben hiervoor 22 observaties en 21 interviews (waaronder een dubbelinterview) op 21 bovengemiddeld presterende scholen uitgevoerd. Geselecteerde scholen verschilden wat betreft startmoment (eibo/vvto) en percentage leerlingen met een formatiegewicht. Dit om een zo herkenbaar mogelijk beeld van de praktijk te verkrijgen. Doel was het verzamelen van inspirerende lespraktijken in deze bovengemiddeld presterende klassen/scholen en het verkrijgen van inzicht in de (vak)didactische invulling van de lessen Engels in groep 8. In deze paragraaf gaan wij eerst in op gevonden overeenkomsten, zowel wat betreft de visie over Engels in het PO als de concrete uitwerking daarvan in de lespraktijk. Binnen deze overeenkomsten komen gevonden verschillen tussen leerkrachten aan bod. Daarna wordt beschreven waar wij ruimte voor verbetering zien – gekoppeld aan aanbevelingen voor lerarenopleidingen (Pabo's), beleid en de lespraktijk, gevolgd door de beperkingen van de studie.

Voor het onderzoek zijn twee instrumenten gebruikt. De interviews kenden een semigestructureerde interviewleidraad en gingen in op 3 thema's: a) visie op Engels in het PO, b) de eigen lespraktijk en c) scholing en opleiding. Voor de observaties is gebruik gemaakt van een observatie-instrument dat uit twee delen bestond: de ICALT PO en een vakdidactisch deel. De observaties laten zien dat het algemeen pedagogisch-didactische handelen van de leerkracht<sup>103</sup> in de lessen Engels zich over het algemeen kenmerkt door het aanbieden van een veilig klasklimaat, het zorgen voor een goede klassenorganisatie, het verzorgen van duidelijke instructies en het (tot op zekere hoogte) activeren van leerlingen. Gemiddeld stemden leerkrachten hun lessen echter niet af op verschillen tussen leerlingen en

---

<sup>103</sup> Het pedagogisch didactische handelen is gemeten met behulp van ICALT PO.



leerden zij de leerlingen geen leerstrategieën aan. Daarnaast is gekeken naar de aanwezigheid in de les van vakdidactische leerkrachtactiviteiten die van belang zijn voor de verwerving van een vreemde taal. Leerkrachten bleken gemiddeld van de 15 activiteiten 9 à 10 relevante activiteiten toe te passen, maar er waren verschillen zichtbaar in de hoeveelheid activiteiten per les.

### 10.11.1 Overeenkomsten (en verschillen) tussen leerkrachten

De bevindingen uit de interviews en observaties laten zien dat er duidelijke overeenkomsten zijn tussen de leerkrachten over datgene wat zij belangrijk vinden voor Engels in het PO en wat er in de lessen is gezien. Binnen deze overeenkomsten kunnen leerkrachten qua visie en uitvoering van de lessen echter wel van elkaar verschillen.

#### *Belang van Engels in het PO*

Het overgrote merendeel van de leerkrachten (86%) is van mening dat Engels een belangrijk vak is op de basisschool. Gewezen wordt op het belang voor de verdere schoolloopbaan en de toekomst van de leerlingen, in een globaliserende wereld. Maar dit belang lijkt niet altijd door alle leerkrachten op de school te worden gedeeld: De geïnterviewde leerkrachten hebben over het algemeen een bovengemiddelde voorliefde voor de taal. Ook het belang van voldoende Engelse beheersing om in het VO succesvol te starten wordt genoemd, door bijna een kwart van de leerkrachten. Zij zoeken de dialoog met het VO om de overgang te versoepelen. Binnen één bestuur wordt derhalve aangestuurd op samenwerking tussen leerkrachten van PO en VO: gezamenlijk wordt een doorgaande leerlijn in termen van het ERK vastgesteld en PO-leerkrachten leren van hun vakcollega's in het VO om hun Engelse lessen te verbeteren.

Het belang dat aan Engels wordt gehecht lijkt echter te worden beïnvloed door de leerlingpopulatie. Op scholen met een relatief hoog percentage leerlingen met een formatiegewicht noemen drie leerkrachten dat het voor 'hun leerlingen' veel belangrijker is dat zij Nederlands leren en kunnen rekenen, waardoor een beperkte aandacht voor Engels volstaat. Overwegingen over het belang van Engels ten opzichte van de vakken Nederlands en rekenen spelen ook een rol wanneer het gaat om de wenselijkheid om Engels in de eindtoets op te nemen. Leerkrachten verschillen hierover van mening, ook omdat de aangeleerde vaardigheden in het PO volgens sommige leerkrachten niet goed aansluiten bij wat er in een toets gevraagd wordt. Gezien de positie van Engels in het VO, waar Engels voor alle leerlingen een verplicht examenvak is, zouden heldere en concrete richtlijnen voor de na te streven doelen (en voor het te bereiken Engelse niveau) basisscholen en -leerkrachten kunnen helpen vergelijkbare keuzes te maken over aanbod en doelstellingen.

#### *Visie op Engels en uitwerking daarvan in de les*

Gevraagd naar de visie van de leerkrachten op Engels in het PO is het opmerkelijk dat allen ingaan op aspecten die vallen onder een communicatieve focus. Leerkrachten willen vooral bereiken dat leerlingen in het Engels kunnen communiceren, dat ze plezier beleven in het leren van een vreemde taal, dat ze veel met Engels in contact komen en daardoor een basis leggen en dat er durf wordt gekweekt om de taal (mondeling) te gebruiken. Het kunnen uitvoeren van eenvoudige communicatieve handelingen, waarbij zowel taalbegrip als –productie van belang zijn, staat voor de meeste leerkrachten centraal. Leerlingen moeten iets met de taal kunnen doen en dat wat ze doen moet relevant voor ze zijn. Dit betekent dat leerkrachten zowel inzetten op input en output, terwijl grammatica vooral een dienende functie heeft. Zij wordt vooral besproken om de boodschap te kunnen overbrengen. De bevinding van Corda en collega's (2014) dat grammatica populair is bij basisschoolleerkrachten wordt daarmee niet bevestigd. Dit betekent

dat de grammatica-vertaalmethode, die in de 20<sup>ste</sup> eeuw lange tijd dominant was – in ieder geval in het voortgezet onderwijs – niet prominent aanwezig is in de lespraktijk van Engels in het PO. Vanuit taaltheoretische overwegingen kan dit worden beschouwd als een positieve uitkomst van dit onderzoek (Westhoff, 2008; Bodde-Alderlieste & Schokkenbroek, 2011; Pigott, 2018).

In de uitwerking van de visie zeggen de meeste leerkrachten (86%) lange en korte termijndoelen voor ogen te hebben. In meer dan de helft van de gevallen wordt er gewerkt aan methodedoelen, maar ook Anglia-doelen, ERK-doelen en kerndoelen worden genoemd, ook al vinden enkele leerkrachten juist deze kerndoelen weinig richtinggevend. De taaldoelen voor de les werden in de geobserveerde lessen door slechts de helft van de leerkrachten met de leerlingen gedeeld en de relevantie van die doelen werd slechts in 9% van de lessen besproken. Aan het eind van de geobserveerde lessen werd bovendien slechts in 19% nagegaan of de taaldoelen waren behaald: de meeste leerkrachten lieten dit niet zien. Het behalen van taaldoelen op de langere termijn wordt door driekwart van de leerkrachten vastgesteld met behulp van toetsing, meestal met behulp van methodetoetsen, maar ook met eigen gemaakte toetsen. De frequentie van afname verschilt echter. Bovendien blijkt uit het hoofdonderzoek dat niet alle vaardigheden waar veel aandacht aan besteed wordt, zoals mondelinge vaardigheid, evenredig vaak getoetst worden.

Om het communicatieve doel te bereiken zegt 95% van de leerkrachten de meeste aandacht aan mondelinge vaardigheid te besteden, vaak in combinatie met luistervaardigheid en woordenschat. Aan lezen, schrijven, spelling en grammatica zeggen zij aanzienlijk minder aandacht te besteden. De observaties bevestigen dit beeld. Uit de observaties en interviews blijkt daarnaast dat in de meeste lessen gezorgd wordt voor veel afwisseling, zowel in geoefende vaardigheden, werkvormen als opdrachten. Leerlingen worden in de les over het algemeen veelvuldig geactiveerd en leerkrachten gebruiken die afwisseling ook als middel om leerlingen te motiveren en de lesmethode aantrekkelijker te maken. Lesmethoden worden namelijk veelvuldig gebruikt. Van de 21 bezochte scholen is er één school waar niet met een lesmethode wordt gewerkt en bij twee derde van de leerkrachten is de methode leidend. Over de gebruikte methode zijn de meeste leerkrachten (gematigd) enthousiast. De helft van de leerkrachten zet echter ook nog additionele leermiddelen en werkvormen in. Vvto-leerkrachten/-scholen maken vaker gebruik van eigen materialen of leerlijnen. Zowel bij methodegebruik als bij het gebruik van additionele middelen speelt ICT een grote rol: het digibord wordt veelvuldig gebruikt om instructie te geven (bijvoorbeeld via de digital agent van de methode), verwerkingsopdrachten uit te voeren en filmpjes, teksten, videoclips/karaoketeksten en spelletjes te doen.

### *Gebruik van de doeltaal en andere aspecten van de les*

Een ander punt waarop de leerkrachten overeenstemming vertonen is dat ze het belangrijk vinden om de doeltaal in de les te gebruiken, al dan niet doelbewust afgewisseld met het Nederlands om begrip bij leerlingen te bevorderen. Het bewust gebruikmaken van de doeltaal om de Engelse vaardigheden van leerlingen te bevorderen wordt breed onderschreven (Westhoff, 2008; Oskam, 2017; Verspoor et al., 2010; Dönszelmann, 2016). Van leerlingen verwacht echter ongeveer de helft van de leerkrachten van de leerlingen dat zij Engels (proberen te) gebruiken. Dit zijn met name leerkrachten van vvto-scholen. De observaties bevestigen ook hier de ideeën van de leerkrachten grotendeels: in 86% van de lessen wordt Engels (in meer of mindere mate) gebruikt en in bijna de helft van de lessen verwacht de leerkracht van leerlingen dat zij Engels spreken. Wanneer zij echter terugvallen in het Nederlands onderneemt 71% van de leerkrachten echter niets. Omdat voldoende gelegenheid tot output een belangrijk onderdeel is van het taalleerproces, zouden leerkrachten bewuster, actiever en consequenter van leerlingen moeten verwachten dat zij de doeltaal gebruiken. Overigens veronderstelt een dergelijke maatregel dat leerkrachten zelf ook in



staat zijn de doeltaal flexibel te hanteren. De interviews laten echter zien dat leerkrachten zelf aangeven vaak niet voldoende beheersing van het Engels te hebben om heel soepel in de doeltaal te functioneren. Zij schatten hun eigen taalniveau weliswaar in als voldoende voor het basisonderwijs, maar een veel grotere beheersing zeggen zij meestal niet te hebben.

### *Overige overeenkomsten – differentiatie en evaluatie*

De observaties laten verder zien dat de meeste leerkrachten tevens aandacht besteden aan andere aspecten die bevorderlijk zijn voor het aanleren van een vreemde taal. Zo bevorderen zij de voorkennis van leerlingen, besteden zij ruime aandacht aan begrip en passen daarbij, indien nodig, hun taalgebruik aan, maken zij gebruik van positieve feedback, laten zij leerlingen samenwerken, geven zij correctieve feedback en laten zij leerlingen oefenen met gegeven regels/woorden. De werkvormen die gekozen worden sluiten bovendien over het algemeen aan bij de taaldoelen en hebben een communicatieve focus.

Twee overeenkomsten in de geobserveerde lessen zijn opmerkelijk: Leerkrachten evalueren de taaldoelen amper en ook is slechts in één les differentiatie geobserveerd. Dat leerkrachten nauwelijks differentiëren is opvallend, omdat leerkrachten in de interviews een sterk bewustzijn van niveauverschillen tussen leerlingen tentoonspreiden. Echter, ook de interviews laten zien dat de meeste leerkrachten deze kennis niet omzetten in aanpassingen: slechts 24% van de leerkrachten zegt structureel in te gaan op deze verschillen. En als dit het geval is lijken zij met name hun aanbod beter aan te sluiten bij de goede presteerders. De aansluiting van het onderwijs bij de in de klas aanwezige niveaus van leerlingen lijkt dan ook (zeer) beperkt, hetgeen eerdere onderzoeksbevindingen bevestigen (Thijs, et al., 2011). Thijs en collega's halen SLO onderzoek aan waarin slechts 17% van de leerkrachten remediërende materialen en activiteiten inzet en noemen dat 37% van de leerkrachten niet differentieert, ondanks ervaren problemen met aansluiting van lesaanbod bij de grote vaardigheidsverschillen. Het gebrek aan handreikingen in de methode wordt hiervoor door de leerkrachten aangehaald. In het huidige verdiepende onderzoek verwijzen enkele leerkrachten die wel structureel differentiëren juist naar mogelijkheden voor differentiatie die de methode biedt, zoals verschillende versies voor teksten.

### *Scholing en opleiding*

Een opvallende, maar niet onverwachte, bevinding was dat de leerkrachten vertelden dat Engels een beperkt aandachtsgebied in de opleiding was. Ongeveer 20% van de leerkrachten heeft in de opleiding zelfs helemaal geen Engels gehad. Van het aanbod dat wel was genoten konden de leerkrachten niet duidelijk benoemen welke elementen relevant waren. Wel konden zij goed aangeven wat zij hadden gemist: er was geen aandacht voor de eigen vaardigheid en enkel beperkte aandacht voor didactische aspecten. Dit betekent dat leerkrachten in het PO wat betreft hun eigen Engelse vaardigheid moeten vertrouwen op de Engelse competenties die zij in het VO hebben geleerd. Al in hun rapport uit 2011 problematiseren Thijs en collega's of er wel vanuit gegaan kan worden dat studenten van de Havo (en vooral ook het MBO) over het toen nog geëiste B1-niveau beschikten, laat staan het huidige verwachte B2-niveau (Kennsbasis Pabo Engels, 2018).

De geïnterviewde leerkrachten hadden over het algemeen zelf een bovenmatige interesse in de Engelse taal en hebben hun eigen taalniveau met name zelf ontwikkeld door in hun privéleven veelvuldig in contact te komen met Engels. Zij voelden zich daarom voldoende toegerust om Engels te verzorgen in het basisonderwijs. Echter, de leerkrachten merkten wel op dat niet alle collega's op hun scholen dit ook zo ervoeren: niet elke leerkracht voelt zich competent genoeg om lessen Engels te geven, er heerst onzekerheid. Vergroting van de vaardigheden om Engels te geven kan natuurlijk plaatsvinden middels na-

/bijbscholing. Van de geïnterviewde leerkrachten heeft 38% aan extra scholing op het gebied van Engels deelgenomen, met name vvtto-leerkrachten. Dit percentage is aanzienlijk hoger dan dat in het hoofdonderzoek. Dit kan het gevolg zijn van de grote interesse voor Engels van de leerkrachten en de overrepresentatie van vvtto-leerkrachten in dit verdiepende onderzoek.

Naast de beperkte aandacht voor eigen taalvaardigheid in hun opleiding noemen de leerkrachten ook dat zij aandacht hebben gemist voor differentiatie en dat zij behoefte zouden hebben aan praktijkervaring in het geven van onderwijs van Engels.

### 10.11.2 Geconstateerde verbeterpunten

Het verdiepende onderzoek laat zien dat met name de visie en ideeën over hoe Engels op het PO vorm zou moeten krijgen aansluiten bij de taaltheoretische overwegingen over hoe beheersing van een vreemde taal het beste bevorderd kan worden (zie b.v. Westhoff, 2008). Een deel daarvan is ook teruggezien in de lessen van het merendeel van de leerkrachten. Echter, de praktijk blijkt weerbarstig. Er zijn namelijk meerdere verbeterpunten geconstateerd waar het gaat om de implementatie van het onderwijs in Engels in het PO. Deze verbeterpunten zijn over het algemeen al jaren bekend in het veld.

#### *Gebrek aan differentiatie en cyclisch werken*

Uit het hierboven beschrevene komt naar voren dat leerkrachten amper ingaan op verschillen tussen leerlingen, zoals ook in andere onderzoeken reeds naar voren is gebracht (bijvoorbeeld Thijs, et al., 2011). Leerkrachten zijn zich bewust van verschillen in de vaardigheidsniveaus van leerlingen, maar handelen hier niet of nauwelijks structureel naar. Een planmatige aanpak vindt met name plaats door extra stof voor goede leerlingen in een verrijkingsklas of extra werkbladen, opdrachten of (moeilijker/makkelijker) leesboekjes. Als leerkrachten verschil maken, dan heeft dat vooral een incidenteel karakter, is het gericht op de groepsamenstelling bij opdrachten en focust het zich op goede presteerders. Doordat de invloed van buitenschoolse omgang met Engels op het leerproces alom bekend is (Munoz, 2014), zijn niveauverschillen tussen leerlingen een bekend en reëel verschijnsel. Door hier niet systematisch op in te gaan wordt het leerpotentieel van (bijna) alle leerlingen onvoldoende bevorderd. Daarnaast dringt de vraag zich op voor welke leerlingen extra aandacht het meest wenselijk is. Gezien het belang van Engels in het VO zou het zinvol zijn meer aandacht te besteden aan remediërende activiteiten voor zwakke presteerders, opdat ook zij in het VO over voldoende basale vaardigheden beschikken om goed mee te kunnen komen.

Door cyclisch te werken kan een zinvolle omgang met verschillen tussen leerlingen worden geïmplementeerd. De inspecteurs gaven aan dat er van cyclisch werken, waarbij taaldoelen geëxpliciteerd én structureel geëvalueerd worden - leidend tot benodigde aanpassingen in aanbod, instructie en verwerking -, zeer beperkt sprake was. Als leerkrachten weinig zicht hebben op de mate waarin *alle beoogde taaldoelen* behaald worden, is het voor leerkrachten moeilijk daadwerkelijk aan te sluiten bij de instructiebehoeften van leerlingen.

Daarnaast blijkt uit de observaties dat het onderwijs Engels per school en per docent (sterk) verschillend is ingericht, ondanks het gebruik van een aantal methodes. Ook in verband met de aansluiting op het VO lijkt het daarom van belang om duidelijke doorlopende leerlijnen te ontwikkelen voor de ontwikkeling van het Engels en dat hierbij het eindniveau PO helderder wordt gedefinieerd. Op die manier heeft de leerkracht meer duidelijkheid naar welke minimale standaarden moet worden toegewerkt én welke vervolgoelen voor de goede presteerders relevant zijn.

### *Duidelijker focus op interactie*

De interviews laten zien dat bijna alle leerkrachten de nadruk leggen op mondelinge vaardigheid, al dan niet in combinatie met luistervaardigheid en/of woordenschat. Wanneer deze opvattingen worden gerelateerd aan de observaties, wordt dit beeld grotendeels bevestigd (zie hierboven). Echter, de weergegeven focus op communicatie kent wel een specifieke uitwerking in de geobserveerde lessen. De aandacht voor mondelinge vaardigheid richtte zich namelijk met name op uitspraak in plaats van op vrije spreek- of interactietaken. Leerlingen stelselmatig en veelvuldig uitdagen zich vrij uit te drukken lijkt daarom een verbeterpunt. Deze bevinding sluit aan bij de inventarisatie van het aanbod uit de vorige peiling, waarin een ruime aandacht werd gevonden voor het naspreken van woorden en het voeren van standaardgesprekjes (Geurts & Hemker, 2013; Rose, 2016). Maar ook in het VO is de tendens dat gesproken interactie nauwelijks aan bod komt (van Batenburg, 2018).

Binnen het huidige hoofdonderzoek geeft 55% van de leerkrachten dat zij 60-100% van de tijd leerlingen uitnodigt Engels met ze te spreken en 35% geeft aan dat leerlingen onderling van 30-60% van de les Engels met elkaar spreken. Hoe deze uitkomst moet worden geïnterpreteerd in het licht van dit verdiepende onderzoek is niet helemaal duidelijk, aangezien niet specifiek is gevraagd in hoeverre er daadwerkelijk sprake is van vrije interactie. Het lijkt relevant om in een volgende peiling sterker in te steken op de vraag in hoeverre er daadwerkelijk sprake is van vrije interactie en of leerlingen veelvuldig worden uitgenodigd tot 'vrij spreken' – hetzij met elkaar of met de leerkracht. In de bezochte lessen werd dit slechts in een kwart van de lessen gezien.

Willen leerkrachten bovendien bereiken dat leerlingen hun mondelinge vaardigheid optimaal ontwikkelen, dan zouden leerkrachten consequenter van leerlingen moeten verwachten dat zij in hun antwoorden, samenwerkingsopdrachten, etc. gebruik maken van de doeltaal. Op deze manier is er immers optimale inzet op (mondelinge) output (Westhoff, 2008, Dönszelmann, 2019).

### *Niveau van de doeltaal van leerkrachten*

Zoals vermeld gebruiken bijna alle geobserveerde leerkrachten de doeltaal in hun lessen Engels in meer of mindere mate. Niet alleen kan het gebruik van de doeltaal het aanleren van de doeltaal stimuleren (zie bijvoorbeeld Westhoff, 2008), ook is het niveau van dit doeltaalgebruik gerelateerd aan de leerlingprestaties: hoe hoger het Engelse niveau van de leerkracht, hoe hoger de taalvaardigheid van de leerlingen (Unsworth, Persson, Prins & De Bot, 2015). Bovendien blijkt een weloverwogen inzet van de doeltaal als *leertaal* gepaard te gaan met betere leerprestaties van de leerlingen (Dönszelmann, 2019). Voldoende beheersing van het Engels door de leerkracht is daarom erg belangrijk. De inspecteurs is gevraagd een inschatting te geven van de vloeiendheid en correctheid van het door de leerkrachten gebruikte Engels (in termen van onvoldoende-voldoende). Voor 77% van de leerkrachten werd het Engels als voldoende voor de doelgroep ingeschat. Dit betekent echter dat het Engelse niveau van ongeveer een kwart van de leerkrachten niet als zodanig kon worden betiteld. Deels was dit omdat de leerkracht überhaupt geen Engels sprak, maar ook omdat enkele leerkrachten relatief veel fouten maakten of vernederlandst Engels gebruikten.

Het op peil krijgen en bestendigen van het eigen taalniveau van de leerkracht is daarmee een duidelijk verbeterpunt, dat reeds vaker naar voren is gebracht (bijv. Thijs et al., 2011). Temeer daar het huidige peilingsonderzoek laat zien dat een aanzienlijk deel van de leerlingen een vaardigheidsniveau B1 of hoger beheerst (9% voor gespreksvaardigheid en hoger voor lezen en luisteren). Ook Bos en Van Baalen (2017) noemen het A2-B1-niveau als uitstroomprofiel van leerlingen. Dit vereist van leerkrachten dat hun eigen taalvaardigheidsniveau deze niveaus overstijgt. Gezien de selectieve keuze voor de onderzochte scholen (scholen met bovengemiddelde prestaties op Engels), is het twijfelachtig in hoeverre dit in de

Nederlandse basisscholen het geval is. Met het oog op het grote belang van voldoende taalvaardigheid van de leerkracht wijzen zowel de observaties als de interviews van het verdiepend onderzoek op de noodzaak van meer aandacht voor Engels in de opleiding van leerkrachten.

### *Engels in de opleiding*

Gekoppeld aan het vorige punt is natuurlijk de bevinding dat de geïnterviewde leerkrachten weinig scholing in Engels op de Pabo hebben gehad en dat deze scholing niet gericht was op het eigen taalniveau. Dit laatste weten de onderzochte leerkrachten te compenseren door veelvuldig gebruik van Engels in het privéleven, maar dat geldt zeker niet voor alle leerkrachten. De leerkrachten kunnen ook niet aanwijzen wat zij hebben geleerd tijdens de lessen Engels op de Pabo. Als Engels een volwaardig vak in het PO zou moeten zijn, zou een duidelijker inbedding van Engels in het curriculum essentieel zijn (Thijs et al., 2011). Een duidelijke aanbeveling die blijkt uit dit onderzoek is daarom de versterking van Engels in de opleiding van leerkrachten die Engels gaan verzorgen in het PO. Het stellen van aanvullende eisen voor deze groep docenten lijkt een zinvolle stap in de verbetering van het onderwijs van het Engels, zowel waar het gaat om hun eigen taalvaardigheid als om de specifieke vakdidactische vaardigheden.

### *Schoolbrede inzet op Engels*

Tenslotte lijkt de systematische omgang van scholen met het Engels voor verbetering vatbaar. Op het merendeel van de scholen is er volgens de inspecteurs geen sprake van een weloverwogen schoolbrede visie op Engels, waardoor doorgaande leerlijnen in gevaar komen. Binnen scholen worden regelmatig verschillen in handelen en denken beschreven. Slechts in drie gevallen kon een uitgesproken schoolvisie op Engels worden opgetekend. Juist in de gevallen waarin een school een dergelijke, specifieke koers uitzette leek er meer doelgerichtheid en eenheid in het handelen van leerkrachten zichtbaar. Dit wordt bevorderd doordat de leerkrachten gezamenlijk nadenken over keuzes en wat ze van leerlingen verwachten. Zoals reeds eerder vermeld, kunnen specifieke richtlijnen voor aanbod en doelstellingen, alsmede een vastgesteld niveau eind-PO, ook handvatten op schoolniveau geven voor het maken van zinvolle keuzes wat betreft het aanbod. De aanbeveling om heldere doelen voor het PO vast te stellen is ook van belang op het niveau van de school en het team.

Samengevat heeft de verdiepende studie geleid tot een vrij coherent beeld waarin voorop staat dat de leerkrachten hun lessen Engels (willen) focussen op een communicatieve aanpak, waarin veel aandacht is voor afwisseling en mondelinge vaardigheden. De praktijk is echter wat weerbarstiger, waardoor niet alle ideeën optimaal uit de verf komen. De studie levert daarmee een aantal, meer dan eens onderstreepte verbeterpunten op, met name gericht op het cyclisch werken, aandacht voor differentiatie en het daadwerkelijk inzetten op (vrije) interactie in de les, om leerlingen in de gelegenheid te stellen hun spreekvaardigheid te ontwikkelen. Tenslotte, maar zeker niet het minst belangrijk, is een gedegen opleiding op het gebied van de taalvaardigheid van de leerkrachten van groot belang voor succesvol onderwijs Engels in het PO.

### *Beperkingen van het onderzoek*

Deze studie kende een aantal beperkingen. Allereerst was het streven om de 32 best-presterende scholen uit het peilingsonderzoek te bezoeken. Door non-respons, gedeeltelijk als gevolg van de afname ná de zomervakantie en vanwege logistieke redenen, bleek dit niet mogelijk. Er zijn – in overleg met de opdrachtgever – minder scholen bezocht en slechts een deel van de bezochte scholen behoorde tot de categorie best-presterend. Dit heeft tot gevolg dat de bevindingen mogelijk niet (helemaal) representatief

zijn voor de praktijken in de zeer goede scholen en weliswaar *good* maar niet *best practices* weergeven. De verdeling van de scholen naar startmoment was evenredig, met 10 vvto-scholen en 11 eibo-scholen en 11 scholen met relatief weinig en 10 scholen met relatief veel gewichtenleerlingen. Doordat alle onderscheiden typen scholen zijn bezocht zijn de onderzoeksresultaten afkomstig van een breed palet aan scholen en daarmee relevant vanuit het oogpunt van herkenbaarheid. De generaliseerbaarheid van de resultaten is gezien de beperkte omvang van de onderzoeksgroep echter zeer beperkt

Ten tweede zijn, conform de opdracht, de schoolbezoeken uitgevoerd door onderwijsinspecteurs. Ook al is expliciet vermeld (door zowel onderzoekers als de inspecteurs zelf) dat zij in de rol van onderzoekers kwamen, toch zou de aanwezigheid van een inspecteur de getoonde lespraktijk en derhalve onze bevindingen kunnen hebben beïnvloed. De representativiteit van de uitkomsten is daarmee niet volledig gewaarborgd. Overigens zouden ook lesobservaties uitgevoerd door onderzoekers invloed kunnen hebben op de geobserveerde lespraktijk: In observatieonderzoek ligt het Hawthorne-effect altijd op de loer.

Een derde beperking is gelegen in het tijdstip van het onderzoek. Doordat scholen voorafgaand geselecteerd moesten worden en met name het beoordelen en analyseren van gespreksvaardigheid intensief en tijdrovend was, was afname pas aan het begin van het volgende schooljaar mogelijk. Dit betekende dat niet dezelfde klas geobserveerd werd, maar vooral ook dat de leerkrachten de klas vaak nog niet goed kenden en bepaald didactisch handelen misschien niet lieten zien, zoals bijvoorbeeld differentiatie. Ook zorgde dit tijdstip van afname voor het afvallen van bepaalde geselecteerde scholen.



## Hoofdstuk 11. Overzicht van databestanden en variabelen

In dit hoofdstuk geven we een beschrijving van de bestanden van de peiling Engels en de variabelen die in de verschillende bestanden zijn opgenomen. In de eerste paragraaf van dit hoofdstuk geven we een toelichting van de opbouw van de verschillende bestanden. Op basis daarvan kan een potentiële gebruiker bepalen welke bestanden voor hem of haar relevant zijn. In de tweede paragraaf geven we voor het hoofdbestand een overzicht van alle variabelen met een beschrijving van elke variabele. Dit overzicht kan een gebruiker helpen bij de selectie en interpretatie van relevante variabelen. Een overzicht van alle variabelen in de bestanden van de vragenlijstbestanden zijn gegeven in bijlage 15, 16 en 17. Daarnaast is een beschrijving van de variabelen die gebruikt zijn in het bestand voor de meerniveau-analyses gegeven in bijlage 11. Tenslotte is de bestandsopbouw van het bestand met daarin de paar-informatie voor de gespreksscores (interactie per paar), gegeven in bijlage 18.

### 11.1 Opbouw van databestanden

De peiling Engels heeft een set aan bestanden opgeleverd. Er is gekozen voor een hoofdbestand met een aantal deelbestanden vanwege het zeer grote aantal variabelen dat in de peiling is verzameld. Het samenvoegen van alle variabelen tot één bestand had geleid tot een niet werkbaar breed bestand.

Het belangrijkste bestand is het hoofdbestand. In dit bestand zijn de belangrijkste gegevens samengebracht die in de peiling Engels zijn verzameld. In het overzicht wat is opgenomen in paragraaf 11.2 is een extra kolom opgenomen die verwijst naar de bron waaruit de variabele is voortgekomen (bijvoorbeeld administratie van aangemelde leerlingen, toetsen, leerlingvragenlijst). Daarnaast wordt een afzonderlijk bestand opgeleverd voor de scores van de gespreksparen.

Een belangrijk onderdeel van de verschillende bestanden betreft de identificatie variabelen. Deze identificatie variabelen zijn beschikbaar op verschillende niveaus, namelijk op leerlingniveau in de vorm van 'LlId' en 'VrId'. Het eerste is het leerlingnummer dat aan de leerling is toegekend na aanmelding van de leerling door de school. Het tweede is de barcode van de door de leerlingen ingevulde vragenlijst. Op schoolniveau is School\_ID beschikbaar. Door middel van dit School ID-nummer kunnen gegevens van de schoolvragenlijst gekoppeld worden aan de gegevens van de leerlingen. Ook kan deze schoolidentificatie variabele gebruikt worden voor zogenaamde meerniveau analyse, aangezien deze weergeeft welke leerlingen samen op een school leskrijgen. Het derde niveau van identificatievariabelen betreft de klas (klas). In het hoofdbestand is gebruik gemaakt van een geanonimiseerde klasnaam. Deze variabele 'klas' kan gekoppeld worden aan de gegevens van de leerlingen en is tevens gebruikt in de meerniveau analyses ter identificatie van niveau 2. Voorts zijn er twee variabelen voor de identificatie van de leerkracht toegevoegd ('barc' en 'lkr\_klas'). De variabele 'barc', de identificatie van de leerkracht in het vragenlijstbestand, kan gebruikt worden om de gegevens uit de leerkrachtvragenlijst aan het totaalbestand te koppelen. Als op een school met meerdere groepen 8 slechts één leerkracht de vragenlijst heeft ingevuld, zijn deze leerkrachtgegevens ook voor de leerlingen in een of meer andere groep(en) op school gebruikt, vanwege de verwachte relatieve gelijkenis in antwoorden op een school. Uitzondering zijn scholen met meerdere groepen 8, waarbij minimaal twee leerkrachten de vragenlijst hadden ingevuld. In dat geval is de variabele 'barc' alleen ingevuld bij de leerlingen in de 'eigen klas'. Dit om te voorkomen dat er moest worden gekozen tussen leerkrachten. Omdat de vragenlijst niet door alle leerkrachten in de school is ingevuld, is tevens een

tweede leerkracht-identificatie toegevoegd ('lkr\_klas'). Hiermee kan de leerkrachtinformatie uit de vragenlijst specifiek aan de leerlingen in zijn/haar klas gekoppeld worden. De variabele 'lkr\_klas' is gelijk aan de variabele 'barc' maar toont alleen voor de leerlingen waarvan zeker is dat ze in de klas van de leerkracht zitten de barcode van de leerkracht.

## 11.2 Overzicht van variabelen in het hoofdbestand

Variabele	Label	Bron
LlId	Leerling ID	Admin.
VrId	Vragenlijst ID	Admin.
klas	Klasnaam	Admin.
School_ID	School ID	Admin.
Barc	Leerkracht ID (zoals in vragenlijst)	Admin.
Lkr_klas	Leerkracht ID (=barcode, maar alleen bij zekere koppeling leerkracht – leerlingen in de klas)	Admin.
Wegingsfactor	Gewicht op basis van Leerlingen (Populatie 2017-2018 Aantallen Alle 2088 lln, alle taken)	Admin.
CC4_SCH_POP18_ALL_VL_LE_WS	Gewicht op basis van Scholen Populatie 2017-2018 Aantallen Alle 95 scholen (VL,LE,WS)	Admin.
Leeftijd	Leeftijd leerling	School
LF_jaar_1okt2017	Leeftijd (in jaren) op 1 okt 2017	Admin.
Leertijd	Doorlopen schoolloopbaan op basis van leeftijd (regulier, vertraagd, vervroegd)	Admin.
geboortedag	Geboortedag	School
geboortemaand	Geboortemaand	School
geboortjaar	Geboortjaar	School
Geslacht	Geslacht	School
Leerlinggewicht	Leerlinggewicht	School
VO_adv_code	Advies VO (code)	School
VO_advies_cat	Voorlopig schooladvies, gecategoriseerd	Admin.
Ho_Mi_Zw	Door leerkracht ingeschat niveau Engels van de leerling (codes)	Admin.
VdhBgl_cat	Score begrijpend lezen Cito M7 – gecategoriseerd	School
vrl_dyslexie	Diagnose dyslexie	school
VVTO_code	VVTO-school (ja/nee)	School
Stratum_code	Stratum Code (percentage gewichenleerlingen: S1, S2, S3, S4)	DUO
Combi_code_8	Code clustering: VVTO (2) - Stratum (4)	Admin.
Combi_code_4	Code clustering: VVTO (2) - Stratum (2)	Admin.
stedelijk	Mate van stedelijkheid	DUO
Denom_1	Denominatie school	DUO
PROV_1	Provincie waar de school in staat	DUO
schoolregio	Regio waar de school in staat	DUO
d_TO	Minimaal 1 toets gedaan?	Toetsen



xE	Versie voor LE (SEL)	Toetsen
yE	Versie voor LE (<100)	Toetsen
PmLE	LE % ontbrekend	Toetsen
rLE	LE: Max # op een rij ontbrekend	Toetsen
WML_Lezen_x_serie	Vaardigheidsscore lezen, x-serie	Toetsen
WML_Lezen_y_serie	Vaardigheidsscore lezen, y-serie	Toetsen
SEWMLLEx	Vaardigheidsscore lezen SE, x-serie	Toetsen
LE_SEWMLy	Vaardigheidsscore lezen SE, y-serie	Toetsen
LE_ERK_wmlX	ERK-niveau lezen (A1, A1, A2, B1), x-serie	Toetsen
LE_ERK_wmlY	ERK-niveau lezen (A1, A1, A2, B1), y-serie	Toetsen
xU	Versie voor LU (SEL)	Toetsen
yU	Versie voor LU (<100)	Toetsen
PmLU	LU % ontbrekend	Toetsen
rLU	LU: Max # op een rij ontbrekend	Toetsen
WML_LUx	Vaardigheisscore luisteren, x-serie	Toetsen
LU_WMLy	Vaardigheisscore luisteren, y-serie	Toetsen
SEWMLLx	Vaardigheisscore luisteren SE, x-serie	Toetsen
LU_SEWMLy	Vaardigheisscore luisteren SE, y-serie	Toetsen
LU_ERK_wmlX	ERK-niveau luisteren (A1, A1, A2, B1), x-serie	Toetsen
LU_ERK_wmlY	ERK-niveau luisteren (A1, A1, A2, B1), y-serie	Toetsen
xW	Versie voor WS (SEL)	Toetsen
yW	Versie voor WS (<100)	Toetsen
PmWS	WS % ontbrekend	Toetsen
rWS	WS: Max # op een rij ontbrekend	Toetsen
WML_WSx	Vaardigheidsscore woordenschat, x-serie	Toetsen
WS_WMLy	Vaardigheidsscore woordenschat, y-serie	Toetsen
SEWMLWSx	Vaardigheidsscore woordenschat SE, x-serie	Toetsen
WS_SEWMLy	Vaardigheidsscore woordenschat SE, y-serie	Toetsen
GELIJKits	%woorden (zoals gerepresenteerd door de huidige items) - alle items even representatief	Toetsen
aPAR	%woorden (zoals gerepresenteerd door de huidige items) - representatief op a-parameter	Toetsen
GESPR_IIInr	Leerling met uitgevoerde gesprekstaken	Admin.
Toestemming_gesprekstaken	Toestemming van ouders voor gesprekstaken	School
gtaak_deel	Deelname aan de gesprekstaken	Taken
gtaak_anker	Deelname aan de ankertaak	Taken
GESPR_Bklet	Boekje (hangt af van Design): 1=alleen Taak 1, 2=1+2A, 3=1+2B, 4=1+2A+3A, 5=1+2A+3B, 6=1+2B+3A, 7=1+2B+3B	Taken
GESPR_NIts	Aantal opgaven (bk1: 57; bk2-3: 69; en bk4-7: 87)	Taken
GESPR_UScore	Ongewogen score	Taken
GESPR_WML	Vaardigheidsschatting GESPREEKKEN	Taken
GESPR_SEWML	Standaardfout vaardigheidsschatting GESPREEKKEN	Taken
GESPR_record	Locaal leerling ID (voor het record;	Taken

	gegarandeerd geen dubbelingen)	
GESPR_DS	ingeschat niveau Engels (pas invullen als alle taken van leerling gezien): 1=<A1, ..., 7=>B1 (zie invulformulier)	Taken
GESPR_DESIGN	Gemaakte taken (basis voor Boekje)	Taken
in_gesprek	Leerling heeft gesprek uitgevoerd (en een WML score)	Taken
GESPR_WML_ERK	ERK-niveau gespreksvaardigheid (A1, A1, A2, B1)	Taken
GESPREKSID	GESPREKS ID	Taken
Nkeer	Aantal keer beoordeeld	Taken
Beoordelaar	Beoordelaar van de gesprekstaken	Taken
Ntakengedaan	Aantal taken gedaan	Taken
NtakenOK	Aantal taken paar beoordeeld	Taken
Welke	Welke taken	Taken
Taak1	Taak 1	Taken
Taak2	Taak 2	Taken
Taak3	Taak 3	Taken
GT1_NlIn	Aantal leerlingen beoordeeld op taak 1	Taken
GT1_N_oor	Aantal oordelen taak 1	Taken
T1Sc1	Score 1 Taak1: mate van interactie	Taken
T1Sc2	Score 2 Taak1: mate van engels - mild	Taken
T1Sc3	Score 3 Taak1: mate van engels - strict	Taken
GT2_NlIn	Aantal leerlingen beoordeeld op taak 2	Taken
GT2_N_oor	Aantal oordelen taak 2	Taken
T2Sc1	Score 1 Taak2: mate van interactie	Taken
T2Sc2	Score 2 Taak2: mate van engels - mild	Taken
T2Sc3	Score 3 Taak2: mate van engels - strict	Taken
GT3_NlIn	Aantal leerlingen beoordeeld op taak 3	Taken
GT3_N_oor	Aantal oordelen taak 3	Taken
T3Sc1	Score 1 Taak3: mate van interactie	Taken
T3Sc2	Score 2 Taak3: mate van engels - mild	Taken
T3Sc3	Score 3 Taak3: mate van engels - strict	Taken
Attitude_best	Attitude van de leerlingen (1,3,4,10.1/2/3/6R)	Leerling
SelfConf_best	Linguistic Self Confidence van de leerlingen (LSC: 2,5,6,7,11)	Leerling
VL_BsG15	Buitenschools gebruik Engels, door de leerlingen, 15 items (12-22; zonder 17)	Leerling
VL_Bsg13	Buitenschools gebruik Engels door de leerlingen, 13 items (VL_BsG15 zonder 14.4 en 14.5)	Leerling
IV_LEZ	Ingeschatte vaardigheid Lezen door de leerlingen	Leerling
IV_LUI	Ingeschatte vaardigheid Luisteren door de leerlingen	Leerling
IV_SPR	Ingeschatte vaardigheid Spreken door de leerlingen	Leerling

IV_Sch	Ingeschatte vaardigheid Schrijven door de leerlingen (zonder item 29.09)	Leerling
IV_Engels	Ingeschatte vaardigheid (alle vaardigheden; excl. it.29.09)	Leerling
VL_ALL	Totaalscore vragenlijst	Leerling
scale_buitenschoolsgebruik	Buitenschools gebruik Engels door leerkracht (3_1-3_7)	Leerkracht
scale_buitenschoolsgebruik_zonder3_2	Buitenschools gebruik Engels door leerkracht (3_1-3_7, MINUS 3_2)	Leerkracht
schaal_zelfinschatting_spreken	Zelfinschatting leerkracht spreken	Leerkracht
schaal_zelfinschatting_schrijven	Zelfinschatting leerkracht schrijven	Leerkracht
schaal_zelfinschatting_lezen	Zelfinschatting leerkracht lezen	Leerkracht
schaal_zelfinschatting_luisteren	Zelfinschatting leerkracht luisteren	Leerkracht
ERKinschatting	Gemiddelde ERK inschatting	Leerkracht
methoden_ingedikt_nw_v5_1	Gebruikte methoden - ingedikt	Leerkracht
CC8_type	Type school EIBO/VVTO	School
CC8_GewStrat	Gewichtenstratum	DUO
CC8_LLN_POP18_ALL	Gewicht op basis van Leerlingen Populatie 2017-2018 Aantallen Alle 2088 Iln	Admin.
CC4_type	Type school EIBO/VVTO	School
CC4_GewStrat	Gewichtenstratum	Admin.



## Literatuurlijst

- Andrews, R. (2006). The effect of grammar teaching on writing development. *British Educational Research Journal*, 32, 39-55.
- Bachman, L. F. (1990). *Fundamental consideration in language testing*. Oxford: Oxford University Press.
- Barneveld, van, F., Sanden, van der, B. (1987). Engels in het basisonderwijs: een leerlinggericht begin! *Levende Talen*, 421, 339-344.
- Batty, A. O. (2015). A comparison of video-and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing*, 32(1), 3-20.
- Bodde, M., & Schokkenbroek, J. (2013). Scenario's voor Engels in het primair onderwijs: gevolgen voor de brugklas. Opgehaald van *Levende Talen Magazine*: <http://www.lt-tijdschriften.nl/ojs/index.php/lrm/article/view/561>
- Bos, L., & Baalen, E. (2017). Hoe vaardig zijn onze groep 8-leerlingen in Engels? Inzicht in de ERK-niveaus aan de hand van uitslagen van de IEP Eindtoets Engels. *Levende Talen Tijdschrift*, 18(2), 31-38. Opgehaald van *Levende Talen Tijdschrift*: <http://www.lt-tijdschriften.nl/ojs/index.php/lrm/article/view/1689>
- Buck, G. (1991). The testing of listening comprehension: an introspective study. *Language Testing*, 8(6791).
- Butler, Y. G., & Zeng, W. (2014). Young foreign language learners' interactions during task-based paired assessments. *Language Assessment Quarterly*, 11(1), 45-75.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1. 1-47.
- Celce-Murcia, M. (2008). Rethinking the role of communicative competence in language teaching. In *Intercultural language use and language learning* (pp. 41-57). Springer, Dordrecht.
- Chan, H., Lowie, W., & de Bot, K. (2016). Cognitive perspectives on bilingualism. *Trends in Applied Linguistics*, 182-203.
- Clément, R., & Kruidenier, B. G. (1985). Aptitude, Attitude and Motivation in Second Language Proficiency: A Test Of Clément's Model. *Journal of Language and Social Psychology*, 4(1), 21-37. doi:10.1177/0261927X8500400102
- Clément, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, Self-confidence, and Group Cohesion in the Foreign Language Classroom. *Language Learning*, 44(3), 417-448. doi:10.1111/j.1467-1770.1994.tb01113.x
- Council of Europe (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) - A Manual*. Opgehaald van <https://rm.coe.int/1680667a2d>
- Council of Europe (2011). *Manual for Language Test Development and Examining for use with the CEFR*. Opgehaald van <https://rm.coe.int/manual-for-language-test-development-and-examining-for-use-with-the-ce/1680667a2b>

- Council of Europe, Language Policy Programme (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Companion Volume with New Descriptors*. Opgehaald van <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Council of Europe, Bank of Supplementary Description. (2018). *Collated Representative Samples of Descriptors of Language Competences Developed for Young Learners - Resource for Educators*. Opgehaald van: <https://www.coe.int/en/web/common-european-framework-reference-languages/bank-of-supplementary-descriptors>
- Corda, A., Kraay, T. de, & Feuerstake, M. (2012). *And yet all different: Handreiking voor een curriculum Engels op de pabo*. Leiden: Expertisecentrum Moderne Vreemde Talen.
- Corda, A., Philipsen, K., & Graaff, R. de. (2014). *Handboek VVTO*. Bussum: Coutinho.
- Damhuis, R., & Litjens, P. (2003). *Mondelinge communicatie. Drie werkwijzen voor mondelinge taalontwikkeling*. Nijmegen: Expertisecentrum Nederlands.
- De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34, 5 – 34.
- Eggen, T.J.H.M. (2004). Contributions to the theory and practice of computerized adaptive testing. (Doctoral thesis Twente University Enschede).
- Ellis, R. (1998). Teaching and Research: Options in Grammar Teaching. *TESOL Quarterly*, 32, 39-60.
- European Commission: Education and Training. (sd). *First European Survey on Language Competences: Final Report*. Opgehaald van [http://ec.europa.eu/dgs/education\\_culture/repository/languages/policy/strategic-framework/documents/language-survey-final-report\\_en.pdf](http://ec.europa.eu/dgs/education_culture/repository/languages/policy/strategic-framework/documents/language-survey-final-report_en.pdf)
- Evers, A., Lucassen, W., Meijer, R.R. & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de Kwaliteit van Test*. COTAN ( <https://www.psynip.nl/wp-content/uploads/2016/07/COTAN-Beoordelingssysteem-2010.pdf> )
- Dörnyei, Z., & Scott, M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language Learning*, 47, 173-210.
- Fasoglio, D., & Smulders, C. (2015). *Handreiking schoolexamen Chinese Taal en Cultuur vwo – Variant elementair*. Enschede: SLO
- Fasoglio, D., & Tuin, D. (2017). *Eindniveau gespreksvaardigheid Engels in het voortgezet onderwijs*. Enschede: SLO. Opgehaald van <http://downloads.slo.nl/Repository/eindniveau-gespreksvaardigheid-engels-in-het-voortgezet-onderwijs.pdf>
- Flowerdew, J. M. & Miller, L. (2005). *Second language listening: theory and practice*. Cambridge University Press.
- Fulcher, G. (2010). *Practical Language Testing*. Londen : Hodder Education
- Galaczi, E. D. (2008). Peer–peer interaction in a speaking test: the case of the First Certificate in English examination. *Language Assessment Quarterly* 5. 89–119.

- Galaczi, E. D. (2014). Interactional competence across proficiency levels: how do learners manage interaction in paired speaking tests? *Applied Linguistics* 35. 553–74.
- Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, 19, 3 – 17.
- Geurts, B., & Hemker, B. (2013). *Balans van het Engels aan het einde van de basisschool 4: uitkomsten van de vierde peiling in 2012*. Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.
- Graham, S., Santos, D., & Francis-Brophy, E. (2014). Teacher beliefs about listening in a foreign language. *Teaching and Teacher Education*, 40, 44-60.
- Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing*, 32(3), 317-336.
- Hedges, L. V. (2007). Effect sizes in cluster randomized designs. *Journal of Behavioral and Educational Statistics*, 32, 341-370
- Keuning, J., Straat, H., and R.C.W. Feskens (2017). The Data-Driven Direct Consensus (3DC) Procedure: A New Approach to Standard Bepaling. In *Standard Bepaling – International State of Research and Practices in the Nordic Countries* (Eds: Sigrid Blömeke and Jan-Eric Gustafsson). Springer series Methodology of Educational Measurement and Assessment, pp 263-278).
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: University Press.
- Koert-Hoogervorst, van, M., Rispens, J., Leona, N., van der Molen, M., & Snellings, P. (09-06-2017). *Het ORWELL project: voorspellen en stimuleren van Engels leren op de basisschool*. Universiteit van Amsterdam, Rudolf Berlin Center. Opgehaald van ORWELL: <https://www.lbbo.nl/userdata/Brochure-ORWELL.pdf>
- Kormos, J. (2014). *Speech production and second language acquisition*. Routledge.
- Krom, R., Ouborg, M., & Kamphuis, F. (2001). *Wetenschappelijke verantwoording van de toetsseries Luisteren 1, 2 en 3. Leerlingvolgsysteem*. Arnhem: Citogroep.
- Kuysk, M., & Socket, G. (2012). From Informal Resource Usage to Incidental Language Acquisition: Language Uptake from Online Television Viewing in English. *ASp*, 45-65.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19, 255–271. Retrieved from <http://applied.oxfordjournals.org/cgi/reprint/19/2/255>
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16, 33-51.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399–436.
- Lee, J., & Schallert, D. L. (1997). The Relative Contribution of L2 Language Proficiency and L1 Reading Ability to L2 Reading Performance: A Test of the Threshold Hypothesis in an EFL Context. *TESOL Quarterly*, 31, 713-739. <http://dx.doi.org/10.2307/3587757>
- Levelt, W. (1989). *Speaking. From Intention to Articulation*. Cambridge, Mass.: MIT.

- Lightbown, P.M. (2014). *Focus on Content-based Language Teaching*. Oxford: Oxford University Press.
- Lightbown, P.M., & Spada, N., (2006) *How Languages are Learned: An introduction to the main theories of first and second language acquisition* (3e editie). Oxford: Oxford Handbooks for Language Teachers.
- Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, 105-129.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of language acquisition*. Vol. 2: Second language acquisition (pp. 413 – 468 ). New York, NY : Academic Press.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of Form in Communicative Classrooms. *Studies in Second Language Acquisition*, 19(1), 37-66. Meijerink, H., e.a. (2012). *Een goede basis: Advies van de Commissie Kennisbasis Pabo*. Den Haag: HBO-raad.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing Willingness to Communicate in a L2: A Situational Model of L2 Confidence and Affiliation. *The Modern Language Journal*, 82(4), 545-562. doi:10.1111/j.1540-4781.1998.tb05543.x
- Mackey, A. (1999). Input, interaction, and second language development. *Studies in Second Language Acquisition*, 21, 557-587.
- Masgoret, A. —, & Gardner, R. C. (2003). Attitudes, Motivation, and Second Language Learning: A Meta-Analysis of Studies Conducted by Gardner and Associates. *Language Learning*, 53(1), 123-163. doi:10.1111/1467-9922.00212
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Meara, P. (1990). A note on passive vocabulary. *Second Language Research*, 6, 150-154.
- Muñoz, C., Casas, I., & Cadierno, T. (2018). Different starting points for English language learning: a comparative study of Danish and Spanish young learners. *Language learning: a journal of research in language studies*, 1-34.
- O'Sullivan, B., & Dunlea, J. (2015). *Aptis General Technical Manual*. Version 1.0. Londen: British Council.
- Oskam, S. (2008). *Praktische didactiek voor Engels in het basisonderwijs*. Bussum: Coutinho.
- Palmberg, R. (1987). Patterns of vocabulary development in foreign language learners. *Studies in Second Language Acquisition*, 9, 202-221.
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167–208). New York, NY: Oxford University Press. (PDF) *Word Knowledge in a Theory of Reading Comprehension*. Available from: [https://www.researchgate.net/publication/263247534\\_Word\\_Knowledge\\_in\\_a\\_Theory\\_of\\_Reading\\_Comprehension](https://www.researchgate.net/publication/263247534_Word_Knowledge_in_a_Theory_of_Reading_Comprehension) [accessed Nov 20 2018].
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10, 355-371.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The Acquisition of Reading Comprehension Skill doi:10.1002/9780470757642.ch13



- Rose, P. (2016). *Engels in het basisonderwijs: domeinbeschrijving ten behoeve van peilingsonderzoek*. Enschede: SLO. Opgehaald van SLO: <http://mvt.slo.nl/engels-in-het-basisonderwijs-domeinbeschrijving-ten-behoeve-van-peilingsonderzoek>
- Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar. In J. M. Norris & L. Ortega (Eds.), *Synthesizing Research on Language Learning and Teaching*. Amsterdam/Philadelphia: Benjamins.
- Saffran, J. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, 12(4), 110-114.
- Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations. *Studies in Second Language Acquisition*, 38(4), 677-701.
- Sandlund, E., Sundqvist, P., & Nyroos, L. (2016). Testing L2 Talk: A Review of Empirical Studies on Second-Language Oral Proficiency Testing. *Language and Linguistics Compass*, 10(1), 14-29.
- Sijtsma, K. & Molenaar, I.W. (2002). *Introduction to nonparametric item response theory*. London: Sage.
- SLO (sd). *Kerdoelen Engels voor het primair onderwijs*. Opgehaald van <http://tule.slo.nl/Engels/F-KDEngels.html>
- SLO (2011). *Toetsen en beoordelen met het ERK*. Opgehaald van: <http://www.erk.nl/docent/toetsing/toetsen-en-beoordelen-met-het-ERK.pdf>
- SLO (2015). *Taalprofielen 2015*. Opgehaald van <http://downloads.slo.nl/Repository/taalprofielen-2015.pdf>
- SLO (2018). *Startnotitie Engels: meewerken aan het onderwijs van morgen*. Opgehaald van Curriculum.nu: <https://curriculum.nu/wp-content/uploads/2018/03/Startnotitie-Engels-MVT.pdf>
- Thijs, A., Tuin, D., & Trimbo, B. (2011). *Engels in het basisonderwijs. Verkenning van de stand van zaken*. Enschede: SLO
- Thijs, A. & Tuin, D. (2015). Verkenning van competenties voor leerkrachten Engels in het basisonderwijs. SLO: Enschede.
- Tomlinson, C.A., Brighton, C., Hertberg, H., Callahan, C.M., Moon, T.R., Brimijoin, K., Conover, L.A., & Reynold, T. (2003). Differentiating Instruction in Response to Student Readiness, Interest, and Learning Profile in Academically Diverse Classrooms: A Review of Literature. *Journal for the Education of the Gifted*, Vol. 27, No. 2/3, pp. 119-145.
- Unsworth, S., Persson, L., Prins, T., & De Bot, K. (2014). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, 36(5), 527-548.
- Van de Grift, W. J. C. M., Van der Wal, M., & Torenbeek, M. (2011). Ontwikkeling in de pedagogische didactische vaardigheid van leraren in het basisonderwijs. *Pedagogische Studiën*, 88, 416 - 432.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language teaching*, 40(3), 191-210.

- Vandergrift, L., & Goh, C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. New York: Routledge.
- Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, 65(2), 390-416.
- Vandergrift, L., & Baker, S. C. (2018). Learner variables important for success in L2 listening comprehension in French immersion classrooms. *Canadian Modern Language Review*, 74(1), 79-100.
- Verhelst N.D., Glas C.A.W. (1995) The One Parameter Logistic Model. In: Fischer G.H., Molenaar I.W. (eds) Rasch Models. Springer, New York, NY.
- Verhelst, N.D., Glas, C.A.W., and Verstralen, H.H.F.M. (1995). *OPLM: One Parameter Logistic Model: Computer Program and Manual*. Arnhem: Cito.
- Verspoor, M.H., Schuitemaker-King, J., Rein, E.M.J. van, Bot, K. de & Edelenbos, P. (2010). *Tweetalig onderwijs: vormgeving en prestaties*. Groningen: Rijksuniversiteit Groningen.
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79–95.
- Westhoff, G. J. (1981). Voorspellend lezen. Een didactische benadering van de leesvaardigheidstraining in het moderne-vreemdetalenonderwijs. Groningen: Wolters Noordhoff.
- Westhoff, G.J. (2008). *Een schijf van vijf voor het vreemdetalenonderwijs (revisited)*. Enschede: NaB-MVT.
- Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faisca, L., ... Blomert, L. (2010). Orthographic Depth and Its Impact on Universal Predictors of Reading: A Cross-Language Investigation. *Psychological Science*, 21(4), 551–559. <https://doi.org/10.1177/0956797610363406>

## Bijlage 1. Scoreformulier gesprekstaken

### Taak 1: Toelichting Realisatie Kwaliteit en I&F

<b>TOELICHTING Realisatie kwaliteit (woordenschat, grammatica , samenhang)</b>	
<b>A1</b>	Heeft een zeer elementaire woordenschat die bestaat uit geïsoleerde woorden en eenvoudige uitdrukkingen met betrekking op persoonlijke gegevens en bepaalde concrete situaties.
	Toont slechts beperkte beheersing van enkele eenvoudige grammaticale constructies en zinspatronen in een geleerd repertoire
	Kan woorden of woordgroepen met elkaar verbinden met heel elementaire lineaire verbindingswoorden als 'en' of 'dan'

<b>Realisatie interactie en fluency</b>	
<b>A1</b>	Kan vragen stellen en beantwoorden over persoonlijke <b>TOELICHTING</b> gegevens.
	Kan op een simpele manier interactief zijn maar de communicatie is totaal afhankelijk van herhaling, herformulering en correcties. Kan aangeven wat hij of zij bedoelt door het aan te wijzen.
	Kan zeer korte, geïsoleerde, voornamelijk kant-en-klare uitingen hanteren, met veel pauzes om naar uitdrukkingen te zoeken, minder vertrouwde woorden uit te spreken en de communicatie te herstellen

## Taak 2: Toelichting Realisatie Kwaliteit en I&F

TOELICHTING Realisatie kwaliteit (woordenschat, grammatica, samenhang)		TOELICHTING Realisatie interactie en fluency	
A2 en A2+	Beschikt over voldoende <b>woordenschat</b> om zich te redden bij primaire levensbehoeften. Kan een beperkt repertoire hanteren met uit het hoofd geleerde uitdrukkingen om te voorzien in concrete alledaagse behoeften	A2 en A2+	kan op eenvoudige maar <u>doeltreffende wijze</u> sociale contacten aangaan met gebruikmaking van de eenvoudigste gangbare uitdrukkingen en door elementaire routines te volgen. (A2+) Kan zich in <u>korte</u> bijdragen verstaanbaar maken, ondanks <u>korte stiltes</u> , valse starts en herformuleringen (A2+).
	Beschikt over voldoende <b>woordenschat</b> om alledaagse handelingen uit te voeren die betrekking hebben op vertrouwde situaties en onderwerpen (A2+).		Kan voor korte interacties <u>zeer korte zinsdelen</u> over vertrouwde onderwerpen met voldoende gemak gebruiken, ondanks heel duidelijke aarzelingen en valse starts. Kan <u>zeer korte</u> sociale contacten hanteren en daarbij gebruik maken van alledaagse beleefdheidsvormen.
	<b>Grammatica</b> -Gebruikt een aantal eenvoudige constructies correct, maar maakt nog stelselmatig elementaire fouten – bijvoorbeeld door verschillende tijden door elkaar te gebruiken en niet te letten op congruentie; toch is meestal wel duidelijk wat hij of zij probeert te zeggen.		Kan vragen beantwoorden en reageren op eenvoudige uitspraken. Kan aangeven wanneer hij of zij het gesprek volgt maar is zelden in staat genoeg te begrijpen om uit zichzelf de conversatie gaande te houden.
	<b>Samenhang</b> - kan woordgroepen verbinden met eenvoudige verbindingswoorden als 'en', 'maar' en 'omdat'		☒ interactiestrategieën Kan communicatie in stand houden door: ☒ om aandacht te vragen; ☒ te zeggen dat hij of zij iets in het gesprek niet kan volgen; ☒ te vragen om een langzamer spreektempo, herhaling of uitleg, eventueel met behulp van gebaar en mimiek. ☒ gebruik te maken van 'fillers', zoals 'na ja', 'ehm', en stopwoorden zoals 'soort van', 'kijk' enz.
	<b>Samenhang</b> - kan de meest voorkomende verbindingswoorden gebruiken om enkelvoudige zinnen te verbinden om een verhaal te vertellen of iets te beschrijven als een eenvoudige opsomming van punten. (A2+)		Kan de betekenis van een niet passend woord met gebaren verduidelijken. Kan redelijk gebruik maken van een overkoepelend begrip ('fruit' voor 'orange'). Kan een woord uit de moedertaal 'verbuitenlandsen' en om bevestiging vragen. (A2)

**Taak 3: Toelichting Realisatie Kwaliteit en I&F**

TOELICHTING Realisatie kwaliteit (woordenschat, grammatica, samenhang)		TOELICHTING Realisatie interactie en fluency	
<b>B1</b>	Beschikt over een voldoende <b>woordenschat</b> om zich, met enige omhaal van woorden, te uiten over onderwerpen die betrekking hebben op het dagelijks leven, zoals familie, vrijetijdsbestedingen, werk, reizen en actualiteiten	<b>B1</b>	Is goed te volgen, hoewel in langere stukken tekst pauzes voorkomen om na te denken over grammatica en woordkeuze en fouten te herstellen.
	Heeft een goede beheersing van elementaire <b>woordenschat</b> , al doen zich nog wel grote fouten voor bij meer complexe gedachten of niet vertrouwde onderwerpen en situaties.		Kan een eenvoudig rechtstreeks gesprek beginnen, gaande houden en besluiten over onderwerpen die vertrouwd of van persoonlijk belang zijn. Kan gedeeltelijk herhalen wat iemand anders heeft gezegd om wederzijds begrip te bevestigen.
	<b>Grammatica</b> - Maakt met een redelijke mate van nauwkeurigheid gebruik van een repertoire van veelgebruikte 'routines' en patronen die bekend zijn van meer voorspelbare situaties.		Kan een eenvoudig woord gebruiken waarvan de betekenis lijkt op het begrip dat hij of zij wil overbrengen, en vraagt daarbij om 'correctie'. Kan opnieuw beginnen met een andere tactiek wanneer de communicatie mislukt.
	<b>Samenhang</b> - Kan een reeks kortere, op zich zelf staande eenvoudige elementen verbinden tot een samenhangende lineaire opeenvolging van punten.		Kan anderen vragen te verhelderen of uit te werken wat zij zojuist gezegd hebben. Kan een woord omschrijven als hij/zij zelf niet op het woord kan komen, eventueel gebruik makend van gebaren en mimiek (bijvoorbeeld: vrachtwagen voor mensen = bus). Kan gedeeltelijk herhalen wat iemand heeft gezegd om wederzijds begrip te bevestigen of om (denk)tijd te winnen.

Schoolnummer

--	--	--	--

Leerlingnummer

--	--	--	--

(in opname en/ in bestandsnaam)

Gesprek

1 2 3

Initialen beoordelaar

--	--	--

Bij de leerling zijn de volgende taken (deels) afgenomen:

		Samen met andere leerling:	Alleen				
1 – Wie ben ik?	Ja <input type="checkbox"/> <input type="checkbox"/> nee	<table border="1"><tr><td></td><td></td><td></td><td></td></tr></table>					<input type="checkbox"/>
2 – Wat spreken we af?	Ja <input type="checkbox"/> <input type="checkbox"/> nee	<table border="1"><tr><td></td><td></td><td></td><td></td></tr></table>					<input type="checkbox"/>
3 – Wat vind jij er van?	ja <input type="checkbox"/> <input type="checkbox"/> nee	<table border="1"><tr><td></td><td></td><td></td><td></td></tr></table>					<input type="checkbox"/>

Geslacht                      jongen                           meisje

Positie leerling op video    links                           rechts

Rol leerling                      Roy/Rosa                           Leo/Lisa

Leerlingnummer in opname    ja                           nee                       niet aanwezig op opname

Datum opname (bestand)                      

--	--

--	--

 2018

**Taak 1 – Wie ben ik? (observatie)**

   ja    nee    Niet aanwezig op de  
                        opname

Instructiefilm gebruikt           

.....                         

Opmerkingen: 

--

Taak 1: Wie ben ik?		Niet kunnen	Inhoud				Kwaliteit			Interactie & Fluency		
			0	1	2		0	1	2	0	1	2
A1-1av	vragen naar naam	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A1-1bv	vragen naar leeftijd	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A1-1cv	vragen naar verjaardag	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A1-1aa	vertellen hoe je heet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A1-1ba	vertellen hoe oud je bent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A1-1ca	vertellen wanneer je verjaardag is	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A1-2a	vertellen over familie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A1-2b	vertellen over huisdieren	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A1-3av	vragen wat je lekker vindt en wat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A1-3ba	vertellen wat je lekker vindt en	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A2-4av	vragen naar sport/hobby	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A2-4bv	vragen hoe vaak/wanneer de	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A2-4cv	vragen naar reden voor	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A2-4dv	vragen naar bijzonderheden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A2-4aa	vertellen over sport/hobby	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A2-4ba	Vertellen hoe vaak/wanneer je	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A2-4cc	vertellen waarom sport/hobby	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
A2-4dc	vertellen meer details	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

**Inhoudselement:** 0= niet gerealiseerd (stilte, volledig NL) óf onvoldoende; 1 = onvolledig (half NL, half EN en/of niet alles verteld); 2 = volledig

**Realisatie kwaliteit & interactie en fluency:** 0= onvoldoende; 1= voldoende; 2 = goed (zie toelichting)

**Taak 2 – Wat spreken we af? (observatie)** ja ne Niet aanwezig op de  
e opnamze

Instructie volgens protocol

Opmerkingen:

Taak 2. Wat spreken we af?		Niet kunnen	Inhoud			kwaliteit			Interactie & Fluency		
			0	1	2	0	1	2	0	1	2
Roy/Rosa	voorstellen om af te spreken	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Roy/Rosa	schooltijden, voorstel vanmiddag	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Roy/Rosa	reageert leuk. Hoe laat avondeten?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Roy/Rosa	kan niet. Hond uitlaten en daarna oma	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Roy/Rosa	reageert jammer. Oppassen. Kan wel bij mij.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Leo/Lisa	instemmen met afspreken. Vragen wanneer. Druk!	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Leo/Lisa	schooltijden, gitaarles	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Leo/Lisa	half 6 eten, trainen, daarna afspreken?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Leo/Lisa	reageert leuk. Bij mij thuis, na gitaarles?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Leo/Lisa	goed idee. Spreek af 4 uur	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Inhoudselement:** 0= niet gerealiseerd (stilte, volledig NL) óf onvoldoende; 1 = onvolledig (half NL, half EN en/of niet alles verteld); 2 = volledig

**Realisatie kwaliteit & interactie en fluency:** 0= onvoldoende; 1= voldoende; 2 = goed (zie toelichting)



**Taak 3 – Wat vind jij ervan? (observatie)** ja    nee    Niet aanwezig op de opname

Instructie volgens protocol                             

Opmerkingen:

Taak 3. Wat vind jij ervan?		Niet kunnen waarnemen	Inhoud			Kwaliteit			Interactie & Fluency		
			0	1	2	0	1	2	0	1	2
<b>Rol 1: vóór</b>											
1A. argument	Gamen is leerzaam want: snelheid, probleem oplossen, creatief/fantasie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2B. reactie	Valt best mee	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2B. tegenargument	Valt best mee want: licht aandoen, bij het raam zitten, paar uur per dag kan	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3B. reactie	Reactie: Onzin	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3B. tegenargument	Genoeg beweging naar school fietsen, gymnastiek	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4A. argument	Groot voordeel: online vrienden, makkelijk mensen ontmoeten, contacten over hele wereld, letten niet op uitdijkt, bla bla bla	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Rol 2: tegen</b>											
1B. reactie	Niet mee eens	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1B. tegenargument	Gebruikt geen fantasie, beter een boek lezen, eigen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2A. argument	Gehoord dat: slecht voor je ogen, hoofdpijn	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3A. argument	Argument: ongezond. Niet alleen ogen, ook slecht voor armen/rug, lui, slechte	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4B. reactie	Helemaal niet waar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4B. tegenargument	Geen tijd voor je vrienden op school, altijd alleen, beter afspreken met echte vrienden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Inhoudselement:** 0= niet gerealiseerd (stilte, volledig NL) óf onvoldoende; 1 = onvolledig (half NL, half EN en/of niet alles verteld); 2 = volledig

**Realisatie kwaliteit & interactie en fluency:** 0= onvoldoende; 1= voldoende; 2 = goed (zie toelichting)

## Bijlage 2. Beoordelingsformulieren voor parenbeoordeling

### Taak 1:

Kenmerken van het gesprek - 1 formulier per taak invullen								
Code: xxxx_g#_yyyy_en_zzzz (schoolnummer;gespreknummer;leerlingnrA; leerlingnummerB)								
Taak uitgevoerd door twee leerlingen?	Invullen: Ja =1, nee=0							
De observatie heeft uitsluitend betrekking op het gedrag van de kinderen als ze allebei deelnemen aan de gesprekstaak. Niet scoren als een kind met de testleider de gesprekstaak uitvoert.								
<b>GESPREKSTAAK 1</b>								
		<b>Subtaak / beurtenwisseling</b>						
<b>Tweetal in gesprek</b>	<b>1</b>	<b>ST 1.01</b>	<b>ST 1.02</b>	<b>ST 1.03</b>	<b>ST 1.04</b>			
		aantal personen	aantal personen	aantal personen	aantal personen	<b>Somscore (data)</b>	<b>"Items"</b>	<b>score</b>
(1) Volledig Engelstalig						0	GT1_1	0 t/m 4
(2) Deels Engelstalig, deels non-verbaal						0	GT1_2	0 t/m 4
(3) Deels Engelstalig, deels Nederlandstalig						0	GT1_3	0 t/m 4
(4) Deels Engelstalig, deels Nederlandstalig, deels non-verbaal						0	GT1_4	0 t/m 4
(5) Deels Nederlandstalig, deels non-verbaal						0	GT1_5	0 t/m 4
(6) Volledig non-verbaal						0	GT1_6	0 t/m 4
(7) Volledig Nederlandstalig						0	GT1_7	0 t/m 4
(8) Geheel afwezig						0	GT1_8	0 t/m 4
		0	0	0	0	<b>CHECK</b>		
		check!	check!	check!	check!			

## Taak 2:

Kenmerken van het gesprek - 1 formulier per taak invullen									
Code: xxxx_g#_yyyy_en_zzzz (schoolnummer;gespreknummer;leerlingnrA; leerlingnummerB)									
Taak uitgevoerd door twee leerlingen? Invullen: Ja =1, nee= 0									
De observatie heeft uitsluitend betrekking op het gedrag van de kinderen als ze allebei deelnemen aan de gesprekstaak. Niet scoren als een kind met de testleider de gesprekstaak uitvoert.									
<b>GESPREKSTAAK 2</b>									
		<b>Subtaak / beurtenwisseling</b>							
tweetal in gesprek	<b>1</b>	<b>ST 2.01</b> (afspreken & druk)	<b>ST 2.02</b> (vanmiddag & gitaar)	<b>ST 2.03</b> (avondeten & trainen)	<b>ST 2.04</b> (hond & bij mij)	<b>ST 2.05</b> (opassen & 4 uur)			
		aantal personen	aantal personen	aantal personen	aantal personen	aantal personen	<b>Somscore (data)</b>	<b>"Items"</b>	<b>score</b>
(1) Volledig Engelstalig							0	GT1_1	0 t/m 5
(2) Deels Engelstalig, deels non-verbaal							0	GT1_2	0 t/m 5
(3) Deels Engelstalig, deels Nederlandstalig							0	GT1_3	0 t/m 5
(4) Deels Engelstalig, deels Nederlandstalig, deels non-verbaal							0	GT1_4	0 t/m 5
(5) Deels Nederlandstalig, deels non-verbaal							0	GT1_5	0 t/m 5
(6) Volledig non-verbaal							0	GT1_6	0 t/m 5
(7) Volledig Nederlandstalig							0	GT1_7	0 t/m 5
(8) Geheel afwezig							0	GT1_8	0 t/m 5
		0	0	0	0	0			
		check!	check!	check!	check!	check!	<b>CHECK</b>		

### Taak 3:

Kenmerken van het gesprek - 1 formulier per taak vullen

Code: xxxx\_g#\_yyyy\_en\_zzzz (schoolnummer;gespreknummer;leerlingnrA; leerlingnummerB)

Taak uitgevoerd door twee leerlingen? ja=1, nee=0

De observatie heeft uitsluitend betrekking op het gedrag van de kinderen als ze allebei deelnemen aan de gesprekstaak. Niet scoren als een kind met de testleider de gesprekstaak uitvoert.

#### GESPREKSTAAK 3

**Merk op:** per subset komt telkens 1 van de sprekers 2 keer aan het woord (Argument en Tegenargument), en 1 spreker 1 keer (Reactie); wie twee keer aan het woord is en wie 1 keer aan het woord is wisselt per subtaak

		Subtaak / beurtenwisseling				Somscore (data)	"Items"	score
Tweetal in gesprek	1	ST 3.01 (Ar-Re-Tr) aantal personen	ST 3.02 (Ar-Re-Te) aantal personen	ST 3.03 (Ar-Re-Te) aantal personen	ST 3.04 (Ar-Re-Te) aantal personen			
(1) Volledig Engelstalig						0	GT3_1	0 t/m 4
(2) Deels Engelstalig, deels non-verbaal						0	GT3_2	0 t/m 4
(3) Deels Engelstalig, deels Nederlandstalig						0	GT3_3	0 t/m 4
(4) Deels Engelstalig, deels Nederlandstalig, deels non-verbaal						0	GT3_4	0 t/m 4
(5) Deels Nederlandstalig, deels non-verbaal						0	GT3_5	0 t/m 4
(6) Volledig non-verbaal						0	GT3_6	0 t/m 4
(7) Volledig Nederlandstalig						0	GT3_7	0 t/m 4
(8) Geheel afwezig						0	GT3_8	0 t/m 4
		0	0	0	0	<b>CHECK</b>		
		check!	check!	check!	check!			

## Bijlage 3. Beoordelingsformulier spreekvaardigheid

### Taak 1. De Brug

Vul hieronder het school- en leerlingnummer in. Dit nummer is gelijk aan de naam van het bestand.

Schoolnr.       - Leerlingnr.

Beoordelaar  
1  2  3  4  5  6

#### INDIEN NIET TE BEOORDELEN

(Kruis 1x 'ja' aan.)

Er kunnen verschillende redenen zijn waarom de gehele respons van de leerling niet beoordeeld kan worden volgens dit formulier. Kruis s.v.p. hieronder aan wat daarvan de reden is en ga door naar de volgende leerling.

- Is de opnamekwaliteit van het geluid te slecht?  ja
- Ontbreken de beschrijvingen bij alle plaatjes?  ja
- Anders?  ja

#### AFBEELDING 1

- Drukt de leerling in het Engels uit dat Jim van de brug in het water valt?  ja  nee (Ga door naar afbeelding 2.)
- Gebriikt de leerling de volgende woorden?
  - is falling/falls  ja  ja, maar vervoeging fout
  - from the bridge  ja  ja, maar voorzetsel fout
  - down/down from/from the rail  ja  ja, maar voorzetsel fout
  - into the water/river/canal  ja  ja, maar voorzetsel fout
- Geen beoordeling want respons is  afwezig  onverstaanbaar  NLs

#### AFBEELDING 2

- Drukt de leerling in het Engels uit dat de kinderen hulp vragen?  ja  nee (Ga door naar afbeelding 3.)
- Gebriikt de leerling de volgende woorden?
  - his friends/ the boy and the girl  ja
  - the children  ja  ja, maar meervoudsvorm fout
  - are stopping/stop  ja
  - a car  ja
  - are asking/ask (for help)  ja
  - people  ja  ja, maar meervoudsvorm fout
  - call  ja  ja, maar vervoeging fout
  - one-one-two/an ambulance  ja
- Geen beoordeling want respons is  afwezig  onverstaanbaar  NLs

#### AFBEELDING 3

- Drukt de leerling in het Engels uit dat de kinderen naar het water rennen?  ja  nee (Ga door naar afbeelding 4.)
- Gebriikt de leerling de volgende woorden?
  - his friends/the boy and the girl  ja
  - the children  ja  ja, maar meervoudsvorm fout
  - are running/run to  ja
  - river/bank/water/canal  ja
  - is swimming/is in the water/is almost drowning  ja
- Geen beoordeling want respons is  afwezig  onverstaanbaar  NLs

**AFBEELDING 4**

- Drukt de leerling in het Engels uit dat de kinderen Jim uit het water halen?  ja  nee (Ga door naar afbeelding 5.)
- Gebruikt de leerling de volgende woorden?*
- his friends/ the boy and the girl  ja
- the children  ja  ja, maar meervoudsvorm fout
- are trying/try to  ja
- rescue  ja
- are pulling/pull  ja
- out (of the water)/onto the bank  ja  ja, maar voorzetsel fout
- Geen beoordeling want respons is  afwezig  onverstaanbaar  NLs

**AFBEELDING 5**

- Drukt de leerling in het Engels uit dat de ambulance aangekomen is/dat Jim op een brancard wordt gelegd?  ja  nee (Ga door naar afbeelding 6.)
- Gebruikt de leerling de volgende woorden?*
- the ambulance has arrived/come  ja  ja, maar vervoeging fout
- two/three  ja
- the people/men/nurses  ja
- are putting Jim/put Jim/Jim is put  ja
- Jim is lying  ja
- on a stretcher  ja  ja, maar voorzetsel fout
- (the ambulance) is driving off/drives off/is leaving/leaves  ja  ja, maar vervoeging fout
- they are taking/the ambulance takes/Jim is taken  ja  ja, maar 'bring' i.p.v. 'take'
- (to) hospital  ja
- Geen beoordeling want respons is  afwezig  onverstaanbaar  NLs

**AFBEELDING 6**

- Drukt de leerling in het Engels uit dat Jim in een ziekenhuis(bed) ligt / dat Jim ok is?  ja  nee (Ga door naar uitspraak.)
- Gebruikt de leerling de volgende woorden?*
- hospital  ja
- is feeling/feels  ja  ja, maar vervoeging fout
- fine/better/ safe/ okay  ja
- his friends/ the boy and the girl/the children  ja
- are visiting/visit  ja
- Geen beoordeling want respons is  afwezig  onverstaanbaar  NLs

**GLOBALE BEOORDELING UITSPRAAK***(Kruis 1x 'ja' aan.)*

- Het Engels is volstrekt onverstaanbaar.  ja
- Het Engels klinkt erg Nederlands.  ja
- Het Engels klinkt behoorlijk Engels/Amerikaans.  ja
- De uitspraak is heel natuurlijk en zeer Engels/Amerikaans.  ja

**ALGEMENE INDRUK INVLOED TOETSLEIDER (indien van toepassing)**

- De leerling krijgt veel hulp van de toetsleider.  ja
- (De toetsleider 'trekt de antwoorden uit de leerling', 'grijpt in' en/of 'verbetert de leerling'.)*

## Bijlage 4. Brief naar besturen van scholen in de steekproef

Ons kenmerk:  
Peil.Engels\_B1

Geselecteerde scho(o)l(en):

Groningen, 12 juni 2017

Geacht bestuur,

In deze brief vragen wij uw aandacht voor Peil.Engels 2017. Het is belangrijk om in brede zin te weten wat kinderen in Nederland leren op school. Daarom vinden er voor alle inhoudsgebieden periodiek peilingsonderzoeken plaats. Deze onderzoeken geven zicht op de kennis, vaardigheden en/of houding van leerlingen en bieden aanknopingspunten voor (verdere) verbetering van het onderwijs.

Met Peil.Engels brengen we in kaart hoe het er voor staat met het onderwijs in de Engelse taal en de Engelse taalvaardigheid van leerlingen aan het einde van de het basisonderwijs. Peil. Engels wordt in opdracht van de Inspectie van het Onderwijs uitgevoerd door de Rijksuniversiteit Groningen in samenwerking met Cito, en de Hogeschool van Amsterdam (HvA).

Voor het uit te voeren onderzoek is een aselechte steekproef getrokken. Eén of meerdere van de scholen van uw bestuur zijn geselecteerd om aan deze peiling deel te nemen. Welke scholen dat zijn vindt u bovenaan deze brief. Binnenkort zullen we deze scholen benaderen voor deelname. Wij hopen dat u de scholen kunt motiveren om aan de peiling deel te nemen. Hun deelname is van groot belang om een goed beeld te kunnen schetsen van de stand van het Engels in het basisonderwijs! Bovendien levert deelname hen bruikbare inzichten op over de verschillende Engelse taalvaardigheden van hun leerlingen, mede in vergelijking met de andere deelnemende scholen.

### **Over Peil.Engels**

Goede kennis van het Engels wordt steeds belangrijker in onze maatschappij. Daarom is Engels naast Nederlands en Wiskunde een kernvak in ons onderwijssysteem: alle leerlingen doen in Nederland eindexamen in deze drie vakken. In het basisonderwijs is het vak Engels al enige jaren sterk in beweging. Er bestaat een tendens naar eerder beginnen en intensievere blootstelling.

**Hoe staat het er voor met het Engels in het basisonderwijs? En zien we verschillen tussen leerlingen die alleen Engels in groep 7-8 krijgen en leerlingen die al eerder zijn gestart met Engels (het zogenoemde vroeg vreemdetalen onderwijs, VVTO)? Dat zijn de vragen waarop met Peil.Engels een antwoord wordt gezocht.**

In januari en februari 2018 worden de prestaties met betrekking tot het Engels in kaart gebracht op een representatieve groep van 100 basisscholen (groep 8). De belasting voor de scholen bij deelname aan de peiling is gering. De peiling zelf wordt uitgevoerd binnen **één ochtend, waarbij 4 taken van 30 minuten worden afgenomen**. Een voorbeeld van zo'n taak is de luistervaardigheid Engels.

Op de website [www.peil-engels.nl](http://www.peil-engels.nl) vindt u meer informatie over het onderzoek. Op deze website heeft u tevens toegang tot een kort filmpje, waarin doel en uitvoering van de peiling nader worden toegelicht.

### **Terugrapportage**

Scholen die meedoen aan Peil.Engels ontvangen een terugrapportage. In de terugrapportage kunnen de scholen zien hoe hun leerlingen hebben gepresteerd op de afgenomen onderdelen. Tevens wordt een vergelijking gemaakt met de andere scholen die aan de peiling hebben deelgenomen. Naast de terugrapportage stellen we tevens een aantal **'good practices'** beschikbaar op basis van interviews en observaties van 'voorbeeld' leerkrachten. De terugrapportage en de 'good practices' kunnen behulpzaam zijn bij een verdere ontwikkeling van het Engels op uw scholen.

### **Uitkomsten onderzoek**

Peil.Engels maakt het mogelijk de Engelse taalvaardigheden in Nederland in kaart te brengen en na te gaan of er sprake is van een voor- of achteruitgang ten opzichte van een eerdere meting uit 2012. De resultaten van de peiling bieden een basis voor een brede discussie over de kwaliteit van het onderwijs in het Engels, meer specifiek ook gericht op de effectiviteit van vroegtijdige blootstelling aan de Engelse taal. Deelname aan Peil.Engels staat los van het instellingstoezicht op de betreffende scholen. De resultaten zullen met andere woorden op geen enkele manier worden gebruikt voor de kwaliteitsbeoordeling van scholen.

### **Hoe verder?**

Wij zullen binnenkort persoonlijk contact opnemen met de scholen die voor Peil.Engels zijn geselecteerd. De geselecteerde scholen kunnen zich ook zelf aanmelden, en wel door een e-mail te sturen naar [peil.Engels@rug.nl](mailto:peil.Engels@rug.nl) of door telefonisch contact op te nemen met Harm Naayer (050- 363 6678) of Lieneke Ritzema (050 363 6997).

Ook als u vragen heeft, kunt u via bovenstaande gegevens contact met ons opnemen.

Met vriendelijke groet,

Prof. dr. Roel Bosker  
hoogleraar Onderwijskunde  
Rijksuniversiteit Groningen

Dr. Arnold Jonk  
Plaatsvervangend Inspecteur-  
generaal  
Inspectie van het Onderwijs



## Bijlage 5. Wervingsmail richting scholen

Geachte mevrouw .....,

Goede kennis van het Engels wordt steeds belangrijker in onze maatschappij en daarom is het Engels naast Nederlands en wiskunde een kernvak in ons onderwijssysteem. In het basisonderwijs is het vak Engels al enige jaren sterk in beweging. Er bestaat een tendens naar eerder beginnen en intensievere blootstelling. Hoogste tijd dus voor onderzoek naar hoe het ervoor staat met het Engels in het basisonderwijs.

In opdracht van de Inspectie van het Onderwijs brengt [Peil! Engels](#) de Engelse taalvaardigheid van leerlingen uit groep 8 in beeld en verzamelt het gegevens over het aanbod Engels van scholen en leerkrachten. Het peilingsonderzoek wordt uitgevoerd op 100 basisscholen. Daarbij is een onderscheid gemaakt tussen scholen met leerlingen die alleen Engels in groep 7-8 krijgen en scholen waar al eerder is gestart met Engels (het zogenoemde vroeg vreemdetalen onderwijs, VVTO).

Uw school is geselecteerd als 1 van de 100 in de landelijke steekproef!

In het bijgaande filmpje wordt nader toegelicht wat het belang van Peil! Engels is, wat deelname aan het onderzoek zou betekenen en wat het voor u als school kan opleveren.

[Klik hier om de film te starten.](#)

Binnenkort neemt één van onze medewerkers contact met u op. Hij/zij kan u nader informeren over zowel de inhoud van het onderzoek als de verdere gang van zaken. Daarnaast kunt u met eventuele vragen bij deze persoon terecht. U kunt zich als school ook zelf aanmelden. Dat kan door een reactie op dit bericht te sturen.

De informatie uit de film treft u ook aan op [www.peil-engels.nl](http://www.peil-engels.nl)

Wij hopen op uw (enthousiasme en) medewerking!

Met vriendelijke groeten  
De onderzoekers van Team Peil Engels.

## Bijlage 6. Intakeformulier scholen

### Introductie

Goede kennis van het Engels wordt steeds belangrijker in onze maatschappij en daarom is het Engels naast Nederlands en wiskunde een kernvak in ons onderwijssysteem. In het basisonderwijs is het vak Engels al enige jaren sterk in beweging. Engels is verplicht voor groep 7 en 8, maar er bestaat een tendens naar eerder beginnen en intensievere blootstelling.

Peil. Engels is een groot landelijk onderzoek naar de Engelse vaardigheden van leerlingen in groep 8, aan het einde van het basisonderwijs. Specifiek gaat het om de vaardigheden begrijpend lezen, woordenschat, luistervaardigheid en gespreksvaardigheid. Ook brengt Peil Engels het aanbod Engels van scholen en groepsleerkrachten in kaart en vindt er in het voorjaar van 2018 een verdiepend onderzoek plaats

Uw school is van de 100 scholen die deelneemt aan het onderzoek. In een intakegesprek, waarvan dit formulier een verslaglegging is, wordt de opzet van het onderzoek toegelicht, worden de testmaterialen besproken, en worden afspraken gemaakt over de verdere voorbereiding en de datum waarop een testteam de school zal bezoeken voor de afname van de onderzoeksmaterialen.

Het onderzoek wordt in opdracht van de Inspectie voor het Onderwijs uitgevoerd door GION onderwijs/onderzoek & de vakgroep Applied Linguistics (beide Rijksuniversiteit Groningen), Cito en de Hogeschool van Amsterdam.

We benadrukken dat dit onderzoek niet als doel heeft om de school te testen of te beoordelen. Peil Engels is een landelijk representatief onderzoek waarmee in kaart wordt gebracht hoe het er voor staat met Engels in het Nederlandse basisonderwijs en om vast te stellen waar de *best practices* (als ook mogelijke onvolkomenheden) zitten in de lespraktijk. Deelname aan het onderzoek heeft als voordeel voor de school dat zij een terugkoppeling ontvangen van de prestaties van de leerlingen. Ook zullen de *best practices* met de school worden gedeeld.

### 1. Achtergrondgegevens:

Naam en plaats van de school .....

Volgnummer school: .....

*Vul hier de viercijferige code in (10xx, 20xx, 30xx voor eerste keuze, 1<sup>e</sup> reserve, 2<sup>e</sup> reserve etc..)*

Brin en vestigingsnr school: ..... [bv. 01AB 00]

Datum van intakegesprek: .....

- Telefonisch
- Bezoek aan de school

**Aanwezigen :**

- Schoolleider: dhr/mevr.....
  - Groepsleerkracht 1: dhr/mevr.....
  - Groepsleerkracht 2: dhr/mevr.....
  - Hoofdtestleider: dhr/mevr.....
- 

De contactpersoon op de school is *(bij voorkeur de groepsleerkracht die Engels geeft)*:

- Naam: .....
- Telefoonnummer: .....
- E-mail: .....

Naast de hierboven genoemde leerkracht zijn de volgende (vak)leerkracht(en) ook betrokken bij het geven van Engels op school *(zij zullen ook de leerkrachtenvragenlijst ontvangen)*:

- Naam 2<sup>e</sup> groepsleerkracht : .....
- Telefoonnummer: .....
- E-mail: .....
  
- Naam vakleerkracht Engels: .....
- Telefoonnummer: .....
- E-mail: .....

## 2. Afspraken

### Datum, tijd en deelnemende klassen

De peiling zal uitgevoerd worden op:

- Datum: .....
- Tijdstip: van ..... tot .....  
(*evt. bij middagpauze:*) van ..... tot .....
- Lokaal: ..... (*indien anders dan regulier lokaal groep 8*)

*NB1: de testleiders (twee per klas) zullen ruim van te voren aanwezig zijn om de benodigde materialen klaar te zetten. NB2: bij twee of drie klassen moeten wellicht afwijkende tijdstippen en data worden vastgelegd.*

Het gaat om de volgende klassen:

Klasnaam	Aantal leerlingen

### Aanleveren gegevens

Bij het maken van de afspraak voor de intake is de contactpersoon verzocht leerlinggegevens aan te leveren van alle leerlingen die in schooljaar 2017-2018 in groep 8 zitten. Het gaat om de volgende administratieve leerlinggegevens:

- voornaam, tussenvoegsel, achternaam
- sekse (**M**eisje of **J**ongen)
- geboortedatum (maand, dag, geboortjaar)
- naam van de klas waarin de leerling zit
- leerlinggewicht/formatiegewicht (0; 0,3; 1,2)
- indicatie van het niveau van de Engelse vaardigheden (**H**oog, **M**idden, **Z**wak)
- de scores voor begrijpend lezen in groep 7 (E7)
- het schooladvies of voorlopig schooladvies (begin groep 8)
- eventueel: indicatie van gediagnosticeerde dyslexie

Is dit gedaan?

- Ja, de hoofdtestleider heeft de lijst op voorhand ontvangen
- Nee, maar de leerlingenlijst is tijdens het intakegesprek besproken en de gegevens zullen nog digitaal naar [peil.engels@rug.nl](mailto:peil.engels@rug.nl) worden gestuurd.

De school geeft hierbij toestemming voor het gebruik van de achtergrondgegevens van de leerlingen voor de afname van de onderdelen van de peiling. De gegevens worden alleen gebruikt voor de tenaamstelling van leerlingvragenlijsten en antwoordformulieren. Na de afname worden de leerlingnamen alleen gebruikt voor de schoolrapportage. Persoonsgegevens noch schoolnaam worden opgenomen in databestanden.

## Onderdelen en afnamecondities

Voor de klas bestaat de afname uit vier onderdelen die elk een half uur (30 minuten) in beslag nemen. Het gaat om:

1. Het invullen van een vragenlijst
2. De afname van een gecombineerde lees- en woordenschattoets
3. De afname van een luistertaak via het digibord of een dvd
4. De afname van tweede gecombineerde lees- en woordenschattoets

Tijdens taak 2, 3 en 4 worden telkens 2 leerlingen uit de klas gevraagd om een gesprekstaak uit te voeren. Deze leerlingen selecteren we van tevoren en deze selectie zal worden voorgelegd aan de groepsleerkracht. De gesprekstaken worden opgenomen op video en naderhand door de betrokken onderzoekers geanalyseerd.

Voor de verschillende onderdelen zijn afnameprotocollen opgesteld. Deze gaan over standaard afnamecondities (klasopstelling, afkijken, vragen stellen e.d.) als ook over de instructie die aan leerlingen wordt gegeven voor elk onderdeel ('wat ga je maken', 'hoe vul je het antwoordformulier in' e.d.). Deze protocollen kunnen in het intakegesprek aan bod komen of worden kort voorafgaande aan het eerste onderdeel op de afnamedag met de groepsleerkracht besproken. Ook aan de leerlingen zal voor het eerste onderdeel (de vragenlijst) de procedure en de 'spelregels' worden uitgelegd. **Een essentiële afnameconditie is dat de groepsleerkracht tijdens alle onderdelen van de afname in de klas aanwezig is!**

## Leerkrachtvragenlijst

Op de afnamedag worden ook (per klas) leerkrachtvragenlijsten meegenomen. De ervaring leert dat deze tijdens één blok - waarin de leerlingen een toets maken - door de leerkracht is in te vullen. Mochten er meerdere (vak)leerkrachten betrokken zijn bij het vak Engels op de school dan zal de hoofdtestleiders meerdere exemplaren meenemen. Bij afwezigheid van deze extra leerkrachten zal ook een retourenveloppe met antwoordnummer worden achtergelaten.

## Overzicht onderdelen en benodigde tijd

	<b>Onderdeel 1</b> (30 min)	<b>Onderdeel 2</b> (30 min)	<b>Onderdeel 3</b> (30 min)	<b>Onderdeel 4</b> (30 min)
<b>Leerlingen</b>	Vragenlijst	Lezen/woordenschat toets deel A	Luistertoets	Lezen/woordenschat toets deel B
		Gesprekstaak koppel 1	Gesprekstaak koppel 2	Gesprekstaak koppel 3
<b>Leerkracht</b>	Leerkrachtvragenlijst			

NB: tussen de onderdelen zullen in overleg met de leerkracht korte pauzes, of reguliere schoolpauzes worden ingelast.

### **Informereren van ouders**

Voor de onderzoekers als ook de opdrachtgever is het van essentieel belang dat ook de ouders van deelnemende leerlingen voorafgaande aan het onderzoek goed worden geïnformeerd. Hiertoe heeft de contactpersoon voorafgaand aan het intakegesprek een voorbeeld flyer en toestemmingsbrief ontvangen. Deze is reeds of dient alsnog door de school te worden verspreid (geprint of digitaal). In het onderzoek wordt gebruik gemaakt van passieve toestemming. Dit houdt in dat ouders in principe geen actie hoeven te ondernemen tenzij zij (ernstige) bezwaren hebben tegen deelname van het kind. Dan kunnen zij dit kenbaar maken aan de groepsleerkracht via een antwoordstrook die is toegevoegd aan de brief. Ouders hebben de volgende opties om bezwaar aan te tekenen:

1. het kind mag wel meedoen aan de peiling, maar niet aan de (op video op te nemen) gesprekstaken;
2. het kind moet helemaal worden uitgesloten van deelname aan de peiling.

### **(eventuele) Uitsluiting van leerlingen**

Tijdens het intakegesprek wordt met behulp van de (eerder of op de dag verkregen) leerlingenlijst nagegaan of bepaalde leerlingen in de klas(sen) moeten worden uitgesloten van deelname, hetzij vanwege gediagnosticeerde problematiek (bijvoorbeeld taalontwikkelingsstoornis, visuele beperking) of vanwege de niet verkregen toestemming van ouders (indien de brief al eerder naar de school is gestuurd en uitsluiting om die reden al bij de leerkracht bekend is). NB: leerlingen met een auditieve beperking, gedragsstoornissen als ook dyslexie kunnen gewoon met de afname meedoen. Slechthorende kinderen kunnen wellicht de luistertoets overslaan.

### **Medewerking en terugkoppeling**

De school verleent medewerking aan het peilingsonderzoek tot en met juli 2018. De school ontvangt na de afname van de peiling, maar voor de zomervakantie 2018 een retourrapportage waarin de prestaties van de leerlingen in het onderzoek vergeleken worden met de landelijke scores. De rapportages zullen op individueel niveau scores bevatten van de afgelegde onderdelen (vaardigheden). De wijze van rapportage is vergelijkbaar met de bekende I tot V indeling die Cito ook hanteert bij andere toetsen. Voorts wordt de omvang van de woordenschat gerapporteerd alsmede een ERK-niveau voor de domeinen gespreksvaardigheid, luistervaardigheid en leesvaardigheid. De rapportage bevat derhalve namen van leerlingen maar is alleen bedoeld voor de school. De rapportages worden niet gedeeld met andere scholen of onderzoeksinstituten binnen de Peiling. Op een deel van de scholen zal in mei-juni een verdiepingsonderzoek plaatsvinden. Hiervoor kan de school nog een vrijblijvende uitnodiging ontvangen. Elke school ontvangt op basis van dat verdiepende onderzoek, na de zomervakantie, een rapport 'best practices'. Tevens wordt de school blijvend op de hoogte gehouden van alle overige publicaties die uit de Peiling volgen.

### **Privacy en geheimhoudingsplicht**

Het onderzoek consortium verklaart hierbij dat zij zeer zorgvuldig met de gegevens zal omgaan. Daarbij wordt benadrukt dat a) geheimhoudingsplicht zal worden betracht ten aanzien van alle ontvangen gegevens, b) dat gegevens alleen versleuteld digitaal worden verstuurd, en c) dat zij de opgestuurde toetsresultaten uitsluitend zal gebruiken voor analysedoeleinden in het kader van dit specifieke onderzoek.

### **Contact mogelijkheden**

Voor vragen, op- en aanmerkingen kan altijd contact opgenomen worden. Dit kan via het e-mailadres [peil.engels@rug.nl](mailto:peil.engels@rug.nl) of door te bellen naar Lieneke Ritzema (050-3636997), Harm Naayer (050-3636678) of Denise Leidelmeijer (050-3632373).

De school wordt in de toekomst op de hoogte gehouden van het verloop van de peiling de resultaten via e-mails, nieuwsbrieven en de website [www.peil-engels.nl](http://www.peil-engels.nl).

### 3. Vragen over het aanbod van Engels op de school.

In de peiling worden de groepsleerkrachten van groep 8 uitgenodigd voor het invullen van een leerkrachtenvragenlijst. Ook op schoolniveau hebben we een aantal korte vragen. Die zijn hieronder opgenomen.

#### Organisatie en inrichting van het onderwijsaanbod in de Engelse taal op uw school

1. Hieronder staan drie aanbodprofielen geschetst. Welk profiel is van toepassing op uw school?

Kruis s.v.p. aan welk profiel **het beste** het onderwijsaanbod voor Engels op uw school beschrijft: A, B of C. Per profiel kunt u eventueel aangeven wat specifiek is voor het aanbod op uw school, bijvoorbeeld:

- school gestuurde projecten waarin (aanvullend) aandacht aan Engels wordt geschonken (musical, Taaldorp)
- het meedoen aan landelijke initiatieven
- het gebruik van Engels bij andere onderdelen van het curriculum, zoals lichamelijke opvoeding, muziek, wereldoriëntatie, etc.

<input type="checkbox"/> Profiel A	<input type="checkbox"/> Profiel B	<input type="checkbox"/> Profiel C
Regulier EIBO	Vervroegd EIBO	Een vorm van vroeg vreemde talenonderwijs
Ruimte voor schoolspecifieke informatie:		

2. Wat is het startmoment voor lessen Engels bij de leerlingen in de **huidige groep 8 (2017-2018)**?

- Jaargroep:

- Kalenderjaar:



3. Hoeveel onderwijstijd Engels heeft de huidige groep 8 genoten in de voorgaande jaargroepen? *Hoeveel onderwijstijd hebben de leerlingen in de huidige groep 8 op uw school gemiddeld **per week** besteed aan Engels in elk van de genoemde groepen? Hoeveel weken in het jaar is Engels gegeven? Vul 0 minuten in als in een groep GEEN Engels is gegeven.*

Groep	Minuten per week	Aantal weken	Groep	Minuten per week	Aantal weken
1	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>	5	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>
2	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>	6	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>
3	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>	7	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>
4	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>	8	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>

4a. Is er op uw school een coördinator of werkgroep voor het vak Engels?

- Ja  
 Nee

4b. Is er op uw school sprake van:

- een coördinator voor Engels  
 een werkgroep voor Engels

5a. Welke faciliteiten voor Engels biedt de school aan leerkrachten.

*Kruis s.v.p. aan wat van toepassing is of omschrijf de faciliteiten die u aanbiedt.*

- Budget voor aanvullende of ondersteunende materialen voor Engels  
 Stimuleren van deelname aan internationale partner- en uitwisselingsprojecten  
 Stimuleren van activiteiten waarbij Engels voertaal is (Engelstalige gast in de klas, Engelse lunch, Engelse schooltelevisie, etcetera)

- GEEN faciliteiten.
- Anders, namelijk  
.....  
.

5b. Hoeveel leerkrachten hebben in de afgelopen 3 jaar specifiek nascholing gevolgd op het gebied van Engels?

Aantal leerkrachten: .....

5c. Welke insteek had deze nascholing?

- taalvaardigheid Engels
- didactiek Engels
- taalvaardigheid en didactiek Engels
- anders, nl.....

6. De volgende vragen gaan om een inschatting van de thuissituatie van de leerlingen.

6a. Allereerst een vraag over de dominante thuistaal van leerlingen. Hoe zijn de leerlingen van uw school ongeveer over deze categorieën verdeeld? (samen 100%)

	Percentage leerlingen met Nederlands als de dominante thuistaal
	Percentage leerlingen met Nederlands dialect of streektaal als de dominante thuistaal
	Percentage leerlingen waar de dominante thuistaal niet-Nederlands is
100%	

6b. In hoeverre hebben leerlingen een taalrijke thuistaalsituatie? Binnen een taalrijke thuistaalsituatie wordt veel (voor-)gelezen, gepraat en gediscussieerd, ongeacht de taal waarin dat gebeurt. In een taalarme thuistaalsituatie gebeuren die dingen juist heel weinig. Hoe zijn de leerlingen van uw school ongeveer over deze categorieën verdeeld? (samen 100%)

	Percentage leerlingen met een uitgesproken taalrijke thuistaalsituatie
	Percentage leerlingen met een gemiddelde thuistaalsituatie
	Percentage leerlingen met een uitgesproken taalarme thuistaalsituatie

---

100%

7. Worden leerlinggegevens voor Engels overgedragen aan het voortgezet onderwijs?  
*Kruis s.v.p. één van onderstaande mogelijkheden aan of beschrijf de wijze van overdracht.*

- Onze school draagt GEEN resultaten voor Engels over
- Resultaten voor Engels worden overgedragen in de vorm van rapportcijfers
- Gegevens voor Engels worden overgedragen in het overdrachtdossier
- Er is sprake van mondelinge afstemming voor Engels met de betreffende scholen voor voortgezet onderwijs
- anders, nl.....

Dit waren de vragen die wij wilden stellen over het Engels op uw school. Hartelijk dank voor de informatie.

## 6. Checklist.

In het intakegesprek zijn de volgende zaken aan bod gekomen:

- Er is een afspraak gemaakt voor de datum en het tijdslot voor de afname
- De vragen uit de algemene schoolvragenlijst (deel 3 van dit document) zijn ingevuld
- De vragenlijst voor de groepsleerkracht(en) is getoond. Afgesproken is dat deze op de dag van de afname aan de groepsleerkracht wordt gegeven en dat ter plekke wordt besloten wanneer deze wordt ingevuld.
  - Op de afname dag tijdens een van de onderdelen voor leerlingen
  - De vragenlijst wordt achtergelaten en kan met de retourenveloppe worden teruggestuurd. (bij eventueel verlies kan (kunnen) de vragenlijsten(en) worden teruggestuurd naar GION/Harm Naayer, Antwoordnummer 172, 9700 VB, Groningen
- Er is een voorbeeld van de leerlingvragenlijst getoond (of besproken)
- Er is besproken hoe de retourrapportage er (ongeveer) uitziet
- Er is afgesproken dat er op de testdag een aparte ruimte zal worden gereserveerd waar de gesprekstaken uit het onderzoek kunnen worden afgenomen. Deze ruimte is in principe vooraf en gedurende de gehele periode van klassikale taken beschikbaar voor de onderzoekers.
  - Indien al bekend: ..... (naam ruimte)
- Er is een kopie van dit document gemaakt voor de school/contactpersoon
- Er is besproken dat tijdens de klassikale taken de bevoegde groepsleerkracht in de klas aanwezig is.
- De flyer + toestemmingsbrief voor de ouders is reeds verspreid of wordt nog uitgedeeld voorafgaande aan de afnamedag. Voorafgaand of op de afnamedag worden eventuele ouderlijke bezwaren meegenomen in de dagplanning (bv. geen deelname aan gesprekstak)

- Er is aangekondigd dat er een verdiepend onderzoek plaats gaat vinden bij een deel van de scholen.

## 7. Eventuele aanvullende afspraken en/of bijzondere omstandigheden:

.....

.....

.....

.....

.....

.....

## 8. Ondertekening

Vertegenwoordiger school

Hoofdtestleider *Peil Engels*

Naam:

Naam:

Datum:

Datum:

Handtekening:

Handtekening:

**Contact**

Voor vragen, op- en aanmerkingen kan altijd contact opgenomen worden. Dit kan via het e-mail-adres [peil.engels@rug.nl](mailto:peil.engels@rug.nl) of door te bellen naar Lieneke Ritzema (050-363XXXX), Harm Naayer (050-363XXXX) of Denise Leidelmeijer (050-36XXXX).

De school wordt in de toekomst op de hoogte gehouden van het verloop van de peiling de resultaten via e-mails, nieuwsbrieven en de website [www.peil-engels.nl](http://www.peil-engels.nl).

## Aanbodvragenlijst 2018

# Peiling Engels

einde basisonderwijs 2017/2018

Vragen aan de leerkracht die het onderwijs Engels verzorgt in groep 8

### Toelichting

Peiling Engels onderzoekt de vaardigheden van leerlingen in de Engelse taal aan het eind van het basisonderwijs. Daarnaast brengt deze peiling een aantal aspecten van het onderwijsaanbod voor Engels in het basisonderwijs in beeld. Het aanbod wordt geïnventariseerd met onderstaande lijst van vragen aan de groepsleerkracht(en) van (de) groep(en) 8 en met behulp van een aantal korte vragen aan de schoolleiding. De verkregen gegevens worden geanonimiseerd verwerkt en geanalyseerd.

*Het is de bedoeling dat alle leerkrachten die het onderwijs Engels in groep 8 verzorgen de vragenlijst hieronder invullen. Als er twee leerkrachten zijn die het vak Engels voor dezelfde groep 8 verzorgen, ontvangen beide leerkrachten de vragenlijst.*

Ik vul de vragenlijst in als groep-8 leerkracht/LIO van:

Volgnummer school:

Leerkracht

Klasnaam:

*NB: alleen bij meerdere klassen*

## Deel 1: Achtergrond

### Vraag 1. Algemeen

Bent u de groepsleerkracht van deze groep

Ja  Nee

Wat is uw geslacht?

Man  Vrouw  zeg ik liever niet/anders

Wat is uw leeftijd?

20 – 30 jr.  30 – 40 jr.  40 – 50 jr.  50 – 60 jr.  > 60 jr.

Bent u in Nederland geboren?

Ja  Nee

Beschouwt u Nederlands als uw moedertaal?

Ja  Nee

Beschouwt u Engels als uw moedertaal?

Ja  Nee

In welke taal/talen communiceert u regelmatig?

Nederlands  Engels  Anders

*U kunt hier de andere taal/talen specificeren*

Hoelang werkt u als leerkracht in het basisonderwijs?

0-5 jaar  6-10 jaar  11-15 jaar  16–20 jaar  meer dan 21 jaar

Hoelang geeft u Engels als leerkracht in het basisonderwijs?

0-5 jaar  6-10 jaar  11-15 jaar  16–20 jaar  meer dan 21 jaar

Wat is uw behaalde diploma?

reguliere PABO  verkorte PABO  academische PABO  leraren opleiding Engels  diploma nog niet behaald

(1<sup>e</sup> of 2<sup>e</sup>-graads)

### Vraag 2. Aanbod Engels in de door u genoten opleiding (PABO)

*Deze vraag hoeft alleen te worden ingevuld door leraren die de Pabo hebben afgerond of nog bezig zijn met de Pabo. Indien u niet de Pabo heeft gedaan, kunt u doorgaan met vraag 3.*

Welke scholing heeft u gehad voor het geven van Engels?

Engels op de Pabo – verplichte module(s)

Ja  Nee

Keuzemodule Engels op de Pabo (bijvoorbeeld een minor)

Ja  Nee

Anders namelijk:

Op welke aspect lag de nadruk tijdens de lessen Engels op de Pabo?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	taal en didactiek	alleen taal	alleen didactiek	n.v.t.
Bent u tevreden over omvang en inhoud van het vak Engels op de Pabo?	<input type="checkbox"/>	<input type="checkbox"/>		
	Ja	Nee		
Heeft u na-/bijscholing op het gebied van Engels ontvangen?	<input type="checkbox"/>	<input type="checkbox"/>		
	Ja	Nee	<i>(indien nee: ga door met vraag 3)</i>	
Op welk aspect was deze na-/bijscholing gericht?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	taal en didactiek	alleen taal	alleen didactiek	

### Vraag 3. Uw eigen gebruik van het Engels in uw dagelijkse leven

*U kunt onderscheid maken tussen:*

1 = (Bijna) nooit

2 = minstens 1 keer per jaar

3 = minstens 1 keer per maand

4 = minstens 1 keer per week

5 = dagelijks

Hoe vaak kijkt u naar Engelse TV of Engelstalige films/series, <u>zonder</u> Nederlandse ondertiteling?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
Hoe vaak kijkt u naar Engelstalige films/series, <u>met</u> Nederlandse ondertiteling?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
Hoe vaak luistert u naar Engelstalige muziek?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
Hoe vaak leest u een tekst in het Engels, zoals een boek, tijdschrift of blog?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
Hoe vaak leest u zinnen in het Engels tijdens het gebruik van 'social media' of 'online gamen'?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
Hoe vaak spreekt u Engels buiten school?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
Hoe vaak schrijft u een tekst in het Engels? (social media, chat, e-mail, <u>brief</u> , <u>kaart</u> )	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5



## Deel 2: onderwijsaanbod op uw school

### Vraag 4. Gehanteerde doelen voor de lessen Engels

Welke doelen zijn voor u het meest bepalend voor de inrichting van uw onderwijs in de Engelse taal?

Kruis *maximaal twee antwoorden aan*.

- De kerndoelen Engels voor het primair onderwijs
- Doelen die zijn beschreven in het ERK (Europees Referentiekader)
- Doelen voor vroeg vreemdetalenonderwijs (van het Europees Platform/Nuffic of van Early Bird)
- Doelen die binnen school of door het schoolbestuur bepaald zijn
- Doelen die ikzelf vaststel. Daarbij maak ik gebruik van uiteenlopende bronnen.
- De doelen die in de gebruikte methode worden aangegeven
- Anders namelijk:

### Vraag 5. Gehanteerde methode voor de lessen Engels

Welke methode gebruikt u in groep 8? Noteer a.u.b. welke methode u gebruikt: titel en uitgever.

Methode (titel)

Naam Uitgever

Welke methode gebruikte uw school vorig jaar in groep 7?

Methode (titel)

Naam Uitgever

### Vraag 6. Gebruik van de methode in de lessen Engels

Gebruikt u uitsluitend de methode? En hoe was dat voor de huidige leerlingen van groep 8 vorig jaar in groep 7?

Groep 8 (2017 - 2018)

- Ja, uitsluitend de methode
- Nee, er worden ook andere leermiddelen gebruikt

Groep 7 (2016 - 2017)

- Ja, uitsluitend de methode
- Nee, er zijn ook andere leermiddelen gebruikt
- Dat weet ik niet

## Vraag 7. Gebruik andere leermiddelen

7a. Welke andere leermiddelen zet u in? *N.B. dit zijn alle materialen die niet zijn gekoppeld aan de door u gebruikte methode. U kunt onderscheid maken tussen:*

(Bijna) nooit = Dit gebeurt niet of bijna niet

Af en toe = Dit komt één of twee keer per maand voor

Vaak = Dit is minimaal één keer per week aan de orde

Hoe vaak zet u authentiek Engelstalig beeld- en geluidsmateriaal in?

(Bijna) nooit

Af en toe

Vaak

Hoe vaak zet u Engelstalig tekstmateriaal in (krant, tijdschrift, leesboeken, internet)?

(Bijna) nooit

Af en toe

Vaak

Zet u een ander leermiddel in ... en hoe vaak zet u dit middel in?

(Bijna) nooit

Af en toe

Vaak

(Bijna) nooit

Af en toe

Vaak

(Bijna) nooit

Af en toe

Vaak

7b. Was het gebruik van de leermiddelen voor de huidige leerlingen van groep 8 vorig schooljaar in groep 7 hetzelfde?

Ja

Nee

Dat weet ik niet

## Vraag 8. Waardering van de methode Engels

*In hoeverre bent u het eens met de volgende stellingen*

	Helemaal eens	Eens	Neutraal	Oneens	Helemaal oneens
De leeractiviteiten in de methode spreken mijn leerlingen aan	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De activiteiten in de methode sluiten aan bij het niveau van mijn leerlingen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De methode besteedt aandacht aan de relevante aspecten van Engels voor mijn leerlingen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mijn indruk is dat mijn leerlingen door de activiteiten in de methode veel leren	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De methode biedt goede mogelijkheden voor differentiatie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Vraag 9. Activiteiten

### 9a. Aan welke vaardigheid besteedt u de meeste aandacht?

Hieronder staan zeven vaardigheden. Deel deze zeven vaardigheden in naar de mate waarin deze aandacht krijgen in uw lessen Engels, door ze een cijfer van 1 (meest) tot 7 (minst) te geven. Plaats de cijfers 1 t/m 7 in de vakjes vóór de vaardigheden. NB: gebruik elk cijfer maar één keer, en begin bij de vaardigheid die de meeste aandacht krijgt.

<input type="text"/>	Lezen
<input type="text"/>	Luisteren
<input type="text"/>	Spreken (solo)
<input type="text"/>	Schrijven
<input type="text"/>	Woordenschat
<input type="text"/>	Grammatica
<input type="text"/>	Gespreksvaardigheid (interactief)

**LET OP: bij de vragen 9b tot en met 9g kunt u onderscheid maken tussen:**

(Bijna) nooit = Dit gebeurt niet of bijna niet

Af en toe = Dit komt één of twee keer per maand voor

Vaak = Dit is minimaal één keer per week aan de orde

### 9b. Spreekactiviteiten

Hoe vaak oefenen de leerlingen op school het spreken in het Engels op onderstaande wijzen? Het gaat hier om activiteiten waarin leerlingen Engels spreken in de vorm van een monoloog (= geen interactie).

⊕

	(Bijna) nooit	Af en toe	Vaak
Naspreken van nieuwe woorden in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Naspreken van standaard gesprekken in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In het Engels vertellen wat er op plaatjes te zien is	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In het Engels vertellen over een gebeurtenis (b.v. verjaardag, sportactiviteit, weekend, vakantie)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Een tekst, boek of film(ogje) navertellen in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Het zingen van liedjes in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anders, namelijk: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

□

### 9c. Gespreksactiviteiten

Hoe vaak oefenen de leerlingen op school het **voeren van een gesprek** in het Engels op onderstaande wijzen? Het gaat hier om activiteiten waarbij leerlingen interacteren met één of meerdere anderen.

	(Bijna) nooit	Af en toe	Vaak
Een rollenspel spelen in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Voeren van een gesprek in het Engels met één of meer medeleerlingen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Voeren van een gesprek in het Engels met de leerkracht	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Voeren van een gesprek in het Engels met een ( <i>near</i> ) <i>native speaker</i> (iemand anders dan de leerkracht)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Voeren van een gesprek in het Engels met leeftijdgenoten via internet (b.v. Skype)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Een spelletje spelen of activiteit doen in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anders, namelijk: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### 9d. Luister-/Kijkactiviteiten

Hoe vaak oefenen de leerlingen op school met **luisteren** (en kijken) in het Engels op onderstaande wijzen?

	(Bijna) nooit	Af en toe	Vaak
Luisteren naar een tekst of gesprek in het Engels uit de methode	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Luisteren naar een tekst of gesprek in het Engels van tv, internet of uit een film ( <u>zonder ondertiteling</u> )	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Luisteren naar een tekst of gesprek in het Engels van tv, internet of uit een film ( <u>met ondertiteling</u> )	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Luisteren naar de leerkracht die alleen maar Engels spreekt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>Nederlandse</u> vragen beantwoorden bij een Engelse luistertekst	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>Engelse</u> vragen beantwoorden bij een Engelse luistertekst	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anders, namelijk: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### 9e. Leesactiviteiten

Hoe vaak oefenen de leerlingen op school met **lezen** in het Engels op onderstaande wijzen?

	(Bijna) nooit	Af en toe	Vaak
Korte Engelse teksten uit de methode lezen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Korte Engelse teksten uit een tijdschrift of op de tablet/computer lezen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Langere Engelse teksten uit een boek of tijdschrift of op de tablet/computer lezen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>Nederlandse</u> vragen beantwoorden bij een Engelse leestekst	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>Engelse</u> vragen beantwoorden bij een Engelse leestekst	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Woorden invullen die in een Engelse leestekst zijn weggelaten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Een Nederlandse samenvatting geven van een Engelse leestekst	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anders, namelijk: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### 9f. Woordenschatactiviteiten

Hoe vaak oefenen de leerlingen op school met **woordenschat** in het Engels op onderstaande wijzen?

	(Bijna) nooit	Af en toe	Vaak
De betekenis van woorden afleiden uit een Engelse context	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engelse woorden opzoeken in een woordenboek / -lijst	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engelse woorden proberen te omschrijven	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Een tegenovergesteld woord zoeken (Engels-Engels)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engelse woorden bij afbeeldingen (tekeningen / foto's) zoeken	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De Nederlandse vertaling geven van een Engels woord	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engelse woorden classificeren (in categorieën plaatsen)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engelse woorden uit het hoofd leren	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engelse woorden oefenen met de computer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anders, namelijk: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## 9g. Schrijfactiviteiten

Hoe vaak oefenen de leerlingen op school met **schrijven** in het Engels op onderstaande wijzen?

	(Bijna) nooit	Af en toe	Vaak
Engelse woorden bij afbeeldingen (tekeningen/ foto's) schrijven	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engelse zinnen bij afbeeldingen schrijven	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De Engelse vertaling van Nederlandse woorden opschrijven	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engelse woorden en/of zinnen overschrijven	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engelse woorden en/of zinnen opschrijven in een dictee	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Antwoorden op vragen bij een tekst opschrijven in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Een brief, e-mail, WhatsApp, Instagrambericht of SMS schrijven in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Een verhaaltje, gedichtje of liedje schrijven in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Een werkstuk schrijven in het Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anders, namelijk: <input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Vraag 10. Gebruik van Engels als voertaal tijdens de lessen Engels

Welk deel van de lessen Engels wordt gevoerd in de Engelse taal?

	0 tot 5%	5 tot 30%	30 tot 60%	60 tot 85%	85 tot 100%
Ik spreek Engels met de leerlingen voor:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ik nodig de algehele klas uit om met mij Engels te spreken voor:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De leerlingen spreken Engels met elkaar voor:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Vraag 11. Werkvormen tijdens de lessen Engels

Op welk deel van de lessen Engels zijn de volgende werkvormen van toepassing?

	0 tot 5%	5 tot 30%	30 tot 60%	60 tot 85%	85 tot 100%
Ik werk klassikaal met de leerlingen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De leerlingen werken in tweetallen of groepjes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De leerlingen werken individueel	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### Vraag 12. Gebruik van huiswerk

Geeft u de leerlingen huiswerk voor Engels?

- (Bijna) nooit
- Af en toe
- Vaak

*Indien u gebruik maakt van huiswerk voor Engels, kruis dan a.u.b. hieronder de activiteiten aan die als huiswerk worden gedaan en/of omschrijf de (eventuele) andere activiteiten.*

- Lezen
- Luisteren
- Spreken (solo/monoloog)
- Schrijven
- Woordenschat
- Grammatica
- Gespreksvaardigheid (interactief)
- Anders namelijk:

### Vraag 13. Integratie van Engels met andere vakken

Wordt Engels gegeven in samenhang of geïntegreerd met andere vakken?

- Ja
- Nee

*Zo ja, kruis a.u.b. hieronder aan in samenhang waarmee Engels wordt gegeven.*

- Oriëntatie op jezelf en de wereld (aardrijkskunde, geschiedenis, biologie)
- Burgerschapsvorming
- Kunstzinnige oriëntatie (muziek, drama, beeldende vorming)
- Internationaliseringsprojecten
- Lichamelijke opvoeding
- Anders namelijk:

### Vraag 14. Overdracht vorderingen Engels

Op welke wijze heeft overdracht van groep 7 naar groep 8 plaatsgevonden m.b.t. de vorderingen van leerlingen voor Engels? *Kruis a.u.b. hieronder de mogelijkheden aan die van toepassing zijn of beschrijf hoe de overdracht plaatsvond.*

- Dat was niet nodig, want ik was vorig jaar groepsleerkracht van dezelfde leerlingen in groep 7.
- Vorderingen voor Engels worden overgedragen in de vorm van rapportcijfers.
- Vorderingen voor Engels worden overgedragen via een overdrachtdossier (met daarin resultaten uit methodetoetsen).
- Er is sprake van mondelinge afstemming voor Engels met de betreffende leerkracht van vorig jaar.
- Er is geen overdracht.
- Anders namelijk:

### Vraag 15. Toetsing

#### 15a. Welke toetsen gebruikt u?

- Ik gebruik de toetsen bij de methode
- Ik stel zelf toetsen samen
- Ik zet toetsen in van Engelstalige instituten, zoals *Anglia* en *Cambridge*
- Ik zet Me2! Engels (Cito) in
- Ik zet de IEP eindtoets Engels in
- Engels wordt op onze school NIET getoetst
- Anders namelijk:

#### 15b. Krijgen leerlingen met achterblijvende prestaties voor Engels gericht ondersteuning om de achterblijvende prestaties in te halen?

- (Bijna) nooit
- Ja, incidenteel
- Ja, structureel

*Wat is de aard van deze ondersteuning?*

- Verlengde instructie
- Extra oefening in groepjes
- Remedial teaching
- Inzet extra oefenmateriaal
- Aangepaste taken/opdrachten
- Anders namelijk:

#### 15c. Waar is de meeste aandacht voor bij toetsing?

*Geef in volgorde van meest (1) naar minst (7) aan wat u het meest toetst. Plaats de cijfers 1 t/m 7 in de vakjes vóór de vaardigheden.*

	Lezen
	Luisteren
	Spreken (solo)
	Schrijven
	Woordenschat
	Grammatica
	Gespreksvaardigheid (interactief)



### Deel 3: Attitudes t.o.v. Engels

#### Vraag 16. Waardering Engels in het primair onderwijs

Kruis a.u.b. aan in hoeverre u het eens bent met de stellingen hieronder.

Ga uit van uw eigen ervaringen en de situatie op uw school.

	Helemaal eens	Eens	Neutraal	Oneens	Helemaal oneens
De doelen voor Engels in het primair onderwijs zijn duidelijk over wat leerlingen moeten kunnen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De meeste leerlingen kennen al aardig wat Engels voordat het op school gegeven wordt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De meeste leerlingen zijn gemotiveerd voor Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De meeste leerlingen volgen de lessen Engels met plezier	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De niveaoverschillen tussen leerlingen zijn voor Engels groot	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engels moet worden aangeboden vanaf groep 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engels moet worden aangeboden vanaf groep 5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engels moet worden opgenomen in de eindtoets	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De aandacht voor Engels gaat ten koste van het niveau voor taal en rekenen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Leerkrachten zijn gemotiveerd voor Engels	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Om Engels te kunnen geven moet de taalvaardigheid van leerkrachten verbeterd worden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Om Engels te kunnen geven moet de didactische vaardigheid van leerkrachten verbeterd worden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Deel 4. Inschatting vaardigheid Engels van uzelf en de leerlingen

Hieronder volgen twee schema's waarin vier vaardigheden Engels (lezen, luisteren, spreken en schrijven) met daaronder beschrijvingen zijn opgenomen.

Bij vraag 17a en 17b vragen we u een inschatting te maken van uw eigen vaardigheden Engels. Kruis a.u.b. per beschrijving aan of u vindt dat deze beschrijving op u zelf van toepassing is (ja of nog niet volledig).

Bij vraag 18 vragen we u een inschatting te maken van de vaardigheden Engels van uw leerlingen. Kruis a.u.b. per beschrijving aan of deze beschrijving op uw leerlingen van toepassing is. Ga uit van de 20% best presterende leerlingen in de groep, voor Engels.

### Vraag 17a. Inschatting eigen taalvaardigheid in het Engels

Lezen van Engels	Ja	Nog niet volledig
<ul style="list-style-type: none"><li>Ik kan Engelse teksten lezen die hoofdzakelijk bestaan uit hoogfrequente, alledaagse of aan mijn werk gerelateerde taal. Ik kan de beschrijving van gebeurtenissen, gevoelens en wensen in persoonlijke brieven begrijpen.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Ik kan zeer korte, eenvoudige Engelse teksten lezen. Ik kan specifieke, voorspelbare informatie vinden in Engelstalige advertenties, folders, menu's en dienstregelingen en ik kan korte, eenvoudige persoonlijke brieven begrijpen.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Ik kan Engelstalige artikelen en verslagen lezen die betrekking hebben op eigentijdse problemen, waarbij de schrijvers een bepaalde houding of standpunt innemen. Ik kan eigentijds Engelstalig literair proza begrijpen.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Ik kan lange en complexe Engelse teksten begrijpen, en het gebruik van verschillende stijlen in de Engelstalige literatuur waarderen. Ik kan gespecialiseerde artikelen en lange technische instructies begrijpen, zelfs wanneer deze geen betrekking hebben op mijn terrein.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>

Luisteren naar Engels	Ja	Nog niet volledig
<ul style="list-style-type: none"><li>Ik kan Engelse zinnen en de meest frequente woorden begrijpen die betrekking hebben op gebieden die van direct persoonlijk belang zijn (bijvoorbeeld basisinformatie over mezelf en mijn familie, winkelen, plaatselijke omgeving, werk). Ik kan de belangrijkste punten uit korte, duidelijke eenvoudige berichten en aankondigingen halen.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Ik kan de hoofdpunten begrijpen wanneer in duidelijk Engels wordt gesproken over vertrouwde zaken op mijn werk, school, vrije tijd enz. Ik kan de hoofdpunten van veel Engelstalige radio- of tv-programma's over actuele zaken of over onderwerpen van persoonlijk of beroepsmatig belang begrijpen, wanneer er betrekkelijk langzaam en duidelijk gesproken wordt.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Ik kan een langer betoog in het Engels begrijpen, zelfs wanneer dit niet duidelijk gestructureerd is en wanneer verbanden niet expliciet worden aangegeven. Ik kan zonder al te veel inspanning Engelstalige tv-programma's en films (zonder ondertiteling) begrijpen.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Ik kan een langer betoog en lezingen in het Engels begrijpen en zelfs complexe redeneringen volgen, wanneer het onderwerp redelijk vertrouwd is. Ik kan de meeste nieuws- en actualiteitenprogramma's op de tv begrijpen. Ik kan het grootste deel van films in standaard-Engels begrijpen.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>

### Vraag 17b. Inschatting eigen taalvaardigheid in het Engels

Engels spreken	Ja	Nog niet volledig
<ul style="list-style-type: none"> <li>Ik kan mezelf vloeiend en spontaan uitdrukken zonder naar uitdrukkingen te hoeven zoeken. Ik kan het Engels flexibel en effectief gebruiken voor sociale en professionele doeleinden. Ik kan ideeën en meningen precies formuleren en gedetailleerde beschrijvingen geven over complexe onderwerpen met specifieke standpunten en een passende conclusie. Bovendien kan ik mijn bijdrage vaardig aan die van andere sprekers relateren.</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> <li>Ik kan duidelijke, gedetailleerde beschrijvingen presenteren over een breed scala van onderwerpen die betrekking hebben op mijn interessegebied. Ik kan een standpunt over een actueel onderwerp verklaren en de voordelen en nadelen van diverse opties uiteenzetten.</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> <li>Ik kan de meeste situaties aan die zich kunnen voordoen op reis in een Engelstalig gebied. Ik kan onvoorbereid deelnemen aan gesprekken over onderwerpen die vertrouwd zijn, mijn persoonlijke belangstelling hebben of die betrekking hebben op het dagelijks leven (b.v. familie, hobby's, werk, reizen en actuele gebeurtenissen).</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> <li>Ik kan een reeks Engelse uitdrukkingen en zinnen gebruiken om in eenvoudige bewoordingen mijn familie en andere mensen, leefomstandigheden, mijn opleiding en mijn huidige of meest recente baan te beschrijven.</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>

Schrijven in het Engels	Ja	Nog niet volledig
<ul style="list-style-type: none"> <li>Ik kan een duidelijke, gedetailleerde Engelse tekst schrijven over een breed scala van onderwerpen die betrekking hebben op mijn interesses. Ik kan een opstel of verslag schrijven, informatie doorgeven of redenen aanvoeren ter ondersteuning vóór of tégen een specifiek standpunt. Ik kan brieven schrijven waarin ik het persoonlijk belang van gebeurtenissen en ervaringen aangeef.</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> <li>Ik kan me in duidelijke, goed gestructureerde Engelse tekst uitdrukken en daarbij redelijk uitgebreid standpunten uiteenzetten. Ik kan in een brief, een opstel of een verslag schrijven over complexe onderwerpen en daarbij de voor mij belangrijke punten benadrukken. Ik kan schrijven in een stijl die is aangepast aan de lezer die ik in gedachten heb.</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> <li>Ik kan korte, eenvoudige notities en boodschappen in het Engels opschrijven. Ik kan een zeer eenvoudige persoonlijke brief schrijven, bijvoorbeeld om iemand voor iets te bedanken.</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> <li>Ik kan eenvoudige, samenhangende Engelse tekst schrijven over onderwerpen die vertrouwd of van persoonlijk belang zijn. Ik kan persoonlijke brieven schrijven waarin ik mijn ervaringen en indrukken beschrijf.</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>

### Vraag 18. Inschatting taalvaardigheid in het Engels van de leerlingen

Ga uit van de 20% best presterende leerlingen in de groep, voor Engels.

Lezen van Engels	Ja	Nog niet volledig
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen Engelse teksten lezen die hoofdzakelijk bestaan uit hoogfrequente, alledaagse of aan school en hobby's gerelateerde taal. Zij kunnen de beschrijving van gebeurtenissen, gevoelens en wensen in persoonlijke brieven begrijpen.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen vertrouwde namen, woorden en zeer eenvoudige Engelse zinnen begrijpen, bijvoorbeeld in mededelingen, op posters en in catalogi.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen zeer korte, eenvoudige Engelse teksten lezen en specifieke, voorspelbare informatie vinden in Engelstalige advertenties, folders, menu's en dienstregelingen. Zij kunnen korte, eenvoudige persoonlijke brieven begrijpen.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>

Luisteren naar Engels	Ja	Nog niet volledig
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen Engelse zinnen en de meest frequente woorden begrijpen die betrekking hebben op gebieden die van direct persoonlijk belang zijn (bijvoorbeeld basisinformatie over henzelf en hun familie, winkelen, plaatselijke omgeving, werk). Ze kunnen de belangrijkste punten in korte, duidelijke eenvoudige boodschappen en aankondigingen volgen.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen vertrouwde Engelse woorden en basiszinnen begrijpen die henzelf, hun familie en hun directe concrete omgeving betreffen, wanneer er langzaam en duidelijk gesproken wordt.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen de hoofdpunten begrijpen wanneer in duidelijk Engels wordt gesproken over vertrouwde zaken op school, hobby's, vrije tijd, enz. Ze kunnen de hoofdpunten van veel Engelstalige radio- of tv-programma's over actuele zaken of over onderwerpen van persoonlijk belang begrijpen, wanneer er betrekkelijk langzaam en duidelijk gesproken wordt.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>

Engels spreken	Ja	Nog niet volledig
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen uitingen op een simpele manier aan elkaar verbinden, zodat ze ervaringen en gebeurtenissen, hun dromen, verwachtingen en ambities kunnen beschrijven. Ze kunnen in het kort redenen en verklaringen geven voor hun meningen en plannen. Ze kunnen een verhaal vertellen, of de plot van een film of een boek weergeven en hun reacties beschrijven.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen een reeks Engelse uitdrukkingen en zinnen gebruiken om in eenvoudige bewoordingen familie en andere mensen, leefomstandigheden en bijvoorbeeld schoolactiviteiten te beschrijven.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Mijn leerlingen zijn in staat deel te nemen aan een eenvoudig gesprek, wanneer de gesprekspartner bereid is tot hulp, herhaling of herformulering. Zij kunnen eenvoudige vragen stellen en beantwoorden m.b.t. directe behoeften of zeer vertrouwde onderwerpen. Zij zijn in staat tot een eenvoudige en directe uitwisseling van informatie over vertrouwde onderwerpen en activiteiten (familie, vrienden en kennissen, woonomgeving, school, hobby's).</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>

Schrijven in het Engels	Ja	Nog niet volledig
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen een korte, eenvoudige ansichtkaart schrijven, bijvoorbeeld voor het zenden van vakantiegroeten. Ze kunnen op formulieren persoonlijke details invullen, bijvoorbeeld naam, nationaliteit en adres.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen eenvoudige, samenhangende Engelse teksten schrijven over onderwerpen die vertrouwd of van persoonlijk belang zijn. Zij kunnen persoonlijke brieven schrijven waarin ik mijn ervaringen en indrukken beschrijf.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li>Mijn leerlingen kunnen korte, eenvoudige notities en boodschappen in het Engels opschrijven. Ze kunnen een zeer eenvoudige persoonlijke brief schrijven, bijvoorbeeld om iemand voor iets te bedanken.</li></ul>	<input type="checkbox"/>	<input type="checkbox"/>

Ruimte voor opmerkingen:

Dit is het einde van de vragenlijst voor leerkrachten van groep 8.

**Hartelijk dank voor het invullen!**

# Your English Quiz

## Peiling Engels

einde basisonderwijs 2017/2018

Vragenlijstnummer: 1001

## Your English Quiz

### Peiling Engels einde basisonderwijs 2017/2018

We willen graag weten hoe jij over Engels denkt. Ook zijn we benieuwd of je buiten school met Engels in aanraking komt.

Lees de vragen goed, maar denk niet te lang over de antwoorden na.  
Kies op je gevoel het antwoord dat het beste bij jou past.

Let op: je mag meestal maar één antwoord kiezen.

Ieder antwoord is goed, als het maar past bij jou.

Gebruik een grijs potlood en maak het hokje helemaal zwart/grijs.

1. Vind je Engels op school een gemakkelijk vak?
  - Ik vind het niet zo moeilijk, maar ook niet heel gemakkelijk.
  - Ja, ik vind Engels gemakkelijk.
  - Nee, ik vind Engels moeilijk.
  
2. Stel je voor: Je bent op vakantie. Kinderen vragen in het Engels of je mee wilt spelen. Wat doe je?
  - Je knikt ja, maar maakt in je beste Engels duidelijk dat je ze niet zo goed verstaat.
  - Je schudt nee, want je denkt dat je Engels niet goed genoeg is.
  - Je zegt "yes" en speelt in het Engels mee.
  
3. Vind je Engels op school een leuk vak?
  - Ja, ik vind Engels een leuk vak.
  - Nee, ik vind Engels niet echt leuk.
  - Ik vind het niet zo'n heel leuk, maar ook geen vervelend vak.
  
4. Vind je het belangrijk voor jouw toekomst om goed Engels te leren?
  - Ik heb er geen mening over.
  - Ja, ik vind Engels belangrijk.
  - Nee, ik vind Engels niet zo belangrijk.
  
5. Stel je voor: er is op tv een Engelse tekenfilm die je graag wilt zien, maar de film heeft geen ondertiteling. Wat doe je?
  - Ik kijk de film gewoon zonder ondertiteling.
  - Ik probeer of ik de film kan volgen.
  - Ik zet de film uit.
  
6. Durf jij Engels te spreken?
  - Ik durf het wel, maar vind het vervelend om fouten te maken.
  - Ja, ik durf best Engels te spreken, ook al maak ik soms fouten.
  - Nee, ik durf niet zo goed, ik ben bang om fouten te maken.
  
7. Stel je voor: je krijgt een boek dat je graag wilt lezen, maar het boek is Engelstalig. Wat doe je?
  - Ik begin er meteen in, het Engels snap ik wel.
  - Ik probeer gewoon of ik het boek in het Engels kan lezen.
  - Ik vind het moeilijk om Engels te lezen, dus ik ga het boek niet lezen.



8. Vind jij taal (Nederlands) op school gemakkelijk?
- Ik vind taal niet zo moeilijk, maar ook niet heel gemakkelijk.
  - Ja, ik vind taal gemakkelijk.
  - Nee, ik vind taal moeilijk.
9. Spreek je wel eens Engels tijdens de les Engels?
- Vaak
  - Af en toe
  - (Bijna) nooit
10. Wat vind je leuk tijdens de Engelse les?  
*(je mag meerdere hokjes zwart/grijs maken)*
- Engelse tekstjes lezen
  - een kort Engels berichtje of tekstje schrijven
  - in het Engels praten
  - luisteren naar liedjes of verhalen in het Engels
  - Engelse liedjes zingen
  - Niets, ik vind de Engelse les niet leuk
11. Stel je voor: op internet zoek je informatie over een onderwerp dat je erg interesseert. Je vindt een Engelstalige website. Wat doe je?
- Die website sla ik over, want ik begrijp toch niet wat er staat.
  - Ik gebruik een automatisch vertaalprogramma.
  - Ik lees die informatie, daarvoor ken ik genoeg Engels.
  - Ik probeer die informatie te lezen, ik kan wel een beetje Engels lezen.
  - Ik vraag iemand om mij te helpen.
12. Heb je wel eens in het Engels een e-mailtje / briefje geschreven?  
*(alleen of samen met je ouders)*
- Ja, dat doe ik wel vaker in het Engels.
  - Ja, één of twee keer.
  - Nee, dat heb ik nog nooit gedaan in het Engels.
13. Heb je wel eens in het Engels gechat, bijvoorbeeld via een chatprogramma of via een online game?
- Ja, dat doe ik wel vaker in het Engels.
  - Ja, één of twee keer.
  - Nee, dat heb ik nog nooit gedaan in het Engels.

14. Hoe vaak ben jij buiten school met Engels bezig?

	<b>Nooit</b>	<b>Soms</b>	<b>Best vaak</b>	<b>Heel vaak</b>
		Eén keer per week of minder	Elke week twee of drie keer	Bijna elke dag
Maak bij elke activiteit het juiste hokje zwart/grijs				
1. <u>Engelstalige</u> filmpjes op internet kijken <u>zonder Nederlandse ondertiteling</u> (bijv. <i>Youtube</i> )	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Engelstalige computerspelletjes spelen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Een boek of tijdschrift in het Engels lezen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Op tv, dvd, internet of in de bioscoop <u>Engelstalige</u> films, series of andere programma's kijken <u>met Nederlandse ondertiteling</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Luisteren naar Engelse liedjes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Engelse websites bekijken om informatie te zoeken	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

15. Hoeveel tijd besteed je per dag ongeveer aan het lezen, luisteren of schrijven in het Engels? Denk aan *Youtube*, gamen, social media, *Spotify*, *Netflix*, series op tv of Engels praten met bijvoorbeeld je Engelstalige buurjongen of klasgenootje.

- 0 minuten
- Tussen 1 en 10 minuten
- Tussen 10 en 30 minuten
- Tussen 30 en 60 minuten
- Meer dan 1 uur

16. Begrijp je Engelstalige liedjes die je bijvoorbeeld op de radio/op internet hoort?

- Ja, dat lukt vaak al na één of twee keer luisteren.
- Ja, maar ik begrijp ze pas als ik vaak luister.
- Nee, ook als ik vaak luister, snap ik Engelse liedjes meestal niet.

17. Welke uitspraak klopt voor jouw gevoel? Mijn klasgenootjes....

- luisteren minder vaak naar Engelse liedjes dan ik.
- luisteren net zo vaak naar Engelse liedjes als ik.
- luisteren veel vaker dan ik naar Engelse liedjes.
- Ik weet niet hoe vaak mijn klasgenootjes naar Engelse muziek luisteren.

18. Spreek je wel eens Engels buiten school?
- Ja, best vaak.
  - Ja, soms.
  - Nee, bijna nooit.
19. Spreek je wel eens Engels met een vriend/vriendin of kennis?
- Ja, best vaak.
  - Ja, soms.
  - Nee, bijna nooit.
20. Spreek je wel eens Engels met familieleden?
- Ja, best vaak.
  - Ja, soms.
  - Nee, bijna nooit.
21. Spreek je wel eens Engels met mensen die je op vakantie tegenkomt?
- Ja, best vaak.
  - Ja, soms.
  - Nee, bijna nooit.
22. Spreek je wel eens Engels met onbekenden (*bijvoorbeeld op straat, in een winkel*)?
- Ja, best vaak.
  - Ja, soms.
  - Nee, bijna nooit.
23. Welke talen spreek je thuis? (*je mag meerdere hokjes zwart/grijs maken*)
- Nederlands
  - Engels
  - een andere taal
24. Welke taal spreekt je moeder/verzorgster met jou?  
(*je mag meerdere hokjes zwart/grijs maken*)
- Nederlands
  - Engels
  - anders, namelijk:
25. Is je moeder/verzorgster in Nederland geboren?
- Ja
  - Nee

26. Welke taal spreekt je vader/verzorger met jou?

*(je mag meerdere hokjes zwart/grijs maken)*

Nederlands

Engels

anders, namelijk:

27. Is je vader/verzorger in Nederland geboren?

Ja

Nee

28. Wat kun jij al in het Engels? *Maak steeds het hokje met 'ja' of 'nog niet' zwart/grijs.*

#### Lezen van Engels

- Ik kan informatie vinden in menu's en in folders  ja  nog niet
- Ik kan borden en posters met eenvoudige woorden en zinnen begrijpen  ja  nog niet
- Ik kan beschrijvingen van gebeurtenissen of gevoelens in brieven, mails of chats begrijpen  ja  nog niet
- Ik kan korte, eenvoudige e-mails, sms-berichten en WhatsApp berichten lezen  ja  nog niet

#### Luisteren naar Engels

- Ik kan simpele zinnen begrijpen in een gesprek over familie en het dagelijks leven  ja  nog niet
- Ik kan de belangrijkste punten in omroepberichten volgen  ja  nog niet
- Ik kan woorden en zinnen verstaan in een gesprek over school en hobby's  ja  nog niet
- Ik kan de hoofdzaken begrijpen van Engelstalige radio- en tv-programma's  ja  nog niet
- Ik kan de hoofdzaken begrijpen van Engelstalige filmpjes op internet  ja  nog niet

29. Wat kun jij al in het Engels? *Maak steeds het hokje met 'ja' of 'nog niet' zwart/grijs.*

#### Engels spreken

- Ik kan mijn mening in het Engels geven en daarbij een uitleg geven  ja  nog niet
- Ik kan in eenvoudige woorden en zinnen vertellen over mijn familie of omgeving  ja  nog niet
- Als mensen langzaam spreken en dingen herhalen, kan ik deelnemen aan een gesprek in het Engels  ja  nog niet
- Ik kan eenvoudige vragen stellen en korte antwoorden geven  ja  nog niet
- Ik kan de inhoud van een boek of film navertellen in het Engels  ja  nog niet

#### Schrijven in het Engels

- Ik kan een vakantiekaart, e-mail, WhatsApp- of chatbericht schrijven in het Engels  ja  nog niet
- Ik kan een tekst schrijven over bijvoorbeeld mijn huis, sport of hobby's  ja  nog niet
- Ik kan aantekeningen maken in het Engels  ja  nog niet
- Ik kan mijn naam, adres en nationaliteit invullen op een formulier  ja  nog niet
- Ik kan mijn ervaringen beschrijven in het Engels  ja  nog niet

30. Je gaat volgend jaar naar de middelbare school (voortgezet onderwijs). Naar welke brugklas ga je waarschijnlijk?

*Als het waarschijnlijk een combinatiebrugklas wordt, bijvoorbeeld havo/vwo, kun je meer dan één antwoord hokje zwart/grijs maken (dus havo en vwo).*

- vwo (gymnasium of atheneum)
- havo
- vmbo – gemengd of theoretisch
- vmbo kader
- vmbo basis
- praktijkonderwijs

**Dit is het einde van de vragenlijst. Dank je wel voor het invullen!**

## Bijlage 9. Resultaten voor de opgaven in de vragenlijstschalen

Afkorting	aantal		Mi		Gemiddelde		SCHAAL	
	Valid	ss	Max	Score	P'			
i01	2073	2	2	1.27	0.64	1: Attitude	1. Vind je Engels op school een gemakkelijk vak.	
i03	2073	2	2	1.38	0.69	1: Attitude	3. Vind je Engels op school een leuk vak?	
i04	2073	2	2	1.88	0.94	1: Attitude	4. Vind je het belangrijk voor je toekomst om goed Engels te leren?	
i10.1	2071	4	1	0.38	0.38	1: Attitude	10.1. Wat vind je leuk tijdens Engels: teksten lezen	
i10.2	2071	4	1	0.26	0.26	1: Attitude	10.2. Wat vind je leuk tijdens Engels: schrijven Engels	
i10.3	2071	4	1	0.59	0.59	1: Attitude	10.3. Wat vind je leuk tijdens Engels: praten Engels	
i10.6	2071	4	1	0.93	0.93	1: Attitude	10.6R. Wat vind je leuk tijdens Engels: niets (REVERSED)	
i02	2073	2	2	1.59	0.80	2: LSC	2. Stel je voor: Je bent op vakantie. Kinderen vragen in het Engels of je mee wilt spelen. Wat doe je?	
i05	2074	1	2	1.46	0.73	2: LSC	5. Stel je voor: Er is op tv een Engelse tekenfilm die je graag wilt zien, maar de film heeft geen ondertiteling. Wat doe je?	
i06	2074	1	2	1.55	0.78	2: LSC	6. Durf je Engels te spreken?	
i07	2071	4	2	1.04	0.52	2: LSC	7. Stel je voor: Je krijgt een boek dat je graag wilt lezen, maar het boek is Engelstalig. Wat doe je?	
i11	2042	33	4	2.60	0.65	2: LSC	11. (was 9) Stel je voor: Op internet zoek je informatie over een onderwerp dat je erg interesseert. Je vindt een Engelstalige website. Wat doe je?	
i12	2064	11	2	0.61	0.31	3:VL_BsG	12 (was 11). Heb je in het Engels wel eens een e-mailtje geschreven of gechat, bijvoorbeeld tijdens het gamen?	
i13	2070	5	2	1.21	0.61	3:VL_BsG	13. (new) Heb je in het Engels wel eens gechat	
i14.1	2069	6	3	1.84	0.61	3:VL_BsG	14.1. (was 16.1...) Engelstalige filmpjes op internet kijken zonder Nederlandse ondertiteling (bijvoorbeeld Youtube)	
i14.2	2066	9	3	1.97	0.66	3:VL_BsG	14.2. Engelstalige computerspelletjes spelen	
i14.3	2050	25	3	0.64	0.21	3:VL_BsG	14.3. Een boek of tijdschrift in het Engels lezen	
i14.6	2054	21	3	0.77	0.26	3:VL_BsG	14.6. Engelse websites bekijken om informatie te zoeken	
i15	2067	8	4	2.88	0.72	3:VL_BsG	15. Hoeveel tijd besteed je per dag aan Le, Lu, Schr, Spre, Engels	
i16	2066	9	2	1.40	0.70	3:VL_BsG	16. Begrijp je Engelstalige liedjes die je bijvoorbeeld op de radio/op internet hoort?	
i18	2047	28	2	0.86	0.43	3:VL_BsG	18. Spreek je wel eens Engels buiten school?	

i19	2043	32	2	0.55	0.28	3:VL_BsG	19. Spreek je wel eens Engels met een vriend(in)?
i20	2049	26	2	0.57	0.29	3:VL_BsG	20. Spreek je wel eens Engels met familie?
i21	2047	28	2	1.11	0.56	3:VL_BsG	21. Spreek je wel eens Engels met mensen op vakantie?
i22	2046	29	2	0.32	0.16	3:VL_BsG	22. Spreek je wel eens Engels met onbekenden?
i14.4	2059	16	3	1.96	0.65	3:VL_BsG15	14.4. Op tv, dvd, internet of in de bioscoop Engelstalige films, series of andere programma's kijken met Nederlandse ondertiteling
i14.5	2065	10	3	2.42	0.81	3:VL_BsG15	14.5. Luisteren naar Engelse liedjes
i28.1	2047	28	1	0.68	0.68	4.1:IV_LE	Vr_28_1 (was 24.1...) Lezen: Ik kan informatie vinden in menu's en in folders
i28.2	2048	27	1	0.92	0.92	4.1:IV_LE	Vr_28_2 Lezen: Ik kan borden en posters met eenvoudige woorden en zinnen begrijpen
i28.3	2049	26	1	0.58	0.58	4.1:IV_LE	Vr_28_3 Lezen: Ik kan beschrijvingen van gebeurtenissen of gevoelens in brieven, mails of chats begrijpen
i28.4	2044	31	1	0.90	0.90	4.1:IV_LE	Vr_28_4 Lezen: Ik kan korte, eenvoudige e-mails, sms-berichten en what's app-berichten lezen
i28.5	2048	27	1	0.88	0.88	4.2:IV_LU	Vr_28_5 Luisteren: Ik kan simpele zinnen begrijpen in een gesprek over familie en het dagelijks leven
i28.6	2042	33	1	0.63	0.63	4.2:IV_LU	Vr_28_6 Luisteren: Ik kan de belangrijkste punten in omroepberichten volgen
i28.7	2036	39	1	0.85	0.85	4.2:IV_LU	Vr_28_7 Luisteren: Ik kan woorden en zinnen verstaan in een gesprek over scholen hobby's
i28.8	2034	41	1	0.57	0.57	4.2:IV_LU	Vr_28_8 Luisteren: Ik kan de hoofdzaken begrijpen van Engelstalige radio- en tvprogramma's
i28.9	2041	34	1	0.83	0.83	4.2:IV_LU	Vr_28_9 Luisteren: Ik kan de hoofdzaken begrijpen van Engelstalige filmpjes op internet
i29.06	2019	56	1	0.58	0.58	4.3:IV_SC	Vr_29_6 Schrijven: Ik kan een vakantiekaart, e-mail, what's app of chatbericht schrijven in het Engels
i29.07	2023	52	1	0.62	0.62	4.3:IV_SC	Vr_29_7 Schrijven: Ik kan een tekst schrijven over bijvoorbeeld mijn huis, sport of hobby's
i29.08	2010	65	1	0.51	0.51	4.3:IV_SC	Vr_29_8 Schrijven: Ik kan aantekeningen maken in het Engels
i29.10	2006	69	1	0.43	0.43	4.3:IV_SC	Vr_29_10 Schrijven: Ik kan mijn ervaringen beschrijven in het Engels
i29.01	2020	55	1	0.46	0.46	4.3:IV_SP	Vr_29_1 (was 25.1...) Spreken: Ik kan mijn mening in het Engels geven en daarbij een uitleg geven
i29.02	2031	44	1	0.81	0.81	4.3:IV_SP	Vr_29_2 Spreken: Ik kan in eenvoudige woorden en zinnen vertellen over mijn familie of omgeving
i29.03	2030	45	1	0.86	0.86	4.3:IV_SP	Vr_29_3 Spreken: Als mensen langzaam spreken en dingen herhalen, kan ik deelnemen aan een gesprek in het Engels
i29.04	2031	44	1	0.94	0.94	4.3:IV_SP	Vr_29_4 Spreken: Ik kan eenvoudige vragen stellen en korte antwoorden geven
i29.05	2011	64	1	0.28	0.28	4.3:IV_SP	Vr_29_5 Spreken: Ik kan de inhoud van een boek of film navertellen in het Engels

## Bijlage 10. Toelichting op selectie van voorbeelden ten behoeve van de standaardbepaling bij gesprekken

De scores die leerlingen op de verschillende taken hebben gehaald vormden de basis voor de standaardbepaling. Voor iedere taak zijn voorbeelden geselecteerd behorend bij een behaalde schaalscore. Bij scores waar er een keuzemogelijkheid was bij de selectie van voorbeelden van leerlingprestaties is met een aantal factoren rekening gehouden bij de keuze van de voorbeelden. De eerste factor is de bruikbaarheid van het filmpje, de tweede factor de beoordelaar, de derde factor is de school en de laatste factor zijn de kenmerken van de leerling.

De eerste factor was een knock-out criterium. Als de *kwaliteit van het filmpje* dusdanig was dat een beoordeling relatief lastig zou zijn, dan is een dergelijk filmpje niet als voorbeeld gebruikt. Er waren verscheidene redenen waarom de kwaliteit niet voldoende was, zoals geluid dat te zacht was of dat de leerlingen de helft van de tijd slecht in beeld waren. Deze kandidaten waren wel beoordeelbaar, maar als voorbeeld voor de een standaardbepaling waren ze niet bruikbaar.

De tweede factor *beoordelaar* is relevant, omdat de score een reflectie zou moeten zijn van de leerlingprestatie en niet van de beoordelaar. Bij de beschikbare data is een deel van de taken door twee beoordelaars beoordeeld om de interbeoordelaarsbetrouwbaarheid te kunnen evalueren. Deze gegevens waren ook zeer geschikt om de voorbeelden te kiezen: wanneer twee (of drie zoals in een aantal gevallen) beoordelaars exact dezelfde score hebben gegeven aan een kandidaat, is het aannemelijker dat deze kandidaat ook werkelijk deze score heeft dan wanneer de beoordelaars scores hebben gegeven die ver uit elkaar lagen. Wat het maximaal acceptabele verschil was per voorbeeld hing ook af van de scorering. Bij taak 1 waar de scorering een stuk breder was (in theorie van 0 tot en met 108) was een verschil van 3 punten relatief minder groot dan bij de andere twee taken. Bij taak 2 en taak 3 waren er minder dubbel beoordeelde voorbeelden om uit te kiezen aangezien deze door minder personen waren gemaakt. Bij taak 1 was 90% van de gebruikte voorbeelden door minstens 2 personen beoordeeld, met een gemiddeld absoluut verschil tussen twee beoordelaars van onder de 2 (1,89). Bij taak 2 en taak 3 was 60% van de gebruikte voorbeelden dubbel beoordeeld, met een gemiddelde absoluut verschil van precies 1 punt bij taak 2 en 2 punten bij taak 3.

In sommige gevallen was het niet mogelijk goede voorbeelden te kiezen waarover twee beoordelaars hun oordeel hadden gegeven. Dan werd gekeken naar de bruikbare filmpjes van de kandidaten met slechts één beoordelaar. Er is hierbij gelet op de ervarenheid en de strengheid van de beoordelaar. De beoordelaar moest relatief veel leerlingen van verschillende scholen gezien hebben. Een beoordelaar moest daarnaast niet te streng of te mild zijn. Bij een te strenge beoordelaar lag de score lager dan de prestatie zou verdienen, en bij een milde beoordeling was deze juist te hoog. Bij deze voorbeelden was ook de expertblik van de toets-deskundige die de voorbeelden selecteerde van belang. De reeksen voorbeelden moesten ook altijd voldoen aan de eis dat de volgorde zoals gegeven ook daadwerkelijk de volgorde in de vaardigheid representeerde als de reeks door expert werd aanschouwd.

De factoren school en leerling zijn meegenomen om ervoor te zorgen dat de leerlingen niet allemaal van dezelfde school of paar scholen zouden komen. Dit onder andere om regionale spreiding van



voorbeelden te verkrijgen. Daar waar er leerlingkenmerken bekend waren –zoals geslacht– zijn die van tevoren ook meegenomen, zodat de voorbeelden niet alleen maar afkomstig waren van jongens of alleen maar van meisjes. Als specifieke gegevens niet bekend waren zijn deze achterhaald door het bekijken van de opnamen. Tijdens het bekijken van de opnamen is tevens naar de kwaliteit van de opname gekeken. Ook is, om opnieuw rekening te houden met de factor beoordelaar, gekeken of de ordening van de gekozen voorbeelden bij hogere scores inderdaad oplopend was in vaardigheid. De geselecteerde voorbeelden zijn een illustratie van de scoreschaal (met een relatie tot de vaardigheidsschaal) waarop standaarden zijn bepaald. Op basis van die voorbeelden kon achterhaald worden welke vaardigheid net voldoende is om indicatief te zijn voor het niveau A1 en het niveau A2. De standaardbepaling A1 is gestart met een bespreking van de eisen die in het referentiekader gesteld worden aan niveau A1. Voor de A2 en B1 standaard is dit ook gebeurd.

## Bijlage 11. Operationalisatie van de predictoren in de meerniveau-analyses

### **Model 2**

*Startmoment\*stratum.* Deze categorische variabele is een interactie-term, te weten het product van de dichotome variabele startmoment (VVTO vs. EIBO) en de dichotome variabele stratum dat staat voor het percentage leerlingen met een formatiegewicht op een school (strata 1 en 2: 0 - 10 % gewichten leerlingen vs. strata 3 en 4: > 10% gewichten leerlingen).

### **Model 3a**

*Jongen.* Het geslacht van de leerling is gecodeerd als 1 = jongen, 0 = meisje.

*Leerlinggewicht.* Het formatiegewicht van een leerling volgt de indeling zoals die door de Rijksoverheid tussen 2006 en 2017 is gehanteerd (<https://duo.nl/open Onderwijsdata/images/toelichting-01-leerlingen-bo-gewicht%2C-leeftijd.pdf>).

Het is gecodeerd als 0.3 voor laag opgeleide ouders en 1.2 voor erg laag opgeleide ouders. Leerlingen zonder formatiegewicht vormen de referentiecategorie.

*Leeftijd.* De leeftijd van de leerling is uitgedrukt in jaren met 1 decimaal.

*Advies vo.* Het advies dat een leerling in groep 8 krijgt voor het best passende schooltype in het voortgezet onderwijs is een ordinale variabele met vmbo-laag (vmbo-bbl en/of vmbo-kbl) als referentie categorie (= 1), vmbo-hoog (vmbo-tl en/of vmbo-gl, mavo) als midden (= 2) en havo en/of vwo als hoogste categorie (= 3).

### **Model 3b**

*Taalgebruik thuis.* De taal die de leerling thuis spreekt is een nominale variabele en gecategoriseerd als 1 = Nederlands, 2 = Nederlands en Engels, 3 = Nederlands en een andere taal, 4 = Nederlands, Engels en een andere taal en 5 = overig.

*Buitenschools gebruik van Engels.* De schaal VL\_BsG\_15 (zie paragraaf 5.3.2 en Bijlage 9).

*Linguistic self confidence.* De schaal SelfConf\_best (zie paragraaf 5.3.2 en Bijlage 9).

*Attitude.* De schaal attitude\_best (zie paragraaf 5.3.2 en Bijlage 9).

*Eigen inschatting van de betreffende vaardigheid.* (zie paragraaf 5.3.2 en Bijlage 9) .

### **Model 4**

*Onderwijstijd.* De hoeveelheid onderwijstijd Engels is een schatting van de schoolleider van het aantal minuten Engelse les die de groep 8 leerlingen hebben gehad in de groepen 1 tot en met 8. Het is berekend als de som over de jaargroepen van het gemiddelde aantal minuten Engels per week keer het aantal weken voor iedere jaargroep.

*Didactische methode.* De gehanteerde methoden voor de lessen Engels in groep 8 zijn verwerkt in de nominale variabele didactische methode. De oorspronkelijke 18 verschillende methoden zijn gehercodeerd naar de vijf meest voorkomende methoden. Het gaat om 'Just do it', 'Step to', 'Groove me' en 'overige methoden'. De meest frequent gebruikte methode 'Take it easy' is de referentiecategorie.

*Extra materiaal.* Het gebruik van andere leermiddelen dan de methode Engels voor de leerlingen in groep 8 is een dichotome variabele (1 = ja, 0 = nee).

*Differentiatie.* Het gericht ondersteunen van leerlingen met achterblijvende prestaties voor Engels is ordinale variabele en gecodeerd als 1 = (bijna) nooit, 2 = incidenteel, 3 = structureel.

*Faciliteiten Engels.* De diverse faciliteiten voor Engels die de school biedt aan leerkrachten zoals budget voor aanvullende of ondersteunende materialen voor Engels, stimuleren van deelname aan internationale partner- en uitwisselingsprojecten en stimuleren van activiteiten waarbij Engels de voertaal is, is gehercodeerd naar een dichotome variabele (1 = wel faciliteiten, 0 = geen faciliteiten).

*Bijscholing Engels.* De leerkracht van groep 8 heeft wel (=1) of geen (=0) bij- of nascholing gehad op het gebied van Engels.

*Inzet authentiek Engels materiaal.* De mate waarin de leerkracht authentiek Engels materiaal inzet voor de lessen Engels is een ordinale variabele en geconstrueerd als de somscore van twee items, namelijk het inzetten van authentiek Engelstalig beeld- en geluidsmateriaal (1 = (bijna) nooit, 2 = af en toe, 3 = vaak) en Engelstalig tekstmateriaal (krant, tijdschrift, leesboeken, internet) (1 = (bijna) nooit, 2 = af en toe, 3 = vaak). De range van de schaal is 1 tot en met 6.

*Engels als voertaal.* Het percentage van de lessen Engels waarin de leerkracht Engels als spreektaal gebruikt is een ordinale variabele (1= 0-5%, 2 = 5-30%, 3 = 30-60%, 4= 60-85%, 5 = 85-100%).

*Werken in groepen.* Het percentage van de lessen Engels waarin de leerlingen in tweetallen of groepjes werken is een ordinale variabele (1= 0-5%, 2 = 5-30%, 3 = 30-60%, 4= 60-85%, 5 = 85-100%).

*Huiswerk Engels.* De frequentie waarmee de leerkracht huiswerk geeft voor Engels is een ordinale variabele op een driepuntsschaal (1 = (bijna) nooit, 2 = af en toe, 3 = vaak).

*Integratie Engels in andere vakken.* De leerkracht geeft Engels wel (=1) of niet (=0) in samenhang met of geïntegreerd in andere vakken.

*Toetsing Engels.* De diversiteit aan toetsvormen die de leerkracht gebruikt is gehercodeerd naar een dichotome variabele (1 = een vorm van toetsing voor Engels, 0 = Engels wordt op onze school niet getoetst).

*Overdracht VO.* De diverse vormen van overdracht aan het voortgezet onderwijs van leerlinggegevens voor Engels zijn gehercodeerd naar een dichotome variabele: wel (=1) of niet (=0).

*Coördinator Engels.* Op school is wel (= 1) of niet (=0) sprake van een werkgroep of coördinator Engels.

*Prioritering in de les.* Het rangnummer die de leerkracht heeft gegeven voor de mate waarin hij of zij aandacht in de lessen Engels heeft voor de vaardigheden lezen, luisteren, spreken (solo), schrijven, woordenschat, grammatica en gespreksvaardigheid (interactief) is het uitgangspunt voor de prioritering in de les. De range loopt uiteen van 1 (meeste aandacht) tot 7 (minste aandacht) en is als domeinspecifieke predictor in de meerniveau modellen meegenomen. Dit impliceert dat in de modellen voor de samenhang met leesvaardigheid van de leerlingen het rangnummer voor lezen is meegenomen, voor luistervaardigheid het rangnummer voor luisteren, voor woordenschat het rangnummer voor woordenschat en voor gespreksvaardigheid de rangnummers voor spreken en gespreksvaardigheid.

*Prioritering bij toetsing.* De prioritering bij de toetsing voor Engels is het rangnummer die de leerkracht heeft gegeven voor frequentie waarmee hij of zij de vaardigheden lezen, luisteren, spreken (solo), schrijven, woordenschat, grammatica en gespreksvaardigheid (interactief) toetst. De range loopt uiteen van 1 (meest) tot 7 (minst) en is als domeinspecifieke predictor in de meerniveau modellen meegenomen, conform de prioritering in de les.

*Lesactiviteiten.* De mate van toepassen van diverse onderwijsactiviteiten zijn domeinspecifiek in de meerniveau modellen als volgt meegenomen. De antwoordschaal van een item is steeds 1 = (bijna) nooit, 2 = af en toe, en 3 = vaak.

- Voor leesvaardigheid zijn twee somscores meegenomen als predictoren. De eerste somscore is de frequentie van impliciet lezen van korte en langere Engelse teksten uit een boek, tijdschrift, tablet of computer. De tweede somscore geeft de frequentie van gericht lezen door de leerlingen weer. Ze lezen een Engelse leestekst en beantwoorden daarover Nederlandse of Engelse vragen.
- Ook voor luistervaardigheid gaat het om twee somscores die als predictoren worden meegenomen. De eerste somscore is de frequentie van impliciet luisteren van de leerlingen. Dit is het luisteren naar een tekst of gesprek in het Engels van tv, internet of uit een film met en zonder ondertiteling. De tweede somscore is de frequentie van gericht luisteren. Leerlingen luisteren gericht naar een Engelse luistertekst en beantwoorden daarover Nederlandse of Engelse vragen.
- Voor woordenschat nemen we ook twee somscores mee in de modellen: receptieve en productieve woordenschat. Receptieve woordenschat geeft weer hoe vaak leerlingen Engels woorden bij afbeeldingen zoeken en hoe vaak ze Engelse woorden classificeren. Productieve woordenschat betreft de frequentie waarmee leerlingen een tegenovergesteld woord in het Engels zoeken en Engelse woorden uit het hoofd leren.
- Voor gespreksvaardigheid gaat het om één somscore, namelijk hoe vaak leerlingen een gesprek in het Engels voeren met één of meer medeleerlingen en met de leerkracht.

### **Model 5**

*Opleidingsniveau.* Het opleidingsniveau van de leerkracht is een nominale variabele met drie categorieën. Het gaat om het hoogste niveau dat is afgerond met een diploma: reguliere PABO, verkorte PABO of academische PABO. Reguliere PABO is de referentiecategorie in de modellen.

*Ervaring in het geven van Engels.* Het aantal jaren dat een leerkracht Engels geeft in het basisonderwijs is gecategoriseerd tot een ordinale variabele: 1 = 0-5 jaar, 2 = 6-10 jaar, 3 = 11-15 jaar, 4 = 16-20 jaar en 5 = meer dan 21 jaar.

*Moedertaal.* De moedertaal van de leerkracht is een nominale variabele met de drie categorieën: Nederlands, Engels of een andere taal. Nederlands is de referentiecategorie.

*Buitenschools gebruik Engels.* scale\_buitenschoolsgebruik (zie paragraaf 5.3.2).

*Ingeschatte vaardigheid.* De inschatting van het eigen vaardigheidsniveau Engels van leerkrachten is domeinspecifiek meegenomen in de modellen.

- Voor leesvaardigheid is de de schaal\_zelfinschatting\_lezen meegenomen (zie paragraaf 5.3.2).
- Voor luistervaardigheid betreft het twee schalen schaal\_zelfinschatting\_luisteren en schaal\_zelfinschatting\_spreken (zie paragraaf 5.3.2).
- Voor woordenschat hebben we de schaal\_zelfinschatting\_lezen meegenomen (zie paragraaf 5.3.2).
- Voor gespreksvaardigheid gaat het ook om twee schalen: de schaal voor zelfinschatting bij luisteren en de schaal voor zelfinschatting bij spreken.

#### **Model 6**

*Thuis taal.* De dominante thuistaal van leerlingen op school is bevraagd bij de directeur en uitgedrukt als een percentage. Er zijn drie categorieën onderscheiden voor deze nominale variabele: Nederlands, Nederlands dialect of een andere taal als thuistaal. Het percentage leerlingen met Nederlands als dominante thuistaal is de referentiecategorie.

*Stedelijkheid locatie.* De urbanisatiegraad van de schoollocatie is uitgedrukt als een ordinale variabele met vijf categorieën (1 = niet, 2 = weinig, 3 = matig, 4 = sterk, 5 = zeer sterk). Een niet-stedelijke schoollocatie is de referentiecategorie.

*Denominatie.* De denominatie van de school is een nominale variabele. Op geleide van de data zijn de vijf meest frequent voorkomende categorieën onderscheiden: algemeen bijzonder, openbaar, protestants christelijk, rooms katholiek en overig (=antroposofisch). Algemeen bijzonder is de referentiecategorie in de modellen.

*Percentage gewichten leerlingen.* Het percentage leerlingen in groep 8 met een formatiegewicht (0.3 of 1.2) geldt als indicator voor het percentage gewichten leerlingen op een school.

*Schoolgemiddelde advies VO.* Het gemiddelde schooladvies van de leerlingen in groep 8 is gebruikt als indicator voor het schoolgemiddelde advies VO.

## Bijlage 12. Lesobservatieformulier Peil Engels (verdiepend onderzoek)

### Algemene informatie

Informatie over de school		
Schoolnaam:	Vul hier in	
Vestigingsnaam:	Vul hier in	
Leerkacht id/klas id	Vul hier in	
Informatie over de leerkracht en de klas waarin geobserveerd wordt		
Datum observatie (dd-mm-jjjj)	Vul hier in	
Naam en geslacht leerkracht	<input type="checkbox"/> Man <input type="checkbox"/> Vrouw	
Tijdstip observatie	<input type="checkbox"/> begin van de ochtend (voor 10 uur) <input type="checkbox"/> begin van de middag (12-14 uur) <input type="checkbox"/> einde van de ochtend (10 – 12 uur) <input type="checkbox"/> einde van de middag (na 14 uur)	
Starttijd les	Kies een item.	Kies een item.
Eindtijd les	Kies een item.	Kies een item.
Combinatieklas	<input type="checkbox"/> ja <input type="checkbox"/> nee	
Aantal leerlingen in de klas tijdens de observatie	Typ getal.	
Is er tijdens de les een tweede persoon aanwezig die onderwijskundige taken uitvoert	<input type="checkbox"/> ja <input type="checkbox"/> nee	
Het Engels van de leerkracht is grotendeels vloeiend	<input type="checkbox"/> ja <input type="checkbox"/> nee	
Het Engels van de leerkracht is grotendeels correct	<input type="checkbox"/> ja <input type="checkbox"/> nee	

### Observatieschema (ICALT-PO)

De leerkracht			Score 1-4
1.	Klimaat	Toont in gedrag en taalgebruik respect voor de leerling	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
2.	klimaat	Zorgt voor een ontspannen sfeer	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
3.	klimaat	Ondersteunt het zelfvertrouwen van de leerling	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
4.	klimaat	Zorgt voor wederzijds respect	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
5.	organisatie	Zorgt voor een ordelijk verloop van de les	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
6.	organisatie	Gaat tijdens de verwerking na of leerlingen de opdrachten op een juiste manier uitvoeren	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
7.	organisatie	Zorgt voor een doelmatig klassenmanagement	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
8.	organisatie	Betrekt alle leerlingen bij de les	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
9.	organisatie	Gebruikt de leertijd efficiënt	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
10.	instructie	Geeft duidelijke uitleg van leerstof en opdrachten	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
11.	instructie	Bevordert dat leerlingen hun best doen	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
12.	instructie	Geeft feedback aan de leerlingen	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
13.	instructie	Gaat tijdens de instructie na of leerlingen de leerstof hebben begrepen	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
14.	instructie	Geeft duidelijke uitleg van het gebruik van didactische hulpmiddelen en opdrachten	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
15.	activeren	Hanteert werkvormen die leerlingen activeren	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
16.	activeren	Laat leerlingen hardop denken	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
17.	activeren	Zorgt voor interactieve instructie- en werkvormen	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
18.	afstemmen	Stimuleert het zelfvertrouwen van zwakke leerlingen	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
19.	afstemmen	Biedt zwakke leerlingen extra leer- en instructietijd	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
20.	afstemmen	Stemt instructie af op verschillen tussen leerlingen	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
21.	afstemmen	Stemt verwerking van leerstof af op verschillen tussen leerlingen	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
22.	leerstrategie	Leert leerlingen hoe zij complexe problemen kunnen vereenvoudigen	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
23.	leerstrategie	Moedigt kritisch denken van leerlingen aan	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
24.	leerstrategie	Vraagt leerlingen na te denken over oplossingsstrategieën	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
De leerlingen			Score 1-4

	Betrokkenheid	Zijn betrokken bij de les	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
	betrokkenheid	Tonen zich geïnteresseerd	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
	betrokkenheid	Zijn actief op leren gericht	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4



## Vakdidactisch gedeelte van het observatie-instrument

	De leerkracht....	Dit kun je zien aan	Ja	Nee	Wat doet de leerkracht? - steekwoorden
Taaldoelen	1. Benoemt de taaldoelen		<input type="checkbox"/>	<input type="checkbox"/>	
	2. Motiveert de taaldoelen		<input type="checkbox"/>	<input type="checkbox"/>	
	3. Hanteert SMART-taaldoelen		<input type="checkbox"/>	<input type="checkbox"/>	
Uitleg & instructie	4. Activeert (talige) voorkennis bij de leerlingen	De leerkracht blikt terug op voorgaande les(sen)	<input type="checkbox"/>	<input type="checkbox"/>	
		De leerkracht vraagt naar wat leerlingen al weten over het onderwerp	<input type="checkbox"/>	<input type="checkbox"/>	
	5. Controleert of <i>de meeste</i> leerlingen begrijpen wat zij lezen of horen	De leerkracht stelt controlerende vragen voor, tijdens of na het lezen of luisteren – <u>de leerlingen</u>	<input type="checkbox"/>	<input type="checkbox"/>	
		.... Geven verbaal antwoord	<input type="checkbox"/>	<input type="checkbox"/>	
		.... Maken gebruik van kaartje, b.v. rode/groene kaartjes	<input type="checkbox"/>	<input type="checkbox"/>	

	.... Maken gebruik van wisbordjes	<input type="checkbox"/>	<input type="checkbox"/>	
	.... Maken gebruik van digitale middelen	<input type="checkbox"/>	<input type="checkbox"/>	
	Anders, namelijk	<input type="checkbox"/>	<input type="checkbox"/>	
6. Gebruikt het Nederlands wanneer de situatie daarnaar vraagt		<input type="checkbox"/>	<input type="checkbox"/>	
Interactie	7. Stimuleert leerkracht-leerling(en) interactie in de doeltaal en elicitert daarbij met name vrije interactie	<input type="checkbox"/>	<input type="checkbox"/>	
	8. Stimuleert dat leerlingen onderling Engels spreken	<input type="checkbox"/>	<input type="checkbox"/>	
	9. Onderneemt actie als leerlingen Nederlands spreken	<input type="checkbox"/>	<input type="checkbox"/>	
(Vak)didactisch repertoire	De leerkracht verlaagt zijn tempo	<input type="checkbox"/>	<input type="checkbox"/>	
	De leerkracht gebruikt gebaren en of andere lichaamstaal	<input type="checkbox"/>	<input type="checkbox"/>	
	10. Maakt <u>zijn/haar eigen spreken</u> in het Engels begrijpelijk voor de leerlingen	<input type="checkbox"/>	<input type="checkbox"/>	
	De leerkracht gebruikt visuele ondersteuning van bijvoorbeeld plaatjes of voorwerpen of digibord	<input type="checkbox"/>	<input type="checkbox"/>	
	De leerkracht herhaalt zichzelf in andere bewoording	<input type="checkbox"/>	<input type="checkbox"/>	

	11. Kiest werkvormen en activiteiten voor communicatieve doeleinden in de doeltaal		<input type="checkbox"/>	<input type="checkbox"/>	
		Door niet aandacht te besteden aan elke fout, maar alleen aan veelgemaakte fouten door de individuele leerling of de hele klas	<input type="checkbox"/>	<input type="checkbox"/>	
		Door de leerlingen <i>tijdens</i> het spreken te corrigeren	<input type="checkbox"/>	<input type="checkbox"/>	
		Door de leerlingen <i>na</i> het spreken te corrigeren	<input type="checkbox"/>	<input type="checkbox"/>	
	12. Besteedt aandacht aan (veelgemaakte) taalfouten van leerlingen	Door de leerlingen te herhalen, maar dan het in correcte bewoording te geven, zonder expliciet aandacht te vestigen op de gemaakte fout	<input type="checkbox"/>	<input type="checkbox"/>	
		Door de grammaticaregel uit te leggen	<input type="checkbox"/>	<input type="checkbox"/>	
		Door de leerlingen de fout zelf te laten herstellen door bijvoorbeeld een vraag te stellen, of de fout te herhalen in de vorm van een vraag	<input type="checkbox"/>	<input type="checkbox"/>	
		Door een andere leerling te vragen het correcte Engelse antwoord te geven op het moment dat een leerling alleen in het Nederlands antwoordt, omdat hij/zij het niet in het Engels kan zeggen	<input type="checkbox"/>	<input type="checkbox"/>	
	13. Geeft leerlingen positieve feedback op taalgebruik		<input type="checkbox"/>	<input type="checkbox"/>	
	14. Past de leerstof aan op het niveau van de individuele leerling of groepjes leerlingen		<input type="checkbox"/>	<input type="checkbox"/>	
	15. Past de leerstof aan op interesse en leerprofiel van de individuele leerling of groepjes leerlingen		<input type="checkbox"/>	<input type="checkbox"/>	

(Vak)didactisch repertoire	16. Stimuleert het gebruik van compenserende strategieën	Omschrijven van woorden als een leerling een woord niet weet	<input type="checkbox"/>	<input type="checkbox"/>	
		Het gebruik van synoniemen	<input type="checkbox"/>	<input type="checkbox"/>	
		Het gebruik van gebaren	<input type="checkbox"/>	<input type="checkbox"/>	
		Informatie afleiden uit de context	<input type="checkbox"/>	<input type="checkbox"/>	
	17. Besteedt aandacht aan ...	Leesvaardigheid	<input type="checkbox"/>	<input type="checkbox"/>	
		Luistervaardigheid	<input type="checkbox"/>	<input type="checkbox"/>	
		Schrijfvaardigheid	<input type="checkbox"/>	<input type="checkbox"/>	
		Spreekvaardigheid	<input type="checkbox"/>	<input type="checkbox"/>	
		Gespreksvaardigheid	<input type="checkbox"/>	<input type="checkbox"/>	
		Woordenschat	<input type="checkbox"/>	<input type="checkbox"/>	
		Grammaticale vaardigheid	<input type="checkbox"/>	<input type="checkbox"/>	

	Uitspraak	<input type="checkbox"/>	<input type="checkbox"/>	
	Spellingsvaardigheid	<input type="checkbox"/>	<input type="checkbox"/>	
	Cultuur van Engelstalige landen	<input type="checkbox"/>	<input type="checkbox"/>	
	Verschillen en overeenkomsten tussen het Engels en het Nederlands	<input type="checkbox"/>	<input type="checkbox"/>	
	Anders, namelijk ....	<input type="checkbox"/>	<input type="checkbox"/>	
18. Laat oefenen met het toepassen van een gegeven regel/woorden		<input type="checkbox"/>	<input type="checkbox"/>	
19. Hanteert de volgende werkvorm	leerlingen werken zelfstandig	<input type="checkbox"/>	<input type="checkbox"/>	
	leerlingen werken in tweetallen	<input type="checkbox"/>	<input type="checkbox"/>	
	leerlingen werken in groepjes	<input type="checkbox"/>	<input type="checkbox"/>	
	leerlingen luisteren/werken klassikaal	<input type="checkbox"/>	<input type="checkbox"/>	
	Anders, namelijk ... (naam activerende werkvorm)	<input type="checkbox"/>	<input type="checkbox"/>	

	20. De gekozen werkvormen en activiteiten sluiten aan bij de taaldoelen	<i>Dit geldt voor de hele les</i>	<input type="checkbox"/>	<input type="checkbox"/>	
Evaluatie	21. Sluit de les af door met de klas na te gaan of de taaldoelen zijn bereikt.		<input type="checkbox"/>	<input type="checkbox"/>	
algemeen	22. Gebruikt de les om leerlingen te motiveren voor het Engels		<input type="checkbox"/>	<input type="checkbox"/>	

Opmerkingen:

## Bijlage 13. Interview invulformulier (verdiepend onderzoek)

Informatie over de school	
Schoolnaam	
Vestigingsnaam	
Leerkracht naam	
Functie (vak-/groepskr)	

### Thema 1. Visie op Engels in het PO

1. Wat is uw visie op Engels in het PO: wat is voor u het belangrijkste in de lessen Engels? Wat wilt u bereiken?

2. Komt uw visie op Engels overeen met die van de school?

3. Vindt u Engels een belangrijk vak op school? Zo ja, zou Engels volgens u ook een onderdeel van de eindtoets moeten zijn?

## Thema 2. Een standaardles Engels (en de geobserveerde les)

Uw school is geselecteerd omdat de leerlingen van groep 8 van vorig jaar goed presteerden in relatie tot de leerlingen op vergelijkbare scholen. In uw reguliere lessen Engels leren de leerlingen blijkbaar bovengemiddeld goed Engels. Daarom willen we ons in dit interview vooral richten op een standaardles Engels, zoals u die dagelijks/wekelijks geeft.

4. Kunt u een reguliere les Engels, zeg maar een standaardles Engels, beschrijven? Hoe ziet zo'n les eruit? Wat doet u meestal?

- methode vs eigen materialen
- vooraf vastgestelde taaldoelen
- Aandacht voor alle vaardigheden?
- Moeilijke vaardigheid?
- Balans input-output en afwisseling werkvormen
- Doeltaal=voertaal?
  - Bij leerkracht
  - Bij leerlingen
- Evaluatie doelen (in les, langere termijn)
- Aansluiting zone naaste ontwikkeling/differentiatie niveau, interesse, leerprofiel
- Gebruik ICT
- Motiveren voor Engels



5. Is de zojuist geobserveerde les representatief voor uw gebruikelijke lessen, m.a.w. een

- methode vs eigen materialen
- vooraf vastgestelde taaldoelen
- Aandacht voor alle vaardigheden?
- Moeilijke vaardigheid?
- Balans input-output en afwisseling werkvormen
- Doeltaal=voertaal?
  - Bij leerkracht
  - Bij leerlingen
- Evaluatie doelen (in les, langere termijn)
- Aansluiting zone naaste ontwikkeling/differentiatie niveau, interesse, leerprofiel
- Gebruik ICT
- Motiveren voor Engels

voorbeeld van een standaardles?

6. Zo ja, waarom?

Zo nee, op welke punten verschilt de geobserveerde les van de reguliere lessen?

Ja:

Nee:

**Thema 3. Opleiding en professionele ontwikkeling**

7. Bent u in uw opleiding tot leerkracht met het vak Engels in aanraking gekomen of is dit vooral via na-/bijscholing tot stand gekomen? (of een combinatie)?

8. Heeft u het gevoel dat u voldoende bent toegerust om Engels te geven en waarom wel/niet?

9. Waar heeft u daadwerkelijk iets aan gehad als u terugkijkt op uw scholing aangaande het vak Engels voor de basisschool?

10. Wat heeft u in de opleiding concreet gemist als het gaat om het lesgeven in Engels?

11. Zijn dat aspecten die u nu nog steeds moeilijk/lastig vindt, of heeft u dat via nascholing kunnen compenseren? Heeft u voorbeelden van nascholing of eigen bijscholingsprogramma's waarvan u denkt dat die ook nuttig zouden zijn voor andere leerkrachten?

12. Indien vakleerkracht Engels: kunt u aangeven *wat precies* u onderscheidt van 'gewone' groepsleerkrachten? Ervaart u dat zij met bepaalde aspecten van het vak juist meer of minder moeite hebben? Hoe zou dat volgens u kunnen verbeteren?

Onderstaande aspecten dienen bij **thema 2** aan de orde te komen. Na ieder aspect wordt aangegeven waarnaar doorgevraagd kan worden.

- **methode vs eigen materialen:** gebruik van de methode versus (deels) zelf samengestelde lessen met eigen materialen (vragen naar: concrete voorbeelden van materialen)
- **vooraf vastgestelde taaldoelen:** Gebruik van vooraf vastgestelde *taaldoelen* en manieren waarop deze doelen centraal worden gesteld aan het begin van de les. Wordt er ook rekening gehouden met kerndoelen en ERK doelen?
- **Aandacht voor alle vaardigheden:** Aandacht voor alle vaardigheden? Gespreksvaardigheid?
- **Moeilijke vaardigheid:** En wat zijn moeilijk te realiseren vaardigheden?
- **Balans input-output en afwisseling werkvormen:** vergelijk met het vierfasenmodel: variatie in receptieve en productieve vaardigheden tijdens een les, gebruikmakend van een opbouw waarin nieuwe kennis wordt geconsolideerd. Ter illustratie de volgende lesopbouw: 1) luisteren naar liedje met woorden over night en stars, 2) lezen van tekstje over een spooktocht laat in de avond/nacht, schrijven van kort verhaaltje over verschillen tussen dag en nacht, gesprekje voeren over een gebeurtenis laat op de avond.
- **Doeltaal=voertaal?**
- Bij leerkracht: Gebruik van de doeltaal en doelbewust gebruik van het Nederlands door de leerkracht (wanneer?)
- Bij leerlingen: gebruik van Engels door leerlingen? Belangrijk en waarom?
- **Evaluatie doelen (in les en op langere termijn):** hoe stellen leerkrachten vast waar leerlingen staan? Eind les /over langere periode. Welke toetsen worden gebruikt voor welke vaardigheden. Wat doet de leerkracht met deze gegevens? Formatief gebruik?
- **Aansluiting zone naaste ontwikkeling/differentiatie niveau, interesse, leerprofiel:** Aansluiting aanbod Engelse lessen bij de "zone van naaste ontwikkeling" van de leerlingen, uitmondend in differentiatie naar niveau, interesse of leerprofiel. Onder differentiatie naar leerprofiel wordt verstaan dat de leerkracht rekening houdt met voorkeuren van leerlingen als culturele achtergrond en leervoorkeuren (b.v. voorkeur voor leren in groepjes of zelfstandig - (Tomlinson, 2003).
- **Gebruik ICT:** voorbeelden van ICT/digibord aanbod, zelf gezochte ICT-materialen en digitale hulpmiddelen → Concrete voorbeelden
- **Motiveren voor Engels:** Hoe laat leerkracht relevantie van Engels voor de leerlingen tijdens de les zien? Hoe sluiten lessen aan bij belevingswereld leerlingen? Hoe laat leerkracht leerlingen succeservaringen op laten doen en wordt motivatie bij leerlingen gestimuleerd?

## Bijlage 14. Observatie instrument (lange versie)

	De leerkracht....	Dit kun je zien aan	Ja	Nee	Wat doet de leerkracht? - steekwoorden
Taaldoelen	1. Benoemt de taaldoelen	<i>De leerkracht benoemt wat de leerlingen aan het eind van de les hebben geleerd</i>			
	2. Motiveert de taaldoelen	<i>De leerkracht vertelt leerlingen wat ze aan het bereiken van de taaldoelen hebben buiten de les, waarom deze taaldoelen belangrijk voor de leerlingen zijn.</i>			
	3. Hanteert SMART-taaldoelen				
Uitleg & instructie	4. Activeert (talige) voorkennis bij de leerlingen	De leerkracht blikt terug op voorgaande les(sen)			
		De leerkracht vraagt naar wat leerlingen al weten over het onderwerp			
	5. Controleert of de meeste leerlingen begrijpen wat zij lezen of horen	De leerkracht stelt controlerende vragen voor, tijdens of na het lezen of luisteren – de leerlingen			
		.... Geven verbaal antwoord			
		.... Maken gebruik van kaartje, b.v. rode/groene kaartjes			
		.... Maken gebruik van mini-whiteboards			
		.... Maken gebruik van digitale middelen			
	Anders, namelijk				
6. Gebruikt het Nederlands wanneer de situatie daarnaar vraagt	<i>De docent gebruikt het Nederlands om Iln zich veilig te doen voelen.</i> <i>De docent gebruikt het Nederlands om te controleren of Iln de input hebben begrepen.</i> <i>De docent gebruikt het Nederlands als hij grammatica-uitleg geeft.</i> <i>Anders, namelijk</i>				
Interactie	7. Stimuleert leerkracht-leerling(en) interactie in de doeltaal en elicitert daarbij met name vrije interactie	<i>De leerkracht doet meer dan het stellen van begripsvragen die tot (zeer) korte antwoorden leiden, bestaande uit 1 of 2 woorden.</i> <i>De leerkracht stelt niet alleen gesloten vragen</i> <i>De leerkracht vraagt de leerlingen niet alleen maar taal te reproduceren die ze al geleerd hebben</i> <i>De leerkracht voert een spontaan gesprek met de leerling(en)</i> <i>De leerkracht vraagt door na een antwoord van de leerling(en)</i>			

	8. Stimuleert dat leerlingen onderling Engels spreken	<i>De leerkracht stimuleert dat leerlingen onderling Engels spreken</i>  <i>De leerkracht tolereert Nederlands bij de voorbereiding van de taak maar corrigeert leerlingen die overgaan op het Nederlands bij de uitvoering, of zorgt dat andere leerlingen hulp bieden</i>			
	9. Onderneemt actie als leerlingen Nederlands spreken				
(Vak)didactisch repertoire	10. Maakt zijn/haar eigen spreken in het Engels begrijpelijk voor de leerlingen	De leerkracht verlaagt zijn tempo			
		De leerkracht gebruikt gebaren en of andere lichaamstaal			
		De leerkracht gebruikt visuele ondersteuning van bijvoorbeeld plaatjes of voorwerpen of digibord			
		De leerkracht herhaalt zichzelf in andere bewoording			
	11. Kiest werkvormen en activiteiten voor communicatieve doeleinden in de doeltaal	<i>De leerkracht laat leerlingen iets spreken of schrijven dat relevant voor ze is in de buitenschoolse context</i>			
	12. Besteedt aandacht aan veelgemaakte taalfouten van leerlingen	Door niet aandacht te besteden aan elke fout, maar alleen aan veelgemaakte fouten door de individuele leerling of de hele klas			
		Door de leerling <i>tijdens</i> het spreken te corrigeren			
		Door de leerling <i>na</i> het spreken te corrigeren			
		Door de leerling te herhalen, maar dan het in correcte bewoording te geven, zonder expliciet aandacht te vestigen op de gemaakte fout			
		Door de grammaticaregel uit te leggen			
Door de leerling de fout zelf te laten herstellen door bijvoorbeeld een vraag te stellen, of de fout te herhalen in de vorm van een vraag					
Door een andere leerling te vragen het correcte Engelse antwoord te geven op het moment dat een leerling alleen in het Nederlands antwoordt, omdat hij/zij het niet in het Engels kan zeggen					
13. Geeft leerlingen positieve feedback op taalgebruik	<i>De leerkracht geeft aan wat de leerling al kan met de taal</i>				
14. Past de leerstof aan op het niveau van de individuele leerling of groepjes leerlingen	<i>De leerkracht differentieert in instructie, verwerking, product, etc.</i>				
15. Past de leerstof aan op interesse en leerprofiel van de individuele leerling of groepjes leerlingen	<i>De leerkracht differentieert in onderwerp, materialen, etc.</i>				
16. Stimuleert het gebruik van compenserende	Omschrijven van woorden als een leerling een woord niet weet				
	Het gebruik van synoniemen				

	strategieën	Het gebruik van gebaren				
		Informatie afleiden uit de context				
	17. Besteedt aandacht aan ...	Leesvaardigheid				
		Luistervaardigheid				
		Schrijfvaardigheid				
		Spreekvaardigheid				
		Gespreksvaardigheid				
		Woordenschat				
		Grammaticale vaardigheid				
		Uitspraak				
		Spelling				
		Cultuur van Engelstalige landen				
		Verschillen en overeenkomsten tussen het Engels en het Nederlands				
		Anders, namelijk....				
		18. Laat oefenen met het toepassen van een gegeven regel	<i>Oefeningen</i> <i>Rollenspel</i> <i>Vrije productie</i>			
		19. Hanteert de volgende werkvorm	Leerlingen werken zelfstandig			
	Leerlingen werken in tweetallen					
Leerlingen werken in groepjes						
Leerlingen luisteren/werken klassikaal						
Anders, namelijk ... (naam activerende werkvorm)						
20. De gekozen werkvormen en activiteiten sluiten aan bij de taaldoelen	<i>Dit geldt voor de hele les</i>					
<b>Evaluatie</b>	21. Sluit de les af door met de klas na te gaan of de taaldoelen zijn bereikt.	De leerkracht gaat inhoudelijk na of de leerlingen de taaldoelen hebben bereikt.				
<b>algemeen</b>	22. Gebruikt de les om leerlingen te motiveren voor het Engels	Zorgt ervoor dat lln plezier hebben in de les, door bijvoorbeeld zelf enthousiast te zijn of door activiteiten te kiezen die lln leuk vinden.				

## Bijlage 15. Overzicht van variabelen in het leerlingvragenlijstbestand

Variabele naam	Label
LlId	Leerling ID
VrId	Vragenlijst ID
School_ID	School ID
sted	Mate van stedelijkheid
Denom_1	Denominatie van de school
schoolregio	Regio waarin de school zich bevindt
d_gebd	Geboortedatum bekend?
LF_jaar_1okt2017	Leeftijd (jaar) op 1 oktober 2017
GES_code	Geslacht
VO_adv_code	Advies VO
VVTO_code	VVTO (ja/nee)
Stratum_code	Stratum Code
Combi_code_8	Combi code 8: VVTO (2) - Stratum (4)
Combi_code_4	Combi code 4: VVTO (2) - Stratum (2)
Dyslexie	Notaties scholen omtrent dyslexie
new_gew	Leerlinggewicht
PctMiss54	Percentage missende waarden
ScVr_01	1. Vind je Engels op school een gemakkelijk vak.
ScVr_02	2. Stel je voor: Je bent op vakantie. Kinderen vragen in het Engels of je mee wilt spelen. Wat doe je?
ScVr_03	3. Vind je Engels op school een leuk vak?
ScVr_04	4. Vind je het belangrijk voor je toekomst om goed Engels te leren?
ScVr_05	5. Stel je voor: Er is op tv een Engelse tekenfilm die je graag wilt zien, maar de film heeft geen ondertiteling. Wat doe je?
ScVr_06	6. Durf je Engels te spreken?
ScVr_07	7. Stel je voor: Je krijgt een boek dat je graag wilt lezen, maar het boek is Engelstalig. Wat doe je?
ScVr_08	8. Vind je taal (Nederlands) op school gemakkelijk?
ScVr_09	9. (new) Spreek je wel eens Engels tijdens de les Engels?
ScVr_10_1	10.1. Wat vind je leuk tijdens Engels: teksten lezen
ScVr_10_2	10.1. Wat vind je leuk tijdens Engels: schrijven Engels
ScVr_10_3	10.3. Wat vind je leuk tijdens Engels: praten Engels
ScVr_10_4	10.4. Wat vind je leuk tijdens Engels: luisteren (liedjes/.verhalen)
ScVr_10_5	10.5. Wat vind je leuk tijdens Engels: liedjes zingen
ScVr_10_6	10.6. Wat vind je leuk tijdens Engels: niets
ScVr_11	11. (was 9) Stel je voor: Op internet zoek je informatie over een onderwerp dat je erg interesseert. Je vindt een Engelstalige website. Wat doe je?
ScVr_12	12 (was 11). Heb je in het Engels wel eens een e-mailtje geschreven of gechat, bijvoorbeeld tijdens het gamen?
ScVr_13	13. (new) Heb je in het Engels wel eens gechat
ScVr_14_1	14.1. (was 16.1...) Engelstalige filmpjes op internet kijken zonder Nederlandse ondertiteling (bijvoorbeeld Youtube)
ScVr_14_2	14.2. Engelstalige computerspelletjes spelen

ScVr_14_3	14.3. Een boek of tijdschrift in het Engels lezen
ScVr_14_4	14.4. Op tv, dvd, internet of in de bioscoop Engelstalige films, series of andere programma's kijken met Nederlandse ondertiteling
ScVr_14_5	14.5. Luisteren naar Engelse liedjes
ScVr_14_6	14.6. Engelse websites bekijken om informatie te zoeken
ScVr_15	15. Hoeveel tijd besteed je per dag aan Le, Lu, Schr, Spre, Engels
ScVr_16	16. Begrijp je Engelstalige liedjes die je bijvoorbeeld op de radio/op internet hoort?
ScVr_17	17. Luister je vaker, even vaak of minder vaak Engels liedjes dan klasgenoten?
ScVr_18	18. Spreek je wel eens Engels buiten school?
ScVr_19	19. Spreek je wel eens Engels met een vriend(in)?
ScVr_20	20. Spreek je wel eens Engels met familie?
ScVr_21	21. Spreek je wel eens Engels met mensen op vakantie?
ScVr_22	22. Spreek je wel eens Engels met onbekenden?
ScVr_23_1	Vr23.1: Je spreekt thuis: Nederlands
ScVr_23_2	Vr23.2: Je spreekt thuis: Engels
ScVr_23_3	Vr23.3: Je spreekt thuis: Een andere taal
ScVr_24_1	Vr24.1: Je moeder spreekt met jou: Nederlands
ScVr_24_2	Vr24.2: Je moeder spreekt met jou: Engels
ScVr_24_3	Vr24.3: Je moeder spreekt met jou: Een andere taal
TekstVr24	
ScVr_25	Vr25: Je moeder is in Nederland geboren
ScVr_26_1	Vr26.1: Je vader spreekt met jou: Nederlands
ScVr_26_2	Vr26.2: Je vader spreekt met jou: Engels
ScVr_26_3	Vr26.3: Je vader spreekt met jou: Een andere taal
TekstVr26	
ScVr_27	Vr27. Je vader is in Nederland geboren
ScVr_28_01	Vr_28_1 (was 24.1...) Lezen: Ik kan informatie vinden in menu's en in folders
ScVr_28_02	Vr_28_2 Lezen: Ik kan borden en posters met eenvoudige woorden en zinnen begrijpen
ScVr_28_03	Vr_28_3 Lezen: Ik kan beschrijvingen van gebeurtenissen of gevoelens in brieven, mails of chats begrijpen
ScVr_28_04	Vr_28_4 Lezen: Ik kan korte, eenvoudige e-mails, sms-berichten en what's app-berichten lezen
ScVr_28_05	Vr_28_5 Luisteren: Ik kan simpele zinnen begrijpen in een gesprek over familie en het dagelijks leven
ScVr_28_06	Vr_28_6 Luisteren: Ik kan de belangrijkste punten in omroepberichten volgen
ScVr_28_07	Vr_28_7 Luisteren: Ik kan woorden en zinnen verstaan in een gesprek over scholen hobby's
ScVr_28_08	Vr_28_8 Luisteren: Ik kan de hoofdzaken begrijpen van Engelstalige radio- en tvprogramma's
ScVr_28_09	Vr_28_9 Luisteren: Ik kan de hoofdzaken begrijpen van Engelstalige filmpjes op internet
Scvr_29_01	Vr_29_1 (was 25.1...) Spreken: Ik kan mijn mening in het Engels geven en daarbij een uitleg geven
Scvr_29_02	Vr_29_2 Spreken: Ik kan in eenvoudige woorden en zinnen vertellen over mijn familie of omgeving
Scvr_29_03	Vr_29_3 Spreken: Als mensen langzaam spreken en dingen herhalen, kan ik deelnemen aan een gesprek in het Engels
Scvr_29_04	Vr_29_4 Spreken: Ik kan eenvoudige vragen stellen en korte antwoorden geven
Scvr_29_05	Vr_29_5 Spreken: Ik kan de inhoud van een boek of film navertellen in het Engels



Scvr_29_06	Vr_29_6 Schrijven: Ik kan een vakantiekaart, e-mail, what's app of chatbericht schrijven in het Engels
Scvr_29_07	Vr_29_7 Schrijven: Ik kan een tekst schrijven over bijvoorbeeld mijn huis, sport of hobby's
Scvr_29_08	Vr_29_8 Schrijven: Ik kan aantekeningen maken in het Engels
Scvr_29_09	Vr_29_9 Schrijven: Ik kan mijn naam, adres en nationaliteit invullen op een formulier
Scvr_29_10	Vr_29_10 Schrijven: Ik kan mijn ervaringen beschrijven in het Engels
ScVr_30	Vr30: VO advies
SelfConf2567	Self confidence zonder Vr 11 (Vr 2+Vr5+Vr6+Vr7; controle of de categorieen goed gescoord waren)
SelfConf256711	Self confidence met Vr 11
SelfConf_best	Linguistic Self Confidence (LSC: 2,5,6,7,11comp)
Attitude_best	Attitude (1,3,4,10.1/2/3/6R)
buitenssch_best	Buitenschoolseactiviteit (BSE; 12-22; zonder 14.4 en 14.5 en met 17T)
IV_LEZ	Ingeschatte vaardigheid Lezen (28.1-28.4)
IV_LUI	Ingeschatte vaardigheid Luisteren(28.5-28.9)
IV_SPR	Ingeschatte vaardigheid Spreken (29.1-29.5)
IV_SCH	Ingeschatte vaardigheid Schrijven (alle items: 29.06-29.10)
IV_SCH4	Ingeschatte vaardigheid Schrijven (4 items: zonder 29.9)
IV_Engels	Ingeschatte vaardigheid Engels (Som IV LEZ/LUI/SPR/SCH)
IV_LSC	IV + LSC
IV_LSC_BSE	IV + LSC + BSE
VL_SOM	IV + LSC + BSE + Attitude

## Bijlage 16. Overzicht van variabelen in het intakebestand

Variabele	Label
School_ID	School ID
intake	Heeft er een intake plaatsgevonden?
vorm	Wijze van intake
profiel_sp	profiel verwacht obv steekproef
profiel_gr8	1. profiel na check variabelen en telefonisch
start_jg	2.1 Start Engels in jaargroep (zoals genoteerd door de school)
start_kj	2.2 Start Engels, kalenderjaar (zoals genoteerd door de school)
profiel_info	Opmerkingen van de school omtrent profiel
owt_1_min	3. Onderwijstijd besteed aan Engels in groep 1 (huidige situatie)
owt_1_wk	3. Gemiddeld aantal weken dat Engels is gegeven in groep 1 (huidige situatie)
owt_2_min	3. Onderwijstijd besteed aan Engels in groep 2 (huidige situatie)
owk_2_wk	3. Gemiddeld aantal weken dat Engels is gegeven in groep 2 (huidige situatie)
owt_3_min	3. Onderwijstijd besteed aan Engels in groep 3 (huidige situatie)
owt_3_wk	3. Gemiddeld aantal weken dat Engels is gegeven in groep 3 (huidige situatie)
owt_4_min	3. Onderwijstijd besteed aan Engels in groep 4 (huidige situatie)
owt_4_wk	3. Gemiddeld aantal weken dat Engels is gegeven in groep 4 (huidige situatie)
owt_5_min	3. Onderwijstijd besteed aan Engels in groep 5 (huidige situatie)
owt_5_wk	3. Gemiddeld aantal weken dat Engels is gegeven in groep 5 (huidige situatie)
owt_6_min	3. Onderwijstijd besteed aan Engels in groep 6 (huidige situatie)
owt_6_wk	3. Gemiddeld aantal weken dat Engels is gegeven in groep 6 (huidige situatie)
owt_7_min	3. Onderwijstijd besteed aan Engels in groep 7 (huidige situatie)
owt_7_wk	3. Gemiddeld aantal weken dat Engels is gegeven in groep 7 (huidige situatie)
owt_8_min	3. Onderwijstijd besteed aan Engels in groep 8 (huidige situatie)
owt_8_wk	3. Gemiddeld aantal weken dat Engels is gegeven in groep 8 (huidige situatie)
v4a	4. Is er een coordinator/werkgroep Engels op school?
v4b	4.b Coördinator/werkgroep
v5a_1	5.1 Faciliteiten Engels -budget
v5a_2	5.2 Faciliteiten Engels - partneruitwisseling
v5a_3	5.3 Faciliteiten Engels - activiteiten Eng voertaal
v5a_4	Geen faciliteiten voor Engels
v5a_5	Anders....
v5a_5_anders	namelijk
v5b	5.b Aantal leerkrachten op school dat nascholing heeft gehad
v5c	5.c De insteek van de nascholing Engels
v5c4_anders	5.c.X andere insteek, namelijk
v6a_1	6.a.1 Percentage leerlingen op school met Nederlands als dominante thuistaal
v6a_2	6.a.2 Percentage leerlingen op school met streektaal/dialect als dominante thuistaal
v6a_3	6.a.3 Percentage leerlingen op school waar Nederlands geen dominante thuistaal is
v6b_1	6.b.1 Percentage leerlingen op school met uitgesproken taalrijke thuissituatie
v6b_2	6.b.2 Percentage leerlingen op school met gemiddeld taalrijke thuissituatie
v6b_3	6.b.3 Percentage leerlingen met taalarme thuissituatie

v7	7. Overdracht leerlinggegevens Engels richting VO-school
v7_anders	anders, namelijk

## Bijlage 17. Overzicht van variabelen in het leerkrachtbestand

Variabele naam	Beschrijving variabele
School_ID	school ID
Combi_code_8	Combi code 8: VVTO (2) - Stratum (4)
Combi_code_4	Combi code 4: VVTO (2) - Stratum (2)
lkr_koppelbaar	is leerkracht koppelbaar voor totaalbestand?
N_BREAK	aantal leerlingen per school met data
barc	code leerkracht
Leerkracht_nr	door leerkrachten genoteerd nummer op vragenlijst (in geval van meerdere leerkrachten)
Klasnaam	klasnaam
Sc_vr1_1	Bent u de groepsleerkracht van deze groep
Sc_vr1_2	Wat is uw geslacht?
Sc_vr1_3	Wat is uw leeftijd?
Sc_vr1_4	Bent u in Nederland geboren?
Sc_vr1_5	Beschouwt u Nederlands als uw moedertaal?
Sc_vr1_6	Beschouwt u Engels als uw moedertaal?
Sc_vr1_7	In welke taal/talen communiceert u regelmatig?
Vraag1_7_txt	U kunt hier andere talen specificeren ....
Sc_vr1_8	Hoelang werkt u als leerkracht in het basisonderwijs?
Sc_vr1_9	Hoelang geeft u Engels als leerkracht in het basisonderwijs?
Sc_vr1_10	Wat is uw behaalde diploma?
Sc_vr2_1	Welke scholing, Engels op pabo, verplichte module
Sc_vr2_2	Welke scholing, Engels op pabo, keuzemodule
Vraag2_2_txt	Anders, namelijk
Sc_vr2_3	Nadruk tijdens lessen Engels?
Sc_vr2_4	Tevredenheid omvang en inhoud Engels op Pabo?
Sc_vr2_5	Na/bijbscholing ogv Engels?
Sc_vr2_6	Nadruk in na/bijbscholing Engels?
Sc_vr3_1	Hoe vaak kijkt u naar Engelse TV of Engelstalige films/series, zonder Nederlandse ondertiteling?
Sc_vr3_2	Hoe vaak kijkt u naar Engelstalige films/series, met Nederlandse ondertiteling?
Sc_vr3_3	Hoe vaak luistert u naar Engelstalige muziek?
Sc_vr3_4	Hoe vaak leest u een tekst in het Engels, zoals een boek, tijdschrift of blog?
Sc_vr3_5	Hoe vaak leest u zinnen in het Engels tijdens het gebruik van 'social media' of 'online gamen'?
Sc_vr3_6	Hoe vaak spreekt u Engels buiten school?
Sc_vr3_7	Hoe vaak schrijft u een tekst in het Engels? (social media, chat, e-mail, brief, kaart)?
Sc_vr4	gehanteerde doelen voor de lessen Engels?
Vraag4_txt	Anders, namelijk
Vraag5_1_txt	Gebruikte methode in groep 8

Vraag5_2_txt	Naam uitgever gebruikte methode groep 8
Vraag5_3_txt	Gebruikte methode in groep 7
Vraag5_4_txt	Naam uitgever gebruikte methode groep 7
Sc_vr6_1	uitsluitend gebruik methode groep 8?
Sc_vr6_2	uitsluitend gebruik methode groep 7 vorig jaar?
Sc_vr7a_1	hoe vaak zet u authentiek Engels beeld- en geluidsmateriaal in?
Sc_vr7a_2	hoe vaak zet u Engelstalig tekstmateriaal in?
Sc_vr7a_3	Andere leermiddelen, namelijk
Vraag7a_3_txt	Hoe vaak zit u dit leermiddel in?
Sc_vr7a_4	Andere leermiddelen, namelijk
Vraag7a_4_txt	Hoe vaak zit u dit leermiddel in?
Sc_vr7a_5	Andere leermiddelen, namelijk
Vraag7a_5_txt	Hoe vaak zit u dit leermiddel in?
Sc_vr7b	Gebruik leermiddelen in groep 8 en groep 7 zelfde?
Sc_vr8_1	De leeractiviteiten in de methode spreken mijn leerlingen aan
Sc_vr8_2	De activiteiten in de methode sluiten aan bij het niveau van mijn leerlingen
Sc_vr8_3	De methode besteedt aandacht aan de relevante aspecten van Engels voor mijn leerlingen
Sc_vr8_4	Mijn indruk is dat mijn leerlingen door de activiteiten in de methode veel leren
Sc_vr8_5	De methode biedt goede mogelijkheden voor differentiatie
Sc_vr8_1_rec	Sc_vr8_1_rec
Sc_vr8_2_rec	Sc_vr8_2_rec
Sc_vr8_3_rec	Sc_vr8_3_rec
Sc_vr8_4_rec	Sc_vr8_4_rec
Sc_vr8_5_rec	Sc_vr8_5_rec
Sc_vr9a_1	Prioritering aandacht - lezen
Sc_vr9a_2	Prioritering aandacht - luisteren
Sc_vr9a_3	Prioritering aandacht - spreken
Sc_vr9a_4	Prioritering aandacht - schrijven
Sc_vr9a_5	Prioritering aandacht - woordenschat
Sc_vr9a_6	Prioritering aandacht - grammatica
Sc_vr9a_7	Prioritering aandacht - gespreksvaardigheid
Sc_vr9b_1	Naspreken van nieuwe woorden
Sc_vr9b_2	Naspreken van standaard gesprekken in het Engels
Sc_vr9b_3	In het Engels vertellen wat er op plaatjes te zien is
Sc_vr9b_4	In het Engels vertellen over een gebeurtenis (b.v. verjaardag, sportactiviteit, weekend, vakantie)
Sc_vr9b_5	Een tekst, boek of film(pje) navertellen in het Engels
Sc_vr9b_6	het zingen van liedjes in het Engels
Vraag9b_txt	Anders, namelijk
Sc_vr9c_1	Een rollenspel spelen in het Engels
Sc_vr9c_2	Voeren van een gesprek in het Engels met één of meer

	medeleerlingen
Sc_vr9c_3	Voeren van een gesprek in het Engels met de leerkracht
Sc_vr9c_4	Voeren van een gesprek in het Engels met een near native speaker
Sc_vr9c_5	Voeren van een gesprek in het Engels met leeftijdgenoten via internet (bv Skype)
Sc_vr9c_6	Een spelletje of activiteit doen in het Engels
Vraag9c_txt	Anders, namelijk
Sc_vr9d_1	luisteren naar een tekst of gesprek in het Engels uit de methode
Sc_vr9d_2	Luisteren naar een tekst of gesprek in het Engels van tv, internet of uit een film (zonder ondertiteling)
Sc_vr9d_3	Luisteren naar een tekst of gesprek in het Engels van tv, internet of uit een film (met ondertiteling)
Sc_vr9d_4	Luisteren naar de leerkracht die alleen maar Engels spreekt
Sc_vr9d_5	Nederlandse vragen beantwoorden bij een Engelse luistertekst
Sc_vr9d_6	Engelse vragen beantwoorden bij een Engelse luistertekst
Vraag9d_txt	Anders, namelijk
Sc_vr9e_1	Korte Engelse teksten uit de methode lezen
Sc_vr9e_2	Korte Engelse teksten uit een tijdschrift of op de tablet/computer lezen
Sc_vr9e_3	Langere Engelse teksten uit een boek of tijdschrift of op de tablet/computer lezen
Sc_vr9e_4	Nederlandse vragen beantwoorden bij een Engelse leestekst
Sc_vr9e_5	Engelse vragen beantwoorden bij een Engelse leestekst
Sc_vr9e_6	Woorden invullen die in een Engelse leestekst zijn weggelaten
Sc_vr9e_7	Een Nederlandse samenvatting geven van een Engelse leestekst
Vraag9e_txt	Anders, namelijk
Sc_vr9f_1	De betekenis van woorden afleiden uit een Engelse context
Sc_vr9f_2	Engelse woorden opzoeken in een woordenboek / -lijst
Sc_vr9f_3	Engelse woorden proberen te omschrijven
Sc_vr9f_4	Een tegenovergesteld woord zoeken (Engels-Engels)
Sc_vr9f_5	Engelse woorden bij afbeeldingen (tekeningen / foto's) zoeken
Sc_vr9f_6	De Nederlandse vertaling geven van een Engels woord
Sc_vr9f_7	Engelse woorden classificeren (in categorieën plaatsen)
Sc_vr9f_8	Engelse woorden uit het hoofd leren
Sc_vr9f_9	Engelse woorden oefenen met de computer
Vraag9f_txt	Anders, namelijk
Sc_vr9g_1	Engelse woorden bij afbeeldingen (tekeningen/ foto's) schrijven
Sc_vr9g_2	Engelse zinnen bij afbeeldingen schrijven
Sc_vr9g_3	De Engelse vertaling van Nederlandse woorden opschrijven
Sc_vr9g_4	Engelse woorden en/of zinnen overschrijven

Sc_vr9g_5	Engelse woorden en/of zinnen opschrijven in een dictee
Sc_vr9g_6	Antwoorden op vragen bij een tekst opschrijven in het Engels
Sc_vr9g_7	Een brief, e-mail, WhatsApp, Instagrambericht of SMS schrijven in het Engels
Sc_vr9g_8	Een verhaaltje, gedichtje of liedje schrijven in het Engels
Sc_vr9g_9	Een werkstuk schrijven in het Engels
Vraag9g_txt	Anders, namelijk
Sc_vr10_1	Ik spreek Engels met de leerlingen voor:
Sc_vr10_2	Ik nodig de algehele klas uit om met mij Engels te spreken voor:
Sc_vr10_3	De leerlingen spreken Engels met elkaar voor:
Sc_vr11_1	Ik werk klassikaal met de leerlingen
Sc_vr11_2	De leerlingen werken in tweetallen of groepjes
Sc_vr11_3	De leerlingen werken individueel
Sc_vr12_1	Geeft u de leerlingen huiswerk voor Engels?
Sc_vr12_2	Huiswerk voor....
Vraag12_txt	Anders, namelijk
Sc_vr13_1	Integratie Engels met andere vakken?
Sc_vr13_2	Engels in samenhang met .....
Vraag13_2_txt	Anders, namelijk
Sc_vr14	Op welke wijze vindt overdracht m.b.t. Engels plaats van groep 7 naar groep 8
Vraag14_txt	Anders, namelijk
Sc_vr15a	Welke toetsen gebruikt u?
Vraag15a_txt	Anders, namelijk
Sc_vr15b_1	Krijgen leerlingen met achterblijvende prestaties voor Engels gericht ondersteuning om de achterblijvende prestaties in te halen?
Sc_vr15b_2	Wat is de aard van de ondersteuning?
Vraag15b_2_txt	Anders, namelijk
Sc_vr15c_1	Toetsing van lezen
Sc_vr15c_2	Toetsing van luisteren
Sc_vr15c_3	Toetsing van spreken
Sc_vr15c_4	Toetsing van schrijven
Sc_vr15c_5	Toetsing van woordenschat
Sc_vr15c_6	Toetsing van grammatica
Sc_vr15c_7	Toetsing van gespreksvaardigheid
Sc_vr16_1	De doelen voor Engels in het primair onderwijs zijn duidelijk over wat leerlingen moeten kunnen
Sc_vr16_2	De meeste leerlingen kennen al aardig wat Engels voordat het op school gegeven wordt
Sc_vr16_3	De meeste leerlingen zijn gemotiveerd voor Engels'
Sc_vr16_4	De meeste leerlingen volgen de lessen Engels met plezier
Sc_vr16_5	De niveaoverschillen tussen leerlingen zijn voor Engels groot
Sc_vr16_6	Engels moet worden aangeboden vanaf groep 1

Sc_vr16_7	Engels moet worden aangeboden vanaf groep 5
Sc_vr16_8	Engels moet worden opgenomen in de eindtoets
Sc_vr16_9	De aandacht voor Engels gaat ten koste van het niveau voor taal en rekenen
Sc_vr16_10	Leerkrachten zijn gemotiveerd voor Engels
Sc_vr16_11	Om Engels te kunnen geven moet de taalvaardigheid van leerkrachten verbeterd worden
Sc_vr16_12	Om Engels te kunnen geven moet de didactische vaardigheid van leerkrachten verbeterd worden
Sc_vr16_1_rec	Sc_vr16_1_rec
Sc_vr16_2_rec	Sc_vr16_2_rec
Sc_vr16_3_rec	Sc_vr16_3_rec
Sc_vr16_4_rec	Sc_vr16_4_rec
Sc_vr16_6_rec	Sc_vr16_6_rec
Sc_vr16_7_rec	Sc_vr16_7_rec
Sc_vr16_8_rec	Sc_vr16_8_rec
Sc_vr16_10_rec	Sc_vr16_10_rec
Sc_vr16_11_rec	Sc_vr16_11_rec
Sc_vr16_12_rec	Sc_vr16_12_rec
Sc_vr17a_1	ik kan Engelse teksten lezen die hoofdzakelijk bestaan uit hoogfrequente, alledaagse of aan mijn werk gerelateerde taal. Ik kan de beschrijving van gebeurtenissen, gevoelens en wensen in persoonlijke brieven begrijpen.
Sc_vr17a_2	ik kan zeer korte, eenvoudige Engelse teksten lezen. specifieke, voorspelbare informatie vinden en ik kan korte, eenvoudige persoonlijke brieven begrijpen.
Sc_vr17a_3	ik kan Engelstalige artikelen en verslagen lezen die betrekking hebben op eigentijdse problemen, waarbij de schrijvers een bepaalde houding of standpunt innemen. Ik kan eigentijds Engelstalig literair proza begrijpen.
Sc_vr17a_4	Ik kan lange en complexe Engelse teksten begrijpen, en het gebruik van verschillende stijlen in de Engelstalige literatuur waarderen. Ik kan gespecialiseerde artikelen en lange technische instructies begrijpen.
Sc_vr17a_5	Ik kan Engelse zinnen en de meest frequente woorden begrijpen mbt gebieden die van direct persoonlijk belang zijn. Ik kan de belangrijkste punten uit korte, duidelijke eenvoudige berichten en aankondigingen halen.
Sc_vr17a_6	Ik kan hoofdpunten begrijpen als in duidelijk Engels over vertrouwde zaken op werk, school, vrije tijd enz; veel radio- of tv-programma's over actuele onderwerpen van persoonlijk of beroepsmatig belang begrijpen.
Sc_vr17a_7	Ik kan een langer betoog in het Engels begrijpen, niet duidelijk gestructureerd en verbanden niet expliciet worden aangegeven. Ik kan Engelstalige tv-programma's en films (zonder ondertiteling) begrijpen.
Sc_vr17a_8	Ik kan een langer betoog en lezingen in het Engels begrijpen en zelfs complexe redeneringen volgen, wanneer het onderwerp redelijk vertrouwd is. Ik kan de meeste nieuws-



	en actualiteitenprogramma's op de tv begrijpen.
Sc_vr17b_1	Ik kan mezelf vloeiend en spontaan uitdrukken zonder naar uitdrukkingen te hoeven zoeken. Ik kan het Engels flexibel en effectief gebruiken voor sociale en professionele doeleinden.
Sc_vr17b_2	Ik kan duidelijke, gedetailleerde beschrijvingen presenteren over een breed scala van onderwerpen die betrekking hebben op mijn interessegebied.
Sc_vr17b_3	Ik kan de meeste situaties aan die zich kunnen voordoen op reis in een Engelstalig gebied
Sc_vr17b_4	Ik kan een reeks Engelse uitdrukkingen en zinnen gebruiken om in eenvoudige bewoordingen mijn familie en andere mensen, leefomstandigheden, mijn opleiding en mijn huidige of meest recente baan te beschrijven.
Sc_vr17b_5	Ik kan een duidelijke, gedetailleerde Engelse tekst schrijven over een breed scala van onderwerpen die betrekking hebben op mijn interesses
Sc_vr17b_6	Ik kan me in duidelijke, goed gestructureerde Engelse tekst uitdrukken en daarbij redelijk uitgebreid standpunten uiteenzetten.
Sc_vr17b_7	Ik kan korte, eenvoudige notities en boodschappen in het Engels opschrijven.
Sc_vr17b_8	Ik kan eenvoudige, samenhangende Engelse tekst schrijven over onderwerpen die vertrouwd of van persoonlijk belang zijn
Sc_vr17a_1rec	Sc_vr17a_1rec, lezen niveau B1
Sc_vr17a_2rec	Sc_vr17a_2rec, lezen niveau A2
Sc_vr17a_3rec	Sc_vr17a_3rec, lezen niveau B2
Sc_vr17a_4rec	Sc_vr17a_4rec, lezen niveau C1
Sc_vr17a_5rec	Sc_vr17a_5rec, luisteren niveau A2
Sc_vr17a_6rec	Sc_vr17a_6rec, luisteren niveau B1
Sc_vr17a_7rec	Sc_vr17a_7rec, luisteren niveau C1
Sc_vr17a_8rec	Sc_vr17a_8rec, luisteren niveau B2
Sc_vr17b_1rec	Sc_vr17b_1rec, gesprekken niveau C1
Sc_vr17b_2rec	Sc_vr17b_2rec, spreken niveau B2
Sc_vr17b_3rec	Sc_vr17b_3rec, Gesprekken niveau B1
Sc_vr17b_4rec	Sc_vr17b_4rec, Spreken niveau A2
Sc_vr17b_5rec	Sc_vr17b_5rec, schrijven niveau B2
Sc_vr17b_6rec	Sc_vr17b_6rec, schrijven niveau C1
Sc_vr17b_7rec	Sc_vr17b_7rec, schrijven niveau A2
Sc_vr17b_8rec	Sc_vr17b_8rec, schrijven niveau B1
Sc_vr18_1	Mijn leerlingen kunnen Engelse teksten lezen die hoofdzakelijk bestaan uit hoogfrequente, alledaagse of aan school en hobby's gerelateerde taal.
Sc_vr18_2	Mijn leerlingen kunnen vertrouwde namen, woorden en zeer eenvoudige Engelse zinnen begrijpen, bijvoorbeeld in mededelingen, op posters en in catalogi.
Sc_vr18_3	Mijn leerlingen kunnen zeer korte, eenvoudige Engelse

	teksten lezen en specifieke, voorspelbare informatie vinden in Engelstalige advertenties, folders, menu's en dienstregelingen
Sc_vr18_4	Mijn leerlingen kunnen Engelse zinnen en de meest frequente woorden begrijpen die betrekking hebben op gebieden die van direct persoonlijk belang zijn
Sc_vr18_5	Mijn leerlingen kunnen vertrouwde Engelse woorden en basiszinnen begrijpen die henzelf, hun familie en hun directe concrete omgeving betreffen, wanneer er langzaam en duidelijk gesproken wordt.
Sc_vr18_6	Mijn leerlingen kunnen de hoofdpunten begrijpen wanneer in duidelijk Engels wordt gesproken over vertrouwde zaken op school, hobby's, vrije tijd, enz.
Sc_vr18_7	Mijn leerlingen kunnen uitingen op een simpele manier aan elkaar verbinden, zodat ze ervaringen en gebeurtenissen, hun dromen, verwachtingen en ambities kunnen beschrijven.
Sc_vr18_8	Mijn leerlingen kunnen een reeks Engelse uitdrukkingen en zinnen gebruiken om in eenvoudige bewoordingen familie en andere mensen, leefomstandigheden en bijvoorbeeld schoolactiviteiten te beschrijven.
Sc_vr18_9	Mijn leerlingen zijn in staat deel te nemen aan een eenvoudig gesprek, wanneer de gesprekspartner bereid is tot hulp, herhaling of herformulering
Sc_vr18_10	Mijn leerlingen kunnen een korte, eenvoudige ansichtkaart schrijven, bijvoorbeeld voor het zenden van vakantiegroeten
Sc_vr18_11	Mijn leerlingen kunnen eenvoudige, samenhangende Engelse teksten schrijven over onderwerpen die vertrouwd of van persoonlijk belang zijn
Sc_vr18_12	Mijn leerlingen kunnen korte, eenvoudige notities en boodschappen in het Engels opschrijven
Vraag18_txt	algemene opmerkingen vragenlijst
lkr_eruit	geeft de leerkracht Engels aan groep 8?
scale_buitenschoolsgebruik	schaal buitenschools (3_1-3_7, gemiddelde)
scale_buitenschoolsgebruik_zonder3_2	schaal buitenschools (3_1-3_7, MINUS 3_2, gemiddelde)
schaal_zelfinschatting_spreken	zelfinschatting leerkrachten spreken (17b_1rec - 17b_4rec)
schaal_zelfinschatting_schrijven	zelfinschatting leerkrachten schrijven (17b_5rec - 17b_8rec)
schaal_zelfinschatting_lezen	zelfinschatting leerkrachten lezen (17a_1rec - 17a_4rec)
schaal_zelfinschatting_luisteren	zelfinschatting leerkrachten luisteren (17a_5rec - 17a_8rec)
ERKinschatting	gemiddelde ERK inschatting
casus_eruit	leerkrachten die vragenlijst niet hebben ingevuld

## Bijlage 18: Overzicht van variabelen m.b.t. gesprekken op paarniveau

GESPREKSID	GESPREKS ID
Record	Record
Nkeer	N keer
volgnr_ll	Volgnummer leerling
LIID	Leerling ID
In	In uit
Beoordelaar	Code beoordelaar
Ntakengedaan	Aantal taken gedaan
NtakenOK	Aantal taken paar beoordeeld
StandaardGedaanOKampregulier	Standaard: Gedaan = OK & regulier
ReguliergedaanampOK	Regulier
Codegedaan	Code gedaan
CodeOK	Code OK
Welke	Welke taken
Taak1	Taak 1
Taak2	Taak 2
Taak3	Taak 3
GT1_Nlln	aantal ll n beoordeeld tk 1
GT1_N_oor	aantal oordelen tk1
T1Sc1	T1 Sc 1: Tk1: mate van interactie
T1Sc2	T1 Sc 2: Tk1: mate van engels - mild
T1Sc3	T1 Sc 3: Tk1: mate van engels - strict
GT2_Nlln	aantal ll n beoordeeld tk 2
GT2_N_oor	aantal oordelen tk 2
T2Sc1	T2 Sc 1: Tk2: mate van interactie
T2Sc2	T2 Sc 2: Tk2: mate van engels - mild
T2Sc3	T2 Sc 3: Tk2: mate van engels - strict
GT3_Nlln	aantal ll n beoordeeld tk 3
GT3_N_oor	aantal oordelen tk 3
T3Sc1	T3 Sc 1: Tk3: mate van interactie
T3Sc2	T3 Sc 2: Tk3: mate van engels - mild
T3Sc3	T3 Sc 3: Tk3: mate van engels - strict

