

Meerjarige monitoring in de pilot tpo, eindverslag

Vooraf

Het beoordelen van het niveau Engels op de pilotscholen tpo en het volgen van de voortgang heeft al geruime tijd de aandacht van de scholen en de stuurgroep tpo. Voor de lagere leerjaren – en de eerste jaren van onze pilot - zijn door de SLO in samenwerking met de stuurgroep en de scholen leerdoelen, woordenlijsten en observatieformulieren ontwikkeld. Het accent in die jaren ligt op de deeltaalvaardigheden luisteren en spreken. Met steun van het ministerie van OCW heeft de stuurgroep van de pilot tweetalig primair onderwijs onderzoek kunnen doen naar de mogelijkheden voor het meten van de voortgang Engels van de leerlingen vanaf groep 4.

Vanaf groep 4 nemen lezen en schrijven in het Engels een allengs belangrijker plaats in het programma in en neemt de behoefte aan een meer kwantitatief meetinstrument toe. Kinderen beschikken over een grotere woordenschat, kunnen zich beter uiten – in gesprekken en op papier – en voor leerkrachten is het belangrijk om kennis van de voortgang systematisch te verzamelen, te toetsen aan leerdoelen en om het programma te evalueren.

Voor dit onderzoek is ten eerste gekeken naar het aanbod van longitudinale meetinstrumenten in binnen- en buitenland, waarbij de bijzondere plaats van het tpo op het continuüm van Engels in het basisonderwijs van belang is. Het gaat immers om een programma dat zich o.a. naar percentage onderwijstijd bevindt tussen vvtoE en International Schools. De longitudinaliteit is van belang omdat we eerder een formatief dan een summatief instrument willen onderzoeken. Er zijn met diverse partijen gesprekken gevoerd om hun producten nader te verkennen. Het eerste hoofdstuk doet hier verslag van, aangevuld met de al opgedane ervaringen op internationale primaries. We hebben gesprekken gevoerd met deskundigen van IGBO-scholen en er is een analyse van de meeste gangbare monitoringssystemen op die scholen gedaan.

Gaandeweg dit onderzoek kregen de gewenste specificaties van een in te zetten instrument gestalte. Het moet de taalvaardigheid in de volle breedte kunnen meten, het moet bij voorkeur digitaal zijn en het moet makkelijk inzetbaar zijn voor leerlingen en scholen. Leerkrachten moeten al de reguliere vorderingen bijhouden in hun LOVS. Er zou geen tweede systeem naast moeten komen dat veel tijd vergt. Maar dat wel belangrijke informatie oplevert over de ontwikkeling van het Engels dat voor 30 tot 50 % van de onderwijstijd de doel- en voertaal is.

Er is besloten om dit onderzoek met TELL uit te voeren. Test of English Language Learning is een door Pearson USA ontwikkeld instrument voor tweedetaalleerders in K-12, voor leerlingen van 4-18 jaar. In de VS, hun eerste doelgroep, gaat het om 10 % van de totale schoolpopulatie en verwacht wordt dat dit percentage verder zal stijgen. Op het eerste gezicht leek TELL zo geplaatst te kunnen worden tussen vvtoE en Engels op de international primaries. Het leek ook te voldoen aan de gebruikerseisen die we geformuleerd hadden. Na verdere verkenningen is met veel hulp van experts in de VS het onderzoek uitgevoerd door Kim Elzinga, nu MA 'Meertaligheid en Taalverwerving' van de Universiteit Utrecht, onder verantwoording van Karel Philipsen, lid van de stuurgroep tpo. Het verslag is van beider hand.

Het onderzoek is uitgevoerd op scholen met tpo en als controlegroep scholen met vvtoE. We zijn de scholen dankbaar voor de ruimte om het onderzoek te doen, en voor de gesprekken die we in de tweede testfase met directies, leerkrachten en kinderen gevoerd hebben over het programma. In april is met een *dry run* gestart, gevolgd door twee testfases in juni en oktober. Hoofdstuk twee beschrijft de onderzoeksopzet en de uitkomsten.

Gedurende die periode is het netwerk van tpo regelmatig op de hoogte gebracht van de resultaten. Na de eerste testfase is een tussenrapportage geschreven en gedeeld met stuurgroep en scholenveld. Het tussenrapport is ook naar de vvtoE-controlescholen gezonden. Dat zal ook het geval zijn voor het eindrapport voor de scholen. Op de landelijke studiedag tpo op 25 november heeft Karel Philipsen presentaties verzorgd.

Dit verslag wordt afgerond met een conclusie waarin de wens wordt uitgesproken om aan dit onderzoek een vervolg te geven. We hebben door TELL en de vele gesprekken veel geleerd over de manier waarop we de voortgang Engels kunnen meten en wat de impact daarvan op het programma in de hogere leerjaren kan zijn. De pilot tpo wordt in het buitenland met belangstelling gevolgd. In april is er een conferentie belegd met onderzoekers en lerarenopleiders uit Nederland en een aantal EU-landen. De daar uitgesproken wens tot verdere en meer intensieve samenwerking zou gestalte kunnen krijgen in een comparatieve studie naar de opbrengsten van tweetalige programma's.

Hoofdstuk 1 Vergelijkend onderzoek

In de aanloop naar dit onderzoek is er door middel van gesprekken en bezoeken van vele websites gepoogd inzicht te krijgen in het aanbod van (inter)nationaal opererende organisaties met toetsen en monitoringsystemen voor deze doelgroep. Zo zijn er gesprekken geweest met o.a. ETS-Global, Cambridge, Oxford en de British Council. Deze gesprekken, deels al gevoerd in 2015, en bestudering van hun mogelijkheden hebben geholpen bij het formuleren van de specificaties waar naar onze mening een voor het tpo geschikt programma aan zou moeten voldoen.

De meeste aanbieders hebben nog geen dekkend aanbod voor meerjarig volgen van de vorderingen voor tpo's. Voor programma's als vvtoE zijn er behoorlijk wat toetsen – bijvoorbeeld van Anglia, op Nederlandse scholen veel ingezet – maar die schieten tekort waar het gaat om te verwachten opbrengsten en taalvaardigheid van de leerlingen. Het wil wat uitmaken of je wekelijks een of twee of acht tot tien uur les krijgt in het Engels. Soms worden deze vooral summatieve toetsen toch ingezet op scholen voor tpo, maar snel blijkt dan dat leerlingen opgaan voor een niveau dat eerder in klassen voor het v.o. verwacht kan worden. Qua taalvaardigheid bereiken de leerlingen wel dat niveau, maar de onderwerpen waar ze vragen over moeten beantwoorden sluiten logischerwijs aan bij de belevingswereld van oudere kinderen. Deze discrepantie zal naar verwachting verder toenemen. We zoeken naar een systeem dat '*age appropriate*' recht doet aan het feit dat de leerlingen een tweede taal leren, dat ze dat significant sneller en beter zullen doen dan op scholen met vvtoE. Maar ook dat ze geen toetsen voor Engels als eerste taal moeten maken.

Veel aanbieders werken met een papieren versie van toetsen, een vrij omslachtige manier van werken. De tweede eis is dat er o.i. er een digitaal systeem moet zijn waarmee leerkrachten met betrekkelijk weinig moeite inzage krijgen in het niveau van Engels. Die leerkrachten zijn immers al heel wat uren bezig met het LOVS.

Ten derde is het belangrijk dat er een samenhangend systeem komt dat flexibel is en gedurende meerdere jaren gebruikt kan worden. Er zou sprake moeten zijn van formatieve toetsing. Veel aanbieders leggen het accent op summatief toetsen. Die kun je in zekere zin een formatief karakter geven door leerlingen bijvoorbeeld in groep 6 en groep 8 toetsen van dezelfde aanbieder op verschillend niveau aan te bieden, maar daarmee krijgt een school onvoldoende informatie om het programma tussentijds te evalueren en bij te stellen.

De flexibiliteit is van belang, juist in een pilot. Nog onbekend is wat de uiteindelijke leeropbrengsten van het Engels zijn. En hoe de groot de verschillen naar uitstroomniveau zijn: wat kan je van een leerling verwachten die naar het VMBO gaat, op welke deelvaardigheden kun je verschil zien met leerlingen met bijvoorbeeld een VWO-advies? Het is derhalve belangrijk om met zo'n instrument te zorgen voor een grote bandbreedte en voor inzet door jaarlagen heen. Er zijn immers ook verschillen tussen de pilotscholen: naar mogelijk uitstroomniveau, maar ook naar kwantiteit en kwaliteit van het programma. Formatief toetsen volgt de leerlingen op individueel niveau en kijkt naar de individuele, relatieve vooruitgang.

Tabel 1: Engels in het basisonderwijs in Nederland: kenmerken van toetsen

	Eibo/vvto	Tpo	DIPS
Percentage Engels onderwijstijd	0 tot 15%	30 tot 50%	100%
Voorbeelden van toetsen	Anglia Examens Cambridge Examens IEP Eindtoets Cito Me2	Observatieformulieren van SLO (tot en met groep 3) Pilot met TELL	PTE young learners Single Word Spelling test Reguliere toetsen Engels curriculum
Kenmerk van toetsen	Engels als tweede taal Voornamelijk summatief	Engels als tweede taal Voornamelijk formatief	Engels als eerste taal Summatief en formatief

Een vierde punt is dat het systeem de taalvaardigheid in een zo breed mogelijke zin zou moeten meten. Spreekvaardigheid maakt niet altijd deel uit van het aanbod, en als dat wel het geval is, is dat vaak als aparte activiteit (en ook tegen extra kosten).

Tot slot moet gekeken worden naar de relatie met de opbouw van de deelvaardigheden als op dit moment in de pilot uitgewerkt. Lezen en schrijven als gerichte activiteit komen niet direct aan de orde, zeker niet in groep 3 en lager. Monitoring van leerlingen met toetsen waarin deze deelvaardigheden een min of meer prominente plaats hebben is dan niet aan de orde.

Als gezegd zijn deze specificaties in de loop van deze voorstudie duidelijker geworden. Met deze kennis zijn we te rade gegaan bij de international primaries ; veel van onze tpo-scholen kennen zo'n afdeling. Deze werken samen in het netwerk DIPS: Dutch International Primary Schools. Er hebben gesprekken plaats gevonden, de schoolgidsen van scholen zijn bestudeerd (met daarin opvallend veel aandacht voor de toetsing van Engels).

Alle scholen blijken gebruik te maken van testen voor het reguliere basisonderwijs in het VK. Voorbeelden hiervan zijn de PTE Young Learners van Pearson en de Single Word Spelling Test van GL Assessments. De DIPS hebben voornamelijk leerlingen voor wie Engels ook de tweede (of derde) taal is, maar door de opbouw van het programma en natuurlijk door de volledige onderdompeling in de doeltaal blijkt het goed mogelijk om deze toetsen te hanteren. Dat gebeurt gedurende de gehele schoolloopbaan. 'Tough at first' voor de leerlingen, werd in een gesprek opgemerkt, maar snel doenbaar.

Deze logischerwijs gestandaardiseerde toetsbatterij wordt aangevuld met andere toetsen, vaak zelf ontwikkeld, en observaties door de leerkracht. Met het volgen van het curriculum voor eerstetaalleerders komen lezen en schrijven eerder dan bij tpo aan de orde. De DIPS werken veel samen en de indruk uit die gesprekken en bronnenanalyse is niet alleen dat er een eenduidige aanpak is, maar ook dat die goed onderbouwd is door de expertise op die scholen. Daar wordt vaak al enige decennia in deze constellatie gewerkt.

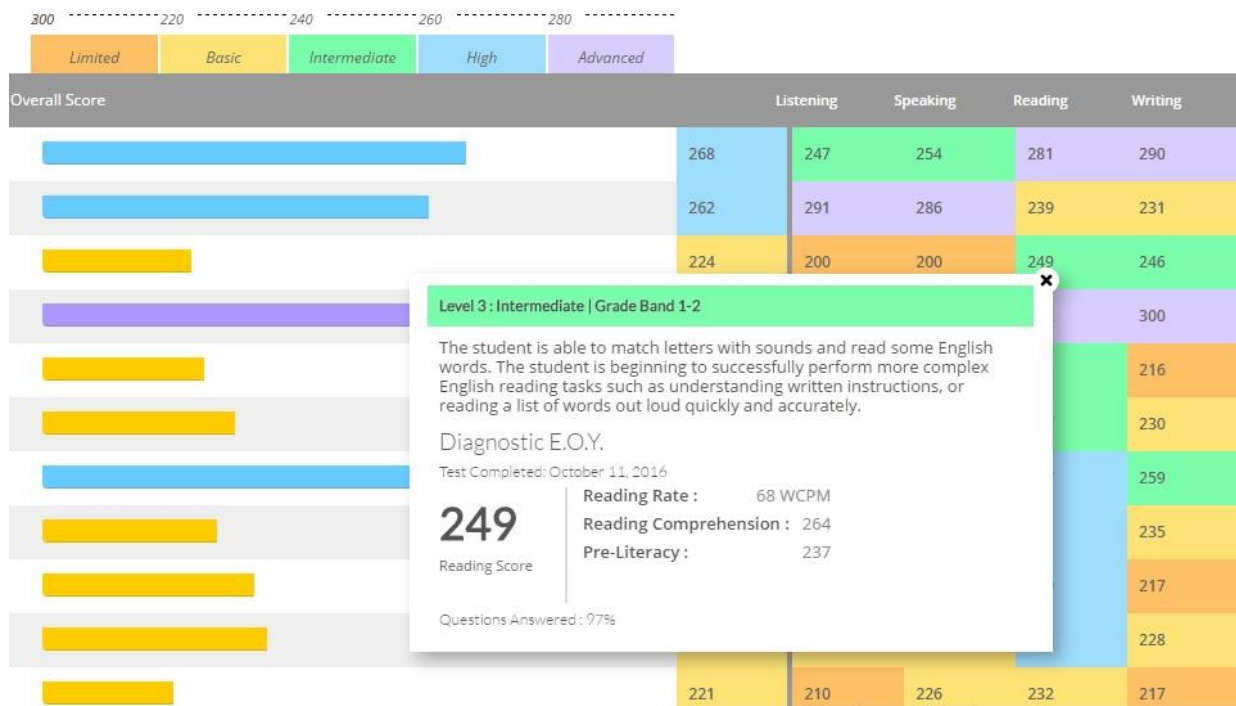
Hoofdstuk 2 Het TELL-onderzoek

TELL is een monitoringssysteem waarbij drie verschillende testen afgenomen kunnen worden: een screeningstest, acht voortgangstesten en twee diagnostische testen. Een toets duurt maximaal 50 minuten en kan zelfstandig op een iPad¹ met headset worden gemaakt. Daarbij komen in de hogere leerjaren telkens meer vaardigheden aan de orde. In de diagnostische test die de leerlingen in dit onderzoek hebben gemaakt, krijgen zij een score op meer dan tien deelvaardigheden. Die staan vermeld in de overzichtstabellen verder in dit verslag. Een belangrijk kenmerk is dat deze deelvaardigheden geïntegreerd worden gemeten. Een gesproken antwoord op een vraag in de test beïnvloedt dan de score op meerdere deelvaardigheden tegelijk.

De resultaten zijn na het maken van de test direct af te lezen in het administratieplatform. Figuur 1 geeft daar een voorbeeld van. De algemene score van 268 van de leerling 1 is het gemiddelde van de vier subscores. De leerlingen maken een test binnen een bepaalde leeftijdscategorie (een *grade band*) en hierbinnen kunnen zij een score behalen met een range van 100. In de leeftijdscategorie van groep 4/5 is dit bijvoorbeeld een score tussen de 200 en de 300. 250 is hierbij de gemiddelde score. Binnen zo'n range is er een verdere onderverdeling in vijf niveaus, lopend van limited tot advanced. Bij elk niveau, zoals in figuur 1 voor *intermediate*, komt een nadere toelichting op een subscore. Andere informatie voor de deelvaardigheid 'reading' betreft de leessnelheid in aantal woorden per minuut (WCPM), het leesbegrip en de 'pre-literacy'. Als een leerling binnen een grade band meerdere toetsen maakt, dan kan de ontwikkeling daarin goed gevolgd worden.

¹ Er is nu ook een Android-versie beschikbaar, zodat TELL op elk touchscreen tablet gedaan kan worden.

Figuur 1: TELL, administratieplatform en scores



In een trial in april is bekeken en vastgesteld dat TELL qua niveau, technische specificaties en toegankelijkheid van een toch voor de Amerikaanse schoolmarkt ontwikkeld product geen problemen voor de leerlingen op zou leveren. In overleg met de deskundigen van Pearson USA is gekozen voor twee grade bands, namelijk grade band 1/2 en 3-5. In tabel 2 wordt weergegeven welke grade band betreft leeftijd past bij de groepen van het Nederlandse basisonderwijs.

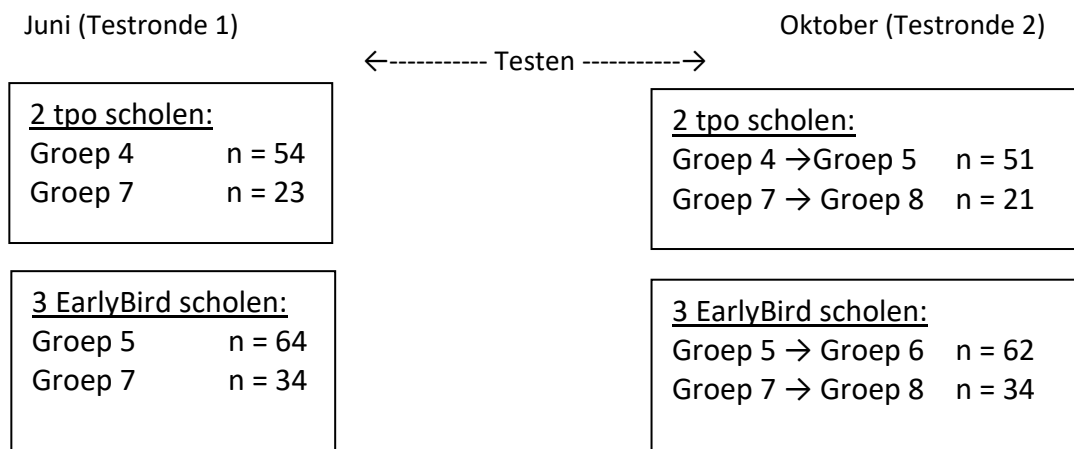
Tabel 2: Grade bands en een eerste indicatie voor het niveau in Nederland

Grade Band	Score range	Vergelijkbare groep in Nederland
K	100-200	1-2
1-2	200-300	3-4
3-5	300-400	5-7
6-8	400-500	8 + brugklas v.o.
9-12	500-600	Klas 3-6 v.o.

De centrale onderzoeksvraag luidde:

‘In welke mate is TELL toepasbaar voor monitoring van de taalontwikkeling van het Engels?’ En wat leert ons dit over de mogelijkheden van meerjarig monitoren in het Engels?

TELL en pilot tweetalig primair onderwijs (tpo)



Figuur 3: tijdpad TELL en pilot tpo

TELL is afgenomen op vijf scholen, waarvan twee tpo- en drie vvto-scholen. De steekproef bestond uit +/-170 kinderen van groep 4 en 7 tpo, en groep 5 en 7 vvto. Naar aanleiding van de trial in april/mei is een keus gemaakt voor groep 4 tpo, omdat er dan wordt begonnen met lees- en schrijfvaardigheden die niet zo makkelijk geobserveerd kunnen worden met de bestaande observatieformulieren. Binnen de vvto-groep is gekozen voor groep 5, omdat er vanaf dat leerjaar wordt begonnen met de geletterde vaardigheden in het Engels.

De trial in april/mei wees uit dat deze groepen en groep 7 geen problemen ondervonden met de toetsen en de omgeving, en dat er vooralsnog geen plafondeffect te zien was. Groep 4/5 en groep 7 hebben in het daadwerkelijke onderzoek een diagnostische test gemaakt. De test is afgenomen in juni, en een tweede keer in oktober. De reden voor het twee keer testen is het nagaan van de consistentie en het gaf ons de gelegenheid te bekijken hoe leerlingen omgingen met de tablet en het programma. Hoeveel instructie zou er nog nodig zijn?

In juni is de Begin-of-Year diagnostische test afgenomen, in oktober de End-of-Year. Pearson gaf aan te verwachten dat er geen grote verschillen in scores zouden zijn. De testen hebben een zelfde moeilijkheidsgraad, kenden soms dezelfde items². Waar het met dit systeem om gaat is om te kijken hoe een leerling aan het begin van een periode scoort, vervolgens de voortgang te monitoren om daarna aan het einde te kijken hoe groot die is. Daarna kan worden bekeken of de leerling (of de klas) met een hogere grade band aan de slag kan gaan.

Uit de resultaten blijkt dat binnen de bandbreedte forse verschillen voorkomen, naar leerling en naar deelvaardigheid. In beide grade bands is de uiteindelijke score rond het gemiddelde van resp. 250 en 350. Een aantal kanttekeningen moet hierbij gemaakt worden.

² Dat was de leerlingen opgevallen. Met normale afname van deze twee toetsen met ruimere tussenpoos is dat geen probleem.

De twee onderzoeksgroepen zijn om meerdere redenen niet representatief. Er is simpelweg gekeken welke tpo-scholen al een tweetalig programma in ontwikkeling in groep 4 hebben. Die zijn benaderd; daar werd direct positief op gereageerd.

De deelnemende scholen hebben integraal tpo ingevoerd, maar daarbij ligt het accent vooral op de onderbouw. De hogere groepen hebben een programma dat zich het best laat omschrijven als intensief vvtoE op weg naar tpo. De leerlingen in groep 7 (en 8) hebben daarnaast niet gedurende hun gehele schoolloopbaan zo'n programma gevolgd.

Ook de deelnemende EarlyBird-scholen kennen verschillen in onderwijstijd, ervaring met vvtoE enz. Het zijn wel scholen die aan de eisen van de landelijke standaard vvtoE voldoen. Ten derde is geen verder onderzoek gedaan naar achtergrondvariabelen. Het doel van het onderzoek was om te onderzoeken of er een monitoringinstrument is en of leerlingen op NL basisscholen met tpo dan wel vvtoE met het gekozen programma uit de voeten kunnen.

Daar is inderdaad sprake van. TELL kan goed gedaan worden door deze leerlingen. Als verwacht zijn de verschillen groot. De standaarddeviatie binnen de groepen is groot, dit laat zien dat er veel spreiding is en veel verschillen tussen de leerlingen onderling. Er is binnen de pilot tpo ruimte voor verscheidenheid, naar bijvoorbeeld leerlingenkenmerken en het percentage onderwijstijd in het Engels. TELL lijkt de verschillen tussen leerlingen en scholen duidelijk te kunnen maken.

Een opvallende constatering is dat TELL ook haalbaar lijkt te zijn voor de vvto-groep. De leerlingen doen op bepaald vlakken niet onder voor de tpo-leerlingen. In de lagere groepen van tpo lijkt in een enkel geval sprake te zijn van een plafondeffect. In de hogere groepen wordt het verschil tussen tpo en vvto groter en worden ook de verschillen tussen leerlingen, zoals verwacht, groter. Opvallend zijn bijvoorbeeld de scores op lezen en schrijven in groep 4 en 5. Die zijn voor leerlingen van vvtoE-scholen gelijk of beter. Dat kan deels verklaard worden door de bewuste keuze van de pilot om niet eerder dan in groep 4 te starten met gerichte aandacht voor deze deelvaardigheden. Een tweede verklaring kan zijn dat oudere kinderen gewoon beter lezen, ook in de tweede onderwijstaal. Voor groep 7 en 8 zijn de scores voor de tpo-leerlingen hoger.

Tabel 3 en 4 laten de resultaten van de toetsen in oktober zien. De in juni behaalde scores zijn ter vergelijking als appendix 1 en 2 toegevoegd. De scores wijken conform de verwachtingen van Pearson niet significant af. Tussen de eerste en de tweede ronde zat betrekkelijk weinig tijd met bovendien nog een vakantie.

Tabel 3: Resultaten frequenties (sub)scores groep 5 tpo and groep 6 vvtoE

Score		N	Gemiddelde	Mediaan	Modus	Standaard deviatie	Min	Max
Overall	tpo	51	248,35	250,00	228	21,189	215	290
	vvto	59	253,32	256,00	263	16,454	209	286
Speaking	tpo	51	242,98	243,00	200	28,734	200	287
	vvto	59	245,61	250,00	200	23,279	200	285
Listening	tpo	51	244,02	241,00	241	25,376	200	292
	vvto	59	240,97	243,00	231	21,862	200	284
Reading	tpo	53	259,86	262,00	276	20,681	205	292
	vvto	63	263,64	266,00	268	15,485	214	288
Writing	tpo	51	246,27	246,00	221	26,567	212	300
	vvto	62	260,85	258,50	300	24,018	213	300
Pre-literacy	tpo	51	254,45	256,00	262	19,073	211	287
	vvto	59	262,78	263,00	278	15,470	223	288
Grammar	tpo	51	252,33	249,00	234	22,644	211	300
	vvto	59	258,95	259,00	248	19,474	213	294
Vocabulary	tpo	51	242,25	239,00	215	25,560	208	296
	vvto	59	249,80	254,00	259	19,629	208	285
Fluency	tpo	51	256,51	260,00	260	23,911	202	291
	vvto	59	260,83	270,00	284	24,968	200	291
Pronunciation	tpo	51	258,61	257,00	257	24,284	206	300
	vvto	59	264,63	272,00	251	25,048	201	300
Reading Comprehension	tpo	51	263,16	266,00	253	23,118	200	291
	vvto	59	263,65	266,00	278	17,250	205	290
Reading Rate	tpo	51	58,78	64,00	0	28,540	0	119
	vvto	59	72,19	70,00	64	22,645	0	139

Tabel 4: Resultaten frequenties (sub)scores groep 8 tpo and vvtoE

Score		N	Gemiddelde	Mediaan	Modus	Standaard deviatie	Min	Max
Overall	tpo	21	358,19	361,00	351	15,361	330	380
	vvto	30	347,90	346,50	331	19,000	313	380
Speaking	tpo	21	365,14	368,00	368	14,958	337	384
	vvto	30	345,93	345,00	337	22,433	301	381
Listening	tpo	21	360,48	364,00	329	16,170	326	378
	vvto	30	343,20	345,50	357	19,871	300	376
Reading	tpo	21	345,57	344,00	336	15,629	320	386
	vvto	30	344,53	344,50	334	17,340	308	383
Writing	tpo	21	360,90	361,00	349	28,256	315	400
	vvto	34	355,26	353,50	400	33,108	311	400
Grammar	tpo	21	359,24	362,00	364	11,144	336	385
	vvto	30	350,13	353,00	366	16,594	313	378
Vocabulary	tpo	21	355,14	353,00	321	19,782	321	387
	vvto	30	342,13	333,50	373	23,697	302	384
Fluency	tpo	21	362,90	364,00	352	18,099	329	388
	vvto	30	344,57	338,00	338	20,761	313	385
Pronunciation	tpo	21	357,90	355,00	348	18,428	332	400
	vvto	30	347,10	350,00	332	20,461	303	387
Reading-Rate	tpo	21	108,33	113,00	105	24,840	53	148
	vvto	30	114,77	115,00	104	13,841	88	139
Expressive-ness	tpo	21	339,19	344,00	344	26,778	302	385
	vvto	30	341,93	344,00	344	21,534	300	385

Na de tweede testronde zijn leerlingen bevraagd naar hun mening over de testen. Ook werd er met de leerkrachten en de coördinator tpo/directie over de prestaties van de leerlingen gesproken, dit werd gedaan aan de hand van hun inschatting van de taalvaardigheid in het Engels van de leerlingen.

De leerlingen hadden na de eerste testronde al een vragenlijst ingevuld. Die waren van meer algemene aard. In de gesprekken toonden zij zich behoorlijk enthousiast. Omgaan met de tablet

leverde geen enkel probleem op. De voor de Amerikaanse schoolkinderen gemaakte introducties en vragen deden dat evenmin. De hogere groepen merkten wel op dat voor sommige items de schrijfpdrachten lastig waren. Ze moesten soms ook veel van de gegeven informatie onthouden. De toegewezen hoeveelheid tijd was voor hen voldoende. Als toezichthouder voor die test kunnen we dat beamen.

De leerkrachten van de scholen met tpo-groepen was vooraf gevraagd een inschatting te maken van het niveau Engels van hun leerlingen. Tabel 5 is daar een voorbeeld van. Met een score van 1 (ver onder het gemiddelde van de klas) tot 5 (ver daarboven) kregen zij en de onderzoekers een beeld dat we af konden zetten tegen de resultaten op de TELL-testen.

Dat werd gedaan in het afrondende gesprek met de leerkrachten en de directie/coördinator tpo. In dat gesprek konden we hen naast de opgaven en de scores ook het administratieplatform, de 'back-end', laten zien. De eerste conclusie, zonder een verdere statistische analyse, was dat er een behoorlijke mate van overeenkomst was tussen het vooraf opgegeven niveau en de scores. De meeste afwijkingen waren die van een prettige verrassing als er hogere scores waren dan zij verwachtten. Men was evenzo verrast door de veelzijdigheid van het programma: op welke deelvaardigheden informatie over het niveau van hun leerlingen beschikbaar was.

Het derde deel van die gesprekken ging vrij snel in op de uitkomsten op groepsniveau. Op een van de scholen leidde dit deel tot een eerste evaluatie van het totale programma.

Tabel 5: Voorbeeld van een inschattingsformulier

School	Leerkracht	Groep	Leerling	Spreken	Luisteren	Schrijven	Lezen
a	1	4 *		2	2	1	2
a	1	4 *		2	3	2	2
a	1	4 *		2	2	2	2
a	1	4 *		4	4	4	4
a	1	4 *		5	5	4	4
a	1	4 *		3	3	3	3
a	1	4 *		3	3	3	2
a	1	4 *		2	2	2	2
a	2	5 *		4	4	3	3
a	2	6 *		3	3	2	2
a	2	6 *		3	3	2	2
a	2	6 *		4	4	3	3
a	2	6 *		5	5	4	4
a	2	7 *		5	5	4	4
a	2	7 *		5	5	3	4
a	2	7 *		5	5	3	3
a	2	7 *		3	3	1	2
a	2	7 *		3	4	1	2
a	2	7 *		5	5	5	5
a	2	7 *		4	4	3	3

Samenvatting en conclusie

Dit onderzoek heeft ons veel geleerd over langdurige monitoring van het Engels op tweetalige basisscholen. Er is veel tijd besteed aan een goede vergelijking met andere systemen. Alle uitgevers reageerden positief op onze vraag en namen ruim de tijd voor overleg en toelichting. De gesprekken

met de collega's op de international schools en de uitvoerige analyse van hun monitoringsysteem zorgde voor verdere en zeer nuttige informatie.

Die werd ook verkregen door de ondersteuning door Pearson. Op hun websites is veel informatie over TELL en andere digitale vormen van monitoring en toetsing, er zijn wetenschappelijk onderbouwde artikelen over hun systematiek en de validiteit; we hebben via de mail en in conference calls hulp gekregen.

Het afnemen van de toetsen en de interactie met de leerlingen vormden nuttige bronnen van informatie. Amerikaans Engels, gesproken door een tweedetaalleerder, en opdrachten met verwijzingen naar het leven op een school in de VS of thuis: het leverde maar in enkele gevallen een probleem op.

Dat is een belangrijke constatering. Het gaat om tweedetaalleerders die een toets maken die specifiek voor hen bedoeld is. De *pronunciation* bijvoorbeeld wordt gewaardeerd op basis van de intelligibiliteit. Hoe verstaanbaar is de leerling, kan die het antwoord zodanig verwoorden dat dit begrijpelijk is? Er lijkt zich in de wereld van toetsontwikkelaars een verschuiving voor te doen van E=L1 en testen daarop gebaseerd naar E=L2 (of zelfs L3). Engels is immers vooral de taal geworden waarmee niet-Engelstaligen met elkaar communiceren.

De resultaten laten zien dat gekozen lijkt te zijn voor de juiste *grade bands*. We zien verschillen tussen vvtoE en tpo, en binnen de twee groepen. De inschatting van de leerkrachten en de scores op de toetsen kwam behoorlijk overeen. Als gezegd moeten we de scores met de nodige terughoudendheid bekijken. Dat heeft te maken met de aard van de pilot, met de behoorlijke diversiteit in de pilot en vooral met de onderzoeksvragen: 'Zou TELL kunnen voldoen aan de eisen van een longitudinaal systeem voor tpo-scholen en wat leert het onderzoek ons in algemene zin over monitoring'?

Het was in die zin een beperkt, verkennend onderzoek, reden ook om niet als voor het FoTo-onderzoek naar achtergrondvariabelen te kijken. Maar we hebben zo veel kennis op kunnen doen, vast kunnen stellen dat voor tpo-leerlingen (en op een ander niveau ook voor vvtoE-leerlingen) zo'n systeem veel nuttige informatie op kan leveren.

Het gaat als gezegd om het in kaart brengen van de vorderingen op een breed terrein van taalvaardigheid Engels, passend bij de leeftijdsgroep bovendien. Dat nu lijkt mogelijk. Het feit dat het programma volledig digitaal is en voor onmiddellijke resultaten zorgt betekent o.i. dat leerkrachten met niet al te veel extra tijd relevante informatie kunnen vergaren. Die hoeft nog niet afgezet te worden tegen 'harde' ERK-eisen voor jonge leerders van Engels. Dat gaat in de loop van de pilot wel een aandachtspunt worden. We komen over de volle breedte van de pilot in hogere groepen, scholen en leerkrachten doen meer ervaring op met tpo. Tijd dan om de focus meer op de opbrengsten te hebben dan in de beginjaren van de pilot. Vorig jaar heeft de pilot de leerdoelen Engels voor groep 1 tot en met 8 uitgewerkt. En als met het observatieformulier in nauwe samenwerking met de SLO. Het lijkt een goed idee om te kijken of die leerdoelen ook behaald worden. De '*can do*' statements van ons overzicht kunnen naast die van de TELL of de ook door Pearson ontwikkelde *Global Scale of English* worden gelegd.

De auteurs van dit eindverslag hopen op een vervolg met TELL. Dat willen we plaatsen in het kader van bovengenoemde ontwikkelingen met naar verwachting hogere niveaus voor Engels op onze scholen. En kunnen we het programma dan inzetten voor het meer primaire doel: de monitoring. Kunnen scholen langer toegang hebben tot TELL, zodat naast de diagnostische toets ook de voortgangstoetsen hun minstens zo belangrijke plaats kunnen krijgen? En kunnen we ook niet gaan kijken of een hogere *grade band* mogelijk is?

Maar we doen dat ook met het oog op een mogelijke Europese vergelijking. Voorafgaand aan de pilot is een literatuurstudie naar vormen van tpo in Europa en daarbuiten gedaan. Er is in april 2016 een conferentie over tpo geweest met onderzoekers uit een aantal landen. In oktober hebben we een bezoek afgelegd aan tpo-scholen in Duitsland. Het streven van de stuurgroep is om tot een Europees netwerk van scholen, lerarenopleiders en onderzoekers te komen. Wij denken dat er veel winst te behalen valt met onderzoek, uitwisseling en wederzijdse schoolbezoeken.

Tot slot willen we de scholen nogmaals hartelijk danken voor hun medewerking. Tpo mag zich verheugen in veel belangstelling, ook wetenschappelijk. Deelname aan het FoTo-onderzoek en het promotie-onderzoek van Jan Willem Chevalking en Claire Goriot op het tweede cohort van de pilotscholen vraagt al heel wat zorg. Zo ook de flankerende onderzoeken als die van de evaluatie van de SLO-documenten, verdere studies naar o.a. translanguaging, CLIL en interculturele competenties. Dat is aan de ene kant logisch, veel onderzoek, want zo'n nieuw programma roept veel vragen op. Maar het moet in die dynamiek allemaal wel te doen zijn en ook nog eens daadwerkelijk bijdragen aan de onderwijskundige ontwikkeling van de pilot zelf. Steun en begrip van scholen is dan onontbeerlijk. De onderzoekers en de stuurgroep waarderen dat zeer!

Utrecht/Rotterdam, 11 januari 2017

Kim Elzinga
Karel Philipsen

Bijlagen: Onderzoeksresultaten ronde 1, juni 2016

Tabel 6 en 7: resultaten onderzoeken juni

Resultaten groep 4 tpo en groep 5 vvto

		N	Mean	Median	Mode	St. deviation	Min	Max
Overall	Tpo	53	245,72	244,00	266	16,72	217	279
	Vvto	63	249,60	251,00	263	15,199	216	283
Speaking	Tpo	53	245,83	256,00	200	26,379	200	283
	Vvto	63	241,11	243,00	200	22,065	200	289
Listening	Tpo	53	244,28	243,00	237	23,210	200	289
	Vvto	63	237,25	239,00	200	21,161	200	284
Reading	Tpo	53	252,60	255,00	266	21,755	200	288
	Vvto	63	262,22	261,00	267	15,652	223	292
Writing	Tpo	54	239,30	235,50	218	18,802	214	287
	Vvto	64	256,94	257,00	236	19,991	215	300
Pre-literacy	Tpo	53	252,60	254,00	257	17,593	200	284
	Vvto	63	259,75	262,00	270	13,997	226	286
Grammar	Tpo	53	249,51	250,00	238	17,964	212	292
	Vvto	63	253,48	251,00	246	17,692	215	291
Vocabulary	Tpo	53	238,94	239,00	219 ^b	19,325	212	288
	Vvto	63	244,16	245,00	232	18,425	209	292
Fluency	Tpo	53	260,13	267,00	284	23,727	209	291
	Vvto	63	255,33	252,00	284	23,469	215	289
Pronunciation	Tpo	53	262,02	264,00	273	22,977	217	300
	Vvto	63	260,33	257,00	289	23,841	220	299
Reading Comprehension	Tpo	54	251,65	258,50	200	27,294	200	289
	Vvto	64	263,75	268,00	255	19,201	200	290
Reading Rate	Tpo	54	54,43	57,00	0	24,491	0	98
	Vvto	64	71,02	71,00	79	20,428	0	134

Resultaten groep 7 tpo en vvto

		N	Gemid delde	Mediaan	Modus	Standaard deviatie	Min	Max
Overall	tpo	23	357,96	357	374	18,639	327	386
	vvto	34	348,97	347,5	333	17,874	318	384
Speaking	tpo	23	357,39	359	370	17,180	317	385
	vvto	34	344,06	342	334	20,638	303	376
Listening	tpo	23	354,96	355	335	17,177	316	381
	vvto	34	341,71	338,5	328	18,825	300	374
Reading	tpo	23	362,74	361	361	23,365	319	400
	vvto	34	352,97	353,5	317	19,053	317	400
Writing	tpo	23	357,22	354	400	29,827	313	400
	vvto	34	356,62	351,5	400	25,383	314	400
Grammar	tpo	23	366,48	373,00	345	14,927	334	385
	vvto	34	359,03	358,50	349	16,176	319	384
Vocabu- lary	tpo	23	348,74	347,00	348	18,626	316	387
	vvto	34	338,74	338,00	326	20,797	305	380
Fluency	tpo	23	361,61	363,00	371	16,008	329	384
	vvto	34	348,38	346,00	326	20,899	307	387
Pronuncia- tion	tpo	23	356,09	355,00	325	19,533	325	392
	vvto	34	350,71	348,00	338	21,923	300	400
Reading- Rate	tpo	23	110,26	113,00	82	22,989	60	162
	vvto	34	110,56	107,00	99	19,629	60	159
Expressive -ness	tpo	23	345,13	344,00	344	27,024	302	385
	vvto	34	343,06	344,00	344	23,380	302	385