

Verslag OTTO-2 aan Europees Platform

13 mei 2013

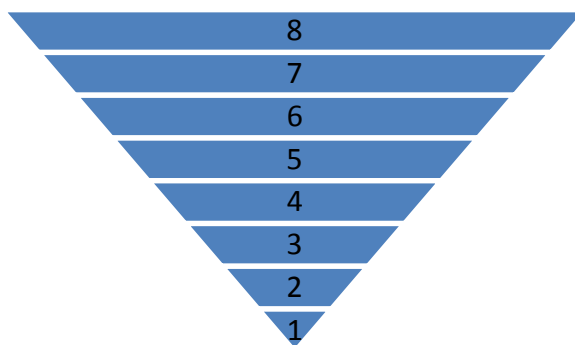
Dr. M.H.Verspoor, Dr. X. Xu, en Prof. Dr. C.J.L. de Bot
Toegepaste Taalwetenschap
Rijksuniversiteit Groningen

T-VWO en T-HAVO. Even effectief?

Achtergrond en doelstelling

In 2009 is het eerste onderzoek naar de effectiviteit van tweetalig onderwijs (OTTO-1) met succes afgerond. De resultaten laten zien dat T-VWO leerlingen in de onderbouw significant meer Engels leren dan hun leeftijdsgenoten in het reguliere onderwijs (Verspoor, de Bot en Xu 2011). Bovendien leren T-VWO leerlingen niet alleen beter Engels, maar ook anders: ze gebruiken eerder meer authentieke uitdrukkingen die hen natuurlijker doen klinken (Verspoor en Edelenbos 2009). Aan het eind van het derde jaar hebben de T-VWO leerlingen het B2 niveau op schrijfvaardigheid behaald. Ook bleek dat er een significante relatie bestaat tussen schoolse aanleg, buitenschools gebruik van het Engels en taalvaardigheid (Verspoor, de Bot en Xu 2012). Tenslotte is onderzocht in hoeverre TTO docenten bewust CLIL didactiek toepassen. Dat bleken ze te doen afgezien van het feit dat ze misschien nog meer gebruik zouden kunnen maken van activerende vormen die de talige output stimuleren (Schuitemaker-King 2012).

Het OTTO-1 project heeft leerlingen gevolgd van klas 1 t/m 3. Het is opmerkelijk dat aan het eind van de derde klas het verschil tussen TTO en reguliere leerlingen relatief klein lijkt. Dit is te verklaren door een breed gedragen aanname die nooit goed is onderzocht: leerlingen van lagere niveaus vorderen sneller naar een niveau hoger dan leerlingen van hogere niveaus omdat naarmate het niveau hoger is, de vaardigheden die men nodig heeft om verder te vorderen met factor 2 zouden groeien (Cummins 2000, Verspoor et al 2012). Met andere woorden, zoals het fictieve voorbeeld in Figuur *a* suggereert, is het verschil tussen bijvoorbeeld niveau 1 en 2 met nog hele simpele woorden en structuren veel kleiner bijvoorbeeld tussen 4 en 5 omdat relatief veel meer talige subsystemen complexer en geavanceerder moeten worden.



Figuur a. Fictief voorbeeld van groei in taalvaardigheid.

Veel TTO docenten zeggen dat het echte TTO effect pas plaatst vindt in de laatste twee jaar wanneer alle taalvaardigheden worden geconsolideerd. Het zou dus interessant zijn te weten in hoeverre TTO leerlingen nog vorderingen maken na de 3^{de} klas en welk niveau ze aan het eind van hun schoolcarrière hebben bereikt.

Bovendien bestaan er nu op een aantal scholen volledige T-HAVO trajecten, maar er is weinig over bekend. De meest prangende vraag is of het net zo veel effect heeft als

T-VWO. Daarnaast weten we weinig over de behaalde niveaus van het Engels, bijvoorbeeld of het gestelde B1 niveau aan het eind van het derde jaar wel realistisch is. En als er verschillen zijn in taalvaardigheid tussen T-HAVO en T-VWO leerlingen, dan is het de vraag of dat ligt aan aanleg, motivatie, “willingness to communicate” of het soort instructie dat hen ligt. Mochten er in dat opzicht verschillen zijn, dan kunnen de gegevens de basis zijn voor een specifieke TTO didactiek op het HAVO.

Daarom is besloten om met financiële ondersteuning van het Europees Platform een nieuw onderzoek op te zetten (OTTO-2) waarin wordt gekeken naar de behaalde niveaus van taalvaardigheid in verschillende schooltypen waar TTO wordt gegeven en mogelijke verschillen in een aantal variabelen die een rol kunnen spelen in de taalverwerving.

De uitkomsten kunnen interessant zijn voor verschillende partijen. Voor het netwerk TTO is het belangrijk te weten of T-HAVO net zo effectief is als T-VWO. Ook is het nodig inzicht te krijgen in de vorderingen in Engelse taalvaardigheid die verwacht kunnen worden van T-HAVO leerlingen om daarmee realistische eindtermen te kunnen stellen voor de standaard. Voor T-VWO scholen is het ook goed om te weten in hoeverre TTO onderwijs na de derde (in de vorm van het IB programma) echt zin heeft en welk ERK niveau de eindexamenklassen uiteindelijk behalen. Bovendien kan inzicht in verschillende factoren zoals motivatie, WTC, of voorkeur voor onderwijsmethodes wellicht leiden tot suggesties voor de algemene aanpak van het Engelse taalonderwijs en de inrichting van het T-HAVO onderwijs.

Samengevat is het doel van OTTO-2 project beter zicht te krijgen op het effect van TTO op HAVO en op de taalvaardigheid niveaus die leerlingen in de verschillende groepen bereiken en in hoeverre factoren zoals motivatie, voorkeuren voor lesmethodes en “willingness to communicate” naast schoolse aanleg een rol spelen in de taalverwerving van de leerlingen. Ten slotte hopen we suggesties te geven voor effectiever onderwijs voor de verschillende groepen, zowel voor TTO als regulier.

Theoretische aannames

Wij werken vanuit een “dynamic usage based approach” (DUB) van taalverwerving (cf. Verspoor et al 2012). Dit houdt in dat taal gezien wordt als een dynamisch, complex systeem met allerlei subsystemen die zich allemaal moeten ontwikkelen, maar dat dit niet tegelijk gebeurt. Bijvoorbeeld het lexicon moet zich eerst ontwikkelen voordat complexere zinnen kunnen ontstaan en receptieve vaardigheden ontwikkelen zich voordat ze productief kunnen worden gebruikt (Caspi 2010). Dit betekent dus dat de verschillende subsystemen zich op het zelfde moment op verschillende niveaus kunnen bevinden (bv. luisteren A2 en spreken A1). Het is dan ook niet verrassend om één subsysteem wel vooruit te zien gaan en een ander voorlopig even niet. Om een totaal beeld van taalvaardigheid te krijgen is het dan wel belangrijk meer dan één vaardigheid te toetsen. Idealiter zou men een heel dossier van een leerling moeten samenstellen met alle vaardigheden zowel receptief als productief, maar voor een onderzoek als dit is dat niet praktisch haalbaar en hebben wij gekozen voor twee deelvaardigheden: de receptieve woordenschat en productieve schrijfvaardigheid. Voor sommige analyses hebben we de scores apart gehouden en voor ander gecombineerd om tot een algemene taalvaardigheid score te komen.

Het onderzoek

Omdat het huidige onderzoek (OTTO-2) in het verlengde ligt van OTTO-1 zijn een aantal instrumenten (vocabulaire toets en schrijfpodracht) gebruikt om als ijkpunten te dienen. Een vergelijking van de scores kan laten zien in hoeverre de leerlingen in OTTO-2 representatief zijn voor T-VWO scholen in het algemeen. Omdat dit onderzoek vooral ook focust op verschillen tussen T-VWO en T-HAVO is in het beginsel gekozen voor één school om zoveel mogelijk andere variabelen in school- en leefomgeving uit te sluiten. Wel zullen de uitkomsten van het onderzoek worden vergeleken met die van het OTTO-1 project dat op andere scholen is uitgevoerd. Bij vergelijkbare uitkomsten is het aannemelijker dat de bevindingen op meer scholen van toepassing zijn.

In april en mei 2011 zijn de meeste gegevens van de 761 leerlingen aan een landelijke school in Brabant verzameld. Omdat enkele groepen in de eerste twee jaren van regulier VWO en HAVO ontbraken in 2011, zijn op dezelfde tijd in het volgende jaar (2012) de gegevens van de ontbrekende groepen verzameld. Statische vergelijkingen zijn op verschillende niveaus uitgevoerd, op leerjaar (1-6), op groep (TTO versus regulier en HAVO versus VWO) en op afdeling (onderbouw en bovenbouw).

Leerjaren	Onderbouw			Bovenbouw			Total
	1	2	3	4	5	6	
T-VWO	55	58	73	44	35	26	291
TTO HAVO	78	35	33	37	16		199
REG VWO	22	25	25	22	14	13	121
REG HAVO	26	23	26	30	45		150
Totaal	181	141	157	133	110	39	761

Tabel a. Aantal leerlingen in groepen, leerjaren en afdelingen .

De onderzoeksvragen zijn als volgt:

1. Hoe vertalen de pure taalvaardigheid scores van OTTO 1 en 2 zich naar de meer communicatieve ERK-niveaus. Deze vraag is belangrijk omdat de meeste onderwijsinstellingen ERK niveaus hanteren.
2. In hoeverre zijn de uitkomsten van leerlingen in OTTO-2 te vergelijken met de OTTO-1 leerlingen op de twee deelvaardigheden (vocabulaire en schrijven) en de combinatie van deze scores (algemene taalvaardigheid). Deze vraag is belangrijk om te bepalen of de huidige school representatief is voor andere TTO scholen in Nederland.
3. (a) Hoe verschillen de leerjaren en groepen in algemene taalvaardigheid en (b) welk taalvaardigheid niveau Engels bereiken de verschillende groepen in de verschillende leerjaren en aan het eind van hun middelbare school carrière? Deze vragen zijn belangrijk om te bepalen (a) of TTO effect heeft op zowel VWO als

HAVO niveau en (b) of de huidige verwachte ERK niveaus aan het eind van de derde klas en aan het eind van de school carrière geëigend zijn.

4. In hoeverre verschillen de groepen in factoren zoals (a) motivatie, (b) willingness to communicate, (c) voorkeur voor onderwijsmethodes en (d) buitenschools taalcontact. Deze vraag is belangrijk om onderzoeksvragen 6 en 7 te beantwoorden.
5. In hoeverre hangen de factoren (a) motivatie, (b) willingness to communicate, (c) voorkeur voor onderwijsmethodes en (d) buitenschools taalcontact samen met de behaalde taalvaardigheid. Deze vraag is ook relevant voor onderzoeksvragen 6 en 7.
6. In hoeverre zou T-HAVO anders ingericht kunnen worden dan T-VWO om het beste aan de behoeften van de T-HAVO leerlingen tegemoet te komen?
7. In hoeverre zouden reguliere VWO en HAVO anders ingericht kunnen worden om tot hogere resultaten te komen?

Resultaten

Hieronder worden de onderzoeksvragen herhaald en de bevindingen kort geschetst. Uitgebreidere gegevens staan vermeld in het “Method and Results” stuk (in het Engels) aan het eind van het Nederlandse verslag. De vragenlijsten en statistische gegevens worden vermeld in de appendix.

1. Hoe vertalen de taalvaardigheid scores van OTTO 1 en 2 zich naar de ERK-niveaus.

In het OTTO-1 onderzoek hebben we onze schrijfvaardigheid scores die zeer betrouwbaar bleken (cf. Verspoor et al 2012) met ERK niveaus gekalibreerd aan de hand van de ERK descriptoren. Om zekerder te zijn van deze kalibraties hebben we in het huidige onderzoek een onafhankelijk gremium (het Talencentrum aan de Rijksuniversiteit Groningen) dat gespecialiseerd is in het vaststellen van ERK niveaus 60 teksten in het OTTO-2 onderzoek laten beoordelen. Ook deze beoordelingen bleken zeer betrouwbaar.

De correlatie tussen de taalvaardigheid scores van 1-9 en de ERK niveau (A1-C2) was zeer hoog (0.87) en significant, maar zoals uit tabel 2 te zien is, is een één-op-één vertaling niet mogelijk, vooral niet op de lagere niveaus. De reden is dat de twee maten niet dezelfde constructen meten. ERK niveaus meten communicatieve competenties waar taalvaardigheid maar één aspect van is (wat kan men communiceren, bijvoorbeeld kan men een eenvoudige brief aan een persoonlijke kennis schrijven) en onze scores meten alleen de taalvaardigheid (hoe geavanceerd is de taal met betrekking tot talige accuraatheid, complexiteit en authenticiteit). Het is bijvoorbeeld mogelijk om in zeer eenvoudige, gebrekkige bewoordingen goed te communiceren naar een persoonlijke kennis over een eenvoudig onderwerp, en dan is het ERK niveau relatief hoog (A2) maar het taalvaardigheid niveau relatief laag (2). Voor de ERK niveaus op C1 en C2 lijkt dat niet het geval: voor complexe,

communicatieve taken met abstractere onderwerpen ligt accurate, complexe en authentieke taal voor de hand.

Voor scholen is het goed ERK niveaus te hanteren omdat communicatieve vaardigheden zeer functioneel zijn en in competenties uit te drukken zijn, maar omdat het blijkt dat onze taalvaardigheid scores vooral op de lagere niveaus (A1 t/m B2) niveaus fijnmaziger zijn dan de ERK niveaus, zijn deze weer nuttiger voor ons onderzoek om taalvaardigheid te vergelijken. De schrijfvaardigheid scores zijn als volgt (zie Tabel b) te vertalen naar de ERK niveaus:

ERK niveaus	A1	A2	B1	B2	C1	C2
OTTO schrijfvaardigheid scores	1-3	2-3	3-5	4-7	8	9

Tabel b. Kalibratie ERK niveaus met OTTO schrijfvaardigheid scores

2. In hoeverre zijn de uitkomsten van leerlingen in OTTO-2 te vergelijken met de OTTO-1 leerlingen op de twee deelvaardigheden (vocabulaire en schrijven) en de combinatie van deze scores (algemene taalvaardigheid).

Om de leerlingen uit het OTTO-1 en OTTO-2 onderzoek te vergelijken hebben we de data van de derde afname genomen aan het einde van het eerste en derde schooljaar uit het OTTO-1 onderzoek en die vergeleken met de leerlingen van het OTTO-2 onderzoek die ook aan het eind van het schooljaar dezelfde toetsen heeft gekregen. Zoals de Tabel c laat zien zijn er enkele significante verschillen (aangeduid met asterisk) te constateren tussen de twee onderzoeksgroepen: De OTTO-2 leerlingen scoren hoger op vocabulaire en de OTTO-1 leerlingen op schrijven in het jaar 1 en algemene taalvaardigheid in jaar 3. Algemene taalvaardigheid is hier geoperationaliseerd als de som van de twee deelvaardigheden vocabulaire en schrijfvaardigheid (zie onderzoeksvraag 3 voor meer detail).

	Jaar 1		Jaar 3	
	TTO	REG	TTO	REG
Schrijven	OTTO-1*	Geen verschil	OTTO-1*	Geen verschil
Vocabulaire	OTTO-2*	OTTO-2*	OTTO-2*	OTTO-2*
Algemene taalvaardigheid	Geen verschil	Geen verschil	OTTO-1*	OTTO-1*

Tabel c. Positieve significante verschillen (met sterretje aangegeven) tussen OTTO-1 en OTTO-2 populatie

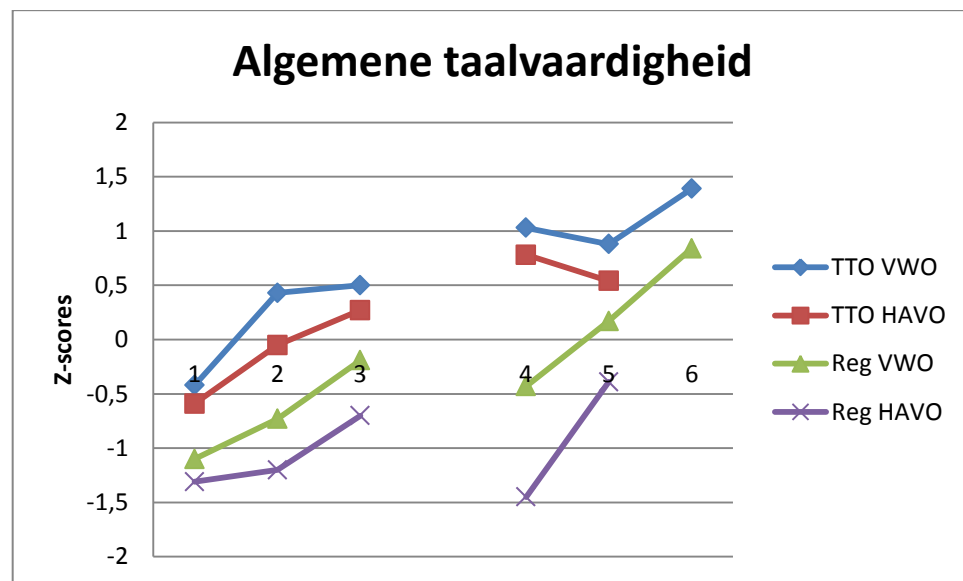
De OTTO-2 leerlingen zijn over het algemeen zwakker in schrijven en sterker in vocabulaire dan de OTTO-1 leerlingen. In jaar 3 is de T-VWO en regulier VWO zwakker in de algemene taalvaardigheid. Dit betekent dat de bevindingen in dit rapport niet zonder meer niet te generaliseren zijn naar andere TTO scholen.

3. Welk taalvaardigheid niveau Engels bereiken de verschillende groepen in ieder jaar en aan het eind van hun middelbare school carrière?

In dit gedeelte bekijken we eerst (a) hoe de leerlingen in de verschillende leerjaren en groepen van elkaar verschillen door de algemene taalvaardigheid scores te vergelijken. (b) Daarna kijken we naar de schrijfscores om een ERK niveau te kunnen inschatten voor de groepen aan het eind van de derde klas en school carrière.

Algemene taalvaardigheid vergeleken tussen leerjaren en groepen

De algemene taalvaardigheid werd berekend door de vocabulaire scores en schrijfvaardigheid scores bij elkaar op te tellen. Omdat deze verschillende soorten maten betreft zijn ze berekend tot z-scores en deze zijn bij elkaar opgeteld. De z-scores geven niet een bepaald niveau aan (zoals A2, B1, of B2) maar laten wel duidelijk de verschillen zien tussen de groepen. Omdat de onderbouw leerlingen allemaal dezelfde toetsen hebben gedaan, zijn deze leerjaren onderling te vergelijken. Zo ook voor de bovenbouw leerjaren. Omdat de toetsen voor onderbouw en bovenbouw niet dezelfde waren, kunnen we niet jaar 3 met jaar 4 vergelijken. Dit wordt aangegeven met de extra ruimte tussen de gegevens van de onderbouw en bovenbouw gegevens in figuur b.



Figuur b: Algemene taalvaardigheid scores vergeleken tussen leerjaren in de onderbouw en de bovenbouw groepen

Figuur b geeft de algemene taalvaardigheid scores weer van de groepen in de onderbouw en bovenbouw. De ANOVA analyses voor groepen 1-5 en de T-Toetsen (omdat het hier maar twee groepen betreft) voor groep 6 laten de volgende significante verschillen zien. Als het niet genoemd wordt dan zijn er geen statistische verschillen gevonden.

Leerjaar 1	TTO groepen > reguliere groepen.
Leerjaar 2	T-VWO groep > T-HAVO en reguliere groepen. T-HAVO groep > reguliere groepen.
Leerjaar 3	T-VWO > reguliere groepen. T-HAVO > regulier HAVO
Leerjaar 4	TTO groepen > reguliere groepen. Regulier VWO > regulier HAVO.
Leerjaar 5	T-VWO > T-HAVO en reguliere groepen. T-HAVO en regulier VWO > regulier HAVO
Leerjaar 6	T-VWO > regulier VWO.

Tabel d: Significante verschillen in algemene taalvaardigheid scores in leerjaren in de onderbouw en de bovenbouw

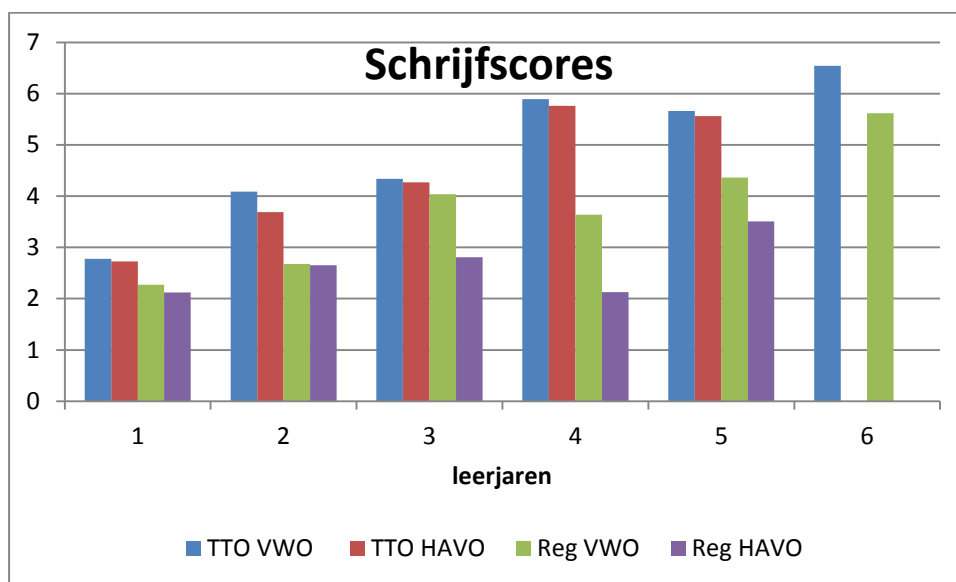
De analyses in Tabel *d* tonen aan dat T-VWO altijd hoger scoort dan de andere groepen. T-HAVO scoort altijd hoger dan regulier HAVO, maar in jaar 5 verdwijnt het verschil tussen T-HAVO en regulier VWO.

Wat opvalt is dat ook al scoren de beide TTO groepen nog steeds significant hoger dan hun reguliere jaargenoten, er voor beide TTO groepen in jaar 5 een terugval lijkt te zijn. De reguliere groepen gaan dan juist hard vooruit. Een verklaring kan zijn dat de TTO leerlingen dan wat minder aandacht hebben voor het Engels vanwege alle extra aandacht aan de Nederlandstalige vakken. Voor de reguliere groepen is er waarschijnlijk meer Engelse input omdat we aannemen dat ze meer lezen ter voorbereiding van het Engels eindexamen.

ERK niveaus aan het eind van de 3^{de} klas en aan het eind van de school carrière. Omdat de vocabulaire toetsen niet goed gekalibreerd kunnen worden met de ERK niveaus, worden alleen de schrijfofdrachten daarvoor gebruikt. Ook al is schrijven maar één aspect van taalgebruik, zijn schrijfproducten bij uitstek geschikt om taalvaardigheid te meten omdat ze alle aspecten van productief taalgebruik tegelijk manifesteren zoals het woordgebruik, de grammatica, de complexiteit van zinnen en de accuraatheid van al deze aspecten. De schrijfofdrachten werden spontaan en snel geschreven zonder voorbereiding en het is mogelijk dat dezelfde leerlingen met meer voorbereiding en tijd een hogere score zouden hebben gekregen.

De geschreven teksten werden net zoals in OTTO-1 holistisch beoordeeld op een groot aantal talige aspecten zoals woordkeuze, complexiteit van de zinsbouw, idiomatisch taalgebruik en eventuele foutjes. Ook al klinken de verschillen tussen cijfers misschien relatief klein, moet in acht worden genomen dat één punt verschil een groot verschil is in taalvaardigheid, vooral op hogere niveaus. (Zie Figuur a.)

Figuur *c* laat de gemiddelde scores zien van de verschillende leerjaren in verschillende groepen. Tabel *e* geeft kort weer waar statistische verschillen zijn per leerjaar.



Figuur *c*. Schrijfvaardigheid scores per groep en leerjaar

- | | |
|---|---|
| 1 | TTO groepen > regulier HAVO |
| 2 | TTO groepen > reguliere groepen |
| 3 | TTO groepen and regulier VWO > regulier HAVO |
| 4 | TTO groepen > reguliere groepen
Regulier VWO > regulier HAVO |
| 5 | TTO groepen > reguliere groepen |
| 6 | TTO groep > reguliere groep |

Tabel *e*: Significante verschillen tussen schrijfvaardigheid scores per leerjaar en groep.

De gegevens in Figuur *c* en de statistische analyses in Tabel *e* laten zien dat over het algemeen de TTO groepen hoger scores dan de reguliere groepen. Toch zijn er enkele verrassende uitkomsten:

Aan het eind van de derde klas halen de verschillende groepen de volgende niveaus:

3 T-VWO is gemiddeld 4.34 (ruim B1).

3 VWO is gemiddeld 4.04 (laag B1).

3 T-HAVO is gemiddeld 4.27 (ruim B1).

3 HAVO is gemiddeld 2.81 (ruim A2).

Tabel f. ERK niveaus aan het eind van de school 3^{de} klas.

In jaar 3 is er geen significant verschil tussen de TTO groepen en regulier VWO in schrijfvaardigheid (wel in algemene taalvaardigheid). Dit is verrassend omdat in het OTTO-1 onderzoek er juist wel een significant verschil bleek te zijn in schrijfvaardigheid tussen de groepen aan het eind van de derde klas. Zoals hierboven gemeld heeft deze 3^{de} T-VWO klas ook een significant lagere score dan de 3^{de} klassen in de OTTO-1 studie. (De reguliere 3^{de} VWO klassen daarentegen scoren hetzelfde als in de OTTO-1 studie op schrijfvaardigheid). De verklaring kan zijn dat deze 3^{de} TTO klas een vrij zwakke groep is of misschien de toets niet serieus heeft genomen.

De tweede vrij verrassende uitkomst is dat er voor geen enkel leerjaar significante verschillen zijn gevonden tussen de T-VWO en T-HAVO leerlingen in schrijfscores. Met andere woorden de T-VWO en T-HAVO leerlingen schrijven even goed. (Wel zijn er verschillen gevonden bij vocabulaire en algemene taalvaardigheid).

Een van de vragen was of de TTO leerlingen na de 3^{de} klas nog veel vordering maken in de bovenbouw. Zoals veel docenten hebben gesuggereerd, blijkt ook uit de analyses dat de TTO bovenbouw nog relatief veel vooruitgang boeken, ondanks het feit dat ze dan minder uren les in het Engels per week hebben. Dit kan verklaard worden door het feit dat de leerlingen in de bovenbouw extra lessen Engels volgen en nog veel in het Engels lezen (voor T-VWO in de vorm van het IB programma).

Tenslotte valt op dat de reguliere VWO leerlingen in hun laatste leerjaar een flinke sprong maken en even hoog scoren als de 5^{de} klas TTO leerlingen. Een verklaring zou kunnen zijn dat ze in het laatste leerjaar veel moeten lezen in voorbereiding op het eindexamen.

Aan het einde van hun studie behalen de leerlingen van deze school de volgende niveaus op schrijfvaardigheid (zei Tabel g):

6 T-VWO is gemiddeld 6.54 (hoog B2).

6 VWO is gemiddeld 5.62 (ruim B2).

5 T-HAVO is gemiddeld 5.56 (ruim B2).

5 HAVO is gemiddeld 3,51 (laag B1).

Tabel g. ERK niveaus aan het eind van de school carrière.

4. In hoeverre verschillen de groepen in factoren zoals (a) motivatie, (b) willingness to communicate, (c) voorkeur voor onderwijsmethodes en (d) buitenschools taalcontact.

In dit gedeelte bekijken we of er verschillen zijn in groepen met betrekking tot motivatie, WTC, voorkeur voor lessen en buitenschools taalcontact. Voor iedere factor werd een vragenlijst samengesteld (zie bijlage). Om te beginnen is een factoranalyse van de antwoorden gedaan om de samenhang tussen de verschillende vragen vast te stellen en verder statistische analyses te kunnen doen. Om deze statistische analyses mogelijk te maken, werden de leerlingen in grotere groepen verdeeld: onderbouw en bovenbouw. De onderbouw betrof leerjaren 1-3 en de bovenbouw leerjaren 4-5 voor Havo en 4-6 voor VWO. Alle verschillen die hieronder genoemd worden zijn statistisch significant.

(a) Motivatie

In de literatuur wordt vaak een verband gesuggereerd tussen motivatie en taalverwerving, maar causaliteit is niet bewezen. Het kan zijn dat leiders die de taal beter spreken meer gemotiveerd zijn en andersom.

Uit de factoranalyse kwamen vier factoren naar voren:

- i. instrumentele motivatie (bv. Engels is goed voor mijn toekomst)
- ii. integratieve motivatie (bv. Engels geeft mij toegang tot de Engelssprekende cultuur en mensen)
- iii. positieve attitude t.a.v. de Engelse taal
- iv. positieve attitude t.a.v. de Engelse klas.

In de onderbouw hebben de TTO groepen significant hogere scores dan de reguliere groepen wat betreft de instrumentele motivatie en de integratieve motivatie. De verklaring kan zijn dat de TTO groepen al zelf-geselecteerde groepen zijn. Omdat ze meer gemotiveerd waren hebben ze voor TTO gekozen. Het kan ook zijn dat ze door het TTO onderwijs nog gemotiveerder zijn geraakt. Met betrekking tot hun attitude naar de Engels taal en lessen zijn er geen verschillen tussen de groepen.

In de bovenbouw zijn er geen verschillen meer tussen de TTO groepen en de reguliere VWO groep, maar regulier HAVO scoort op alle sub-factoren significant lager.

(b) Willingness To Communicate

In de literatuur wordt gesuggereerd dat leiders die niet bang zijn zich uit te drukken in de vreemde taal, de taal meer zullen oefenen en daarom beter worden.

Uit de factoranalyse kwamen duidelijk twee factoren naar voren:

- i. spanning bij het spreken en
- ii. zelfvertrouwen.

Een misschien verrassende uitkomst is dat in de onderbouw, de T-VWO groep en in de bovenbouw, beide TTO groepen zowel meer gespannen zijn als meer zelfvertrouwen hebben bij het spreken van het Engels. Dit is een bekend verschijnsel in de literatuur (MacIntyre et al 2002) en kan verklaard worden door het feit dat de betere sprekers weten wat ze kunnen, maar zich tegelijk ook bewuster zijn van fouten die gemaakt kunnen worden. Bij de bovenbouw hebben zowel de TTO groepen als de reguliere VWO groep meer zelfvertrouwen dan de HAVO groep.

(c) Voorkeur les vormen

In interviews met TTO-docenten is naar voren gekomen dat de T-HAVO leerlingen minder geëngageerd leken dan de T-VWO leerlingen en misschien een ander soortige lesaanpak nodig was voor hen. Dit was aanleiding om de leerlingen te vragen naar hun voorkeuren in lespraktijken. (Zie appendix voor specifieke vragen). De uitkomsten laten inderdaad enkele verschillen zien tussen HAVO en VWO leerlingen.

Uit de factoranalyse kwamen duidelijk drie vormen naar voren.

- i. een traditionele door docent en lesboek gestuurde vorm,
- ii. een input en interactie vorm, en
- iii. een zelf gestuurde vorm.

In de onderbouw hebben de beide HAVO groepen meer dan de T-VWO een voorkeur voor een traditionele stijl van lesgeven. Beide TTO groepen hebben een sterkere voorkeur voor interactie en input dan de reguliere groepen. Dit kan verklaard worden door het feit dat de TTO groepen door de CLIL didactiek al gewend zijn aan input en interactie in de klas en daar dan ook de voorkeur aan geven.

Opvallend is dat in de bovenbouw de HAVO groepen (zowel TTO als regulier) duidelijk meer voorkeur geven aan input en interactie, terwijl de VWO (TTO en regulier) dan ook graag zelfgestuurd werk doen. Kennelijk zijn VWO leerlingen aan het eind van de school carrière () toe aan wat meer zelfgestuurde werkvormen. De reguliere HAVO groep is significant minder enthousiast over alle les vormen dan de andere groepen.

(d) Buitenschoolscontact

Uit onderzoek (o.a. Bonnet et al 2004) is gebleken dat extra input buiten school ook significant invloed heeft op taalverwerving. Uit eerder onderzoek (Verspoor et al 2011) is ook gebleken dat de bovenbouw (die al een hoger niveau van taalvaardigheid Engels heeft) relatief meer input buiten school zoekt en dat deze extra input een significante bijdrage leverde aan de taalvaardigheid score.

Uit de factoranalyse kwamen duidelijk twee factoren naar voren, waarvan de eerste niet helemaal eenduidig te benoemen is:

- i. Lezen, video games en spreken
- ii. Engelse muziek en televisie programma's

Alle groepen krijgen minder Type i input dan Type ii input. Er zijn geen verschillen tussen de groepen m.b.t. Type i input, maar de TTO leerlingen krijgen meer Type ii input dan de reguliere groepen. In dit onderzoek was geen verband te zien tussen input en taalvaardigheid, maar dit is waarschijnlijk te wijten aan het feit dat voor de huidige statistische analyses de groepen relatief klein waren.

5. Welke factoren dragen het meest bij aan taalvaardigheid?

Ten slotte werden er correlatie analyses gedaan om te ontdekken of er een relatie was tussen de behaalde taalvaardigheid scores en de andere gemeten factoren: motivatie, WTC, buitenschools contact, en voorkeuren voor les vormen.

Er werden verschillende positieve relaties gevonden (Zie Tabel OTTO-2: “Correlations between learning factors and test scores” en “Hierarchical multiple regression of predictors on general proficiency” in de appendix, maar de sterkste correlaties waren tussen zelfvertrouwen (één aspect van WTC) en alle talige vaardigheden (vocabulaire, schrijven en algemene taalvaardigheid). Met andere woorden, leerlingen die zelfvertrouwen hadden, hadden een grotere kans hoger te scoren op taalvaardigheid. Dit is niet zo verrassend, want als leerlingen weten dat ze goed zijn, hebben ze waarschijnlijk ook meer zelfvertrouwen.

Verder bleken ook drie van de motivatie factoren (instrumentele motivatie, integratieve motivatie, en positieve attitude t.a.v. de Engelse taal) significant positief met de taalvaardigheden samen te hangen. Een positieve attitude t.a.v. de Engelse les is kennelijk minder belangrijk.

6. In hoeverre zou T-HAVO anders ingericht kunnen worden dan T-VWO om het beste aan de behoeften van de HAVO leerlingen tegemoet te komen?

Uit de analyses blijkt dat qua schrijfvaardigheid de T-HAVO leerlingen niet onder doen voor de T-VWO leerlingen. Wel scoren ze lager op algemene taalvaardigheid. Uit de analyses blijken er verder ook weinig verschillen tussen de T-VWO en T-HAVO groepen in motivatie, WTC, en buitenschools contact. Wel bleken er enkele verschillen in voorkeuren voor les vormen.

De onderbouw T-HAVO gaf de voorkeur aan wat meer traditionele les vormen (docent gestuurd) terwijl de bovenbouw juist graag meer input en interactie heeft. De T-VWO leerlingen daarentegen zijn echter vroeger enthousiast over de activerende werkvormen, maar in de latere leerjaren zouden ze graag minder interactie en meer zelfgestuurd onderwijs krijgen.

7. In hoeverre zouden reguliere VWO en HAVO anders ingericht kunnen worden om tot hogere resultaten te komen?

Aan deze school vorderden de reguliere VWO klassen vrij gestaag in taalvaardigheid en haalden aan het eind vrij hoge scores. De reguliere HAVO klassen daarentegen waren op bijna alle maten significant lager dan de andere groepen en het lijkt samen te hangen met een lage motivatie om de taal te leren en attitude naar de taal. De vraag is of daar niet wat aan te doen is.

Uit onderzoek is bekend dat een van de voornaamste drijfveren in taalontwikkeling input is: hoe vaker men iets hoort of leest, hoe eerder men het leert. Het succes van TTO komt daarom onder andere voort uit de grote hoeveelheid Engelse input die de leerlingen krijgen. Uit het onderzoek van Schuitemaker-King (2012) blijkt dat in de Engelse TTO klassen veel in het Engels gewerkt wordt en er ook aandacht aan taalvormen besteed wordt, maar de hoeveelheid aandacht daarvoor is relatief veel kleiner dan bij de reguliere klassen. In de reguliere Engels lessen wordt veel gewerkt met traditionele lesmethodes: een standaard tekstboek waarin expliciete grammatica een vrij grote plek heeft en waarin zowel het boek als de docent de grammatica in het Nederlands uitlegt. De docenten spreken significant meer Nederlands dan hun TTO collega's en geven meer expliciete aandacht aan vorm. We nemen aan dat dit ook het geval is op de huidige onderzoeksschool.

Om de hoeveelheid input te verhogen is een eerste advies dat docenten consequent Engels spreken in de klas. Om de motivatie te verhogen zou meer authentieke en interessante input (bv in de vorm van films) effectief kunnen zijn. Door interessante lesmaterialen in de doeltaal aan te bieden die de leerlingen niet alleen meer input verschaft maar ook meer engageert, is het wellicht mogelijk juist HAVO leerlingen te boeien.

Conclusies en aanbevelingen

De hoofdvraag was of T-VWO en T-HAVO even effectief waren en het antwoord is volmondig ja. De T-HAVO leerlingen scoren in enkele opzichten even goed als de T-VWO leerlingen en zeker hoger dan de reguliere HAVO groep.

De ERK niveaus aan het eind van de derde klas (B2 voor VWO en B1 voor HAVO TTO) lijken redelijk. De T-HAVO leerlingen in deze onderzoeksgroep scoorden in de B1 bandbreedte. De T-VWO groep ook, maar zoals eerder uitgelegd lijkt het hier om deze specifieke klas te gaan die of vrij zwak was of de toets niet serieus had gemaakt.

De ERK niveaus aan het eind van de school carrière scoren T-VWO, T-HAVO en regulier VWO binnen de B2 bandbreedte (maar wel met duidelijke verschillen in pure taalvaardigheid zoals door ons geoperationaliseerd). Omdat het T-VWO niveau van hoog B2 ons te laag lijkt ingeschat, zou een kort onderzoek met een ander soortige schrijfpodracht (met meer tijd en voorbereiding) voor deze groep nuttig kunnen zijn.

Een sub vraag was of T-VWO en T-HAVO misschien een iets andere didactische aanpak behoeven. Het enige significante verschil tussen T-HAVO en T-VWO groepen blijkt in voorkeuren voor les vormen. De onderbouw HAVO heeft naast een voorkeur voor input en interactie ook een grotere voorkeur voor docent gestuurde lessen dan VWO. De bovenbouw HAVO heeft daarentegen een grotere voorkeur voor input en interactie en VWO een grotere voorkeur voor zelfgestuurd onderwijs. In de TTO standaard is Engels als voertaal en CLIL didactiek met verschillende werkvormen een vereiste. Ons advies zou zijn om de CLIL didactiek in alle TTO stromen zeker te blijven toepassen, maar adviseren een kleine verschuiving van nadruk op de verschillende werkvormen.

- Voor de onderbouw T-HAVO zou aanvankelijk een deel van de werkvormen ook uit wat meer docentgestuurde lessen mogen bestaan en de activerende werkvormen zouden iets geleidelijker kunnen worden ingevoerd.
- Voor de bovenbouw T-HAVO juist activerende groepsworkvormen handhaven.
- Voor de onderbouw VWO leerlingen kunnen de activerende werkvormen zoals gebruikelijk worden ingevoerd
- Voor de bovenbouw T-VWO zou wat meer ruimte kunnen zijn voor ook zelfstandig werken door minder groepsopdrachten en wat meer individuele opdrachten.

Een andere subvraag was of de reguliere stromen iets kunnen leren van de TTO stromen. Het succes van TTO ligt o.a. aan de hoeveelheid input die de leerlingen krijgen. In de reguliere stroom wordt relatief veel Nederlands gesproken en relatief veel aandacht aan grammatica besteed. Behalve dat wij van mening zijn dat deze les vormen te weinig input geven, zijn ze waarschijnlijk ook niet erg motiverend. Vooral de onderbouw HAVO is weinig te spreken over de Engels les. Wel zien we dat de reguliere stromen in de bovenbouw sprongen maken, waarschijnlijk omdat dan de nadruk meer op input komt te liggen (lezen voor eindexamen). Ons advies aan de reguliere HAVO en VWO stromen zou daarom zijn om vooral ook in de begin jaren bijna alle tijd en aandacht te besteden aan input door TTO principes te gebruiken in de reguliere Engelse lessen: Doeltaal = voertaal en geef geen les over de taal, maar gebruik de taal om te communiceren over onderwerpen die de leerlingen werkelijk interesseren.

Method and Results OTTO-2

This is the method and results section of the OTTO-2 study conducted in 2011/2012. For the introduction, conclusion and implications, please see pages 2-15. The main research question was whether T-HAVO is as effective as T-VWO, but seven other related questions have been asked:

1. *How accurately can the OTTO writing scores be compared with the CEFR writing scores?*
This question is important as educators are interested in the results as compared with CEFR scores.
2. *How do OTTO-2 students compare with the OTTO-1 students as far as vocabulary, writing, and general proficiency are concerned?*
This is to see if the current students can be seen as similar to other TTO schools.
3. *What level of proficiency do the different groups attain at the end of each year and at the end of their high school career?*
This is important to answer the main research question. Is T-HAVO as effective as T-VWO? It also will indicate whether the current objectives as set by the Network TTO at the end of 3rd grade are feasible and appropriate.
4. *To what degree do the groups differ in factors such as (a) motivation, (b) willingness to communicate, (c) preferences for instructional modes and (d) out of school contact with the language?*
The answer to this question is relevant to research questions 6 and 7.
5. *How are factors in motivation, willingness to communicate, preferences in instructional methods and out of school contact related to language proficiency.*
The answer to this question is relevant to research questions 6 and 7.
6. *What improvements could be made to T-VWO and T-HAVO English lessons to make instruction more effective?*
7. *What improvements could be made to regular VWO and HAVO English lessons to make instruction more effective?*

Method

Participants

The 761 participating OTTO-2 students from one comprehensive high school in the middle of the country were tested in April and May 2011 or 2012. (Because some first and second grades were missing in 2011, the same test was given to those grades in the months of April and May 2012.) To make comparisons possible between groups and grades, the students were divided into a junior group and a senior division. The students in each division received similar tests.

Grades	Juniors			Seniors			Total
	1	2	3	4	5	6	
T-VWO	55	58	73	44	35	26	291
TTO							
HAVO	78	35	33	37	16		199
REG VWO	22	25	25	22	14	13	121
REG							
HAVO	26	23	26	30	45		150
Total	181	141	157	133	110	39	761

Table 1. Number of participants in the different groups and grades.

Design

The design is cross-sectional (5 or 6 grades) with four groups: T-VWO, T-HAVO, regular VWO and regular HAVO.

Materials and procedures

The students took a vocabulary test, a writing test and a questionnaire to assess the amount of out of school contact, the same as in the previous OTTO study (Verspoor & Edelenbos 2009). The current study also included questions about motivation, out of school contact, willingness to communicate and instructional preferences.

The tests and questionnaire (see Appendix 1) were offered by means of Survey Monkey software. During the testing months, the English teachers took their own students to the computer lab and asked them to do the tests. There were no time constraints, but all finished within the 45 minute class time.

English proficiency was assessed by means of a receptive vocabulary test and a productive informal writing task. To be able to compare the groups with each other students in grades 1-3 (from now on referred to as “juniors”) received the same test and the students in grades 4-6 (from now on referred to as “seniors”) received the same test.

Students were tested on their passive vocabulary knowledge by means of the EFL (English as a Foreign Language) test (Meara & Buxton 1987), which contained words

in different frequency ranges. Students were asked to indicate whether they did or did not know the words. The test also contained a number of pseudo-words. The juniors were all tested on the most frequent words: a 120-word test at the 1000–2000 word frequency level. The seniors were all tested on a 180 –word test at the 3000–4000 word frequency level. To compensate for ‘guessing’, the scores on the EFL-test are calculated by way of the Signal Detection Theory scoring method developed by Huibregtse, Admiraal & Meara (2002). This means that the scores for this vocabulary test can be anywhere between 0 and 1, with higher scores reflecting higher vocabulary knowledge.

The productive informal writing tasks and instructions were also different for juniors and seniors. The juniors received an informal writing task. The seniors could choose between a narrative and an argumentative task. The latter was meant to elicit text at the more advanced level (C1 or C2) as it involved a different register and more formalized text. However, it was also possible to write a narrative at a higher level and in fact the highest score in the study was given to a narrative.

Juniors:

A holiday (vakantie). Please write a short story about a holiday you have had. (It does not have to be true).

Seniors:

*(1) My vacation (doesn't have to be true) or
(2) Please write a short newspaper article (if you can in formal language) stating why you are for or against the following statement. Support your arguments with facts and examples. (200-500 words): Dutch high school students (HAVO and VWO) should have more hours of instruction per year.*

There was no time limit for either group and no word limit for the juniors. For the seniors there was a suggested word limit of 200-500 words. (The shortest text was 12 words and the longest 547 words).

A holistic score was given to each text by a team of 8 experienced EFL teachers. To ensure that the scores in OTTO-2 would be equivalent to those in OTTO-1, the team received a thorough training and assessment procedure with texts from the OTTO-1 study (see Verspoor, Schmid & Xu 2012) before they started evaluating the OTTO-2 texts.

As in the OTTO-1 study, the following training procedure was followed: the group of eight raters first evaluated six OTTO-1 essays individually in order to determine which ones were the strongest and which ones the weakest. Then the group of raters discussed each set of six until consensus was reached. In order to internalize the features to be assessed, the raters were asked to discuss the range of factors that play a role in determining proficiency: vocabulary range, sentence complexity, use of L1, use of tenses, use of authentic expressions, accuracy, etc. When agreement was reached on the rank order, the texts were tentatively classified into nine proficiency levels (0-8). Then the scores were compared to the original OTTO-1 scores. Differences were discussed and adjustments were made to meet the original OTTO-1 scores.

After the training on OTTO-1 texts, the raters continued to work together in a similar manner on the OTTO-2 texts, starting with about 100 essays until they settled on all the proficiency levels. Note that OTTO-1 included only texts from 1st to 3rd grade and OTTO-2 included also texts from 4th to 6th grade so new scores had to be used for the higher levels (from 5 to 9). Assessment criteria were then established for all the proficiency levels, which included the main characteristics of each level that had emerged to help the raters classify the remainder of the essays. After training, the group of raters were divided into two groups of four and assessed the samples per page (first individually) and then compared the scores. The score of the majority (3 out of 4) was taken and differences in opinion were resolved in discussion. If a group was unable to reach consensus, the other group of raters was consulted. The groups of scorers had very high inter-rater reliability scores ($\alpha = .93$ to $.96$). For more detail on the scoring procedure, the proficiency levels, and the text characteristics at each proficiency level, see Verspoor, Schmid & Xu (2012).

Because the vocabulary and writing scores were significantly related to each other ($r = .50$ for the juniors and $r = .38$ for seniors, $p < .01$), they could be combined to generate a “general proficiency” score. To arrive at the general proficiency score, both vocabulary and writing scores were first standardized, making the individual scores directly comparable within groups and then they were summed up and standardized again. Therefore, the mean of the general proficiency score is zero and the standard deviation is one.

To assess the amount of out of school contact learners had with English and in order to find out how often they used English, we used a questionnaire developed by Berns et al. (2006). The questionnaire also contained 15 questions on motivation, 18 questions on willingness to communicate, and 8 questions on instructional preferences. (See appendix.)

Analysis

Several statistical techniques were applied to answer the research questions. For all analyses conducted, the significance level has been set at .05. Either independent-samples T-Tests (two groups involved) or ANOVA (more than two groups) were used to examine both the students’ proficiency development through years and to compare the same grade students’ proficiency across programs.

With respect to the analysis on learning factors (motivation, willingness to communicate, instructional preferences and out-of-school contact), factor analyses were first conducted in order to determine whether or not any of the variables correlated highly enough to be considered representative of a joint, underlying factor. Then the juniors (grades 1-3) and seniors (grades 4-6) were compared respectively across programs on generated factors. If the factors were significantly related, MANOVA was used; otherwise ANOVA. All descriptives and statistical details are given in Appendix 2.

Results

OTTO-1 versus OTTO-2 proficiency scores

In the OTTO-2 study, students were tested in the months of April and May 2011. To ascertain how the students in the current study perform compared to the students in the previous study (OTTO-1), we compared the scores of the current first and third graders (obtained in April-May) with the scores of the previous first and third graders at the third testing point at the end of their academic year (obtained in June-July). Table 2 indicates which groups scored significantly higher than the other and where there were no significant differences.

	First year students		Third year students	
	TTO	REG	TTO	REG
Writing	OTTO-1*	Same	OTTO-1*	Same
Vocabulary	OTTO-2*	OTTO-2*	OTTO-2*	OTTO-2*
General proficiency	Same	Same	OTTO-1*	Same

Table 2. Comparison of skills and general proficiency between students in OTTO-1 (2007) and OTTO-2 (2011 and 2012). The asterisk indicates a significant positive difference.

All groups in the OTTO-2 scored higher in vocabulary than the OTTO-1 students. The OTTO-2 T-VWO groups scored lower in writing and in the 3rd grade the T-VWO group scored lower in general proficiency.

The consequence of these differences is that we cannot assume that findings at this school can be generalized to other schools in the Netherlands.

OTTO-2 writing scores

For the writing scores, we first compare the groups per grade and then discuss how each group developed across grades.

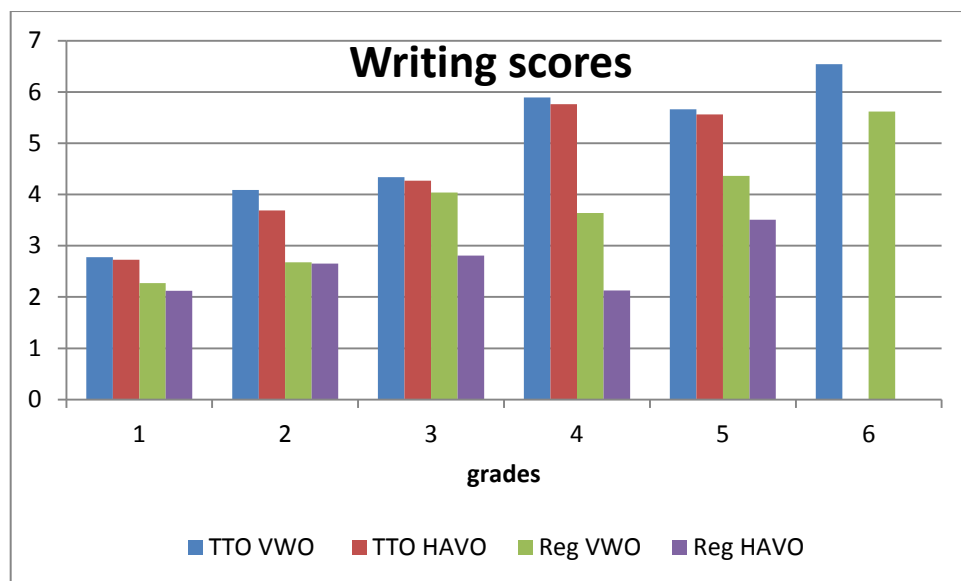


Figure 1: Writing scores across grades in each group

Figure 1 shows the average writing scores of the groups in each grade at the end of the year. The results of the ANOVA analyses for grades 1-5 and the T-Test for grade 6 show the following significant differences. If not mentioned, the differences between groups are not significant.

- Grade 1 TTO groups > regular HAVO.
- Grade 2 TTO groups > regular groups.
- Grade 3 TTO groups and regular VWO > regular HAVO group.
- Grade 4 TTO groups > regular groups.
Regular VWO > regular HAVO.
- Grade 5 TTO groups > regular groups.
- Grade 6 TTO group > regular group.

Table 3. Significant differences in writing scores across grades in each group

Across grades, the writing skills of TTO students seem to develop sooner (between grades 1 and 2) than those of regular students. Table 3 summarizes where significant gains are made in writing scores between consecutive grades.

	1-2	2-3	3-4	4-5	5-6
T-VWO	yes		yes		yes
TTO HAVO	yes		yes		
REG VWO		yes			yes
REG HAVO				yes	

Table 4. Significant gains made in writing scores between consecutive grades for each condition.

On the whole the TTO classes do as expected. They are significantly better in language proficiency as measured by a free writing test than their counterparts. However, one unexpected finding was that at the end of year 3 the T-VWO and regular VWO did not show a difference in writing scores. An explanation could be that the current 3rd grade TTO group was a rather weak group in general as it also scored significantly lower in all three measures (vocabulary, writing and general proficiency) than the OTTO-1 students. In OTTO-1 there was a significant difference between the TTO and regular groups.

The implications of these findings are as follows: The T-VWO students in grade 3 in this study did not reach the required B2 level for writing as indicated in the standard, so the school may need to look at its program to make sure it meets the needs. On writing the T-HAVO students at this school do as well as T-VWO students at this school, so the B1 level in the standard seems appropriate.

OTTO-2 writing scores versus CEFR scores

The writing scores used in the OTTO-1 study scores proved to be very reliable (cf. Verspoor et al 2012). In that study we calibrated our scores with the CEFR scores by comparing the descriptors of the two measures. To be more certain of this calibration we asked an independent testing organization (the Language Center at the University of Groningen) specialized in ascertaining CEFR levels to determine the CEFR level of 60 randomly chosen texts of the OTTO-2 study.

The team of assessors was trained during a two hour session and then scored each text individually. The inter-rater reliability was high (.98).

The correlation between the OTTO-2 writing scores (1-9) and the CEFR levels (A1-C2) was very high ($r = 0.87$, $p < 0.01$) and significant, but as figure 2 shows, a one-on-one match is not possible, especially not at the lower levels (A1-B1).

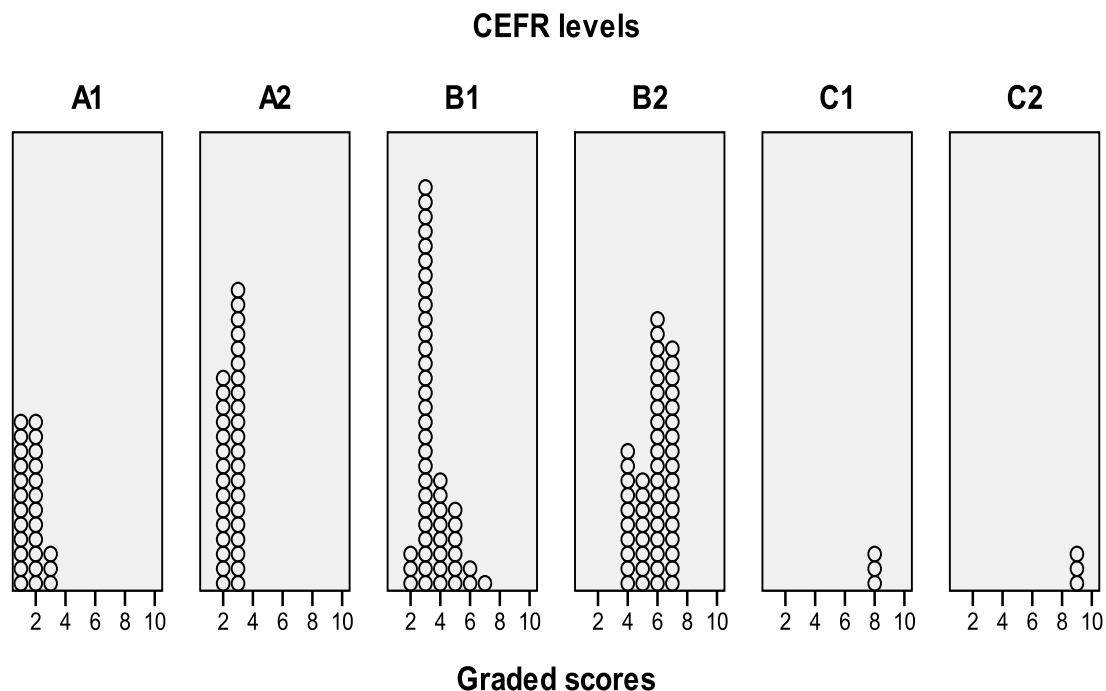


Figure 2: Calibration of the writing scores from the OTTO-2 study with the CEFR writing levels.

A rough translation between OTTO-2 writing scores and the CEFR levels can be seen in Table 5.

CEFR Level	A1	A2	B1	B2	C1	C2
OTTO writing score	1-3	2-3	3-5	4-7	8	9

Table 5: Explanation of the calibration of CEFR writing scores with the writing scores obtained from the OTTO-2 study. N.B. Outliers were not included.

The reason that there is no one-on-one translation possible between OTTO writing scores and CEFR levels is that they measure somewhat different constructs. The OTTO scores reflect pure language skills in terms of commonly applied notions of complexity, accuracy and fluency of the language. The CEFR measures communicative ability and the pure language skills are only a sub-component of that ability. For example, especially at the lower levels, there may be incongruities because it is possible to write a communicatively successful note to a friend in very simple language with some errors. However, at the higher CEFR levels such as C1 and C2 these incongruities are less likely to occur.

For the current study the OTTO-2 scores are useful because they make more fine-grained comparisons possible in pure language skills. For schools and general proficiency levels the CEFR levels are functional because they measure in terms of communicative competencies.

OTTO-2 vocabulary scores

For the vocabulary scores, we first compare the groups per grade and then discuss how each group developed across grades. The juniors (students in grades 1-3) and seniors (students in grades 4-6) took different versions of the vocabulary test. Therefore, the gap in the figure between grades 1-3 and 4-6 is to emphasize that the sets of tests are different and a direct comparison between grades 1-3 and 4-6 is not possible.

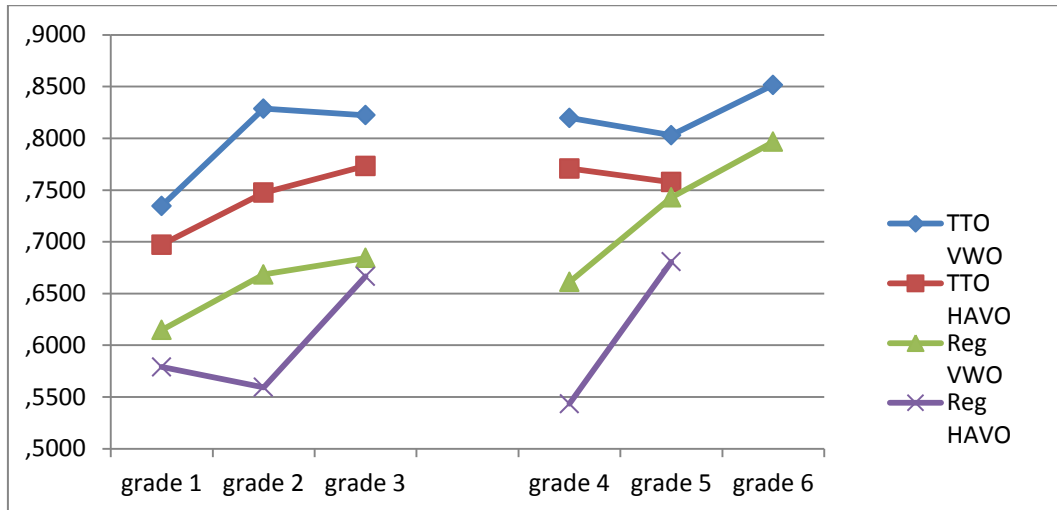


Figure 3: Vocabulary scores across grades in each group.

Figure 3 shows the average vocabulary scores of the groups in each grade. The results of the ANOVA analyses for grades 1-5 and the T-Test for grade 6 show the following significant differences. If not mentioned, the differences between groups are not significant.

Grade 1 TTO groups > regular groups.

Grade 2 T-VWO > T-HAVO and regular groups.
T-HAVO and regular VWO > regular HAVO.

Grade 3 TTO groups > regular groups

Grade 4 TTO groups > regular groups.
Regular VWO > regular HAVO.

Grade 5 T-VWO > regular HAVO

Grade 6 T-VWO > regular VWO

Table 6. Significant differences in vocabulary scores across grades in each group.

The groups seem to make progress in vocabulary at different rates. Table 6 summarizes where significant gains are made in vocabulary scores between consecutive and non-consecutive (gray column) grades.

	1-2	2-3	1-3	4-5	5-6
T-VWO	yes				
T-HAVO			yes		
REG VWO				Yes	
REG HAVO		yes		Yes	

Table 7. Significant gains made in vocabulary scores between (non-) consecutive grades for each condition.

The vocabulary scores give a good indication of the fact that we should not expect development to be linear. Previous research (cf Caspi 2010 and Verspoor et al 2012) has shown that receptive vocabulary usually has to improve before productive writing scores can improve.

Some groups grow steadily, others even decrease before they increase. What is interesting for TTO schools is that T-VWO keeps developing, even in grade 6 (probably because of the high demands of the IB program). T-HAVO has a small dip in year 5, probably because of the central exams and because there is less focus on English. What is interesting for the regular program is that the HAVO regular group seems to make a slower start, but HAVO actually catches up with the VWO regular group at the end of grade 3. Also in grade 5 regular HAVO students make a spurt, most probably because they are preparing for the central exam and have to read a lot.

OTTO-2 General proficiency scores

The general proficiency scores was operationalized as the standardized score of the sum of the two standardized scores of writing and vocabulary. Because the juniors and seniors took different tests, we cannot directly compare third grades with fourth grades.

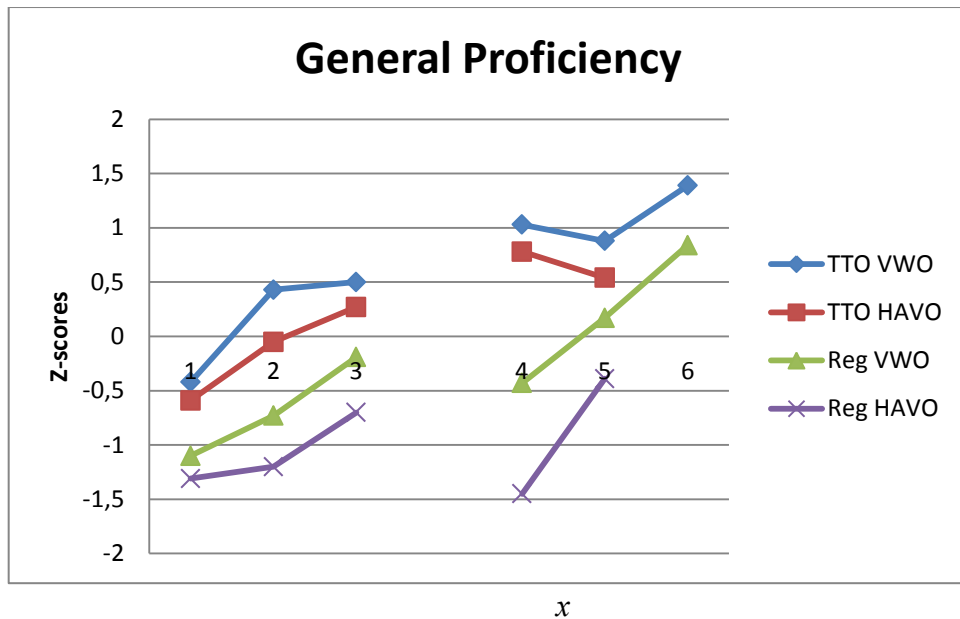


Figure 4: General proficiency scores across grades in each group.

Figure 4 shows the general proficiency scores of the groups in each grade. The results of the ANOVA analyses for grades 1-5 and the T-Test for grade 6 show the following significant differences. If not mentioned, the differences between groups are not significant.

- Grade 1 TTO groups > regular groups.
- Grade 2 T-VWO group > T-HAVO and regular groups.
T-HAVO group > regular groups.
- Grade 3 T-VWO > regular groups.
T-HAVO > regular HAVO
- Grade 4 TTO groups > regular groups.
Regular VWO > regular HAVO.
- Grade 5 T-VWO > T-HAVO and regular groups.
T-HAVO and regular VWO > regular HAVO
- Grade 6 T-VWO > regular VWO

Table 8: Significant differences in general proficiency scores across grades in each group.

As far as general proficiency is concerned, we see that TTO groups both develop steadily but have a slight dip in grade 5. The regular groups develop at a slower rate and have a strong dip in year 4, although the regular VWO group still significantly

outperforms the regular HAVO group. However, they improve quite a bit in year 5 again. As expected the TTO groups have higher scores in almost every grade.

In previous studies (Verspoor et al 2011; Verspoor et al submitted), we found an effect of scholastic aptitude (operationalized as CITO score) and because VWO and HAVO learners are grouped according to CITO scores, we expected differences between the developmental paths in the T-VWO and T-HAVO groups. This study shows that there are indeed some differences between the two groups in some years in vocabulary and general proficiency, but not all the time. One explanation could be that the differences in scholastic ability affect more the rate of acquisition rather than the level that can be attained.

A surprising finding is a dip in general proficiency for both the T-VWO and T-HAVO groups in year 5, even though they still outscored their regular peers. This can be another sign of non-linear development but it could also be explained by the fact that the TTO students after year 3 have to take all subject classes in Dutch (they only get extra English classes) and they may spend more time and energy on activating their Dutch again, which may dynamically affect their English. However in year 6, the T-VWO learners move up again.

Not surprisingly because of the limited input, the regular groups develop at a slower pace than the TTO groups, but in 3 regular VWO students suddenly make a spurt and keep developing thereafter. Regular HAVO students make a spurt at the very end, perhaps because they have to prepare for the central exam which involves a great deal of reading (input).

For regular programs the implication is that students develop steadily, but when they have to read a lot for their central exams they show a great deal of improvement. More meaningful and interesting input (listening and reading) rather than standard textbook lessons would probably increase the rate of development.

Motivation

In the literature it is often suggested (xxx) that there is a relation between motivation, and language development; however, a causal relation is difficult to prove. We assume that they interact dynamically and then it is possible that learners who are motivated learn faster and better, but on the other hand better learners who have good results may become more motivated.

First the responses of all students were used for a factor analysis, with a number of factors emerging. Factor 1 items are related to “instrumental motivation”, Factor 2 to “attitude towards language”, Factor 3 to “integrative motivation” and Factor 4 to “attitude towards English class”. Then the juniors (grades 1-3) and seniors (grades 4-6) were compared respectively across programs on generated factors. If the factors were significantly related, MANOVA was used; otherwise ANOVA.

Factor analysis with 15 motivation items.

	Factors			
	1	2	3	4
Instrumental				
It is important for me to learn English.	.748			
Studying English is important because it will be useful in getting a job.	.684			
English is an important language of communication in the world.	.669			
I will need English after high school for my studies.	.661			
I want to be fluent in English.	.524			
English enables me to communicate with people from different language groups.	.502			
Integrative				
I would like to have many English speaking friends.	.768			
I am interested in the cultures of English speaking countries.	.667			
English helps me to find new friends on holiday.	.623			
I like to talk to native speakers of English.	.590			
Attitude language				
I am good at English compared to my fellow students.			.781	
I understand everything during my English lessons.			.692	
I like the English language.			.635	
I enjoy speaking English.			.624	
Attitude class				
I like my English class.				.900

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

The average scores on the different motivation factors for each group (per cohort of years 1-3) are shown in figure 5.

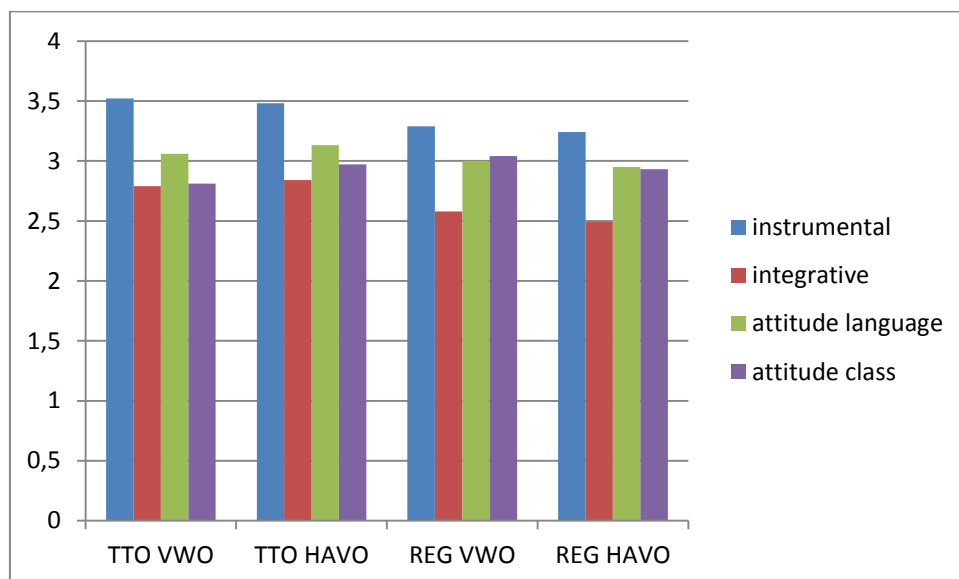


Figure 5. Motivation scores for the juniors (grades 1-3).

MOTIVATION	Grades 1-3
Instrumental	TTO groups > Regular groups.
Integrative	TTO groups > Regular groups
Attitude towards language	T-HAVO > Regular HAVO
Attitude towards English class	No differences

Table 9: Significant differences in motivation factors in groups of grades 1-3.

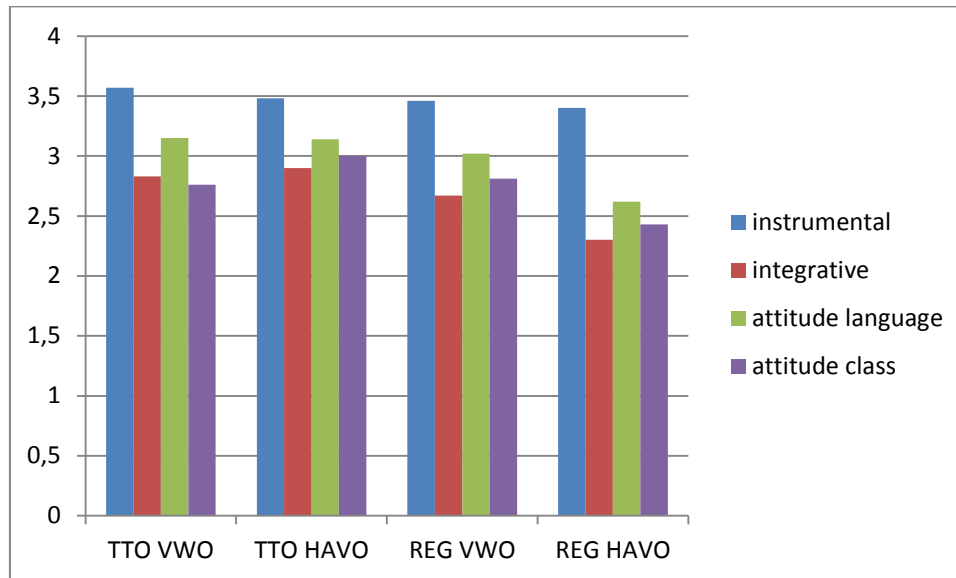


Figure 6. Motivation scores for the seniors (grades 4-6 for VWO and grades 4-5 for HAVO).

MOTIVATION	Grades 4-6
Instrumental	TTO groups and regular VWO > Regular HAVO
Integrative	TTO groups and regular VWO > Regular HAVO
Attitude towards language	TTO groups and regular VWO > Regular HAVO
Attitude towards English class	TTO groups and regular VWO > Regular HAVO

Table 10: Significant differences in motivation in groups of grades 4-6.

In the juniors, the TTO classes differ significantly from the regular classes in both instrumental and integrative motivation, but for attitude towards the language and class there are not many differences, only T-HAVO outscores regular HAVO on attitude towards the language. In the seniors, the regular VWO students become similar to their TTO counterparts and only regular HAVO is on the whole less motivated compared with the other three groups.

The findings are not so surprising as the TTO pupils are self-selected and we do not know if they were more motivated to begin with. However, the overall low motivation for regular HAVO students make one wonder how it could be improved.

Willingness to Communicate

In the literature it is often suggested (xxx) that there is a relation between WTC and language development but there is no empirical evidence as far as we know. It is still interesting to investigate to see to what extent students are willing to communicate in the foreign language.

From the responses in the willingness to communicate questionnaire two complementary factors emerged: Factor 1 showing anxiety and Factor 2 self-confidence.

Factor analysis with 8 willingness to communicate items.

	Factors	
	1	2
Anxiety		
I'm afraid that my classmates will laugh at me when I have to speak English.	.831	
I feel worried when I have to speak English.	.802	
I get nervous speaking English.	.792	
I do not feel comfortable speaking English.	.704	
Self-confidence		
I feel comfortable talking English in a small group of people my age.		.813
I feel comfortable talking English in a large group of people my age.		.746
I feel comfortable speaking English in a group discussion.		.737
I often give answers in English in class.		.726

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Figure 7 and Table 11 shows the results for the juniors and Figure 8 and Table 12 for the seniors.

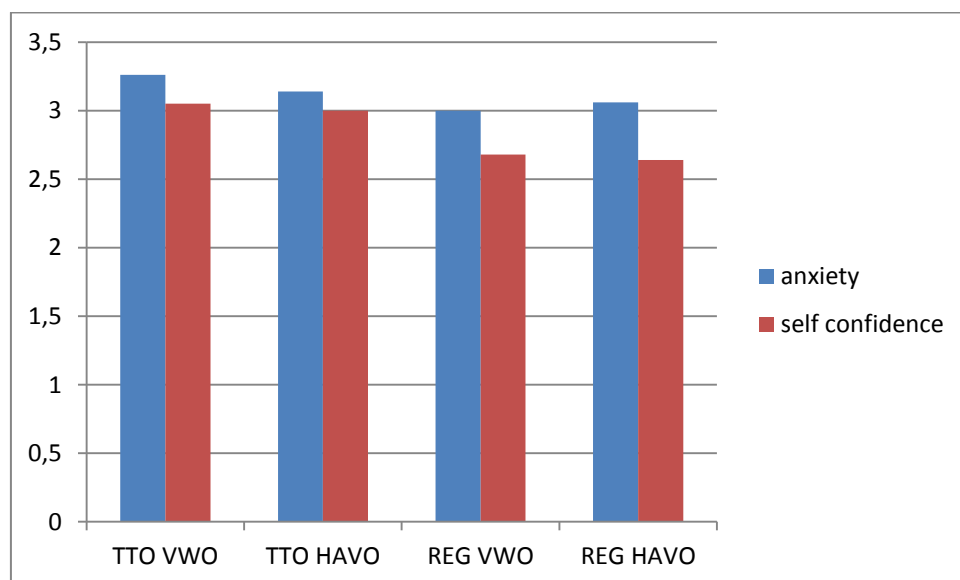


Figure 7. WTC scores for the juniors (grades 1-3).

WTC	Grades 1-3
anxiety	T-VWO > Regular groups
self-confidence	TTO groups > Regular groups

Table 11: Significant differences in WTC in the juniors.

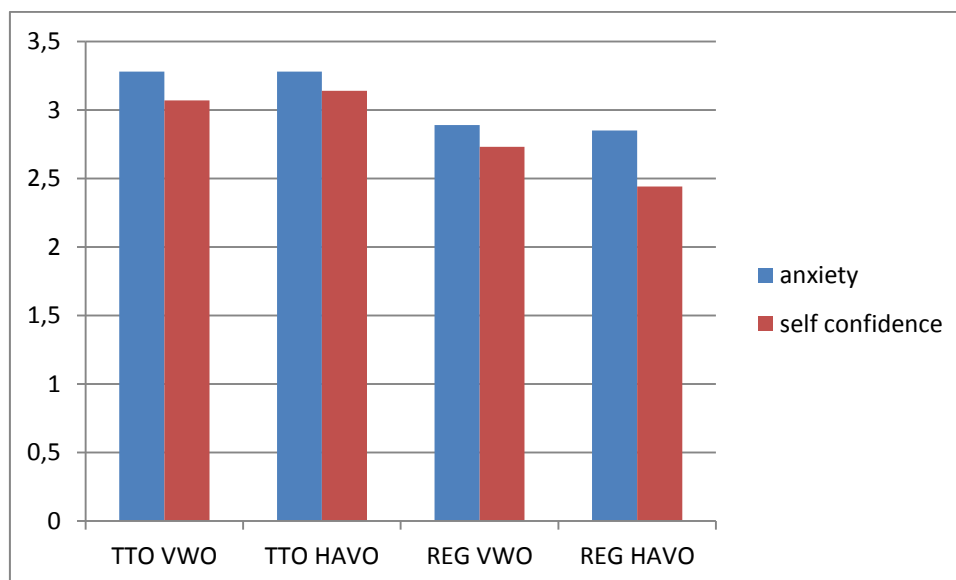


Figure 8. WTC scores for the seniors (grades 4-6 for VWO and grades 4-5 for HAVO).

WTC	Grades 4-6
anxiety	TTO groups > groups.
Self-confidence	TTO groups > Regular VWO TTO groups and regular VWO > Regular HAVO

Table 12: Significant differences in WTC in seniors.

When speaking in the second language, a person can feel either nervous or comfortable. What is surprising is that the younger T-VWO learners and both older TTO groups are both more anxious to speak and more self-confident. However, MacIntyre et al. (2002) point out this is possible because high achievers may know they can speak but still worry about the errors they may make. Both TTO groups in both cohorts are more willing to speak up in groups than their regular peers. In the seniors the regular VWO groups are also more willing to speak than the regular HAVO groups.

Instructional preferences

There is not much literature on instructional preferences, but we included these questions to see if the groups differed in their preferences as we wanted to know if adaptations to instructional modes might be indicated for the T-HAVO groups.

From the responses on the instructional preference items, three factors emerged: Factor 1 seems to indicate a liking of traditional teacher-fronted teaching, Factor 2 a liking of input and interaction, and Factor 3 a liking for individual work.

Factor analysis with 8 items of teaching and learning preferences

	Factors		
	1	2	3
Traditional			
I listen to the teacher in class.	.826		
I listen to the teacher when he/she talks about grammar.	.816		
I listen to the teacher when he/she talks about the meaning of new words.	.804		
I learn a lot from textbook exercises.	.578		
Input and interaction			
I like to watch videos.		.813	
I like to work in a group.		.728	
I learn a lot from listening exercises.		.454	
Individual			
I like to work on my own.			.905

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Although some correlations between the factors are significant, they are low and factors 2 and 3 are not related; therefore, an ANOVA is used here to compare the differences among the factors between groups.

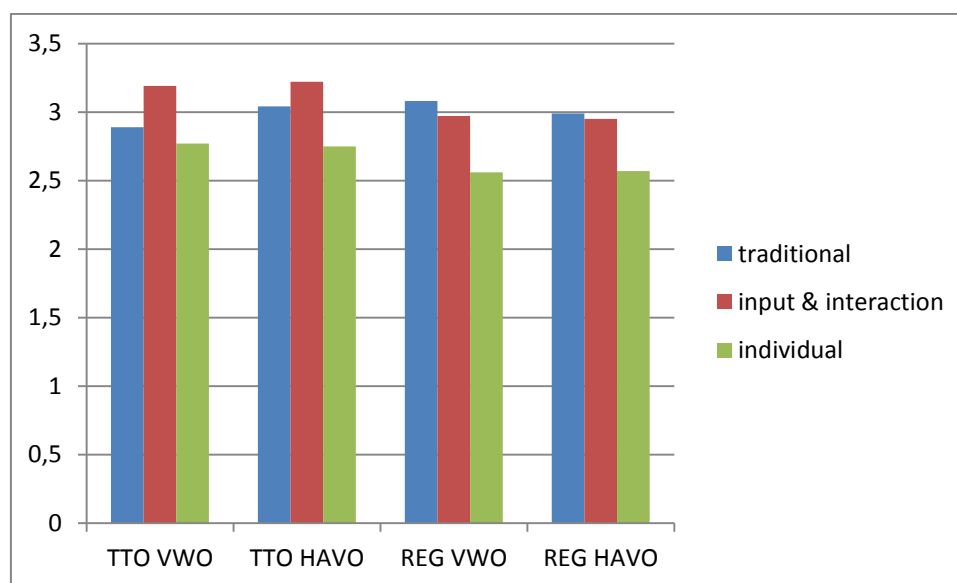


Figure 9. Instructional preferences for the juniors (grades 1-3).

Instructional preferences	Grades 1-3
Traditional	T-HAVO and regular HAVO > T-VWO .
Input and interaction	TTO groups > regular groups
Individual work	No differences

Table 13: Significant differences in instructional preferences for the juniors.

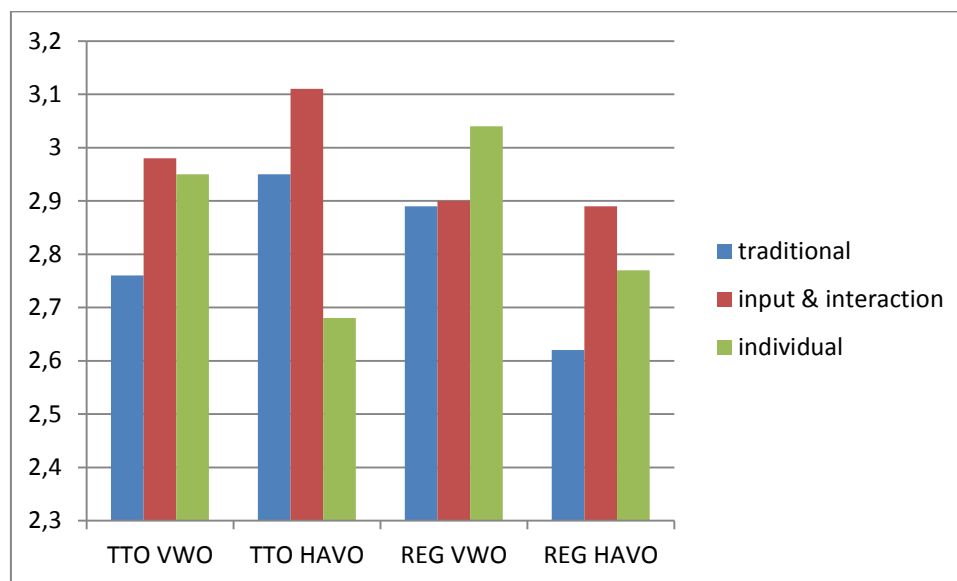


Figure 10. Instructional preference scores for the seniors (grades 4-6 for VWO and grades 4-5 for HAVO).

Instructional preferences	Grades 4-5 or 4-6
Traditional teaching	T-HAVO and Reg VWO > regular HAVO
Input & interaction	T-HAVO > regular HAVO
Individual work	T-VWO > regular HAVO Regular VWO > T-HAVO and regular HAVO

Table 14: Significant differences in teaching and learning preferences in seniors.

Even though not all differences are significant, it seems that some interesting differences emerge in instructional preferences. In the juniors, the HAVO groups are a little more in favor of a traditional style than T-VWO and both TTO groups favor input and interaction, which they are probably used to in their TTO classes. In the seniors T-HAVO groups clearly prefer input and interaction and the VWO groups prefer individual work more than the HAVO groups. In the seniors, the regular HAVO group is less interested in any style than the other groups.

If our assumption is correct that the differences in language development between VWO and HAVO is not so much related to the final levels they may acquire but more

a matter of rate of development, these findings may mean that their preferred paths are also somewhat different. The VWO pupils are apparently ready to interact and work independently from the teacher from the start but then towards the end are ready to work on their own, whereas the HAVO pupils may want to start off a bit slower with more teacher scaffolding earlier.

Out-of-school contact

Previous research has clearly shown a positive effect for out-of-school input (cf. Verspoor et al 2011) on foreign language development, so it is interesting to what degree the groups differ in terms of out-of-school contact.

From the responses on out-of-school contact items, two factors emerged. Type 1 concerns reading, playing video games and speaking, three types of input that are not really related. Type 2 concerns listening to English music and television, popular media.

Factor analysis with 5 items of out-of-school contact with English

	Factors	
	1	2
<i>Type 1</i>		
How many hours per week do you spend reading English (not homework)?	.763	
How many hours per week do you play English video games?	.737	
How much time per week do you spend speaking English out of school?	.684	
<i>Type 2</i>		
How many hours per week do you spend listening to English music?		.771
How many hours per week do you spend watching English television programs?		.722

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Figures 11 and 12 illustrate the types of input the juniors and seniors receive and Tables 15 and 16 summarize the significant differences.

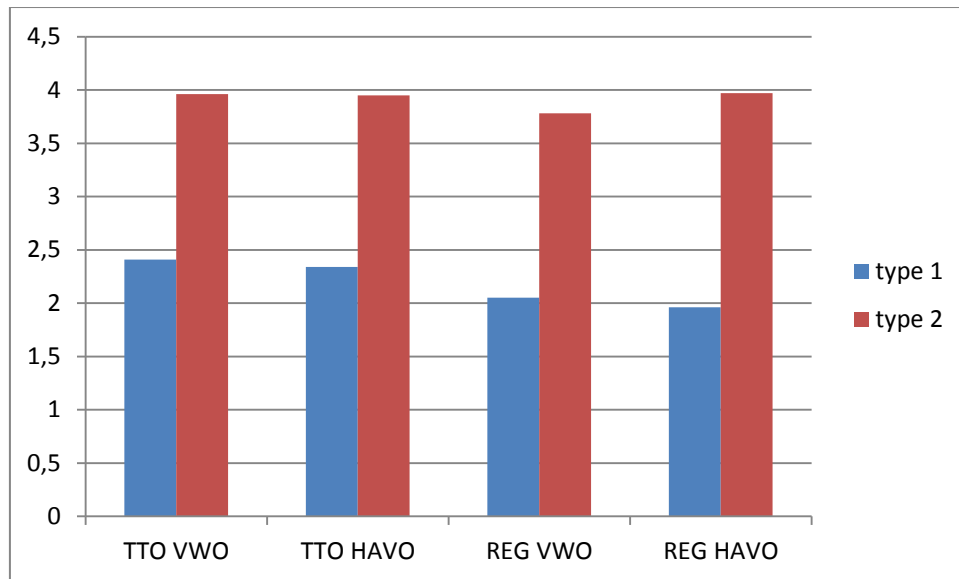


Figure 11. Out of school contact for the juniors (grades 1-3).

Out of school contact	Grades 1-3
Type 1	T-VWO > regular groups T-HAVO > regular HAVO
Type 2	No difference at all.

Table 15. Significant differences in out of school contact in juniors.



Figure 12. Out of school contact scores for the seniors (grades 4-6 for VWO and grades 4-5 for HAVO).

Out of school contact	Grades 4-6
Input 1	TTO groups > regular HAVO
Input 2	No difference at all.

Table 16. Significant differences in out of school contact in seniors.

On the whole all students in both cohorts get more out of school input through type 2 (listening to music and watching television programs) than through type 1 (reading, video games and speaking). There are no differences here among the groups. The TTO groups do get significantly more out of school input through type 1.

Correlation analyses motivation, WTC, instructional preferences, out-of-school contact and language proficiency scores.

To be able to determine which factors seem to contribute most to language proficiency, correlation analyses were conducted to ascertain whether or not there were correlations to be found between instructional preferences, WTC, and out of school contact with the proficiency scores. The full table of correlations with all variables taken into account can be found in the appendix.

Several significantly positive correlations were found, but the strongest positive correlations were found between self-confidence (under WTC) and the proficiency scores across the board (writing, vocabulary, and general proficiency) i.e. students who were more confident were significantly more likely to perform better in all areas. This findings does not indicate any causal relation. It is very well possible that the relation between self-confidence and proficiency is a logical one: the more proficient, the more self-confident.

In addition, a medium strength positive correlation was found between an instrumental or integrative orientation, and a positive attitude to the language (all three under motivation) and writing, vocabulary, and general proficiency scores. The motivational relations could also be logical as for example the TTO learners are taking TTO because they are more motivated, but even among the regulars this motivation relation holds and one wonders if English classes should also try to make learners enthusiastic about the language.

Conclusion and implications

Please see Dutch part pages 14 and 15.

References

- Bonnet, G. & et al. (2004). *The assessment of pupils' skills in English in eight European countries 2002* Paris: European Network of policy makers for the evaluation of education systems.
- Caspi, T. (2010). *A dynamic perspective on second language development*. (Doctoral dissertation). Retrieved from <http://irs.ub.rug.nl/ppn/329338412>.
- Cummins, J. (2000). BICS and CALP. In M. Byram (Ed.), *Encyclopedia of language teaching and learning*. (pp. 76-79). London: Routledge.
- MacIntyre, P.D., Baker, S.C., Clément, R., & Donovan, L.A. (2002). Sex and age effects on willingness to communicate, anxiety, perceived competence, and motivation among junior high school French immersion students. *Language Learning*, 52, 537-564.
- Schuitemaker-King, J. (2012). Teachers' strategies in providing opportunities for second language development. PhD dissertation. University of Groningen.
- Verspoor, M. & Edelenbos, P. (2009). Tweetalig Onderwijs: Beter geschoolde leerlingen in 2024. In De Graaff, R & D.Tuin (Eds). *De Toekomst van het Talenonderwijs: Nodig? Anders? Beter?*. IVLOS: Utrecht (pp147-164)
- Verspoor, M. de Bot, K. & Xu, X. (2011). The role of input and scholastic aptitude in second language development. *Toegepaste Taalwetenschap in Artikelen*. 86.2 (pp.47-60)
- Verspoor. M., Schmid, M.S. & Xu, X. (2012). A dynamic usage based perspective on L2 writing development. *Journal of Second Language Writing*, 21, 3, 239-263.

Appendix 1

Questionnaire: Please indicate how much you agree with these statements.

1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree

Motivation

I am good at English compared to my fellow students.

I enjoy speaking English.

I like the English language.

English enables me to communicate with people from different language groups.

It is important for me to learn English.

I will need English after high school for my studies.

Studying English is important because it will be useful in getting a job.

I like to talk to native speakers of English.

English helps me to find new friends on holiday.

English is an important language of communication in the world.

I like my English class.

I understand everything during my English lessons.

I would like to have many English speaking friends.

I want to be fluent in English.

I am interested in the cultures of English speaking countries.

Willingness to communicate at school and out of school.

I feel comfortable speaking English in a group discussion.

I often give answers in English in class.

I get nervous speaking English.

I'm afraid that my classmates will laugh at me when I have to speak English.

I feel comfortable talking English in a small group of people my age.

I feel comfortable talking English in a large group of people my age.

I do not feel comfortable speaking English.

I feel worried when I have to speak English.

In-class engagement

I learn a lot from textbook exercises.

I like to work in a group.

I like to work on my own.

I like to watch videos.

I learn a lot from listening exercises.

I listen to the teacher in class.

I listen to the teacher when he/she talks about grammar.

I listen to the teacher when he/she talks about the meaning of new words.

Out of school contact

Please indicate how much time you spend on these activities

1. None

2. Less than 2 hours per week

3. 2-4 hours per week

4. 4-6 hours per week

5. More than 6 hours per week

How many hours per week do you spend listening to English music?

How many hours per week do you spend watching English television programmes?

How many hours per week do you spend reading English (NOT English homework)?

How much time per week do you spend speaking English out of school?

How many hours per week do you play English video games?

Appendix 2

OTTO-1 versus OTTO-2: means and standard deviations of test scores

	Grade 1				Grade 3			
	T-VWO		Reg VWO		T-VWO		Reg VWO	
	OTTO-1 N=66	OTTO-2 N=55	OTTO-1 N=42	OTTO-2 N=22	OTTO-1 N=63	OTTO-2 N=73	OTTO-1 N=50	OTTO-2 N=28
Writing	3.58(0.9)	2.78(1.0)	2.43(0.9)	2.27(0.8)	5.56(1.0)	4.34(1.2)	4.42(0.9)	4.04(1.0)
Vocabulary	0.65(0.2)	0.73(0.1)	0.48(0.1)	0.61(0.1)	0.78(0.1)	0.82(0.1)	0.54(0.2)	0.68(0.1)
General proficiency	0.44(1.0)	0.30(0.7)	-0.85(0.9)	-0.45(0.6)	0.66(0.7)	0.19(0.8)	-0.84(0.9)	-0.49(0.9)

OTTO-1 versus OTTO-2: statistical analyses results of tests

Grade 1

T-VWO	Writing	T-Test	t(119)=-4.49	p<.001	eta ² =.14
	Vocabulary	T-Test	t(119)=3.18	p<.01	eta ² =.09
	General proficiency	T-Test	t(119)=-0.93	p=.36	
Reg VWO	Writing	T-Test	t(62)=-0.68	p=.50	
	Vocabulary	T-Test	t(62)=3.93	p<.001	eta ² =.20
	General proficiency	T-Test	t(62)=1.92	p=.06	
Grade 3					
T-VWO	Writing	T-Test	t(134)=-6.23	p<.001	eta ² =.22
	Vocabulary	T-Test	t(134)=2.44	p<.05	eta ² =.04
	General proficiency	T-Test	t(134)=-3.55	p<.01	eta ² =.09
Reg VWO	Writing	T-Test	t(76)=-1.76	p=.08	
	Vocabulary	T-Test	t(76)=4.00	p<.001	eta ² =.17
	General proficiency	T-Test	t(76)=1.65	p=.10	

OTTO-2: means and standard deviations of tests of each grade in each programme

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Writing						
T-VWO	2.78 (1.0)	4.09 (1.0)	4.34 (1.2)	5.89 (0.8)	5.66 (1.2)	6.54 (1.0)
TTO HAVO	2.73 (1.0)	3.69 (0.9)	4.27 (1.5)	5.76 (1.0)	5.56 (1.3)	X
Reg VWO	2.27 (0.8)	2.68 (0.9)	4.04 (1.0)	3.64 (1.0)	4.36 (0.8)	5.62 (0.9)
Reg HAVO	2.12 (0.7)	2.65 (0.9)	2.81 (0.9)	2.13 (1.2)	3.51 (1.2)	X
Vocabulary						
T-VWO	0.73 (0.1)	0.83 (0.1)	0.82 (0.1)	0.82 (0.1)	0.80 (0.1)	0.85 (0.1)
TTO HAVO	0.70 (0.1)	0.75 (0.1)	0.77 (0.1)	0.77 (0.1)	0.73 (0.1)	X
Reg VWO	0.61 (0.1)	0.67 (0.1)	0.68 (0.1)	0.66 (0.1)	0.74 (0.1)	0.80 (0.1)
Reg HAVO	0.58 (0.1)	0.56 (0.2)	0.67 (0.1)	0.54 (0.2)	0.68 (0.1)	X
General proficiency						
T-VWO	-0.42 (0.5)	0.43 (0.5)	0.50 (0.7)	1.03 (0.4)	0.88 (0.7)	1.39 (0.5)
TTO HAVO	-0.59 (0.6)	-0.05 (0.5)	0.27 (0.8)	0.78 (0.5)	0.54 (0.6)	X
Reg VWO	-1.10 (0.5)	-0.73 (0.7)	-0.19 (0.8)	-0.43 (0.7)	0.17 (0.5)	0.84 (0.5)
Reg HAVO	-1.31 (0.6)	-1.20 (0.9)	-0.70 (0.8)	-1.45 (1.2)	-0.39 (0.7)	X

OTTO-2: statistical analyses results of tests

Proficiency comparison across programs

Writing	Grade 1	ANOVA	F(3, 177)=4.5	p<.01	eta ² =.07
	Grade 2	ANOVA	F(3, 137)=20.3	p<.001	eta ² =.31
	Grade 3	ANOVA	F(3, 156)=10.7	p<.001	eta ² =.17
	Grade 4	ANOVA	F(3, 129)=112.9	p<.001	eta ² =.72
	Grade 5	ANOVA	F(3, 106)=26.5	p<.001	eta ² =.43
	Grade 6	T-Test	t(37)=2.77	p<.01	eta ² =.17
Vocabulary	Grade 1	ANOVA	F(3, 177)=18.0	p<.001	eta ² =.22
	Grade 2	ANOVA	F(3, 137)=34.7	p<.001	eta ² =.43
	Grade 3	ANOVA	F(3, 156)=14.3	p<.001	eta ² =.22
	Grade 4	ANOVA	F(3, 129)=30.8	p<.001	eta ² =.42
	Grade 5	ANOVA	F(3, 106)=10.3	p<.001	eta ² =.23
	Grade 6	T-Test	t(37)=2.42	p<.05	eta ² =.14

General proficiency	Grade 1	ANOVA	F(3, 177)=16.9	$p < .001$	$\eta^2 = .22$
	Grade 2	ANOVA	F(3, 137)=46.0	$p < .001$	$\eta^2 = .50$
	Grade 3	ANOVA	F(3, 156)=18.1	$p < .001$	$\eta^2 = .26$
	Grade 4	ANOVA	F(3, 129)=86.16	$p < .001$	$\eta^2 = .67$
	Grade 5	ANOVA	F(3, 106)=26.0	$p < .001$	$\eta^2 = .42$
	Grade 6	T-Test	t(37)=3.39	$p < .01$	$\eta^2 = .24$
Proficiency development across years					
Writing	T-VWO	ANOVA	F(5, 285)=73.0	$p < .001$	$\eta^2 = .56$
	TTO HAVO	ANOVA	F(4, 194)=58.4	$p < .001$	$\eta^2 = .55$
	Reg VWO	ANOVA	F(5, 118)=29.1	$p < .001$	$\eta^2 = .55$
	Reg HAVO	ANOVA	F(4, 145)=11.6	$p < .001$	$\eta^2 = .24$
Vocabulary	T-VWO 1-3 grades	ANOVA	F(2, 183)=17.3	$p < .001$	$\eta^2 = .16$
	TTO HAVO1-3 grades	ANOVA	F(2, 143)=7.0	$p < .05$	$\eta^2 = .09$
	Reg VWO 1-3 grades	ANOVA	F(2, 72)=2.1	$p = .13$	
	Reg HAVO 1-3 grades	ANOVA	F(2, 72)=3.6	$p < .05$	$\eta^2 = .09$
	T-VWO 4-6 grades	ANOVA	F(2, 102)=2.9	$p = .06$	
	TTO HAVO4-5 grades	T-Test	t(51)=1.5	$p = .13$	
	Reg VWO 4-6 grades	ANOVA	F(2, 46)=8.8	$p < .01$	$\eta^2 = .28$
	Rge HAVO 4-5 grades	T-Test	t(73)=-3.6	$p < .01$	$\eta^2 = .15$
General proficiency	T-VWO 1-3 grades	ANOVA	F(2, 183)=44.0	$p < .001$	$\eta^2 = .32$
	TTO HAVO1-3 grades	ANOVA	F(2, 143)=22.1	$p < .001$	$\eta^2 = .24$
	Reg VWO 1-3 grades	ANOVA	F(2, 72)=10.4	$p < .001$	$\eta^2 = .22$
	Reg HAVO 1-3 grades	ANOVA	F(2, 72)=4.6	$p < .05$	$\eta^2 = .11$
	T-VWO 4-6 grades	ANOVA	F(2, 102)=6.7	$p < .01$	$\eta^2 = .12$
	TTO HAVO4-5 grades	T-Test	t(51)=1.5	$p = .14$	
	Reg VWO 4-6 grades	ANOVA	F(2, 46)=19.7	$p < .001$	$\eta^2 = .46$
	Rge HAVO 4-5 grades	T-Test	t(73)=-5.0	$p < .001$	$\eta^2 = .26$

OTTO-2: means and standard deviations of learning factors

	Grades 1-3				Grades 4-6			
	T-VWO	TTO HAVO	Reg VWO	Reg HAVO	T-VWO	TTO HAVO	Reg VWO	Reg HAVO
Motivation factors								
Instrumental	3.52(0.3)	3.48(0.4)	3.29(0.4)	3.24(0.5)	3.57(0.3)	3.48(0.4)	3.46(0.4)	3.08(0.6)
Integrative	2.79(0.5)	2.84(0.5)	2.58(0.6)	2.49(0.6)	2.83(0.5)	2.91(0.4)	2.67(0.5)	2.30(0.6)
Attitude language	3.06(0.4)	3.13(0.4)	3.00(0.5)	2.95(0.5)	3.15(0.4)	3.14(0.4)	3.02(0.5)	2.62(0.8)
Attitude class	2.82(0.7)	2.97(0.6)	3.04(0.7)	2.93(0.7)	2.76(0.6)	3.00(0.6)	2.82(0.8)	2.43(1.0)
Willingness to communicate factors								
Anxiety	3.26(0.6)	3.14(0.5)	3.00(0.5)	3.06(0.6)	3.28(0.6)	3.29(0.6)	2.89(0.5)	2.85(0.8)
Self confidence	3.05(0.5)	3.00(0.5)	2.68(0.5)	2.64(0.6)	3.07(0.6)	3.14(0.5)	2.73(0.5)	2.44(0.7)
Instructional preferences factors								
Traditional	2.89(0.5)	3.04(0.4)	3.04(0.4)	3.08(0.5)	2.76(0.4)	2.95(0.4)	2.89(0.5)	2.62(0.7)
Input/interaction	3.19(0.4)	3.22(0.4)	2.97(0.4)	2.95(0.4)	2.98(0.4)	3.11(0.4)	2.90(0.4)	2.89(0.6)
Individual work	2.77(0.7)	2.75(0.7)	2.56(0.7)	2.57(0.8)	2.95(0.6)	2.68(0.7)	3.04(0.6)	2.40(0.8)
Out-of-school contact factors								
Type 1	2.41(0.9)	2.34(0.9)	2.05(0.9)	1.96(0.8)	2.43(0.9)	2.45(0.9)	2.12(0.9)	1.88(0.9)
Type 2	3.96(0.8)	3.95(0.8)	3.78(0.8)	3.97(0.9)	3.95(0.9)	4.21(0.8)	4.00(0.8)	3.88(0.9)

OTTO-2: statistical analyses results of learning factors

Motivation factors					
Grades 1-3	MANOVA	F(12, 1421)=5.6	$p < .001$	Hotelling's Trace=.14	$\eta^2 = .05$
Grades 4-6	MANOVA	F(12, 821)=8.3	$p < .001$	Hotelling's Trace=.36	$\eta^2 = .11$
Willingness to communicate factors					
Grades 1-3	MANOVA	F(6, 952)=5.6	$p < .001$	Hotelling's Trace=.14	$\eta^2 = .07$
Grades 4-6	MANOVA	F(6, 552)=8.3	$p < .001$	Hotelling's Trace=.26	$\eta^2 = .11$

Instructional preferences factors

Traditional	Grades 1-3	ANOVA	$F(3, 478)=4.47$	$p<.01$	$\eta^2=.03$
	Grades 4-6	ANOVA	$F(3, 278)=5.02$	$p<.01$	$\eta^2=.05$
Input/interaction	Grades 1-3	ANOVA	$F(3, 478)=11.35$	$p<.001$	$\eta^2=.07$
	Grades 4-6	ANOVA	$F(3, 278)=2.78$	$p<.05$	$\eta^2=.03$
Individual work	Grades 1-3	ANOVA	$F(3, 478)=2.57$	$p=.053$	
	Grades 4-6	ANOVA	$F(3, 278)=12.86$	$p<.001$	$\eta^2=.12$

Out-of-school contact factors

Type 1	Grades 1-3	ANOVA	$F(3, 478)=6.71$	$p<.001$	$\eta^2=.04$
	Grades 4-6	ANOVA	$F(3, 278)=6.51$	$p<.001$	$\eta^2=.07$
Type 2	Grades 1-3	ANOVA	$F(3, 478)=1.03$	$p=.337$	
	Grades 4-6	ANOVA	$F(3, 278)=1.63$	$p=.184$	

OTTO-2: Correlations between learning factors and test scores

Predictors		Writing	Vocabulary	General proficiency
Grades		.551**	.173**	.418**
School types		-.413**	-.519**	-.541**
CITO scores		.267**	.384**	.380**
Motivation:	Instrumental	.268**	.341**	.353**
	Integrative	.243**	.224**	.271**
	Attitude language	.295**	.338**	.367**
	Attitude class	-.027	.023	-.002
Willingness to communicate:	anxiety	.195**	.196**	.227**
	Self-confidence	.298**	.325**	.362**
Instructional preferences:	Traditional	-.039	.027	-.007
	Input & interaction	.029	.147**	.103**
	Individual work	.218**	.168**	.224**
Out-of-school contact:	Type 1	.163**	.194**	.207**
	Type 2	.134**	.098**	.134**

**Correlation is significant at the 0.01 level (2-tailed).

Predictors which were shown to be significantly related to the test scores were selected to be entered into a regression model to examine to what extent these factors could explain the variance in the general proficiency (dependent variable) which showed to be related higher to most factors than writing and vocabulary scores.

Hierarchical multiple regression of predictors on general proficiency

	R^2	Adjusted R^2	R^2 change	B	β
<i>Model 1</i>	.501	.499	.501		
Grades				.293	.456**
School types				-.470	-.538**
CITO scores				.009	.040
<i>Model 2</i>	.571	.564	.071**		
Instrumental				.077	.032
Integrative				-.025	-.014
Attitude language					
Anxiety				.464	.233**
Self-confidence				.025	.015
Input and interaction				-.033	-.019
Individual work					
Type 1 contact				.035	.016
Type 2 contact				.054	.040
				.014	.013
				.068	.057*

** $p<.01$; * $p<.05$.