Bucking the trend?

Motivational differences between boys and girls who opt in or out of bilingual education

Tessa Mearns and Rick de Graaff
Leiden University / Utrecht University

Research has suggested that motivation plays a significant role in language learning but that females tend to be more motivated language learners than males. Content and language integrated learning (CLIL) has been suggested as a means of narrowing this motivational gender gap, although there is as yet little empirical evidence to support this claim. In the current study, data regarding the motivation of 581 learners in bilingual and mainstream tracks of Dutch secondary education were analysed for interaction effects in terms of Gender, Education Type, and Year of CLIL study. In this context, it seemed that boys who had chosen bilingual education were the most positive and motivated regarding the learning of English, although girls had more positive attitudes regarding languages in general. No interactions were observed between Gender, Education Type, and Year, suggesting that existing differences may have influenced boys' decision to follow bilingual education rather than the reverse.

Motivatie speelt een belangrijke rol bij het leren van een taal, maar meisjes zijn vaak gemotiveerder dan jongens om talen te leren. *Content and language integrated learning* (CLIL) zou een manier kunnen zijn om het verschil in de taal-motivatie van jongens en meisjes op school te verkleinen, hoewel dit nog niet is bevestigd op basis van empirisch onderzoek. Dit onderzoek maakt gebruik van data over de motivatie van 581 leerlingen in het tweetalig en regulier voortgezet onderwijs in Nederland. Interactie-analyses zijn uitgevoerd op basis van sekse (*gender*), onderwijstype en leerjaar. Het blijkt dat in het voor tweetalig onderwijs jongens gemotiveerder zijn om Engels te leren dan meisjes. De meisjes vertonen echter meer motivatie voor het leren van vreemde talen in het algemeen. Er is weinig interactie tussen onderwijsvariant, sekse en leerjaar. Dit laatste suggereert dat hogere motivatie vooraf een reden zou kunnen zijn voor de keuze van deze jongens voor tweetalig onderwijs, en niet een resultaat van het onderwijs zelf.

1. Introduction

Research has suggested that females tend to be more enthusiastic about learning new languages than males (Csizér & Dörnyei, 2005; Henry, 2009). Considering that motivation has been highlighted as one of the main determinants of successful language learning (Gardner, 1985; Crookes & Schmidt, 1991; Oxford & Shearin, 1994), it follows that disparity between the genders in this area could be cause for concern among educators. Content and language integrated learning (CLIL) at the school level has been identified as a means of counteracting lower levels of motivation among boys on the classroom level, as it is thought to appeal to their greater instrumentality (Lasagabaster, 2008). Approached from a different perspective, Baker and MacIntyre (2000) mention the possibility that females might be better served by the communicative nature of bilingual approaches (in their example, immersion). As Lasagabaster (2008) emphasises, however, little empirical evidence has been published to support either view, especially seeing as the (self-)selection of learners for CLIL programmes can lead to difficulties in identifying the causality of motivational differences in CLIL and non-CLIL settings.

Kissau (2006) has highlighted the importance of considering both classroom-based and contextual influences on language motivation and has suggested that there may be interplay between internal and external factors in determining such motivational differences between genders. In this sense, studies of gender and motivation in language learning may need to consider not only classroom practices but also attitudes, beliefs, and other internal or internalised features of the individual learner as he or she enters the classroom.

A study conducted from 2010–2014 (Mearns, 2015) aimed to investigate motivational differences between learners in bilingual (BE) and Dutch-language mainstream (ME) secondary education in the Netherlands. A central focus of the research was to establish whether motivational differences appeared to have developed during exposure to bilingual (BE) or mainstream (ME) education or whether they had preceded it and perhaps contributed to the choice of educational programme. In the current paper, data from the same study have been re-analysed from the perspective of gender, taking into account the existing literature regarding language motivation, gender, and CLIL. The aim of the re-analysis was to investigate (1) the extent to which a hypothesised gender gap in language motivation surfaced in both the BE and ME contexts and (2) whether the gender gap in BE was narrower among pupils who had been exposed to BE for longer. In investigating gender differences in motivation as not necessarily causally linked to the educational experience, the study directly addresses the possibility that both internal and external factors might influence trends in learner motivation.

2.   CLIL and bilingual education in the Netherlands

CLIL is an approach to teaching and learning that combines and integrates subject content with the learning of an additional language, placing explicit emphasis on both content and language. There is no single methodology for CLIL, whose form tends to be context-bound and can range from small-scale and short-term teacher-led projects to educational paradigms monitored and coordinated on a national level (Coyle, Hood, & Marsh, 2010; Morton & Llinares, 2017). The latter is the case in the Netherlands, where the bilingual education context has been said to "[stand] out particularly prominently as an example of remarkable CLIL investigation" (Pérez-Cañado, 2012, p. 9).

Dutch secondary education begins at the age of approximately twelve and is divided into three main streams: preparatory academic education (VWO), higher general education (HAVO), and preparatory vocational education (VMBO). About 50% of young people attend VMBO, the other 50% being divided between the other two streams (Onderwijs in Cijfers, 2016). Bilingual education (henceforth BE) in the Netherlands was first established in 1989 and was initially provided only for the most academic (VWO) learners. According to the most recent figures (Nuffic, 2017), however, the approach has since spread to a network of over 130 secondary schools and the full range of academic, general and vocational streams. In BE at preparatory academic (VWO) and general secondary (HAVO) levels, 50% of the curriculum is taught in English during lower secondary education (aged 12–15). As in many CLIL contexts, Dutch BE is usually selective, although selection procedures vary across schools. While some schools select on the basis of academic performance or affinity for language, others take only attitude and motivation for the programme into account.

Dutch BE stands out against many other CLIL programmes internationally due to the existence of a national standard and associated quality control measures, established by the Dutch Network of Bilingual Schools. Schools that offer BE at any level are required to work towards and obtain official accreditation from the Network and are subject to inspection and appraisal every five years (de Graaff & van Wilgenburg, 2015).

Many studies into Dutch BE to date have focused on the question of its effectiveness for L2 acquisition (Klaassen, 2001; Verspoor, de Bot, & van Rein, 2010; Verspoor, de Bot, & Xu, 2015; Verspoor & Edelenbos, 2009), although research has also been conducted into the effects on L3 learning (Elzenga & de Graaff, 2015; Rutgers, 2013) and regarding issues pertinent to teaching and teacher training (de Graaff, Koopman, Anikina, & Westhoff, 2007; van Kampen, Admiraal, & Berry, 2016; Koopman, Skeet, & de Graaff, 2014). There has as yet been little attention given to affective factors in BE in the Netherlands, although attitudes and

motivation have appeared as elements of larger studies. For example, Denman's on-going longitudinal study of bilingual pre-vocational education has so far identified positive effects on both attainment and attitudes (Denman, Tanner, & de Graaff, 2013). In another study, Verspoor et al. (2015) investigated the interplay between individual factors among BE learners in their first and third years of secondary school, concluding that motivation and attitude were good predictors of proficiency among the older BE learners, although not among the younger learners. Neither of these studies, however, has taken motivation as its central focus, nor have they addressed the question of gender differences in BE. The current study aims to go some way towards filling that gap in the research.

3.   Motivation in language learning

Motivation is widely considered by both researchers and teachers as a major con-tributor to language learning success (Dörnyei, 2001), although it is also largely understood that motivation is a complex phenomenon that is dependent on a large number of factors. It is therefore perhaps not surprising that language motivation research is an area that has undergone significant change in the last six decades (Dörnyei & Ushioda, 2011). Dörnyei's (2009) L2 Motivational Self System (L2MSS) has been referred to as one of the most influential recent L2 motivation theories (Ryan & Dörnyei, 2013). The L2MSS grew out of the psychological constructs of possible selves (Markus & Nurius, 1987) and self-discrepancy theory (Higgins, 1987). In addition, elements of Gardner's social-psychological models (Gardner & Lambert, 1972), Dörnyei's earlier three-level framework and process model (Dörnyei, 1994, 1998; Dörnyei & Ottó, 1998), and the recent emphasis on the co-inciding roles of individual learner differences, as influenced by society, and the specific teaching and learning context in L2 motivation (Kissau, 2006; Ushioda, 2009) can also be recognised. The tripartite structure of L2MSS recognises the im-portance of individual, societal, and classroom-level features, echoing Kissau's call to consider motivation on multiple levels, while also providing a framework that is easily applicable across a range of teaching and learning contexts.

3.1   The L2 Motivational Self System

The L2MSS consists of three components, namely the Ideal L2 Self, the Ought-to L2 Self and the L2 Learning Experience. The Ideal L2 Self is the learner's positive vision of the person s/he will become if successful in learning the target language (referred to here as L2, although it could be any additional language). Kormos, Kiddle and Csizér's (2011) proposed model of the interrelationships among

attitudes, motivations, and self-related beliefs also highlights the role of language attitudes in influencing the Ideal L2 Self. The Ought-to L2 Self, on the other hand, is the person the learner feels that s/he is *expected* to become through learning the L2. These expectations could come from family, friends, teachers or society as a whole. Dörnyei (2009) posits that harmony between the Ought-to and Ideal L2 selves is an optimal condition for nurturing motivation. The final element of the L2MSS, the L2 Learning Experience, refers to the context in which the language is learned. In most cases, this is the language classroom, although it could be argued that the L2 Learning Experience for English extends into out-of-school environments, in particular in contexts – like the Netherlands – where English has a high level of presence in everyday life (Sylvén & Sundqvist, 2012). A positive and encouraging language learning environment that allows room for the future L2 self-images to develop can also be conducive to language learning motivation (Dörnyei, 2009).

Since Dörnyei (2005) first introduced the concept of the L2MSS, it has formed the basis of several studies in a broad range of cultural and educational contexts. Findings of such studies have been largely positive regarding its applicability, in particular in relation to adolescents, whose self-concept may be more developed than that of younger children, yet more flexible than that of adults (Lamb, 2012; Ryan & Dörnyei, 2013). In a number of such studies, for example Csizér and Kormos's (2009) study of secondary school and university students in Budapest, positive correlations were found between motivation and both the Ideal L2 Self and the L2 Learning Experience. Results for the Ought-to L2 Self have been less convincing, perhaps due to the variability of the role of family across different cultures (Lamb, 2012; Taguchi, Magid, & Papi, 2009).

Other studies have considered the relevance of the L2MSS in specific language learning contexts or in relation to other theories of L2 motivation. Henry (2010, 2012) used the L2MSS as the inspiration for his own approach to L3 learning motivation, returning to Markus and Nurius' (1987) possible selves and the working self-concept to assess the possible plurality of Ideal L2 (or FL) selves. His results suggested that individuals simultaneously possess different self-concepts for different languages, which can interfere with one another. Ryan (2008, 2009) focused specifically on the Ideal L2 Self and its relationship to Gardner's concept of integrativeness (Gardner & Lambert, 1972). Gardner's theories, which dominated the L2 motivation field for much of the twentieth century, were centred on the learner's desire to learn the language either for instrumental goals (e.g., future employment) or in order to integrate with the L2 community. Ryan's results suggested that there was enough overlap between the two concepts to advocate reinterpreting the role of integrativeness as an aspect of the Ideal L2 Self, which is more applicable to a broader, global English-speaking culture (Macintyre, Mackinnon, & Clément, 2009).

3.2   Gender and the L2MSS

Gender differences in language learning have been observed on a number of levels, often with the conclusion that females outperform males. This appears to be no less the case in terms of motivation to learn languages (Henry, 2009), and there may even be a causal link between differences in motivation and in language learning performance (Gardner, 1985). It has been observed, however, that focused research into gender differences in L2 motivation does not abound (Sylvén & Thompson, 2015). The three-level structure of the L2MSS allows for consideration of motivation on the basis of individual and societal expectations as well as features of the language classroom. This structure echoes that recommended by Kissau (2006) in considering gender differences in language motivation, yet while there has been a steady flow of research into the L2MSS, only a handful of these studies have investigated gender differences (Azarnoosh & Birjandi, 2012; Henry, 2009).

In the L2MSS-gender studies that have been published, findings have focused most closely on the concept of the Ideal L2 Self, in all cases suggesting that this element of the L2MSS is stronger among female language learners (Azarnoosh & Birjandi, 2012; Henry 2009; Henry & Cliffordson, 2013; Ryan, 2009). Both Ryan (2009), and Azarnoosh and Birjandi (2012) found a correlation between this element and intended learning effort, which could suggest a positive influence of a well-developed Ideal L2 Self on language performance. Henry's (2009) longitudinal study suggested that the difference in the strength of the Ideal L2 Self among males and females may increase over time, while also highlighting a higher attrition rate for languages among boys. A later study by Henry and Cliffordson (2013), suggested that girls were more concerned with interpersonal relationships, which was hypothesised might indicate a stronger presence of the Ideal L2/L3 self. This hypothesis was borne out in terms of the Ideal L3 self, which was significantly stronger among girls, although no significant difference was observed for the Ideal L2 Self.

The Ought-to L2 Self and the L2 Learning Experience have received relatively less attention than the Ideal L2 Self in published gender-motivation studies. Indeed, doubts regarding the reliability of the Ought-to L2 Self construct across different contexts have led it to be underplayed in some studies (e.g., Ryan, 2009; Sylvén & Thompson, 2015). In the Iranian context, Azarnoosh and Birjandi (2012) found boys to have a stronger Ought-to L2 Self than girls. They posited that this could be associated with the Iranian context, in which boys are under pressure to perform academically, although similar findings from Spain (Heras & Lasagabaster, 2014) may suggest instead that the concept may well be applicable across different social contexts. Similarly, there has been little attention for the L2 Learning Experience in gender studies. In the one published example, Azarnoosh and Birjandi (2012) observed no significant differences between genders with regard to attitudes towards

the experience of learning English.[1] It may be relevant to note, however, that they found this element to be the strongest predictor of intended effort for both genders, highlighting the importance of the L2 Learning Experience within the L2MSS. This gives weight to the argument for investigating specific teaching and learning contexts and approaches such as CLIL in relation to motivation.

3.3   Gender, CLIL, and motivation

Roquet, Llopis, and Pérez-Vidal (2015) argue that boys in school need to be sufficiently aware of the practical application of language learning in order to be motivated for it. According to Lasagabaster (2008), using a language to access subject content and using subject content to apply language learned can increase the perceived relevance and authenticity of the language learning process (Coyle et al., 2010) and might therefore make CLIL particularly effective in raising the motivation of boys.

The evidence with regard to this view has been mixed. In Finland, the CLIL group studied by Merisuo-Storm (2007) generally displayed more positive attitudes towards foreign language (FL) learning, with a narrower gap between boys' and girls' attitudes in CLIL than in non-CLIL. As with much CLIL research, however, it was not clear whether these differences were a result of CLIL or whether they simply reflected the characteristics of the kinds of learners who opted into the CLIL programme.

Doiz, Lasagabaster, and Sierra (2014) attempted to counteract the effect of (self-) selectivity by comparing year-groups (i.e., grade levels) across cross-sectional data. They found CLIL learners to be generally more motivated, although they observed no differences across year-groups. With a similar goal in mind, Rumlich (2014, 2016) took a baseline measurement from learners at the beginning of their CLIL experience in his study of affective factors in CLIL and non-CLIL in German *Gymnasien* (upper secondary schools). His initial findings showed CLIL boys to be significantly more positively disposed than non-CLIL boys. Later measurements, however, indicated little change in interest, suggesting that the narrower gender gap preceded choice/preference for and exposure to CLIL. This finding can be equated with findings from other CLIL research, albeit without a gender focus, which have shown significant differences between learners from the beginning of their BE programme (Mearns, 2015; Sylvén & Thompson, 2015). Two examples of

---

1.   In Azarnoosh and Birjandi (2012), this is referred to as 'Attitudes towards learning English', although the content of this variable is related to the L2 Learning Experience and not to underlying language attitudes.

gender-focused CLIL research in which the L2MSS has formed the theoretical basis are Heras and Lasagabaster (2014) and Sylvén and Thompson (2015). The former was a study in itself, with no baseline measurement to identify whether differences preceded exposure to CLIL, while the latter was drawn from baseline data for a larger longitudinal study.

As in the studies cited above, Heras and Lasagabaster (2014), on the one hand, observed that girls in CLIL scored higher on the Ideal L2 Self. Sylvén and Thompson's (2015) findings, on the other hand, highlighted no interactions between gender and CLIL status in this area. While the Ought-to L2 Self was not included in Sylvén and Thompson's study, Heras and Lasagabaster, like Azarnoosh and Birjandi (2012) found that this element of the L2MSS was stronger among boys. The gap between boys' and girls' motivation was somewhat narrower in the CLIL group, however, neither study identified the CLIL learners as being overall significantly more motivated than the non-CLIL learners.

4.   Research questions

In light of a potentially differential gender effect of L2MSS in CLIL context, it is hoped that the current study will provide new insights, with regard to the following research questions (RQs):

1.   Is a gender gap in motivation for language learning less prominent in bilingual education than in mainstream Dutch-medium education?
2.   Does the gender gap in motivation for language learning appear to differ related to the length of exposure to bilingual education?

Due to the evidence regarding language learning motivation, theoretical assumptions in the CLIL literature, and the research findings of Merisuo-Storm (2007) and Rumlich (2014), it was hypothesised in response to the first RQ that the difference in motivation between girls and boys would be smaller in BE than in ME. Given the findings of Mearns' original study (Mearns, 2015; Mearns, de Graaff & Coyle, 2017) and Rumlich's (2016) later research, however, it seemed likely that patterns in attitudinal and motivational differences between boys and girls in BE and ME would not be significantly different in older and younger years (RQ2).

5.   Method

The larger study from which the current data were drawn was a mixed-methods study with a both cross-sectional and longitudinal data collection (Mearns, 2015; Mearns, de Graaff and Coyle, 2017). For the purpose of the current publication, cross-sectional data were analysed quantitatively with Education Type, Gender, and Year as independent variables. The research process is described below.

5.1   Participants

The data analysed here pertain to 581 learners (aged 11–15) in the first three years of general secondary education (HAVO) across four schools in different regions of the Netherlands. Of those learners, 234 were following BE, entered at the beginning of their secondary education, and 347 were following ME at the same schools. As data were collected early in the academic year, first-year BE participants had only recently entered the bilingual stream. In this sense, their responses can be considered baseline data, as learners had at that time had little experience with CLIL. Distribution across the different Education Types, Genders, and Years is displayed in Table 1. Data were collected anonymously and in accordance with the schools' regulations on (passive) consent and participation in research.

Table 1. Participant numbers per Education Type, Year and Gender

	ME				BE			
	1	2	3	Total	1	2	3	Total
Boys	46	49	65	160	42	23	39	104
Girls	64	51	72	187	48	29	53	130
Total	110	100	137	347	90	52	92	234

It seemed appropriate to focus the research on general secondary education (HAVO) rather than the pre-academic stream (VWO) due to due to teachers' common perception of learners in this stream as being difficult to motivate. Also, the academic level of HAVO, as the middle stream, may be more comparable to students learning in other contexts internationally. Furthermore, research into BE at this level is scarcer than in VWO, for which there is more extensive BE provision.

Due to the existence of the Bilingual Education Standard, it can be assumed that participants' experience of BE was relatively uniform. For HAVO, the Standard stipulates that learners in the first three years should receive a minimum of 50% of their lessons across a range of subject disciplines (including sciences, social sciences, and creative and physical subjects) in English and that they should achieve at least level

B1 on the Common European Framework of Reference (CEFR)[2] for English by the end of their third year. In addition, teachers in BE are expected to have at least B2 level English and to have sound knowledge of CLIL (EP-Nuffic, 2016).

5.2   Questionnaire design

The variables included in the gender analysis were drawn from data originally collected and analysed on the basis of individual scaled items and qualitative open questions for the purpose of a larger study (Mearns, 2015). In that study, comparisons were drawn between BE and ME, across first, second, and third year, and with a second data collection period at the end of the school year. Gender did not form part of the original analysis, although this variable had been included in the demographic data collected.

Theoretical basis for the original questionnaire design was drawn from previous studies using the L2MSS, specifically Ryan (2009) and Csizér and Kormos (2009). Additional items specific to the context and purposes of the research were added to those adapted from previous studies, as described in full in Mearns, de Graaff, and Coyle (2017) and Mearns (2015). These additional items were based largely on preliminary qualitative data collected as part of a participatory research project in one of the participant schools. For a full account of this process and its pitfalls, see Mearns, Coyle, and de Graaff (2014).

In the current analysis, the original items were thematically clustered into ten theoretically-grounded scales, six of which were selected for the purpose of the current analysis. These six variables are displayed in Table 2, alongside reliability scores (ȧ), Shapiro Wilk statistics (Ws) and English translations of the specific items of which they consisted.

The six scales relate to attitudes towards the learning of English and foreign languages (1 & 2), to elements of the Ideal and Ought-to L2 Self (3 & 4), and to aspects of the L2 Learning Experience (5 & 6). Of the six scales, five had a Cronbach's alpha rating of 0.7 or greater, and the sixth scored 0.61. According to Muijs (2011), this can be considered an acceptable level of reliability. The Shapiro-Wilk test confirmed that normality could reasonably be assumed for all variables ($Ws \geq 0.95$). Visual inspection of histograms confirmed this.

---

2.   The CEFR is the primary framework used in Europe to describe levels of foreign language proficiency. The 6 CEFR levels range from A1 (basic user) to C2 (proficient user). The CEFR is available at <https://www.coe.int/en/web/common-european-framework-reference-languages/>

Table 2. Scale variables, including original items and Cronbach's alpha reliability scores (ά)

Category/Factor	Items	ά	Ws
1. Attitude to English (5 items)	– I find the English language useful – I find the English language important – I find the English language boring* – My friends think it's useful to learn English – I need to do well in English because it's useful for communicating with different people	.74	.95
2. Attitude to Foreign Languages (7 items)	– I find languages boring* – I find languages useful – I find languages important – It's important to learn different languages because it's fun – It's important to learn different languages if you want to travel to countries where those languages are spoken – I find languages difficult* – If I make a mistake when speaking another language I try to correct the mistake the next time	.81	.98
3. Vision of Future Self (5 items)	– When I think of myself in 10 years, I think of someone who can speak good English – When I think of myself in 10 years, I think of someone who travels a lot – When I think of myself in 10 years, I think of someone with a good job – When I think of myself in 10 years, I think of someone with friends all over the world – When I think of myself in 10 years, I think of someone successful	.80	.98
4. Family Attitude to English (4 items)	– My family says that English is important for my future – My family would be disappointed if I failed English – My family thinks English is more important than other subjects – I need to do well in English because my family finds it important	.71	.99
5. English Lessons (7 items)	– My English teacher makes learning English fun – I learn a lot during English lessons – My English teacher varies his/her lessons – My English teacher thinks I work hard for English – My English teacher seems to enjoy his/her subject – I find English lessons useful – I find English lessons challenging	.84	.96
6. Extramural English (4 items)	– I watch TV in English outside of lessons – I use English for gaming – I use English for social networking – I seek opportunities to speak English outside of lessons	.61	.99

* Negatively-worded items were recoded before being included in the scale.

Variables 1 and 2 (*Attitude to English* and *Attitude to Foreign Languages*[3]) were included due to the presence of language attitudes as components of motivation in Ryan's (2009) MFQ and evidence to suggest that they can influence motivation and the Ideal L2 Self (Kormos et al., 2011). English and FLs in general were considered in separate variables in light of findings elsewhere that suggest that attitudes to the L2 and L3 in CLIL contexts may differ in either a positive or a negative sense (Elzenga & de Graaff, 2015).

Variables 3–6 relate to elements of the L2MSS. The items categorised under *Vision of Future Self* (3) were based on similar items from Taguchi et al. (2009) and Ryan (2009) and reflect aspects of the Ideal L2 Self. *Family Attitude to English* (4) is related to the Ought-to L2 Self in the sense of family's valuing or otherwise of English exerting a sense of pressure to perform. While they had been included in the original questionnaire, items regarding the influence of friends on the Ought-to L2 Self were omitted from the current analysis as they could not be combined into a statistically reliable factor. The learners involved in the design of the original questionnaire did not overtly recognise the role of friends in their own motivation (Mearns, 2015) and the results for these items were inconsistent, which may have contributed to this outcome.

English Lessons and *Extramural English* (5 & 6) relate to the L2 Learning Experience component of the L2MSS. The decision to include two separate elements here was based on the belief that the English learning experience extends beyond the boundaries of school. Extramural English is defined by Sylvén and Sundqvist (2012, p. 113) as "any type of contact that learners have with English outside the classroom." This is especially relevant given the high level of everyday exposure to English in Dutch society, which has been related to improved attainment in BE English (Verspoor et al., 2010), although it has been identified as a challenge to CLIL in Sweden (Sylvén, 2013).

5.3   Data collection

Respondents completed the online questionnaire in school in early October 2012. They were supervised by one of their teachers, who had received detailed written instructions from the researcher regarding the importance of participants' right to withdraw from the research at any time and of the questionnaire being filled out honestly, anonymously, and independently. The supervising teachers were also

---

3.   In Dutch secondary education, this usually refers to German and French as L3. These languages are much less prominent in Dutch society and education than English, and may therefore be considered less instrumental and less relevant by students.

asked to read out similar information to the respondents and to keep their distance during completion of the questionnaire to prevent students from feeling scrutinised. The questionnaire and instructions to respondents were in Dutch.

Although all scaled items in the questionnaire were compulsory fields, internet connectivity problems led to a small number of incomplete responses. In the majority of these cases, a second, successful attempt had been made, and duplicate (in two cases, triplicate) responses were excluded. Two incomplete responses have been included in the current study, as can be observed in the results section. A repeated measurement was carried out in April/May 2013, but the lower number of responses in the second data collection period made analysis of gender sub-groups unreliable. The second measurement has therefore not been included in the current analysis.

5.4　Data analysis

Data were analysed in SPSS Version 23. Each of the six variables was subjected to factorial analysis of variance (ANOVA) to identify main effects from Education Type (mainstream or bilingual), Gender and Year-group (first, second, or third), and interactions between them. Residual analyses were performed to confirm that the assumptions of ANOVA were met. Cook's distance was not larger than 0.3 for any of the variables, suggesting that outliers would not affect the results of analysis (Cook & Weisberg, 1982). Scatterplots of predicted values and standardised residuals showed that the data met the assumptions of homogeneity of variance and linearity. The Shapiro Wilk statistics for residuals (Ws ≥ 0.98) and visual inspection of histograms confirmed normal distribution.

Partial Eta squared (η_p^2) was used as a measure of effect size for the factorial ANOVAs, within the parameters advised by Cohen (1988) (small 0.01–0.059; medium 0.06–0.13; large > 0.14). Where factorial ANOVAs revealed significant interactions, these were further investigated using independent samples t-tests for the variables with two categories (Gender and Education Type) and ANOVAs for Year-group, which had three categories. Cohen's d and Omega squared (ω^2) were employed as measures of effect size for the t-tests and the one-way ANOVAs respectively, within the parameters (small 0.10–0.29; medium 0.30–0.49; large > 0.50) advised by Field (2013).

Interactions between Gender and Education Type were used to respond to RQ1 and interactions between Gender, Education Type, and Year-group for RQ2. The findings of these analyses will be presented in the following section.

6. Findings

6.1   Gender gap in bilingual and mainstream education

The findings with regard to gender patterns in BE and ME will be presented first for the variables where no significant interaction between Education Type and Gender was observed, and subsequently in terms of significant interactions. Descriptive statistics by Year, Gender, and Education Type are presented in Table 3, with significant interactions between Gender and ET shaded grey.

Table 3. Descriptive statistics by Year, Gender and Education Type (ME/BE)

Dependent variable	Year	Gender	ME			BE		
			M	SD	N	M	SD	N
1. Attitude to English	1	boys	3.73	0.76	46	4.34	0.45	42
		girls	4.18	0.50	64	4.17	0.45	48
	2	boys	3.64	0.68	49	4.19	0.39	23
		girls	3.72	0.71	51	4.15	0.42	29
	3	boys	3.79	0.71	65	4.04	0.61	39
		girls	4.08	0.60	72	4.20	0.48	53
	Total	boys	3.73	0.71	160	4.19	0.52	104
		girls	4.02	0.63	187	4.18	0.45	130
2. Attitude to Foreign Languages	1	boys	3.53	0.62	46	3.96	0.56	42
		girls	3.98	0.54	64	3.88	0.52	48
	2	boys	3.27	0.65	49	3.75	0.56	23
		girls	3.68	0.63	51	3.91	0.81	29
	3	boys	3.21	0.68	65	3.55	0.75	39
		girls	3.73	0.64	72	3.98	0.64	53
	Total	boys	3.32	0.66	160	3.76	0.66	104
		girls	3.80	0.61	187	3.93	0.64	130
3. Vision of Future Self	1	boys	3.23	0.73	46	3.98	0.74	42
		girls	3.35	0.84	64	3.63	0.73	47[†]
	2	boys	3.44	0.87	49	3.91	0.63	23
		girls	3.43	0.71	51	3.68	0.73	29
	3	boys	3.62	0.60	64[†]	4.01	0.57	39
		girls	3.69	0.81	72	3.99	0.71	53
	Total	boys	3.45	0.74	159[†]	3.97	0.65	104
		girls	3.50	0.81	187	3.79	0.73	129[†]

Table 3. (*continued*)

Dependent variable	Year	Gender	ME			BE		
			M	SD	N	M	SD	N
4. Family Attitude to English	1	boys	3.21	0.82	46	3.19	0.95	42
		girls	3.22	0.79	64	3.09	0.82	48
	2	boys	3.43	0.82	49	3.39	0.93	23
		girls	3.18	0.81	51	2.99	0.93	29
	3	boys	3.01	0.85	65	3.39	0.71	39
		girls	3.16	0.91	72	3.11	0.88	53
	Total	boys	3.20	0.84	160	3.31	0.86	104
		girls	3.19	0.84	187	3.08	0.87	130
5. English Lessons	1	boys	3.51	0.62	46	4.01	0.45	42
		girls	3.82	0.55	64	3.87	0.58	48
	2	boys	2.88	1.05	49	3.78	0.37	23
		girls	3.08	0.90	51	3.78	0.53	29
	3	boys	3.27	0.80	64[†]	3.15	0.91	39
		girls	3.59	0.73	72	3.14	0.76	53
	Total	boys	3.22	0.87	159[†]	3.64	0.75	104
		girls	3.53	0.78	187	3.55	0.73	130
6. Extramural English	1	boys	2.76	0.96	46	3.35	0.77	42
		girls	2.57	0.70	64	3.00	0.79	48
	2	boys	2.83	0.89	49	3.61	0.73	23
		girls	2.61	0.83	51	2.89	0.90	29
	3	boys	3.21	0.84	65	3.47	0.71	39
		girls	2.76	0.83	72	3.14	0.84	53
	Total	boys	2.97	0.91	160	3.45	0.74	104
		girls	2.65	0.79	187	3.03	0.83	130

[†] 1 case missing

A significant main effect of Education Type was observed for Variable 6, *Extramural English* ($F(1, 569) = 39.53, p < .001, \eta_p^2 = .065$), with a medium effect size. The mean score in BE ($M = 4.00, SD = .63$) was significantly higher than in ME ($M = 3.88, SD = .68$). A small significant main effect of Gender was also observed for this variable ($F(1, 569) = 27.66, p < .001, \eta_p^2 = .046$). Here, the boys' mean score ($M = 3.91, SD = .68$) was lower than that of girls ($M = 4.08, SD = .57$).

No interactions between Education Type and Gender were observed for *Extramural English* (6). There were no significant main or interaction effects for *Family*

Attitude to English (4). Interaction effects between Gender and Education Type were observed for the remaining variables, as explained below.

For Variable 1, *Attitude to English*, small to medium significant main effects of Education Type (F (1, 569) = 39.73, $p < .001$, $\eta_p^2 = .065$) and Gender (F (1, 569) = 6.19, $p = .013$, $\eta_p^2 = .011$) were observed. There was also a small significant interaction between Gender and Education Type (F (1, 569) = 7.87, $p = .005$, $\eta_p^2 = .014$). Subsequent independent samples t-tests within each Education Type revealed that ME girls' average score ($M = 4.02$, $SE = .05$) was higher than that of ME boys ($M = 3.73$, $SE = .06$). This difference was significant and had a medium effect size: $t(345) = -4.02$, $p < .001$, $d = 0.43$. In BE, boys ($M = 4.19$, $SE = .05$) scored marginally higher than girls ($M = 4.18$, $SE = 0.04$). This difference was not significant: $t(232) = .24$, $p = .800$, $d = 0.03$.

For Variable 2, *Attitude to Foreign Languages*, there was a medium-sized main effect of Education Type (F (1, 569) = 39.73, $p < .001$, $\eta_p^2 = .065$) and a small significant main effect of Gender (F (1, 569) = 32.56, $p < .001$, $\eta_p^2 = .054$). A small significant interaction between Gender and Education Type (F (1, 569) = 6.92, $p = .009$, $\eta_p^2 = .012$) was also observed. Independent samples t-tests revealed that ME girls ($M = 3.80$, $SE = .05$) had a higher average mean score than ME boys ($M = 3.32$, $SE = .05$). The difference was significant and had a large effect size: $t(345) = -7.01$, $p < .001$, $d = .75$. BE girls ($M = 3.93$, $SE = .06$) scored higher on average for this variable than BE boys ($M = 3.76$, $SE = .06$), although this difference was again nonsignificant: $t(232) = -1.96$, $p = 0.051$, $d = 0.26$.

For Variable 3, *Vision of Future Self*, there was a medium-sized main effect of Education Type (F (1, 567) = 39.95, $p < .001$, $\eta_p^2 = .066$), although no main effect of Gender. There was a small significant interaction between Gender and Education Type (F (1, 567) = 4.01, $p = .007$, $\eta_p^2 = .046$). The t-test showed that ME girls' ($M = 3.50$, $SE = .06$) mean score was marginally higher than that of ME boys ($M = 3.45$, $SE = .06$), although this difference was nonsignificant: $t(344) = -0.57$, $p = 0.567$, $d = 0.06$. For the same variable, BE boys ($M = 3.97$, $SE = .06$) scored higher than BE girls ($M = 3.79$, $SE = .06$). This difference was significant, although with a small effect size: $t(231) = 2.03$, $p = .043$, $d = .27$.

For Variable 5, *English Lessons*, there was a small significant main effect for Education Type (F (1, 568) = 16.77, $p < .001$, $\eta_p^2 = .029$) but not for Gender. There was a small significant interaction between Gender and Education Type (F (1, 568) = 6.56, $p = .011$, $\eta_p^2 = .011$). According to the subsequent t-test, ME girls ($M = 3.53$, $SE = .06$) scored higher than ME boys ($M = 3.22$, $SE = .07$). This difference was significant, with a medium effect size: $t(344) = -3.49$, $p = .001$, $d = .37$. In BE, boys' mean score ($M = 3.64$, $SE = .07$) was higher than that of girls ($M = 3.55$, $SE = .06$), although this difference was nonsignificant: $t(232) = 0.89$, $p = 0.376$, $d = 0.12$.

6.2   Motivational differences across Years in BE and ME

Factorial ANOVAs were carried out on the basis of Gender, Year, and Education Type. No significant interactions were observed between Year-group and Gender, nor between Education Type, Year-group, and Gender, for any of the six variables analysed. Significant main effects of Year were observed for all variables with the exception of variable 4, *Family Attitude to English,* as described below.

For (1) *Attitude to English* (F (2, 569) = 3.66, p = .026, η_p^2 = .013), the main effect of Year was small. The Scheffé posthoc test revealed that second years scored significantly lower than either first years (p < .001) or third years (p = .020), while the difference between first and third years was nonsignificant.

There was a small significant main effect of Year for (2) *Attitude to Foreign Languages* (F (2, 569) = 6.84, p < .001, η_p^2 = .023). For this variable, first year scored significantly higher than second (p = .002) or third year (p = .001), while the difference between second and third year was nonsignificant.

Small to medium significant main effects of Year were observed for (3) *Vision of Future Self* (F (2, 567) = 7.94, p < .001, η_p^2 = .027) and (6) *Extramural English* (F (2, 568) = 27.35, p < .001, η_p^2 = .088). For both of these variables, third year scored significantly higher than first (3: p = .001; 6: p = .008) or second years (3: p = .024; 6: p = .050), while the difference between first and second years was nonsignificant.

A medium significant main effect of Year was observed for (5) *English Lessons* (F (2, 568) = 27.35, p < .001, η_p^2 = .088). Here, there was also a medium significant interaction between Education Type and Year-group (F (2, 568) = 23.39, p < .001, η_p^2 = .076). Additional one-way ANOVA within each Education Type revealed significant differences between Year-groups in both ME (F (2, 343) = 21.34, p < .001, ω^2 = .32) and BE (F (2, 241) = 36.531, p < .001, ω^2 = .48), with medium effect sizes. Scheffé posthoc analysis highlighted that ME first years scored significantly higher than both second year (p < .001) and third year (p = .048). Second year ME also scored significantly lower than third year ME in this respect (p < .001). Within BE, posthoc analysis highlighted a significantly higher mean score among third year than either first year (p < .001) or second year (p < .001). The difference between BE first year and BE second year in this respect was nonsignificant. Significant interactions between Education Type and Year were not observed for variables 1, 2, 3, 4, or 6.

7.   Discussion

The findings presented above will now be discussed on the basis of each Research Question in turn.

7.1   RQ1. Is a gender gap in motivation towards language learning less prominent in bilingual education than in mainstream Dutch-medium education?

While previous research has shown that females are generally more positively disposed towards language learning than males (Henry, 2009), literature on CLIL indicates a belief that the more practical nature of the approach will help to motivate boys and to narrow this gender gap (Marsh, 2002). In some respects, the findings reported above support this hypothesis. Girls in the non-CLIL ME stream displayed significantly more motivation than boys with regard to the learning of English (1) and of foreign languages (2), and to their in-school English learning experience (5), while in BE the gap between girls and boys in these respects was not significant.

A more surprising finding was encountered with regard to learners' *Vision of Future Self* (3), where BE boys displayed the highest level of motivation of all four groups, with a significantly larger gender gap in BE than in ME. This is surprising in the light of the findings of Ryan (2009), Henry (2009), Azarnoosh and Birjandi (2012), and Heras and Lasagabaster (2014), in which girls appeared to have a more powerful sense of the Ideal L2 Self, and of Sylvén and Thompson (2015), in which no significant differences between genders were observed in this respect. The stronger vision of future L2 self among BE boys in the current study might reflect Baker and MacIntyre's (2000) finding in an immersion education context that boys had a higher level of job orientation than girls. This could be supported by further data from pupils in the current study, as reported in Mearns, de Graaff, and Coyle (2017), which suggested that employment opportunities were an important consideration in the decision to follow BE. In terms of Lasagabaster's (2008) view that the more authentic and practical nature of CLIL might support boys' motivation in particular, this raises the question of whether a CLIL approach to language learning might further enhance the image of the Ideal L2 Self among boys.

A similar trend can be observed with regard to *Attitudes to English* (1) and *English lessons* (5). Again, CLIL boys responded more positively in these respects than CLIL girls, although the difference was not statistically significant. While confirming the gender gap hypothesis, this to some extent contradicts findings by Merisuo-Storm (2007) and Rumlich (2014), who reported that CLIL girls displayed the most positive attitudes towards language learning. In both of these previous

studies, even where the gender gap was narrow, it was not inverted. The positivity of BE boys regarding English, their future L2 self, and the L2 Learning Experience in school could further support the argument that CLIL can motivate boys (Marsh, 2002). As these data do not show causality, however, it is also possible that they simply highlight inherent enthusiasm for English, which may have influenced boys' choice to follow a bilingual programme (Bruton, 2011).

The strength of BE boys' response regarding the English language was not reflected in their attitude towards foreign languages in general (2). In spite of the narrower (nonsignificant) gap in BE when compared to ME, BE boys were nevertheless markedly less positive than in their attitude towards English, their responses resembling those of ME girls more than those of BE girls. In the studies already cited, CLIL girls' attitudes were the most positive but were closely followed by those of CLIL boys, whereas in the current study both groups of girls expressed more positive attitudes towards foreign languages than either BE or ME boys. Here, the change in trend rather reflects the findings of Henry (2009) and Henry and Cliffordson (2013), who observed more positivity towards languages other than English and stronger Ideal L3 selves among girls. In this sense, it could be argued that while the language motivation gender gap may be narrower within BE than in ME, it is nevertheless still present when 'languages' are considered as going beyond English.

There was no evidence that responses from boys and girls to *Family Attitude to English* (4) or *Extramural English* (6) followed different patterns within BE and ME. With regard to families' attitudes (4), this may be a reflection of the unreliability of the Ought-to L2 Self as a predictor of motivation among adolescents, as has been reported elsewhere (Csizér & Kormos, 2009; Ryan, 2009). This doubt led this element to be omitted from Sylvén and Thompson's (2015) CLIL study, although Heras and Lasagabaster (2014) found it to be stronger among boys. As the variable employed here was limited to the influence of family, however, this interpretation is drawn cautiously. Concerning the role of contact with English outside of school (6), further examination of the responses suggests the large numbers of boys, especially in BE, who reported playing computer games in English outweighed differences between the two educational groups.

7.2   RQ2. Does the gender gap in motivation for language learning appear to differ related to the length of exposure to BE?

There is evidence from the findings discussed above of a difference in the motivations of BE and ME boys and girls, at times contradicting the trend in the literature for girls to be the more motivated language learners (Henry, 2009), even in CLIL

contexts (Heras & Lasagabaster, 2014; Merisuo-Storm, 2007). What is not apparent from these findings, however, is whether the level of motivation among boys was even higher among those who had been exposed to BE for longer.

There were no significant interactions between Gender and Year, nor between Gender, Year and Education Type for any of the factors in this study. This reflects the findings of Doiz et al. (2014), who observed no significant gender differences across year-groups, and provides no evidence of a differential effect of BE on either boys' or girls' motivations.

As already established, the theoretical literature on CLIL would suggest that the more practical and authentic nature of CLIL approaches could make it an effective means of increasing language motivation among boys (Lasagabaster, 2008; Sylvén & Sundqvist, 2012), or at least of closing the motivational gender gap observed by Henry (2009). There is, however, increasing evidence that the (self-) selective nature of most CLIL programmes results in different types of learners following different educational streams and therefore complicates the measurement of value-added gains in motivation (Mearns, 2015; Mearns, de Graaff & Coyle, 2017; Rumlich, 2016; Sieben & van Ginderen, 2014; Sylvén & Thompson, 2015). It appears from these findings that this may also be the case in the current context, where it seems that motivational differences might exist independently, i.e., related to the choice for BE and not to the BE learning experience.

8.   Limitations and looking to the future

8.1   Methodological limitations

This study was not without its limitations in scale, scope, and execution. The motivation questionnaire drew elements from designs tested elsewhere (specifically Csizér & Kormos, 2009 and Ryan, 2009) and was designed with contextualised learner input in mind (Mearns, Coyle & de Graaff, 2014) but it had not been statistically validated as part of the original design. While the variables created for the current analysis were theoretically grounded, a questionnaire designed with specific factors in mind might provide a more empirically solid analysis. It would therefore be valuable to confirm the findings of this study through future research into gender differences in CLIL and non-CLIL settings, either using the same instrument for validation purposes or one previously validated in a different context. A specific improvement that could be made in a future version of this instrument would be to consider aspects of the Ideal and Ought-to L2 Self more broadly, as has been the case in earlier studies.

Another limitation of the data presented here is that they are cross-sectional and drawn from a single quantitative instrument. Motivation being increasingly regarded as multifaceted and dynamic, it might be more suited to a longitudinal approach that takes different perspectives into account as in, for example, complex dynamic systems (Larsen-Freeman, 2012). This approach could include focus not only on learners but also on the effects of actual classroom practices on BE learner motivation.

8.2   Areas for future research

In terms of scale and scope, it would be interesting for future studies to include learners from different educational tracks (pre-academic VWO and/or pre-vocational VMBO) in order to establish whether the conclusions drawn here apply particularly to general (HAVO) education. Given the emphasis on attitudes and motivation prior to embarking on BE and the recent emphasis on background and family influence (Sieben & van Ginderen, 2014), it might also be useful to include more survey questions pertaining to sociocultural status, as in Lasagabaster (2008). Furthermore, in the light of the differences observed between attitudes towards English and towards other foreign languages, future research might incorporate this as a more central focus, as in Elzenga and de Graaff (2015).

For research in a different direction, an area that received little attention in this account was the result for Extramural English. The lack of interaction between Education Type and Gender in this regard appeared to result from the higher levels of gaming reported by boys in both BE and ME, which overpowered the other results, although it was interesting to observe that gaming was popular among BE boys in particular. Sylvén and Sundqvist (2012) highlighted motivation and authenticity in their consideration of the overlap between CLIL and collaborative computer games as vehicles for language learning. The current findings raise the question of whether there may be a connection between gaming and the L2MSS, although this was beyond the scope of this study.

9.   Conclusions

This study sought to address the questions of whether boys and girls in English-Dutch bilingual (BE) and Dutch-medium mainstream (ME) education appeared to possess different traits in terms of their motivation as language learners, and whether that motivation appeared different between pupils at the beginning of BE/ME secondary education compared to in the second or third year.

For both attitudes to English and response to English lessons, the gap between boys' and girls' motivations did appear smaller in BE, although responses did not follow the expected gender pattern. While girls were consistently the more motivated group in ME, for these areas BE boys displayed more motivation than BE girls. This was also the case for the vision of the future L2 self, where the gender gap in BE was actually larger than in ME. In isolation, this might appear to suggest that BE had a positive effect on boys' motivation for English, to the extent that, in some respects, the gender gap was not only reduced, but inverted.

Considered in relation to the interactions between Gender, Year, and Education Type, however, this conclusion becomes less likely. There was no significant difference between the motivations of BE or ME boys and girls in the three separate year-groups for any of the areas investigated. Rather than suggesting a motivating effect from BE, therefore, this finding might imply not only that learners who choose (or are selected) to follow a bilingual stream are naturally more motivated from the outset, but that BE boys in particular possess high levels of motivation on entering secondary bilingual education. The challenge for schools may therefore lie not in sparking motivation among BE boys, but in maintaining existing high levels of motivation for English.

In contrast, conclusions with regard to boys' attitudes towards foreign languages in general followed a different trend. Here, BE girls were the more motivated group, followed by ME girls, although the gap between girls and boys was narrower in BE than in ME. In this regard it might be concluded that higher levels of motivation for English among BE boys are not an indication of higher levels of language motivation in general. This could be a valuable insight for BE schools where, according to the standard for Dutch bilingual education (de Graaff & van Wilgenburg, 2015), not only the English language but also the learning of other languages and broader international orientation should be at the heart of the BE programme.

Acknowledgements

References

Azarnoosh, M., & Birjandi, P. (2012). Junior high school students' L2 Motivational Self System: Any gender differences? *World Applied Sciences Journal*, *20*(4), 577–584.

Baker, S. C., & MacIntyre, P. D. (2000). The role of gender and immersion in communication and second language orientations. *Language Learning*, *50*(2), 311–341. doi:10.1111/0023-8333.00224

Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, *39*(4), 523–532. doi:10.1016/j.system.2011.08.002

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York, NY: Chapman & Hall.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.

Crookes, G., & Schmidt, R. W. (1991). Motivation: Reopening the research agenda. *Language Learning*, *41*(4), 469–512. doi:10.1111/j.1467-1770.1991.tb00690.x

Czisér, K., & Dörnyei, Z. (2005). The internal structure of language learning motivation and its relationship with language choice and learning effort. *The Modern Language Journal*, *89*(1), 19–36. doi:10.1111/j.0026-7902.2005.00263.x

Csizér, K., & Kormos, J. (2009). Learning experiences, selves and motivated learning behaviour: A comparative analysis of structural models for hungarian secondary and university learners of English. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 98–119). Bristol: Multilingual Matters.

Denman, J., Tanner, R., & de Graaff, R. (2013). CLIL in junior vocational secondary education: challenges and opportunities for teaching and learning. *International Journal of Bilingual Education and Bilingualism*, *16*(3), 285–300. doi:10.1080/13670050.2013.777386

Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014). CLIL and motivation: The effect of individual and contextual variables. *The Language Learning Journal*, *42*(2), 209–224. doi:10.1080/09571736.2014.889508

Dörnyei, Z. (1994). Understanding second language motivation: On with the challenge! *Modern Language Journal*, *79*, 505–518.

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching*, *31*(3), 117–135. doi:10.1017/S026144480001315X

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667343

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dörnyei, Z. (2009). The L2 Motivational Self System. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Bristol: Multilingual Matters.

Dörnyei, Z., & Ottó, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics* (Vol. 4, pp. 43–69). London: Thames Valley University.

Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching: Motivation* (Vol. 2). Harlow: Pearson Education.

Elzenga, A., & de Graaff, R. (2015). Motivatie voor Engels en Frans in het tweetalig onderwijs. Zijn leerlingen in het tweetalig onderwijs gemotiveerdere taalleerders? [Motivation for English and French in bilingual education. Are pupils in bilingual education more motivated language learners?] *Levende Talen Tijdschrift* [*Living Languages Journal*], *16*(4), 16–26.

EP-Nuffic. (2016, 7 September). *Standaarden, visitaties en keurmerken* [*Standards, inpections and hallmarks*]. Retrieved from https://www.epnuffic.nl/voortgezet-onderwijs/ tweetalig-onderwijs/het-netwerk-tto/standaarden-visitaties-en-keurmerken

Field, A. (2013). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. MA: Newbury House.

de Graaff, R., Koopman, G. J., Anikina, Y., & Westhoff, G. (2007). An observation tool for effective L2 pedagogy in content and language integrated learning (CLIL). *International Journal of Bilingual Education and Bilingualism*, *10*(5), 603–624. doi:10.2167/beb462.0

de Graaff, R., & van Wilgenburg, O. (2015). The Netherlands: Quality control as a driving force in bilingual education. In P. Mehisto & F. Genesee (Eds.), *Building bilingual education systems: Forces, mechanisms and counterweights* (pp. 167–180). Cambridge: Cambridge University Press.

Henry, A. (2009). Gender differences in compulsory school pupils' L2 self-concepts: A longitudinal study. *System*, *37*, 177–193. doi:10.1016/j.system.2008.11.003

Henry, A. (2010). Contexts of possibility in simultaneous language learning: Using the L2 Motivational Self System to assess the impact of global English. *Journal of Multilingual and Multicultural Development*, *31*(2), 149–162. doi:10.1080/01434630903471439

Henry, A. (2012). *L3 motivation*. (Unpublished doctoral thesis), University of Gothenburg, Gothenburg Sweden.

Henry, A., & Cliffordson, C. (2013). Motivation, gender, and possible selves. *Language Learning*, *63*(2), 271–295. doi:10.1111/lang.12009

Heras, A., & Lasagabaster, D. (2014). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, *19*(1), 70–88. doi:10.1177/1362168814541736

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*, 319–340. doi:10.1037/0033-295X.94.3.319

van Kampen, E., Admiraal, W., & Berry, A. (2016). Content and language integrated learning in the Netherlands: Teachers' self-reported pedagogical practices. *International Journal of Bilingual Education and Bilingualism*, 1–15. doi:10.1080/13670050.2016.1154004

Kissau, S. (2006). Gender differences in motivation to learn French. *The Canadian Modern Language Review*, *62*(3), 401–422. doi:10.1353/cml.2006.0020

Klaassen, R. (2001). *Effective teacher behaviour in English-language higher education*. (Unpublished doctoral thesis), Technical University of Delft, Delft.

Koopman, G. J., Skeet, J., & de Graaff, R. (2014). Exploring content teachers' knowledge of language pedagogy: A report on a small-scale research project in a Dutch CLIL context. *The Language Learning Journal*, *42*(2), 123–136. doi:10.1080/09571736.2014.889974

Kormos, J., Kiddle, T., & Csizér, K. (2011). Systems of goals, attitudes, and self-related beliefs in second-language-learning motivation. *Applied Linguistics*, *32*(5), 495–516. doi:10.1093/applin/amr019

Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning*, *62*(4), 997–1023. doi:10.1111/j.1467-9922.2012.00719.x

Larsen-Freeman, D. (2012). Complex, dynamic systems: A new transdisciplinary theme for applied linguistics? *Language Teaching*, *45*(02), 202–214. doi:10.1017/S0261444811000061

Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, *1*, 30–41. doi:10.2174/1874913500801010030

Macintyre, P. D., Mackinnon, S. P., & Clément, R. (2009). The baby, the bathwater, and the future of language learning motivation research. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 43–65). Bristol: Multilingual Matters.

Markus, H., & Nurius, P. (1987). Possible selves: The interface between motivation and the self-concept. In K. Yardley & T. Honess (Eds.), *Self and identity: Psychosocial perspectives* (pp. 157–172). Suffolk: John Wiley & Sons.

Marsh, D. (2002). *CLIL/EMILE – The European dimension: Actions, trends and foresight potential.* Retrieved from https://jyx.jyu.fi/dspace/handle/123456789/47616?show=full

Mearns, T. (2015). Chicken, egg or a bit of both? Motivation in bilingual education (TTO) in the Netherlands. (Unpublished doctoral thesis), Aberdeen, University of Aberdeen.

Mearns, T., Coyle, D., & de Graaff, R. (2014). Aspirations and assumptions: A researcher's account of pupil involvement in school-based research. *International Journal of Research & Method in Education*, *37*(4), 442–457. doi:10.1080/1743727X.2014.925440

Mearns, T., de Graaff, R., & Coyle, D. (2017). *Motivation for or from bilingual education? A comparative study of learner views in the Netherlands.* Manuscript submitted for publication.

Merisuo-Storm, T. (2007). Pupils' attitudes towards foreign-language learning and the development of literacy skills in bilingual education. *Teaching and Teacher Education*, *23*(2), 226–235. doi:10.1016/j.tate.2006.04.024

Morton, T., & Llinares, A. (2017). Content and language integrated learning (CLIL): Type of programme or pedagogical model? In A. Llinares & T. Morton (Eds.), *Applied linguistics perspectives on CLIL* (pp. 1–18). Amsterdam: John Benjamins. doi:10.1075/lllt.47.01mor

Muijs, D. (2011). *Doing quantitative research in education with SPSS* (Vol. 2). London: Sage.

Nuffic. (2017, February 21 2017). *Tto-scholen* [*Bilingual schools*]. Retrieved from https://www.nuffic.nl/voortgezet-onderwijs/tweetalig-onderwijs/tto-scholen

Onderwijs in Cijfers [Education in Numbers]. (2016). *Leerlingen in het derde leerjaar voortgezet onderwijs* [*Pupils in 3rd year of secondary education*]. Retrieved from https://www.onderwijsincijfers.nl/kengetallen/voortgezet-onderwijs/deelnemersvo/leerlingen-in-het-derde-leerjaar

Oxford, R., & Shearin, J. (1994). Language learning motivation: Expanding the theoretical framework. *The Modern Language Journal*, *78*(1), 12–28. doi:10.2307/329249

Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism*, *15*(3), 315–341. doi:10.1080/13670050.2011.630064

Roquet, H., Llopis, J., & Pérez-Vidal, C. (2015). Does gender have an impact on the potential benefits learners may achieve in two contexts compared: Formal instruction and formal instruction   content and language integrated learning? *International Journal of Bilingual Education and Bilingualism*, *19*(4), 370–386. doi:10.1080/13670050.2014.992389

Rumlich, D. (2014). Prospective CLIL and non-CLIL students' interest in English (classes): A quasi-experimental study on German sixth-graders. In R. Breeze, C. Martínez Pasamar, & C. Tabernero Sala (Eds.), *Integration of theory and practice in CLIL* (pp. 75–95). Amsterdam: Rodopi.

Rumlich, D. (2016). *Evaluating bilingual education in Germany. CLIL students' general English proficiency, EFL self-concept and interest.* Bern: Peter Lang.

Rutgers, D. (2013). *Effects of bilingual education on metacognitive skills in learning a third language.* (Unpublished doctoral thesis), Cambridge University, Cambridge.

Ryan, S. (2008). *The ideal L2 selves of Japanese learners of English.* (Unpublished doctoral thesis), University of Nottingham, Nottingham.

Ryan, S. (2009). Self and identity in L2 motivation in Japan: The ideal L2 self and Japanese learners of English. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 120–143). Bristol: Multilingual Matters.

Ryan, S., & Dörnyei, Z. (2013). The long-term evolution of language motivation and the L2 self. In A. Berndt (Ed.), *Fremdsprachen in der Perspektive lebenslangen Lernens* [*Foreign languages from the perspective of lifelong learning*] (pp. 89–100). Frankfurt: Peter Lang.

Sieben, I., & van Ginderen, N. (2014). De keuze voor tweetalig onderwijs [The choice for bilingual education]. *Mens en maatschappij* [*People and society*], *89*(3), 233–255. doi:10.5117/MEM2014.3.SIEB

Sylvén, L. K. (2013). CLIL in Sweden – why does it not work? A metaperspective on CLIL across contexts in Europe. *International Journal of Bilingual Education and Bilingualism*, *16*(3), 301–320. doi:10.1080/13670050.2013.777387

Sylvén, L. K., & Sundqvist, P. (2012). Similarities between playing World of Warcraft and CLIL. *Apples – Journal of Applied Language Studies*, *6*(2), 113–130.

Sylvén, L. K., & Thompson, A. S. (2015). Language learning motivation and CLIL: Is there a connection? *Journal of Immersion and Content-Based Language Education*, *3*(1), 28–50. doi:10.1075/jicb.3.1.02syl

Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 66–97). Bristol: Multilingual Matters.

Ushioda, E. (2009). A person-in-context relational view of emergent motivation, self and identity. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 215–228). Bristol: Multilingual Matters.

Verspoor, M., de Bot, K., & van Rein, E. M. J. (2010). Binnen- en buitenschools taalcontact en het leren van Engels [Learning English and contact with the language in and out of school, and learning English]. *Levende Talen Tijdschrift* [*Living Languages Journal*], *11*(4), 14–19.

Verspoor, M., de Bot, K., & Xu, X. (2015). The effects of English bilingual education in the Netherlands. *Journal of Immersion and Content-Based Language Education*, *3*(1), 4–27. doi:10.1075/jicb.3.1.01ver

Verspoor, M., & Edelenbos, P. (2009). Tweetalig onderwijs: Beter geschoolde leerlingen in 2024 [Bilingual education: Better-educated pupils in 2024]. In R. d. Graaff & D. Tuin (Eds.), *De toekomst van het talenonderwijs: Nodig? Anders? Beter?* [*The future of language education: Necessary? Different? Better?*] (pp. 147–163). Utrecht; Enschede: IVLOS, Universiteit Utrecht; NaB-MVT.

Authors' addresses

Tessa Mearns
ICLON Graduate School of Teaching
Leiden University
Kolffpad 1
2333 AL Leiden
The Netherlands
t.l.mearns@iclon.leidenuniv.nl

Rick de Graaff
Department of Languages, Literature & Communication
University of Utrecht
Trans 10, Kamer 1.60
3512 JK Utrecht
The Netherlands
r.degraaff@uu.nl