Teacher-Student Interaction in Bilingual Primary Education

Karlijn Alberts (3973379)
February 2021

Master's Thesis
*English Language and Culture: Education & Communication,
Utrecht University*

Supervisor
Prof. dr. Rick de Graaff

Second Reader
Dr. Trenton Hagar

Abstract

The present study took place in the context of a national pilot study on bilingual primary education (BPE) in the Netherlands (Jenneskins et al., 2020). Interaction plays a crucial role in language acquisition; therefore, this study aimed to outline current practices by teachers and relate this to insights from earlier language acquisition research. Initiation-response-feedback (IRF) patterns were studied following distinctions set out by Wierenga (2014), and the teachers' use of the students first language (L1) Dutch was studied. Data included classroom observations at four *groep-5* teachers (students' age 8-9) from four schools, which were transcribed and coded for analysis. Furthermore, interview data was used to interpret the data. Results indicate that the verbal dominance of teachers exceeded that of students, some teachers mostly asked closed/display questions, and others asked mostly open/reference questions. The types of questions were susceptible to change by context; most teachers asked significantly different questions in different lesson parts, and in a prescribed lesson format compared to the other observed lesson. All teachers used positive feedback more often than other types, and focused most of their feedback on content rather than language form. Two out of four teachers used Dutch during the lesson for didactic purposes, to correct students' behaviour, and due to a limited proficiency in the target language. Native language of the teacher and school type did not influence the results, but views on language acquisition and limited proficiency in the target language seemed to explain some of the variation between teachers in the question types, feedback focus, and use of the L1. Therefore, the suggestion was made to adhere to a minimum proficiency level of CEFR-level B2 (Council of Europe, 2001) in English for teachers in this type of education, and to inform and train teachers based on insights from language acquisition research.

Key words: bilingual primary education, foreign language education, content and language integrated learning (CLIL), interaction, IRF-sequence, verbal dominance, question, feedback, language choice, first language (L1), target language (TL).

Table of Contents

1. Introduction

The positive effects on final attainment in language learning that are often ascribed to an early start, combined with the status of English as an important international language in a globalising world, have led to the desire of many parents for their children to start learning English in school at a young age (Phillipson, 2008). As a result, *early foreign language education* (EFLE) has been steadily growing (Nuffic, 2018). Findings in language acquisition research indicate that under the right circumstances an early start in language learning leads to better final attainment, and bilingualism has a positive influence on cognition (Hummel, 2014). However, when the right conditions are not met, learning a second language early on in life usually does not lead to any of these advantages, and can even have a detrimental effect on the development of the first language (L1) and on cognition (Hummel, 2014, p. 45; Muñoz, 2008; Spada, 2015). Still, the positive findings that have been linked to an early start in language learning and bilingualism have supported the bottom-up educational developments, as an increasing number of countries have started offering *content and language integrated learning* (CLIL) in primary education (Van den Broek, De Graaff, Unsworth, & Van der Zee, 2014). Due to the relative novelty of this form of education, knowledge about factors that positively influence learning outcomes are presently limited. In the Netherlands, a longitudinal pilot study into the application and results of bilingual primary education has started in 2014, drawing the participation of 20 primary schools over the country; twelve of these schools participate in bi-yearly measurements (De Graaff & Costache, 2020; Jenniskens et al., 2018; Jenniskens et al., 2020; Van den Broek et al., 2014). The present study takes place in the context of this larger pilot study, and aims to identify characteristics of teachers' practices that might influence pupils' acquisition of the target language (TL). As pointed out by several earlier studies, interaction is a vital part of language learning, both in naturalistic contexts and in the classroom context (Damhuis, 2008). For instance, the verbal dominance of teachers, the types of questions they ask, the feedback they give (Wierenga, 2014), and the role the learners' first language (L1) plays in the classroom can have a significant impact on the language acquisition process and results of their students (Haijma, 2013). Practices in this territory in bilingual primary education have as yet remained unexplored; therefore, the present study attempts to outline current practices by teachers and link these to insights from language acquisition research. Specifically, student-teacher interaction is studied, by analysing the initiation-response-feedback (IRF) sequences and other whole-group teacher talk during CLIL lessons taught by four teachers with varying language

backgrounds in Dutch bilingual primary schools. Components are broken down into teachers' and students' verbal dominance in terms of number of spoken words and number of initiations, the types of questions that are asked by teachers in these initiations, the types of feedback that is given by teachers and whether it is focused on language or on content, and the frequency and context of use of the students' L1 by teachers. Furthermore, the results are compared to the teachers' views on important factors in their pupils' language acquisition, as expressed in interviews following the observations.

The findings will be presented in the following way: First, Chapter 2 will give an outline of earlier findings from language acquisition research, focusing on the effects of bilingualism on the individual and the role of age in language learning. Moreover, the context of language learning in education will be discussed in general, as well as with a particular focus on the Dutch context, and on the national pilot study on bilingual primary education in the Netherlands that forms the context of the present study (Jenniskens et al., 2020). Lastly, Chapter 2 will discuss the role of interaction in language learning, resulting in the research questions that formed the basis of the present study. Chapter 3 will then present the methodology that was used to answer these questions. Results will be presented in Chapter 4. Finally, Chapter 5 discusses the results and links these back to findings in earlier research, answering each of the research questions. Based on the findings, conclusions are drawn, and suggestions are made regarding educational practices and possibilities for future research.

2. Theoretical Background

2.1 Bilingualism

Over the past half century, the ability to speak two or more languages (i.e. *bilingualism* or *multilingualism*, respectively) has become widely recognised as beneficial. Although early research into bilingualism in the beginning of the twentieth century reported disadvantages for the bilingual individual (e.g. a lower IQ and a cognitive delay), these findings have been discredited due to the methodological flaws in these studies (e.g. the omittance of socio-economic status [SES] as a variable for IQ scores) (Hummel, 2014, pp. 228-229).

Positive effects of bilingualism that have been found over the past decades include cognitive advantages such as "concept formation, creativity, [and] visual-spatial abilities" (Cenoz, 2003, p. 81). For instance, bilingual children have a better developed theory of mind than their peers; having to choose which language to speak to whom helps them to understand that different people have different perspectives. Moreover, languages take different

perspectives on things. An example of these perspectives is the difference between verb-framed languages and satellite-framed languages: Verb-framed languages (e.g. English) focus on the path of motion, e.g. '*the bottle floated into the cave*', while satellite-framed languages (e.g. Spanish) focus on the manner of motion, e.g. "'*la botella entró a la cueva (flotando)*' [the bottle moved into the cave (floating)" (Talmy, 2000, p. 49-50; as cited by Schmiedtovà, Von Stutterheim & Carroll, 2011, p. 69). The different perspectives acquired along with the different languages lead to better problem-solving skills in bilingual children, who are better at "selective attention, [at] switching between tasks, [and at] monitoring their behaviour during an activity" (Byalistok; as cited by Tribushinina, 2017). Furthermore, the variation in language structures and phonemes that are acquired by bilinguals provides them with a higher level of metalinguistic awareness. As a result of their advantages regarding metalinguistic awareness, selective attention, and linguistic experience, bilinguals may better be able to learn an additional language, or third language (L3), than are monolinguals, i.e. they have a faster rate of acquisition and a higher ultimate attainment (Sanz, 2000). Bilinguals have a better understanding of the different views that languages can take on the world than their monolingual peers, and they have more languages (i.e. their first language [L1] and second language [L2]) with which they can compare the information of the L3, allowing for a larger quantity of positive transfer (Cenoz, 2003).

Although these positive findings on bilingualism may seem appealing, they require careful interpretation, due to the many different forms bilingualism can take. *Additive bilingualism* (i.e. when speaking two or more languages leads to benefits other than speaking these languages alone, such as the advantages mentioned above) is predominantly found in *balanced bilinguals* (i.e. bilinguals who are equally proficient in both languages) (Hummel, 2014, pp. 226-227). Following Cummins' (1979) Threshold Hypothesis, additive bilingualism is only achieved when the Higher Threshold is reached, i.e. when both languages are at an age-appropriate level. When only one of the bilingual's languages is at a native-like level (i.e. *dominant bilingualism*), the learner will experience "neither positive nor negative cognitive effects" (p. 230). However, when both languages are underdeveloped (i.e. *semilingualism*), the Lower Threshold is not reached. This entails that "bilingual children's competence in a language may be sufficiently weak as to impair the quality of their interaction with their educational environment through that language," (p. 230) which leads to negative effects on cognition.

Apart from the degree of bilingualism, there are various options for the onset of language learning that lead to different types of bilingualism. Balanced bilingualism is most

often achieved by *simultaneous bilinguals* (i.e. bilinguals acquiring both languages "from birth or at least before the age of 3;0 [Paradis, 2008, p. 15], e.g. when both parents speak different languages, or when the child's parents speak a minority language and the child acquires the majority language in a day care), whereas *late bilinguals* (i.e. those "who [have] become a bilingual later that childhood" [Hummel, 2014, p. 227]) rarely achieve the same degree of language proficiency. In the present study, the focus will be on *early sequential bilingualism* (i.e. when "one language is introduced after the other language has become somewhat established" [Paradis, 2008, p. 15]).

2.2 The advantage of early language learning

When it comes to language learning, it seems that the younger a learner starts, the better able they are at acquiring that language (e.g. Hummel, 2014). Learners who start learning a language later than a certain age, whether it is their L1 or L2/L3, rarely reach native-like proficiency. Following this observation, Lenneberg's *Critical Period Hypothesis* states that there is an optimal period for language acquisition early in life, after which the ability for (native-like) language acquisition declines (i.e. *sensitive period*) or ends (i.e. *critical period*) (Hummel, 2014, p. 170). In second language acquisition, studies indeed find different age effects for linguistic areas. For instance, "individuals who begin learning a second language after early childhood […] are characterised by a detectable non-native-like accent in their second language, [while those] who are exposed to another language throughout early childhood, generally grow up having no detectable accent in either language" (Hummel, 2014, p. 175). Yet, despite their non-native accent, adult foreign language (FL) learners may have a good command of vocabulary and grammar (Saxton, 2010).

Although young L2 learners have an *ultimate attainment* advantage over older learners, older learners have an advantage in the early stages of L2 learning (Hummel, 2014, p. 176). This *initial rate* advantage is caused by higher cognitive abilities, a higher metalinguistic awareness, and more learning strategies possessed by older learners. Young learners largely depend on their implicit learning skills, and seem to acquire any new language without much effort. As people get older, their ability to learn new things implicitly declines, and their ability to learn explicitly increases; this helps older learners to learn the L2 at a faster pace – at least initially. Yet, younger learners quickly catch up and even surpass them. Compared to younger learners, older learners will suffer more from negative transfer

from their L1, they are less receptive for certain subtleties in their L2, and they are more likely to fossilise in their L2 learning process (Hummel, 2014).

The ultimate attainment advantage for younger learners has led to a widespread belief that foreign language learning should be introduced early on in the education system (Phillipson, 2008). For instance, in the Netherlands, primary schools are allowed to offer up to 15 percent of their teaching time in a foreign language of their own choice, and there are several grants schools can apply for when teaching FLs to young children (Nuffic, 2018). However, the initial rate advantage for older learners is equally found in classroom settings, as the greater ability to make use of explicit learning strategies and rely on a better developed metalinguistic awareness are highly useful in the FL classroom context, which is often limited to no more than a few hours of instruction per week; conversely, the ultimate attainment advantage for younger learners is only found in naturalistic settings, in which the amount of input is much higher than in most FL classroom contexts (Muñoz, 2008). Learning outcomes in second and foreign language learners are found to correlate with quantity and quality of the language input received by the learner; the input in regular FL classrooms is simply insufficient for young learners to implicitly acquire a new language (Cullen, 1998; Muñoz, 2008). This raises the question, whether a larger quantity of input in the primary school context would be sufficient for young learners to successfully acquire a foreign language, and which qualitative characteristics would aid them in this process. In immersion programmes, a specific type of bilingual education in which "over 50% of instruction [is] provided through the L2" (Hummel, 2014, p. 47), as offered in French-English immersion schools in Canada, young learners seem to have sufficient exposure and time to benefit from their implicit learning skills (Muñoz, 2008); whether this is also the case for bilingual primary schools, where more linguistic input is offered than in mainstream FL education but in which there is less exposure than in immersion education, has not yet been demonstrated.

2.3   Education

The recognition of the numerous advantageous consequences of bilingualism, combined with the awareness of an increasing globalisation, have led to a widespread belief in Europe that the educational system should promote multilingualism, especially targeting English as a foreign language (EFL) (Phillipson, 2008). In 2016, 94% of all upper secondary education pupils in the European Union studied EFL ("Foreign language," 2018). In primary education, 84% of all primary school pupils in the EU studied one or more foreign languages in 2015;

83.5% studied English as a foreign language ("84% of primary school," 2017). Besides FL education, pupils may also acquire languages in school when the language of instruction is not their home language, e.g. when a child attends international education or speaks a minority language.

Considering the magnitude of the school-going EU population studying foreign languages, it is of great importance to determine what characteristics foreign language teaching in primary and secondary school have and need to have in order to optimise language learning. Similarly, the large proportion of pupils who are taught through a language other than their L1, indicates the significance of a proper comprehension of how success in content learning regardless of the language proficiency of pupils could be ensured, and moreover, an understanding of the factors influencing the acquisition of the language of instruction. For instance, the previously discussed absence of an ultimate attainment advantage for young learners and presence of an initial rate advantage for older learners in instructed settings, implies that foreign language teaching should not necessarily be introduced as early as the first years of primary education. It simultaneously raises the question whether an increase in FL input in school, such as in international or bilingual education, would be sufficient to simulate naturalistic settings, such that an earlier start *will* lead to a higher ultimate attainment of the foreign language.

2.3.1 The Netherlands

In the Netherlands, children have been required to study English in the final two years of primary school, i.e. from 10 years of age, since 1986, and continue learning the language throughout secondary education (Wilhelm, 2018). Apart from regular foreign language teaching practices, additional types of education with a great emphasis on FL acquisition have emerged over the past few decades. Nuffic (2018) reports that in 1989, bilingual secondary education was offered for the for the first time to pupils at the pre-university level[1]. In these programmes, half of the subjects are taught in L1 Dutch, while the other half are taught in L2 English. The idea behind this concept is that pupils will learn the L2 more rapidly and with more ease, by having abundant input and by using the language in a meaningful context. The

---

[1] In the Netherlands, secondary education is divided in different programmes, depending on the pupil's aptitude for learning. The highest and most theoretical track is *VWO* (pre-university education) and prepares pupils for studies at university level, the second highest level is *HAVO* (higher general secondary education) and prepares pupils for universities of applied sciences, and the lower, more practical track is *VMBO* (pre-vocational education), which is further subdivided into four levels: Theoretical-, mixed-, cadre-profession-oriented-, and basic-profession-oriented learning path.

simultaneous learning of language and subject content is referred to as *content and language integrated learning* (CLIL). Moreover, "the philosophy behind bilingual education is that language makes other worlds and perspectives accessible" (Nuffic, 2018, p. 20), thus allowing pupils to develop their intercultural awareness and interpersonal skills. Bilingual secondary education has emerged as a result of a bottom-up process, as schools and teachers responded to the wish parents expressed for their children to attain better English proficiency levels by being instructed through the language (Van Gool, 2004). Due to its success, bilingual secondary education was extended to HAVO in 2001 (Vonk, 2007), and to pre-vocational education in 2008, with 119 out of 653 public schools in the Netherlands offering bilingual education in 2017 (Nuffic, 2018).

Besides secondary education, primary schools also increasingly focus on teaching foreign languages to their pupils. Whereas English was previously only taught in the last two years of primary education, a growing number of schools now offer *early foreign language education* (EFLE). EFLE in the Netherlands started in 1996, and has become increasingly popular; it was offered in 18 percent of all primary schools in 2016[2] (Nuffic, 2018). Schools offering early foreign language education are allowed to spend a maximum of 15 percent of the teaching time (i.e. about three hours per week) on the chosen FL, for the whole eight years of primary school (Nuffic, 2018). Despite its popularity, Goriot (2019) points out that benefits are limited, as the only positive effect of EFLE that she found in some age groups compared to their peers in mainstream primary education, was the children's vocabulary size; "executive functions, phonological awareness, and the perception of phonetic contrasts" (p. 166) were unaffected by early English education. On the other hand, De Graaff & Costache (2020) found that EFLE pupils scored significantly higher on all areas of proficiency that were tested in their study, i.e. "listening, reading, use of English, spelling, and speaking" (p. 148), as compared to pupils whose EFL education started in the penultimate year of primary education. It is unclear what caused the contradicting findings in these studies, both of which were observed in the Netherlands; therefore, it is of importance to analyse in more detail what current teaching practices are in various forms of foreign language education at the primary level, and to identify which features positively influence the pupils' language acquisition.

Apart from EFLE, as of 2014, a longitudinal national pilot study (i.e. *Flankerend Onderzoek Tweetalig primair Onderwijs*, or *FoTo*) is being carried out on bilingual primary education (BPE): A total of twenty schools throughout the country currently offers this kind

---

[2] Based on more recent numbers on Nuffic's website, the percentage of primary schools offering EFLE was approximately the same in 2020 (Nuffic, 2020).

of education, twelve of which are being monitored every other year (Jenniskens et al., 2018; Jenniskens et al., 2020; Van den Broek et al., 2014). Based on a selection of studies on the process and results in BPE in various countries, which formed the basis for the pilot study in the Dutch context, Van den Broek et al. (2014) conclude that BPE has positive influences on the language proficiency in the target language for reading, writing, speaking, and listening. The L1 of children enrolled in BPE programmes in the examined research did not suffer and sometimes exceeded that of pupils enrolled in regular primary education, even when controlling for SES, and even for pupils that had a different home language. Moreover, children in BPE programmes did not perform worse on mathematics, a content subject that was taught through the target language, and sometimes even outperformed their peers in regular education in this subject (Van den Broek et al., 2014). These findings seem to indicate that *additive bilingualism* is achieved through reaching the Higher Threshold in Cummins' (1979) Threshold Hypothesis, leading to advantages for language development as well as other cognitive domains.

Results in the *FoTo* study similarly paint a positive picture of current developments in the first cohort of BPE in the Netherlands (Jenniskens et al., 2020). BPE pupils outperformed their peers in EFLE on receptive vocabulary and grammar in *groep 1* (i.e. kindergarten, age 4-5), and after adding the productive component, on both receptive and productive vocabulary and grammar in *groep 3* (age 6-7) and *groep 5* (age 8-9). Moreover, BPE pupils outperformed EFLE pupils on all six aspects of a narrative task, i.e. vocabulary, sentence structure, morphology, pronunciation, fluency, use of English and Dutch, in *groep 3* and *5*. Additionally, in *groep 5*, the BPE students outperformed their EFLE peers on reading comprehension and technical reading skills, and they were better able to formulate complete English sentences in a writing task, as opposed to more frequent use of separate English words or a mixture of both languages by EFLE pupils. However, large within-group differences were found for the various English skills. Generally, children at schools where English was used for a larger number of hours scored higher on receptive vocabulary and grammar tests as well as on productive grammar tests. Children at schools where the English level of the teacher was at least CEFR-level B2 (Council of Europe, 2001) scored higher on receptive vocabulary and grammar tests as well as on productive vocabulary tests (Jenneskins et al., 2020). Costache (2017) furthermore reported that in the same study, both the use of sophisticated vocabulary (i.e. non-high-frequency words) and complex utterances (i.e. multi-clause utterances) by the teacher were positively associated with the children's receptive vocabulary scores, but observed that these factors were not necessarily related to the teachers'

language proficiency. When it comes to factors other than L2-acquisition, enrolment in BPE, EFLE, and regular programmes did not seem to influence the development of Dutch and mathematics skills, as no differences were found between groups in their performance in these areas (Jenniskens et al., 2020).

2.4 Interaction

Interaction is imperative for language acquisition (Damhuis, 2008, p. 20). In a school setting, this entails that learners receive input from their teachers, produce output, and receive feedback on their utterances from their teachers. Yet, the importance of interaction and its components has not always been recognised. Even today, the relative importance of learners' input and output is an ongoing dilemma for teachers, which indicates that there is still much to be gained in the implementation of interaction in the educational system (Damhuis, 2008).

The first component to be recognised for its role in language acquisition was input. Krashen's (1980; as cited by Damhuis, 2008) *Input Hypothesis* assumes that *comprehensible input* is the key ingredient for successful language learning. This means that input is just above the learners' level of competence, such that they are able to understand the teacher while still learning more of the target language. This mirrors first language acquisition, as parents fine-tune their language to the level of their children, which facilitates continuous language acquisition (Haijma, 2013, p. 28). Any feedback that learners are provided with, is viewed as new input (Damhuis, 2008, pp. 30-31). Children who start learning a new language initially go through a *silent period*. During this silent period, they listen to the input and as such develop their receptive skills; from the perspective of the Input Hypothesis, these receptive skills need to be established before developing their productive skills (Damhuis, 2008, p. 20).

Within a few years after the Input Hypothesis was formulated, some acknowledgement for the role of interaction and output emerged (Damhuis, 2008). Firstly, interaction was argued to be important for determining what comprehensible input entailed for learners, and for *negotiation of meaning* (Long, 1981, 1985; as cited by Damhuis, 2008, p. 32). Secondly, input alone did not seem to be enough: Empirical studies into second language acquisition of English-speaking children in French immersion schools in the 1970s and 1980s in Canada found that although the pupils' receptive skills in French resembled those of their native French peers, their productive language skills were observed to be lower than expected, despite the abundance of input they had received (e.g. Swain & Lapkin, 1982). Consequently,

Swain (1985; as cited by Swain, 2008) formulated the *Output Hypothesis*, stating that besides absorbing comprehensible input, learners also need to produce *comprehensible* or *pushed output*; they should be encouraged to produce language that is slightly above their level of competence, such that they produce improved versions of previous erroneous utterances. This type of output then forms a new source of input for language learning: "Speaking serves as a vehicle 'through which thinking is articulated, transformed into an artefactual form, and [as such] is then available as a source of further reflection'" (Smagorinsky, 1998; as cited by Swain, 2008). Additionally, pushed output functions as a means for learners to become aware of gaps in their linguistic knowledge, which they will subsequently try to fill with new knowledge from the input (Damhuis, 2008, p. 34). Furthermore, pushed output enables learners to test (conscious or unconscious) hypotheses they have made about linguistic rules, by observing the response or feedback they receive by their interlocutor (p. 35). Finally, pushed output also serves for learners to reflect on their own language use and that of others and thereby increasing their metalinguistic awareness (p. 36). The combination of input and output that are both adapted through negotiation of meaning is referred to as the *Interaction Hypothesis* (Long, 1996; as cited by Damhuis, 2008, p. 33). As pointed out by Hall (2010), interaction is "simultaneously prerequisite and product, the tools and the results of language learning" (p. 12).

Even today, the debate on what a successful communicative classroom entails persists (e.g. Cullen, 1998; Wierenga, 2014). The earlier assumption that "too much teacher talking time (TTT) deprived students of opportunities to speak" (Cullen, 1998, p. 179) has been disputed, as the efficiency of teacher talk seems to be that which matters most. Cullen (1998) argues that teacher talk is necessary, as it provides learners with comprehensible input, which they need to acquire language. He discerns six contexts for which teacher talk is used, i.e. for questioning, giving feedback on pupils' contribution, explaining, giving instructions, correcting, and socialising. He further explains that "aspects of teacher talk, such as the kind of questions teachers ask, can significantly affect the quantity and quality of student interaction in the lesson" (Brock, 1986; as cited by Cullen, 1998, p. 180). In line with the *Output hypothesis*, Dönszelmann (2018) identifies interaction and production as principal elements for learning through target language use in the classroom. This should be stimulated by teachers, for instance by asking questions and letting pupils respond in full sentences in the target language (Dönszelmann, 2018).

Wierenga (2014) suggests that classroom interaction could be studied by focussing on interaction patterns consisting of teacher (or student) initiation, student response, and teacher

feedback or follow-up (i.e. IRF sequence); teacher initiation mostly entails asking questions, which are answered by pupils, after which the teacher gives the pupil feedback (pp. 5-6). In her study into verbal interaction in *early foreign language education* in the Netherlands, she compared the instructions in lessons by four EFLE teachers in two primary schools on the verbal dominance of teachers versus pupils, the questions asked by teachers, and the feedback provided by teachers. Classroom interaction in bilingual primary education, however, is still a new area (Wierenga, 2014); studying this subject could be highly beneficial for bilingual primary school teachers, as an awareness of factors promoting interaction is likely to result in better language acquisition by their pupils. This study will therefore investigate the aspects of interaction set out by Wierenga (2014) in the context of bilingual primary education instead of early foreign language education, with some additional variables that have been found relevant in the literature (e.g. Damhuis, 2008; Dönszelmann, 2018; Haijma, 2013; Houben, 2017; Wierenga, 2014).

2.4.1 Verbal dominance

As was set out in the previous paragraph, there is some disagreement the relative amount of input and output that should be expected for language learning to take place. However, the general consensus is that lessons should be communicative in order to result in language acquisition (e.g. Cullen, 1998). Teachers need to create opportunities for their pupils to produce output in the target language, to interact in the language, to experiment with using the language, to negotiate meanings, and to ask questions (Dönszelmann, 2018).

Wierenga (2014) investigated verbal dominance of teachers and pupils in EFLE by means of counting the number of words spoken by each, and counting the number of teacher- and pupil-initiations. She found that teachers spoke significantly more than pupils in lessons in English and in Dutch (i.e. 59-85% of all words were spoken by teachers in lessons in English, and 71-82% in lessons in Dutch). Initiation in English-taught lessons was mostly teacher-based, i.e. 50-94% teacher-initiation (Wierenga, 2014). Based on these percentages, there seemed to be much variation between the verbal dominance of different teachers in EFLE. No studies have yet investigated the verbal dominance of teachers and pupils in bilingual primary education.

2.4.2 Question types

Verbal interaction is often initiated with a question, and in the classroom context, this is mostly done by teachers (Wierenga, 2014). The interaction that follows is highly dependent on the question that is asked (Cullen, 1998). As noted by Farahian & Rezaee (2012), "types of teachers' questions play an important role in teaching since they affect students' participation in negotiation of meaning" (p. 161). Graesser, Person, and Hu (2002; as cited by Wierenga, 2014) elaborate that questions asked by teachers often require only single words or short phrases for answers. In her study on verbal interaction in EFLE, Wierenga (2014) discerns three types of questions: *Yes/no questions*, *closed/display questions*, and *open/reference questions*. Closed/display questions can be answered by one of the given options, e.g. '*How did you come to school today, by foot or by car?*' To open/reference questions, on the other hand, there are countless possible answers, e.g. '*What is your favourite food?*' (Wierenga, 2014). Cullen (1998) argues that referential questions often elicit more communicative behaviour, because the teacher is unknowing of the answer. Furthermore, open questions more often requires pupils to apply knowledge rather than only recalling information, which leads to higher levels of thinking and more free language production (Menegale, 2008). "This is in contrast to typical 'display' questions (e.g. comprehension questions on a reading text) to which the teacher already has the answer, and only asks so that the class can display their knowledge" (Cullen, 1998, p. 181). Nonetheless, closed/display questions still serve a function in the classroom, as teachers can use them to "check the students' state of knowledge and provide them with opportunities for practicing language forms" (McCarthy, 1991, as cited by Yang, 2010, p. 187). Yes/no questions in turn can especially be helpful for weaker students, as they do not require much language output and are as such easier for these learners to answer (Gower, Philips, and Walters, 1995; as cited by Yang, 2010).

Yang (2010) and Farahian and Rezaee (2012) both studied the types of questions asked by pre-service or otherwise inexperienced EFL teachers, and found that open/reference were hardly asked, as teachers frequently used closed/display questions and yes/no questions. Furthermore, in contrast to the majority of research (e.g. Cullen, 1998; Menegale, 2008), Farahian and Rezaee (2012) observed that closed/display questions elicited responses longer than three words more often (i.e. 13-20%) than open and reference/questions did (i.e. 0-4%). The authors argue that the results found in their study could be explained by the learners' low proficiency level, as well as the teachers' low proficiency level and inexperience in teaching (Faharian & Rezaee, 2012). Brock (1986; as cited by Yang, 2010) found that explicit training on the incorporation of question types enhanced the degree of communicativeness in the

language classroom, as experienced ESL teachers who had been trained to incorporate referential questions in their lessons used this question type significantly more than a control group of teachers who had not received such training, resulting in "significantly longer and more complex" (p. 186) student responses in the treatment group.

When looking at the question types used in English-taught lessons in early foreign language education, Wierenga (2014) found that three out of four participating teachers mostly used *open/reference questions*, followed by *closed/display questions*, while using *yes/no* questions the least; the other teacher barely made use of *open/reference questions*, and mostly used *closed/display questions* (Wierenga, 2014). Relating this back to the degree of communicativeness, the former three teachers' questions asked for a more active role in answering the questions by the pupils than the latter teachers' questions, as pupils were more free in answering the questions. Active student participation, in turn, leads to better communicative language skills (e.g. Dönszelmann, 2018). In bilingual primary education, the types of questions posed by teachers have yet to be examined.

2.4.3 Feedback

Feedback aids language acquisition (e.g. Dönszelmann, 2018). Dönszelmann (2018) identifies several prerequisites for constructive feedback. Firstly, positive feedback on language use and active participation increases student motivation. Secondly, feedback should be specific, i.e. the teacher needs to let the student know *what* is good or bad about their output; this way the pupil and their classmates will be better able to learn from the feedback and construct new linguistic knowledge (p. 39). Thirdly, Dönszelmann advices to limit explicit corrective feedback to errors related to the material that is being covered at the moment of correction, because despite its effectiveness for language learning, it may mute pupils and discourage them from trying again. Lastly, students should be encouraged to take an active role. This can be done by eliciting the correct answer from the student, or letting the pupil repeat the correct answer. If the learning environment is safe enough, it might be advisable to involve the rest of the group, too, as there are likely to be more pupils struggling with the same linguistic problems (Dönszelmann, 2018, p. 39).

Feedback can be given on *form* or on *meaning* (e.g. Schuitemaker-King, 2013). Feedback on form targets the grammaticality or correctness of an utterance, while feedback on meaning is provided when the speaker's intended message is compromised. Schuitemaker-King (2013) observed that in the Dutch secondary school context, feedback given by CLIL teachers was less often directed at linguistic accuracy than in was the case in EFL classes, a

finding that was explained to be the result of the observed CLIL teachers' "general belief […] that their main priority and objectives are concerned with teaching the content concepts and not with focusing on language form or accuracy" (p. 8). Focusing on the content of the message rather than form diminishes speaking anxiety among pupils (Haijma, 2013). On the other hand, when feedback is *only* directed at meaning, and form is neglected, pupils may not learn the language correctly, as pupils need to receive positive and negative evidence in response to their speech in order for them to test hypotheses they have formed about language (Damhuis, 2008, p. 25). In CLIL lessons in primary education, it is advisable to use corrective feedback methods that do not divert the attention away from the content of the lesson (Houben, 2017, p. 5). To what extent this is carried out in practice, i.e. to what extent feedback actually targets language form or content in bilingual primary education, is presently still an unexplored territory.

In Wierenga's (2014) study in the EFLE context, feedback was categorised as one of five types: Positive feedback, negative feedback, corrective feedback, didactic feedback, and suggestive feedback. Feedback is considered *positive* when a teacher makes a confirmative remark; *negative feedback* and *corrective feedback* both indicate that an error was made, the difference being that the latter includes an explanation of why it is considered to be an error; *didactic feedback* and *suggestive feedback* are both forms of neutral feedback, of which the former is an explanatory remark, and the latter merely indicates something should be done differently (p. 9). In lessons taught in English, Wierenga found that teachers mostly made use of *positive feedback* (i.e. 46-62%). *Didactic feedback* was the second-most used type of feedback (i.e. 17-28%), followed by *corrective feedback* (i.e. 6-19%) and *suggestive feedback* (i.e. 0-16%). Only one teacher used some *negative feedback*, i.e. 5% (Wierenga, 2014). These observations are in agreement with Dönszelmann's (2018) advice, to primarily give abundant positive feedback, and to provide explanations when giving feedback (i.e. corrective feedback and didactic feedback). For the positive feedback, Wierenga (2014) did not make a distinction between positive feedback with and without explanation what was good about the utterance. This study will use the same categories that were used by Wierenga (2014), with an added distinction between *specific positive feedback* and *general positive feedback* (i.e. with or without an explanatory remark), to explore the feedback Dutch bilingual primary school teachers give their pupils in CLIL lessons. Furthermore, since as of yet no research has studies the use of feedback focus in the bilingual primary school context, the frequency of feedback on language form and on content will be compared in the present study.

2.4.4 Use of the L1

Bilingual education is based on the use of *content and language integrated learning*, or CLIL. In the broader sense, this means that language and content are learned simultaneously, which is by definition the case if content is taught through a foreign language. In the narrow sense, it entails that content and language learning are intertwined, in the sense that teachers structurally offer new language alongside the subject content, as language goals and subject goals are simultaneously and equally targeted. In this paper, CLIL is used in the broader sense.

The use of the target language as medium of instruction originates in immersion education (De Graaff & Costache, 2020; Haijma, 2013). Immersion education is based on principles of naturalistic language acquisition, i.e. the way children acquire their L1 (Verhallen & Walst, 2007; as cited by Haijma, 2013). Children initially imitate sounds, words, and sentences they hear around them; then, they search for regularities and produce their own creative linguistic constructions. The interaction that follows is essential for language acquisition (Haijma, 2013, p. 28).

De Graaff (2015) explains that CLIL education should consist of four elements that are interconnected: *Content* of subjects is learned through *Communication* in the language of instruction, which has a positive effect on *Cognition*, and pupils simultaneously develop their *Cultural awareness* (De Graaff, 2015, p. 27). The importance of communication in the target language for language acquisition and subject learning is thus substantial.

Haijma (2013) investigated target language use in FL classrooms in Dutch secondary schools, and found a mismatch between actual practices and what students and teachers considered ideal practices. She found that both teachers and students value the use of the target language as medium of instruction, although its realisation is limited. Teachers often avoid the use of the target language because they fear not to be understood by their pupils (Haijma, 2013), they believe they cannot cover enough teaching material, or because they are insecure about their own language proficiency (Dönszelmann, 2018). A lack of teacher input leads to a lack of student output (Haijma, 2013, p. 33). A third of the students explained that anxiety to speak the target language restrained them from using the target language, although they believed this could be overcome if teachers accentuated speaking skills, and specifically the ability to convey a message (p. 21). Two third of the participating students stated that target language use by teachers serves as an example and stimulates their own use of the target language; they considered their own production of the target language to be difficult, yet informative and challenging (Haijma, 2013, p. 32). Target language use makes pupils

aware that language is a means of communication with conveying a message as its prime purpose, rather than language being a goal in itself (Haijma, 2013, p. 28). In an intervention study, Dönszelmann (2018) found that target language use by Dutch secondary school teachers of French as a foreign language and their pupils increased both students' and teachers' motivation. Moreover, students' anxiety to speak decreased over time, and both receptive and productive language skills were positively influenced by the use of the target language (Dönszelmann, 2018).

A similar underuse of the target language is found in primary schools. In her study into verbal interaction in early foreign language education, Wierenga (2014) observed that three out of four teachers used Dutch during the lessons in English (i.e. 33-90% of their words were spoken in English); while pupils responded for 51-78% of the time in English. Two of the teachers posed some of their questions in Dutch, i.e. 31% and 53% of their questions; the same teachers also provided part of their feedback in Dutch, i.e. 42,5% and 59%. The teachers' use of the L1 could be traced back to their English language proficiency, as teachers using Dutch for questioning and feedback were both assessed at CEFR-B1+ level, and the other teacher who used Dutch had CEFR-B2- level, as opposed to CEFR-C1+ level for the teacher who exclusively spoke English during the CLIL lesson (Wierenga, 2014).

Dönszelmann (2018) points out that the importance of using the target language as a medium of instruction does not necessarily mean the L1 should entirely be banned from the classroom, although its use should be restricted. The target language should be used for 70 to 90 percent of the teaching time; use of the pupils' mother tongue should only happen when it supports the language acquisition process of the pupils, e.g. for the translation of an unknown word, a meta-explanation, to make a connection between the L1 and target language, a reassurance, or a pedagogical intervention (Dönszelmann, 2018, p. 38). Furthermore, in bilingual primary education, it is recommended to offer subject-related vocabulary in both the first and the second language (Houben, 2017, p. 5). This study will attempt to explicate to what extent, and for what purpose, teachers in bilingual primary education use the pupils' first language during lessons taught in English.

2.5   Context present study: FoTo

This study will be performed in the context of a larger longitudinal pilot study on bilingual primary education in the Netherlands, i.e. *Flankerend Onderzoek Tweetalig Primair Onderwijs* (FoTo). The pilot study investigates the progression of L2 English, L1 Dutch, and

mathematics skills of children attending bilingual primary education (control condition), early English as a foreign language primary schools, and mainstream primary schools. Pupils have previously been tested when they were in their first year (age 4-5) and third year (age 6-7) and are now in their fifth year (age 8-9) of primary school.

2.6 Focus present study

The present study will focus on the interaction during classes taught in the L2 (i.e. English). Interaction is considered to be a key element of language learning. Guerrettaz and Johnston (2013) explain this through the concept of *classroom ecology*, i.e. "the totality of participants, relationships, structures, objects, and processes that together constitute the shared experience of classroom language teaching and learning" (p. 779). In their view, important factors for language learning are the creation and use of *affordances* and *emergences*, which establish relationships among the aforementioned elements in a classroom. Affordances are "the potential starting point of the meaning-making process; this meaning-making process is an 'active relationship' that involves engagement between the learner and the environment" (p. 782). In other words, affordances are possibilities for learning that are consciously created or that arise, in which meaning is created through interaction. These possibilities can either be used or neglected. *Emergence* takes place when interaction between a learner and a teacher, peers, or learning materials, leads to learning situations (Guerrettaz & Johnston, 2013).

Within the framework of classroom ecology, different affordances are offered through choices in interaction, such as types of questions and types of feedback offered by teachers. Wierenga (2014) distinguishes three phases in interaction patterns: *initiation*, *response*, and *feedback*. An interaction cycle can be either initiated by the teacher or by a pupil, and starts with a (*yes/no*, *closed*, or *open*) question. Different types of questions asked by teachers will elicit different types of responses by pupils, and thus offer different types of affordances for (language) learning. Finally, teachers can give various types of feedback to their students: *Positive feedback* (confirmation), *negative feedback* (remark about something being wrong), *corrective feedback* (remark about something being wrong, and why), *didactic feedback* (explanation), or *suggestive feedback* (remark or question that stimulates the pupil to do something differently) (Wierenga, 2014). Positive feedback can be general or specific (Dönszelmann, 2018). Feedback can be focused on content or on language form (Schuitemaker-King, 2013). Apart from IRF-cycles, teachers' classroom talk can serve to explain, give instructions, and to correct pupils (Cullen, 1998). Finally, the use of L1 Dutch

by the teacher (and other students) in English-taught lessons might influence the classroom ecology and language use of the learners.

To assess the interaction during CLIL lessons in bilingual primary schools, the aforementioned components will be compared per group and per approach, resulting in the research question: *What are characteristics of teacher-student interaction in bilingual primary education in the Netherlands?* This results in the following sub questions (cf. Wieringa, 2014):

1. What is the relative role of teacher talk and student talk in bilingual primary education?
    a. What is the verbal dominance of teachers versus students during the plenary parts of lessons in English in bilingual primary schools?
    b. Do lesson content, lesson part, or teachers' language background have an influence on verbal dominance?
2. How are questions used during interaction in bilingual primary education?
    a. What types of questions do teachers ask during the plenary parts of CLIL lessons?
    b. Do lesson content, lesson part, or teachers' language background have an influence on question types asked by teachers?
3. How is feedback used during interaction by teachers in bilingual primary education?
    a. What types of feedback do teachers provide during the plenary parts of CLIL lessons?
    b. Do lesson content, lesson part, or teachers' language background have an influence on feedback given by teachers?
4. What is the frequency and function of teachers' use of L1 Dutch during the plenary parts of CLIL lessons in bilingual primary education?

3. Method

3.1   Participants

Four bilingual primary schools located in various places throughout the Netherlands participated in the present study. Of each school, one teacher and all pupils of one *groep 5* class (i.e. children aged 8-9, similar to the American *third grade* and the British *Year 4*) were observed. When there were multiple teachers who taught the same class, the teacher who taught most of the CLIL lessons (i.e. lessons taught in English) participated in the study.

All participating schools were part of a national pilot study on bilingual primary education, i.e. *Flankerend Onderzoek Tweetalig Primair Onderwijs* (*FoTo*). They were selected from the eleven participating schools based on homogeneity (i.e. number of pupils and teaching method). The schools all spent 50 to 70 percent of the time on teaching in Dutch, and 30 to 50 percent on teaching in English. Among the four participating bilingual primary schools, two different forms of content and language integrated learning (CLIL) were used. In some schools ($n = 2$), there were separate teachers for each language of instruction, such that one teacher taught only in Dutch, and one teacher taught only in English, i.e. *one teacher, one language* (OTOL) approach. In other schools ($n = 2$), the class teacher speaks both Dutch and English, without mixing the two languages, such that one lesson may be taught in Dutch, and the next lesson in English, i.e. *one situation, one language* (OSOL) approach (Jenniskens et al., 2018). The teachers' language backgrounds differed between the two types of schools: The teachers at the OTOL schools were both native speakers of English, while the teachers at the OSOL schools were non-native speakers of English with different language proficiencies. The number of students per class, as well as the English proficiency of the teachers, are displayed in Table 1.

3.2   Materials

All teachers taught a prescribed lesson based on a lesson outline developed by the *FoTo* research team, and a free lesson of the teacher's own choice. The lessons took approximately 15 to 30 minutes to complete, apart from the prescribed lesson at School D, which only lasted about 5 minutes, resulting in a total of about 40 to 60 minutes of lessons per teacher. In the prescribed lesson, the children were divided into groups; each group was given a written question to which there could be multiple answers, e.g. '*What can you use water for*?'. They were asked to write down as many answers as they could think of, while taking turns in writing. Then, pupils read their answers out loud, based on which the other children had to guess their question. Some teachers used an introduction activity before starting the prescribed activity, such as an activity in which sounds and pictures related to water were identified and discussed. To maximise comparability between the prescribed lessons, only the prescribed activity itself was coded as belonging to the prescribed lesson; any introduction activity or follow-up was coded as belonging to the free activity. This meant that the prescribed lesson started with the teacher's introduction of the activity to the students, and ended when the discussion of the answers had been rounded off, and the last remarks on the

activity had been made. A complete list of the content of the lessons, as well as the questions that were used during the prescribed lesson activity, are added in Appendix A.

Table 1. Characteristics of the Participants.

	Number of pupils	Teacher's experience primary education (in years)	Teacher's experience teaching CLIL (in years)	Teacher's English proficiency[3]	Bilingual Method	Duration lessons in minutes	
						Prescribed	Free
School A	26	17	5	Native speaker	OTOL	19:17	33:48
School B	25	14	14	Native speaker	OTOL	29:49	32:16
School C	18	28	3	B2+	OSOL	14:58	32:10
School D	27	15	4	A2-B1	OSOL	5:27	35:14

Besides the observations, a structured interview was held with the participating teachers, for which the interview questions were constructed by the *FoTo* research team. Teachers were asked questions about their views and methods regarding bilingual education, such as the way in which they encourage the use and development of the TL, their hopes and views on bilingual primary education, and the representativeness of the lessons that were observed for their overall CLIL teaching style.

3.3 Procedure

At each of the four bilingual primary schools, two lessons were observed, i.e. a prescribed lesson and a 'random' lesson, both of which were CLIL lessons and were thus taught in English. Besides observations, teachers were interviewed on their experiences in teaching

---

[3] Proficiency levels were assessed by members of the *FoTo* research team (Jenniskens et al., 2020) and were based on recordings of the interviews; the last three questions of the interview were asked in English for this purpose. The other questions were asked in the teacher's native language. For the assessment, use was made of classifications following the Common European Framework of References for Languages (CEFR; Council of Europe, 2001).

bilingual primary education. The interviews took place either immediately following the lessons, or later the same day. The lessons and interviews were recorded using Philips DVT1300 Voice Tracer audio equipment. The audio recordings were then used for the transcription of the plenary parts of the lessons, i.e. the introduction and instructions, and the final discussion and ending of the lesson. At some of the schools ($n = 2$), the free lesson also contained plenary activities that could not be categorised as introduction or discussion, but that were nonetheless components of the plenary part of the lesson, and were therefore included in the analyses. The transcriptions were furthermore coded, and the coded data was subsequently statistically analysed using IBM SPSS Statistics versions 25, adhering to a significance-value of $p < .05$.

Table 2. Types of Questions.

Question type	Type of answer	Example from lessons
Yes/no	Yes/no	*Did you say them all?*
	'me' or 'not me'	*Who needs a pencil or a pen?*
Closed/ display	Choose from list of options. Only one correct answer.	*Which text feature are they describing?* *How do you say 'precipitation' in Dutch?* *Everybody say: condensation.*
Open/ reference	Linguistically and/or conceptually free, no options given.	*What do you think you heard?* *Who thinks they know what their question was?*

The transcribed lessons were coded for language (i.e. L1, L2, or both) and speaker (i.e. teacher versus pupil). The interaction was analysed using the distinctions set out by Wierenga (2014) who looked at interaction in lessons taught in English and in Dutch in early foreign language education in the Netherlands. The present study focused on the same topic in a different, but closely related context: Interaction in lessons taught in English in another type of primary education, namely bilingual primary education. Consequently, Wierenga's (2014) methodology is very appropriate to study interaction in lessons for the participating age group.

Similar to Wierenga (2014), the first step was to distinguish the *interaction patterns* (i.e. initiation, response, feedback). This enabled comparison of *verbal dominance*, which was

made up of two components: The initiator (i.e. teacher-initiated versus pupil-initiated interaction), and the number of words spoken by the teacher and by the pupils. For utterances in which languages were mixed, separate word counts were done for the words spoken in Dutch and in English. As noted by Wierenga (2014), it was of importance to code the IRF-patterns before coding the question types, as some forms of feedback might also contain questions, which might influence the results and make them less comparable to previous research. Subsequently, the *types of questions* asked by the teacher and students were evaluated (i.e. yes/no question, closed/display question, or open/reference question; as displayed in Table 2). Questions in the form of '*Who [x]?*', which should be interpreted by the pupils as '*Is [x] the case for you, yes or no?*' were considered to be yes/no questions. Students' responses were correspondingly coded for the type of question they responded to, to enable the calculation and comparison of response lengths to each of the question types. Next, *types of feedback* were determined (i.e. general positive, specific positive, negative, corrective, didactic, and suggestive; as displayed in Table 3); here, the present study deviates slightly from Wierenga's (2014) study, by splitting positive feedback into specific and general positive feedback. Dönszelmann (2018) suggests that positive feedback is most valuable when it is specific about what has been done well. Like in Wierenga's (2014) study, it was possible for feedback utterances to belong to multiple feedback types and foci; an example is provided in Extract 1, where the teacher gives specific positive feedback on content (i.e. the teacher makes it clear what is correct about the utterance), didactic feedback on content (i.e. the teacher explains more about the content), and didactic feedback on language (i.e. the teacher explicitly explains what the correct translation is). Furthermore, feedback was assessed on whether it targeted language form or content, which was another addition to Wierenga's (2014) study. Feedback on the meaning or content expressed in an utterance is considered beneficial for the development of communicative competence (Haijma, 2013), although feedback on form should not completely be withheld (Damhuis, 2008). In cases of use of Dutch during the plenary parts of the lessons, qualitative analyses were carried out to assess what triggered the use of the L1. Teachers' utterances that could not be identified as parts of IRF-sequences were coded for their function, i.e. instruction, explanation, or behaviour correction, following the speech functions of teacher talk as identified by Cullen (1998).

During the coding process, it was discovered that utterances that had the IRF-function of feedback or follow-up could not always be identified as belonging to one of the feedback foci (i.e. language or content) or feedback categories. In these instances, feedback was neutral in the sense that no truth value was provided to the child on either content or language; an

example is provided in Lines 3 and 6 of Extract 2. This type of feedback solely served a communicative (Line 3) or didactic function (Line 6) and was coded as 'neutral' feedback. Furthermore, as it was noticed during coding that recasts (i.e. the teacher's correct repetition of a student's erroneous utterance) and recasts in which the child's Dutch answers were repeated by the teacher in English were prevalent forms of feedback, additional coding categories were made to enable analyses of these factors.

Table 3. Types of Feedback.

Type of feedback	Description	Example from lessons
General positive	Positive remark without specifying what is done well.	"Good."
Specific positive	Positive remark that is specific about what is done well.	"I like how you used 'already'."
Negative	Explicit mention the answer is incorrect, without further specification.	"No, that's not exactly it."
Corrective	Remark correcting a student's utterance.	"[name] hears it raining very hard." (As a response to: "I hear rain very hard")
Didactic	Remark in which an explanation is provided.	"It's Monday, but that's not the date. That's the day."
Suggestive	Suggestion that something should be done differently, or a question that asks pupils to find this out themselves.	"Okay, but now I hear 'the one group can eat, and the other group can eat.' So, is that a difference, then?"

Extract 1.

1 R S *Dauw*.

2 F T *Dauw*, yeah, in English it's called 'dew'. That's also condensation.

Extract 2.

1	I	T	What do you think you heard? Try to say it in English. I think I heard…? [name]
2	R	S1	I think I heard rain.
3	F	T	Okay.
4	I	T	[name]?
5	R	S2	I also think rain.
6	F	T	I also think rain.
7	*instruction*	T	Okay, let's listen to the third one.

Finally, at some schools, much of the English output produced by the children came forth from '*language drills*'. This meant that the students had no linguistic freedom and were not asked to produce meaning with the language they produced; they were solely asked to repeat an utterance as a group, with or without *total physical response*. No individual feedback could be given on such utterances. Examples of these types of drills are displayed in Extracts 3 and 4, where they are preceded by hashtags. Similarly, a few songs that were sung at one of the schools were also counted as *non-content expressing utterances* by pupils. Students' total number of spoken words consisted for 3, 15, 0, and 34% of this type of output at Schools A, B, C, and D, respectively; all data in this category was in English. This type of language production by pupils is arguably qualitatively different from language production by which meaning is constructed and expressed by pupils, and no longer serves a communicative function: The learners cannot use the language freely to test hypotheses (Damhuis, 2008), it does not provide the teacher with information on the pupils' comprehension (Yang, 2010), and no meaning is negotiated (Long, 1981; as cited by Damhuis, 2008). Yet, it is still a form of practice to articulate output in the target language, and as such it is part of the language learning process. The decision was made to include these types of student responses in the analyses of verbal dominance, because they are in fact part of classroom talk. Moreover, as teachers are generally found to do most of the talking in lessons (e.g. Wierenga, 2014), the inclusion of this data will likely lead to more telling verbal dominance results. That is, if these student utterances are included and results suggest that teachers speak significantly more words than their pupils, this result is generalisable to a context in which this data would be excluded, while the opposite would not be possible. Nevertheless, this type of student output was excluded from the analyses of response lengths to question types or for language choice, as they were not deemed telling of the pupils' own abilities and choices in expressing

themselves. Therefore, a separate coding category was made for this type of data, to be able to include as well as exclude it from analyses.

Extraxt 3.

1	Teacher		Someone else? [name].
2	Student		The queen is bigger than the other bees.
3	Teacher		The queen bee is bigger than the other bees. Yes, very well!
4	Teacher		So everybody: 'The queen bee is bigger than the other bees.'
5	All students	#	The queen bee is bigger than the other bees.
6	Teacher		Yes.

Extract 4.

1	Teacher		The group from [name].
2	Student group		Hive.
3	Teacher		Again?
4	Student group	#	Hive.
5	Teacher		Everybody?
6	All students	#	Hive.

3.4   Analyses

3.4.1 Verbal dominance

To assess the verbal dominance of teachers and pupils, the numbers of words spoken by the teacher and by the pupils were calculated for each school in various contexts (i.e. in total, per lesson, per lesson part, and per lesson part of each lesson). Percentages of words spoken by teachers versus students were calculated to aid the interpretation of comparisons between groups. The word counts were compared between teachers and students within groups and contexts using binominal statistical tests, and between groups or contexts using chi-square statistical analyses. Comparisons of contexts that were made within groups included a comparison between lesson parts (i.e. introduction, plenary activity, and discussion), a comparison between the prescribed and the free lesson, and a comparison between the lesson parts within each of the observed lessons. Between groups, a comparison was made between the verbal dominance in terms of spoken words of all participating schools.

Furthermore, for each initiation in IRF-sequences, the initiator was assessed, i.e. teacher-initiation or pupil-initiation. Tokens were counted and compared using chi-square statistical tests. Again, the same within- and between-group comparisons were made.

3.4.2 Types of questions

Questions posed by teachers were assessed on whether they were *yes/no* questions, *open* questions, or *closed* questions. Tokens were counted for each category, and chi-square tests were carried out to compare the use of question types by different teachers (i.e. between-group comparisons) and in different contexts (i.e. within-group comparisons, similar to those described for verbal dominance). Furthermore, students' average response lengths to each question type were calculated and compared statistically using One-Way ANOVAs.

3.4.3 Types of feedback

Tokens for each of the six pre-determined feedback categories and the additional neutral feedback category, as well for feedback focus (i.e. language versus content), were counted and statistically compared using chi-square statistical testing. Again, this was done to make the same within- and between-group comparisons for the use of the feedback types and foci. The use of the feedback categories to focus on language versus on content was similarly compared using chi-square tests, as was the use of neutral feedback.

3.4.4 Use of L1

The number of words spoken by teachers in each language in total and in each of the speaking contexts (i.e. IRF-sequence, instruction, explanation, and correction) were calculated and statistically compared between teachers and between contexts using chi-square tests. Furthermore, the number of words in L1 Dutch and the number of words in L2 English were calculated for pupils and statistically compared between schools, as well as to their teachers, using chi-square tests. In cases in which the teacher used Dutch, qualitative analyses were carried out to determine the context of the L1 use (e.g. specific/difficult topics, translations, disciplining pupils). This was not done for the use of the L1 by the pupils, as this was assumed to be caused by a lack of proficiency or vocabulary in English.

3.4.5 Interview data

Finally, to reach a deeper understanding of the teachers' beliefs regarding CLIL education, qualitative analyses were carried out to find trends in the responses to the interview questions, regarding the teachers' views on verbal dominance, the correction of errors, and the use of the L1.

4. Results

4.1   Verbal dominance

4.1.1. Number of words

The number of words spoken by each teacher and group of pupils are displayed in Table 4. Teachers spoke about five times as many words as their pupils at all schools, i.e. percentages of words spoken by teachers: 82% (School A), 84% (School B), 82% (School C), and 85% (School D). Binomial tests indicate that this difference was statistically significant in all contexts (i.e. $p < .001$), i.e. teachers spoke significantly more words than their pupils in total, in each of the observed lessons (Table 5), in each lesson part (Table 6), and in each lesson part the prescribed and of the free lesson (Table 7). Not all words spoken by teachers were part of IRF-sequences. Yet, Table 8 shows that even when comparing the teachers' total number of words spoken in IRF-sequences to the pupils' total number of words, which were all part of IRF-sequences, teachers still spoke twice to three times as many words as their pupils, i.e. percentages of words spoken by teachers: 74% (School A), 69% (School B), 72% (School C), and 74% (School D).

Table 4. Number of Words Spoken in each Language by Teachers and Pupils in Absolute Numbers and in Percentages of their Total Number of Spoken Words.

	English		Dutch		Total	
	Teacher	Student	Teacher	Student	Teacher	Student
School A	4274 (96%)	647 (66%)	202 (5%)	327 (33%)	4487	980
School B	6174 (100%)	1125 (95%)	0 (0%)	46 (4%)	6190	1176
School C	4169 (100%)	836 (90%)	0 (0%)	84 (9%)	4190	924
School D	3259 (98%)	529 (89%)	60 (2%)	62 (10%)	3332	594

Table 5. Verbal Dominance in Number of Words for each Lesson, in Absolute Numbers and Percentages, and Chi-Square Comparisons between Lessons.

	Prescribed		Free		Chi-Square Test
	Teacher	Student	Teacher	Student	
School A	1384 (80%)	345 (20%) [1]	3093 (83%)	629 (17%)	X^2 (1) = 7.50, p = .006*
School B	3227 (87%)	466 (13%)	2947 (81%)	705 (19%) [1]	X^2 (1) = 61.25, p < .001*
School C	1188 (79%)	309 (21%) [1]	2981 (83%)	611 (17%)	X^2 (1) = 9.41, p = .002*
School D	832 (83%)	169 (17%) [1]	2487 (85%)	422 (15%)	X^2 (1) = 3.28, p = .07

[1]Increase in student output.

Table 6. Verbal Dominance in Number of Words in Lesson Parts, in Absolute Numbers and Percentages, and Chi-Square Comparisons between Lesson Parts.

	Introduction		Plenary activity		Discussion		Chi-Square Test
	Teacher	Student	Teacher	Student	Teacher	Student	
School A	920 (81%)	214 (19%) [1]	1057 (80%)	264 (20%) [1]	2500 (83%)	469 (17%)	X^2 (2) = 8.33, p = .016*
School B	2836 (94%)	193 (6%)	-	-	3338 (77%)	978 (23%) [1]	X^2 (1) = 352.36, p < .001*
School C	2398 (91%)	251 (9%)	-	-	1771 (73%)	669 (27%) [1]	X^2 (1) = 276.10, p < .001*
School D	1463 (92%)	121 (8%)	976 (78%)	268 (22%) [1]	880 (81%)	202 (19%) [1]	X^2 (2) = 119.72, p < .001*

[1]Increase in student output.

Table 7. Verbal Dominance in Number of Words in each Lesson Part of each Lesson, in Absolute Numbers and Percentages, and Chi-Square Comparisons between Lesson Parts of each Lesson.

Lesson	School	Introduction		Plenary activity		Discussion		Chi-Square Test
		Teacher	Student	Teacher	Student	Teacher	Student	
Prescribed	School A	350 (89%)	43 (11%)	-	-	1034 (77%)	302 (23%) [1]	X^2 (1) = 25.86, p < .001*
	School B	2134 (95%)	123 (5%)	-	-	1093 (76%)	343 (24%) [1]	X^2 (1) = 270.53, p < .001*
	School C	480 (100%)	0 (0%)	-	-	708 (70%)	309 (30%) [1]	X^2 (1) = 183.77, p < .001*
	School D	440 (88%)	62 (12%)	-	-	392 (79%)	107 (21%) [1]	X^2 (1) = 14.74, p < .001*
Free	School A	570 (77%)	171 (23%) [1]	1057 (80%)	264 (20%) [1]	1466 (88%)	194 (12%)	X^2 (2) = 61.21, p < .001*
	School B	702 (91%)	70 (9%)	-	-	2245 (78%)	635 (22%) [1]	X^2 (1) = 65.86, p < .001*
	School C	1918 (88%)	251 (12%)	-	-	1063 (75%)	360 (25%) [1]	X^2 (1) = 114.69, p < .001*
	School D	1023 (95%)	59 (5%)	976 (78%)	268 (22%) [1]	488 (84%)	95 (16%) [1]	X^2 (2) = 122.68, p < .001*

[1] Increase in student output.

Table 8. Number of Words Spoken in each Language by Teachers and Pupils in IRF-Sequences in Absolute Numbers and in Percentages of their Total Number of Spoken Words.

	English		Dutch		Total	
	Teacher	Student	Teacher	Student	Teacher	Student
School A	2611 (94%)	647 (66%)	177 (6%)	327 (33%)	2788	980
School B	2571 (100%)	1125 (95%)	0 (0%)	46 (4%)	2571	1176
School C	2324 (100%)	836 (90%)	0 (0%)	84 (9%)	2324	924
School D	1600 (99%)	529 (89%)	17 (1%)	62 (10%)	1617	594

Although all teachers used more words than their students, there was a statistical difference between the schools: $X^2 (3) = 22.25$, $p < .001$. Students at Schools A and C spoke a significantly more during lessons than their peers at Schools B and D. Furthermore, students spoke significantly more in the prescribed lesson than in the free lesson at three schools, although this was only significant at Schools A and C (see Table 5). At School B, pupils spoke significantly more words during the free lesson. Pupils' verbal dominance in terms of spoken words also significantly differed between different lesson parts at all schools, but did not do so in a uniform way (see Table 6). Still, it is noteworthy that pupils took a more active part in the discussions at three out of four schools, and during plenary activities at both schools in which these were present ($n = 2$). When looking at each lesson separately (Table 7), during the prescribed lesson, more student output was observed during the discussion compared to the introduction at all schools. During the free lesson, the same differences are observed as when looking at the two lessons combined.

Table 9. Degree of Initiative-Taking by Teachers and Students during the Prescribed and Free Lesson and in Total, in Absolute Numbers and Percentages.

	Prescribed lesson		Free lesson		Total	
	Teacher	Student	Teacher	Student	Teacher	Student
School A	37 (90%)	4 (10%)	80 (91%)	8 (9%)	117 (91%)	12 (9%)
School B	63 (90%)	7 (10%)	72 (90%)	8 (10%)	135 (90%)	15 (10%)
School C	31 (97%)	1 (3%)	56 (95%)	3 (5%)	87 (96%)	4 (4%)
School D	28 (82%)	6 (18%)	67 (91%)	7 (10%)	95 (88%)	13 (12%)

4.1.2. Initiative-taking

Table 9 displays the amount of initiative taken by teachers and students at each school during the prescribed and the free lesson and in total. Teachers took significantly more initiative than students at all schools in total, in the prescribed and in the free lesson (see Table 11). There were no significant differences in the degree of initiative-taking between the prescribed and the free lesson within schools (see Table 12). There were also no significant differences between the four schools in the degree of initiative-taking of teachers and pupils in total, within the prescribed lesson or within the free lesson (see Table 13).

Table 10. Degree of Initiative-Taking by Teachers and Students during Different Parts of Lessons in Absolute Numbers and Percentages.

	Introduction		Plenary activity		Discussion	
	Teacher	Student	Teacher	Student	Teacher	Student
School A	19 (95%)	1 (5%)	28 (90%)	3 (10%)	70 (90%)	8 (10%)
School B	29 (88%)	4 (12%)	-	-	106 (91%)	11 (9%)
School C	39 (93%)	3 (7%)	-	-	48 (98%)	1 (2%)
School D	30 (83%)	6 (17%)	33 (89%)	4 (11%)	32 (88%)	3 (12%)

In Table 10, the degree of initiative-taking by teachers and students are displayed for each lesson part. All teachers took more significantly more initiative than their pupils during introductions, during plenary activities, and in the discussions (see Table 11). No significant differences were found between the degree of initiative-taking within schools between the various lesson parts (see Table 12). Correspondingly, the schools did not differ significantly in the degree of initiative-taking in the different lesson parts either (see Table 13).

Subsequently, the degree of initiative-taking was compared between the various parts of the prescribed and the free lessons, to find out if there was an interaction effect of lesson part and lesson in the degree of initiative-taking. No significant differences were found in the degree of initiative-taking between various parts of the prescribed lesson of the same teacher, or between the various parts of the free lesson of the same teacher (see Table 12). Likewise, there were no significant differences in the degree of initiative-taking between the introductions of the prescribed and the free lesson of the same teacher, or between the discussions of the prescribed and the free lesson of the same teacher (see Table 12). The degree of initiative-taking of teachers and students thus remained constant both within and

between schools, in all lessons and all lesson parts, as teachers consistently initiated significantly more interaction in the IRF-sequenced than did their students.

Table 11. Chi-Square Comparisons between Initiative Taken by Teachers and Students in for each Lesson, in Total, and for each Lesson Part.

Lesson	Lesson part	Comparison	School	Chi-square test result
Prescribed		*Initiative * teacher/student*	School A	$X^2 (1) = 29.25, p < .001^*$
			School B	$X^2 (1) = 42.23, p < .001^*$
			School C	$X^2 (1) = 22.90, p < .001^*$
			School D	$X^2 (1) = 12.56, p < .001^*$
Free		*Initiative * teacher/student*	School A	$X^2 (1) = 56.85, p < .001^*$
			School B	$X^2 (1) = 43.25, p < .001^*$
			School C	$X^2 (1) = 38.53, p < .001^*$
			School D	$X^2 (1) = 44.29, p < .001^*$
Total		*Initiative * teacher/student*	School A	$X^2 (1) = 86.65, p < .001^*$
			School B	$X^2 (1) = 85.23, p < .001^*$
			School C	$X^2 (1) = 61.39, p < .001^*$
			School D	$X^2 (1) = 56.21, p < .001^*$
	Introduction	*Initiative * teacher/student*	School A	$X^2 (1) = 27.49, p < .001^*$
			School B	$X^2 (1) = 16.04, p < .001^*$
			School C	$X^2 (1) = 25.05, p < .001^*$
			School D	$X^2 (1) = 15.86, p < .001^*$
	Plenary activity	*Initiative * teacher/student*	School A	$X^2 (1) = 12.63, p < .001^*$
			School D	$X^2 (1) = 19.13, p < .001^*$
	Discussion	*Initiative * teacher/student*	School A	$X^2 (1) = 52.08, p < .001^*$
			School B	$X^2 (1) = 69.20, p < .001^*$
			School C	$X^2 (1) = 37.67, p < .001^*$
			School D	$X^2 (1) = 21.87, p < .001^*$

Table 12. Chi-Square Comparisons between Degree of Initiative-Taking and Observed Lesson for each Lesson Part, and Degree of Initiative-Taking and Lesson Part for each Observed Lesson.

Lesson	Lesson part	Comparison	School	Chi-square test result
	Introduction	*Lesson * initiative-taker*	School A	$X^2 (1) = 0.35, p = .55$
			School B	$X^2 (1) = 0.01, p = .91$
			School C	$X^2 (1) = 0.08, p = .78$
			School D	$X^2 (1) = 1.85, p = .17$
	Discussion	*Lesson * initiative-taker*	School A	$X^2 (1) = 0.05, p = .82$
			School B	$X^2 (1) = 0.04, p = .83$
			School C	$X^2 (1) = 0.59, p = .44$
			School D	$X^2 (1) = 0.02, p = .89$
	Total	*Lesson * initiative-taker*	School A	$X^2 (1) = 0.02, p = .90$
			School B	$X^2 (1) = 0.00, p = 1.00$
			School C	$X^2 (1) = 0.19, p = .66$
			School D	$X^2 (1) = 0.02, p = .90$
Prescribed lesson		*Lesson part * initiative-taker*	School A	$X^2 (1) = 0.62, p = .43$
			School B	$X^2 (1) = 0.25, p = .62$
			School C	$X^2 (1) = 0.03, p = .86$
			School D	$X^2 (1) = 1.44, p = .23$
Free lesson		*Lesson part * initiative-taker*	School A	$X^2 (2) = 0.13, p = .94$
			School B	$X^2 (1) = 0.01, p = .91$
			School C	$X^2 (1) = 1.39, p = .24$
			School D	$X^2 (2) = 0.22, p = .90$
Total		*Lesson part * initiative-taker*	School A	$X^2 (2) = 0.53, p = .77$
			School B	$X^2 (1) = 0.21, p = .65$
			School C	$X^2 (1) = 1.40, p = .24$
			School D	$X^2 (2) = 1.18, p = .56$

Table 13. Chi-Square Comparisons between Degree of Initiative-Taking at different Schools within each Lesson and within each Lesson Part.

Lesson	Lesson part	Comparison	Chi-square test result
Prescribed		*Initiative-taker * school*	$X^2 (3) = 3.83, p = .28$
Free		*Initiative-taker * school*	$X^2 (3) = 1.22, p = .75$
Total		*Initiative-taker * school*	$X^2 (3) = 3.67, p = .30$
	Introduction	*Initiative-taker * school*	$X^2 (3) = 2.65, p = .45$
	Plenary activity	*Initiative-taker * school*	$X^2 (1) = 0.02, p = .88$
	Discussion	*Initiative-taker * school*	$X^2 (3) = 0.31, p = .38$

4.2   Types of questions

Figure 1 displays the types of questions asked by teachers at each of the participating schools. A chi-square test revealed that there were significant differences between the four schools in the types of questions being asked by teachers: $X^2 (6) = 85.62, p < .001$. *Yes/no questions* were asked significantly more frequently by Teacher B and significantly less frequently by Teacher D compared to the other teachers. *Closed/display questions* were asked significantly more often by Teacher D and significantly less often by Teacher C as compared to the other teachers. In her interview, Teacher D also named examples of display questions as a way she fosters her pupils' English language acquisition, e.g. reading English texts to each other and by singing. *Open/reference questions* were significantly more frequently asked by Teacher C and significantly less frequently by Teacher D. The biggest differences were thus observed between the two non-native speaker teachers at the OSOL schools, rather than between the two types of schools or mother tongue of the teachers.

The types of questions that were asked by teachers during the prescribed and the free lesson are displayed in Table 14. Three out of four teachers asked significantly different questions in the prescribed lesson than in the free lesson (i.e. Teachers A, C, and D). Teachers A and C asked significantly more *open/reference questions* and fewer *closed/display questions* during the prescribed lesson than during the free lesson; Teacher D asked significantly more *open/reference questions* and significantly fewer *yes/no questions* during the prescribed lesson compared to the free lesson. For Teacher B, there were no significant differences between the types of questions asked in both lessons.

As the activity during the prescribed lesson was similar across all schools, the questions asked during this lesson at the different schools were also compared using a chi-

square test, revealing that there was again a significant difference between the teachers in the questions asked in each lesson: X^2 (6) = 51.56, *p* < .001. Again, *yes/no questions* were less often asked by Teacher D and more often by Teacher B, and *closed/display questions* were more often asked by Teacher D, and less often by Teacher C. *Open/referential questions* were again mostly asked by Teacher C, but in this lesson, Teacher B asked significantly fewer *open/referential questions* than her colleagues at other schools. The latter finding is not surprising, given the nature of the questions asked that pupils answered in the prescribed activity and which the other pupils subsequently needed to guess: Each group was asked to describe one of the text features they had been discussing, which entailed that the rest of the students were asked to guess which of the text features was being described by the group of pupils, which is a type of closed question. At other schools, questions that needed to be guessed were open questions (see Appendix A for the complete lists of questions).
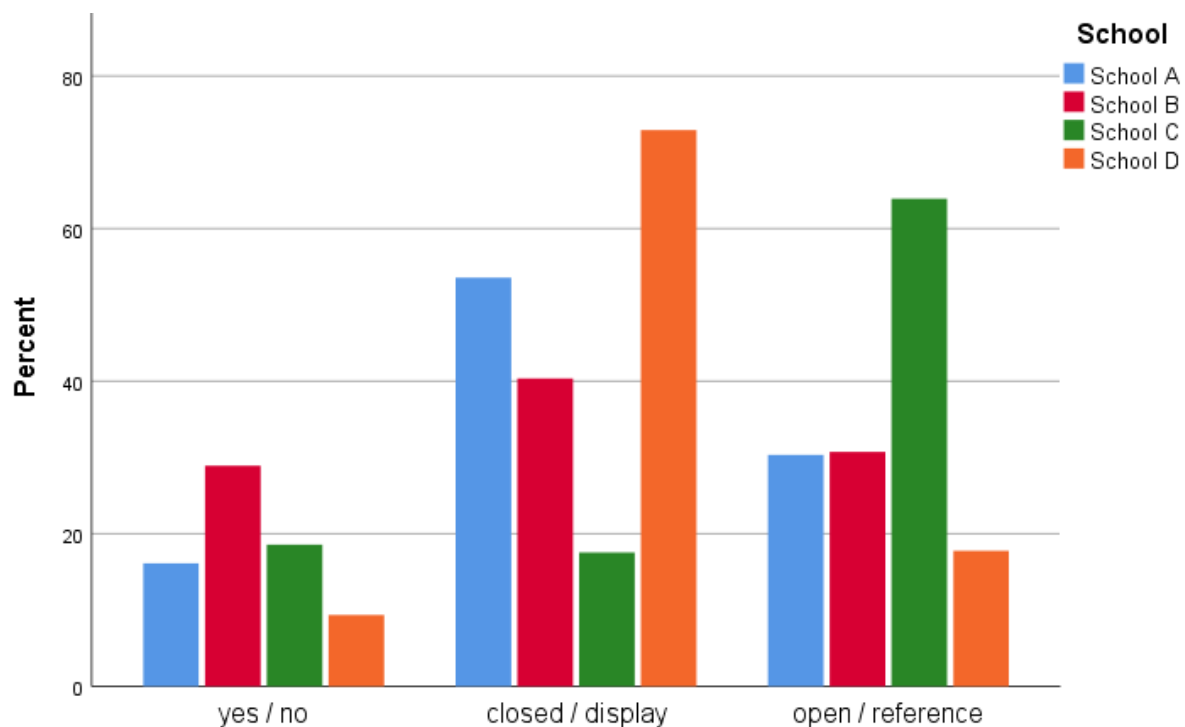
Figure 1. Teachers' Percentual Use of each Question Type.

As the activities during the free lesson might reflect the usual teaching practices of teachers more than the prescribed lesson activity, the questions asked in this lesson were likewise compared using a chi-square test, revealing significant differences between the teachers in the questions asked during this lesson: X^2 (6) = 60.35, *p* < .001. In this lesson, *yes/no questions* were again asked significantly more often by Teacher B than by any of the other teachers, but now there is no significant difference between Teachers A and C and Teacher D.

Closed/display questions were again significantly more often asked by Teacher D, but this time also by Teacher A, compared to Teachers B and C. *Open/reference questions* were again significantly more often asked by Teacher C as compared to the other teachers, and again significantly less often by Teacher D, and this time also by Teacher A.

Next, the questions asked by teachers were compared between different lesson parts. Three out of four teachers (i.e. Teachers A, C, and D) asked significantly different questions in different parts of the lessons, as can be seen in Table 15. Interestingly, these three teachers show mostly similar patterns in the way they change the type of questions they ask over the lesson. All three teachers asked significantly more *closed/display questions* and fewer *open/reference questions* during the introduction part of the lessons. Additionally, Teacher D asked fewer *yes/no questions*, but due to the low expected count, this observation cannot be interpreted as statistically relevant. During plenary activities, Teacher A asked statistically fewer *yes/no questions*. During the discussions, all three teachers asked statistically fewer *closed/display questions*. Furthermore, Teachers A and D asked significantly more *yes/no questions*, and Teachers C and D asked significantly more *open/reference questions* compared to the other parts of the lesson. Another notable finding is that the only teacher who was consistent in the types of questions she asked across the two different lessons, i.e. Teacher B, was the same teacher who also lacked a significant difference between the types of questions asked in the different lesson parts, meaning that the lesson she taught or the part of the lesson had no effect on the types of questions posed by her.

4.2.1 Student responses to question types

Figure 2 displays the mean response length of pupils to each question type. A One-Way ANOVA revealed that the type of questions students responded to, influenced the length of their response: $F(2, 597) = 51.00$, $p < .001$. A Tukey's post-hoc test revealed that *yes/no-questions* elicited significantly shorter answers than *closed/display questions* ($p = .030$) and *open/reference questions* ($p < .001$), and that *open/reference questions* elicited significantly longer answers than *closed/display/questions* ($p < .001$). At all schools, there was a significant difference between the response lengths to the various types of questions, as reported in Table 16. However, there were some differences between the schools. Tukey's post-hoc tests revealed that although *yes/no questions* elicited shorter answers than *closed/display questions* at all schools, this difference was only significant at School A ($p = .045$), and not at School B ($p = .09$), School C ($p = .98$), or School D ($p = .44$). Answers to *yes/no questions* were significantly shorter than those to *open/reference questions* at all schools ($p = .036$, $p < .001$,

$p = .002$, and $p < .001$, for Schools A, B, C, and D, respectively). Answers to *open/reference questions* were longer than answers to *closed/display questions* at all schools, in the case of Schools B ($p < .001$), C ($p < .001$), and D ($p < .001$).
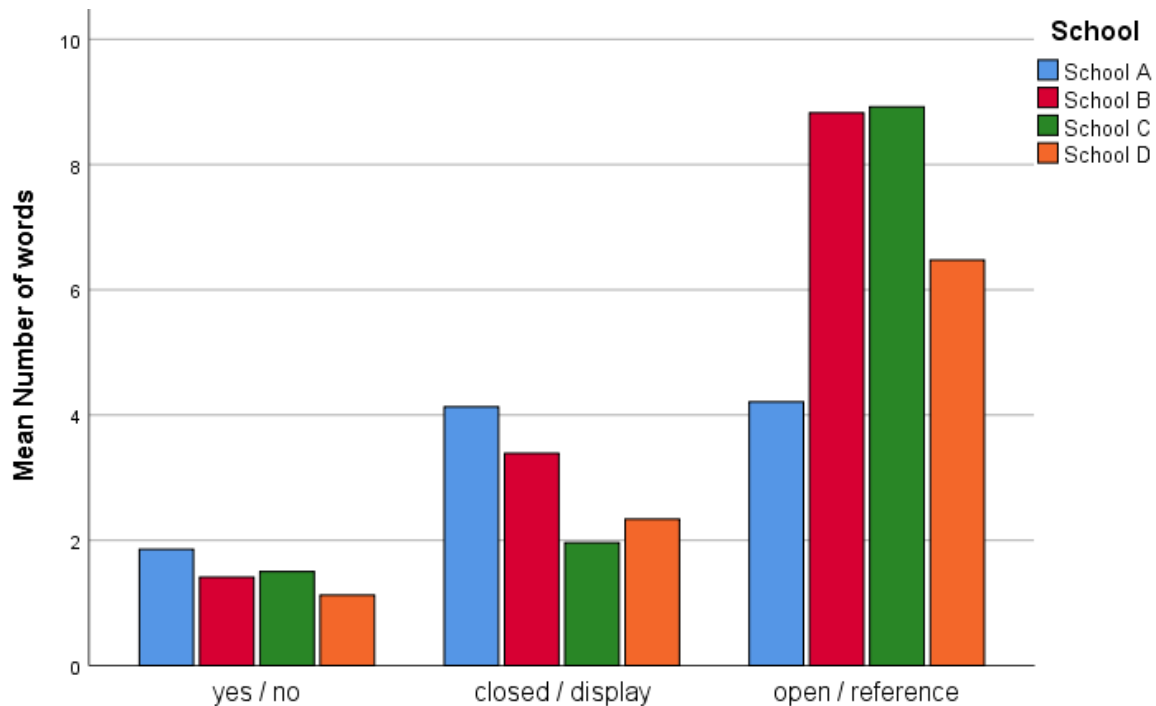
Figure 2. Students' Mean Response Length to each Question Type in Number of Words.

Subsequently, the response lengths to each of the question types between the different schools were compared. A One-Way ANOVA revealed no significant differences between the length of student responses to *yes/no questions* between the four schools: $F(3, 82) = 0.53$, $p = .66$. However, there was a significant difference between response lengths to *closed/display questions*: $F(3, 238) < .001$. A Tukey's post-hoc test revealed that students at School A gave significantly longer answers to this type of question than their peers at Schools C ($p = .006$) and D ($p = .001$); there were no significant differences between Schools C and D, or between School B and the other schools. For *open/reference questions*, there was also a significant difference in the students' response lengths between the four schools: $F(3, 268) = 8.80$, $p < .001$. A Tukey's post-hoc test revealed that students at School A gave significantly shorter answers to this type of question than their peers at Schools B ($p < .001$) and C ($p < .001$); there were no significant differences between any of the other schools.

Table 14. Types of Questions Asked by Teachers in the Prescribed and the Free lesson in Absolute Numbers and Percentages, and Chi-Square Comparisons between the Observed Lessons.

	Yes/no		Closed/display		Open/reference		
	Prescribed	Free	Prescribed	Free	Prescribed	Free	Chi-Square Test
School A	8 (17%)	17 (16%)	12 (26%)	71 (65%)	26 (57%)	21 (19%)	$X^2 (2) = 24.08, p < .001$*
School B	22 (27%)	26 (31%)	38 (47%)	29 (34%)	21 (26%)	30 (35%)	$X^2 (2) = 3.04, p = .22$
School C	6 (19%)	12 (19%)	1 (3%)	16 (25%)	25 (78%)	37 (57%)	$X^2 (2) = 7.16, p = .028$*
School D	0 (0%)	10 (14%)	24 (73%)	54 (73%)	9 (27%)	10 (14%)	$X^2 (2) = 6.89, p = .032$*

Table 15. Types of Questions Asked by Teachers in the Different Lesson Parts in Absolute Numbers and Percentages, and Chi=Square Comparisons between the Lesson Parts.

|	Introduction			Plenary activity			Discussion			Chi-Square Test
	Yes/no	Closed/ display	Open/ reference	Yes/no	Closed/ display	Open/ reference	Yes/no	Closed/ display	Open/ reference	
School A	3 (13%)	17 (71%)	4 (17%)	2 (5%)	24 (63%)	12 (32%)	20 (22%)	42 (45%)	31 (33%)	$X^2 (4) = 9.63, p = .047$*
School B	13 (33%)	14 (35%)	13 (33%)	-	-	-	35 (28%)	53 (42%)	38 (30%)	$X^2 (2) = 0.66, p = .72$
School C	9 (21%)	16 (38%)	17 (41%)	-	-	-	9 (16%)	1 (2%)	45 (82%)	$X^2 (2) = 24.58, p < .001$*
School D	1 (3%)	32 (94%)	1 (3%)	2 (5%)	31 (80%)	6 (15%)	7 (21%)	15 (44%)	12 (35%)	$X^2 (4) = 23.24, p < .001$*

Table 16. Student Responses to different Types of Questions: Number and Percentage of Responses, Mean (and Standard Deviation) of Response Length in Number of Words, Range of Response Length in Number of Words for each School, and One-Way ANOVA Comparisons between Response Lengths and Question Types.

	Yes/no			Closed/display			Open/reference			
	Number	Mean (SD)	Range	Number	Mean (SD)	Range	Number	Mean (SD)	Range	One-Way ANOVA
School A	21 (10%)	1.86 (2.31)	1 – 10	93 (44%)	4.13 (3.71)	1 – 26	98 (46%)	4.21 (4.35)	1 – 31	$F(2, 207) = 3.31, p = .038*$
School B	39 (24%)	1.41 (1.21)	1 – 6	54 (33%)	3.39 (2.70)	1 – 12	69 (43%)	8.82 (6.28)	1 – 28	$F(2, 158) = 41.78, p < .001*$
School C	18 (13%)	1.5 (1.69)	1 - 8	27 (20%)	1.96 (1.34)	1 – 6	92 (67%)	8.92 (9.91)	1 – 52	$F(2, 131) = 11.46, p < .001*$
School D	8 (8%)	1.13 (0.35)	1 – 2	69 (72%)	2.34 (2.56)	1 – 12	19 (20%)	6.47 (3.36)	1 – 13	$F(2, 92) = 20.59), p < .001*$

4.3   Types of feedback

4.3.1 Feedback focus

The relative feedback focus at all schools is displayed in Figure 3. At all schools, more feedback was given on content than on language. Chi-square tests revealed that there was a significant difference between the focus of feedback of the observed teachers, as displayed in Table 17. Although significant between-group differences between the native English speaking Teachers A and B and the non-native Teachers C and D were observed, significant within-group differences seem to indicate that school type or mother tongue of the teacher is not a good predictor of the teachers' feedback focus (see Table 17). This assumption was confirmed by statistical analyses between all four schools. There was a significant difference between the teachers in whether they focused their feedback on language, and on content. Neutral feedback was counted as focusing on neither language nor content, and feedback could also be counted as both. Teacher A focused a significantly larger proportion of her feedback on language and Teacher D did significantly less so than the other two teachers. Teacher B targeted significantly more of her feedback on content, while Teachers A and D did significantly less so. There was also a significant difference between the teachers in their use of neutral feedback. Teacher D gave significantly more neutral feedback than the other teachers, and Teacher B did significantly less so.
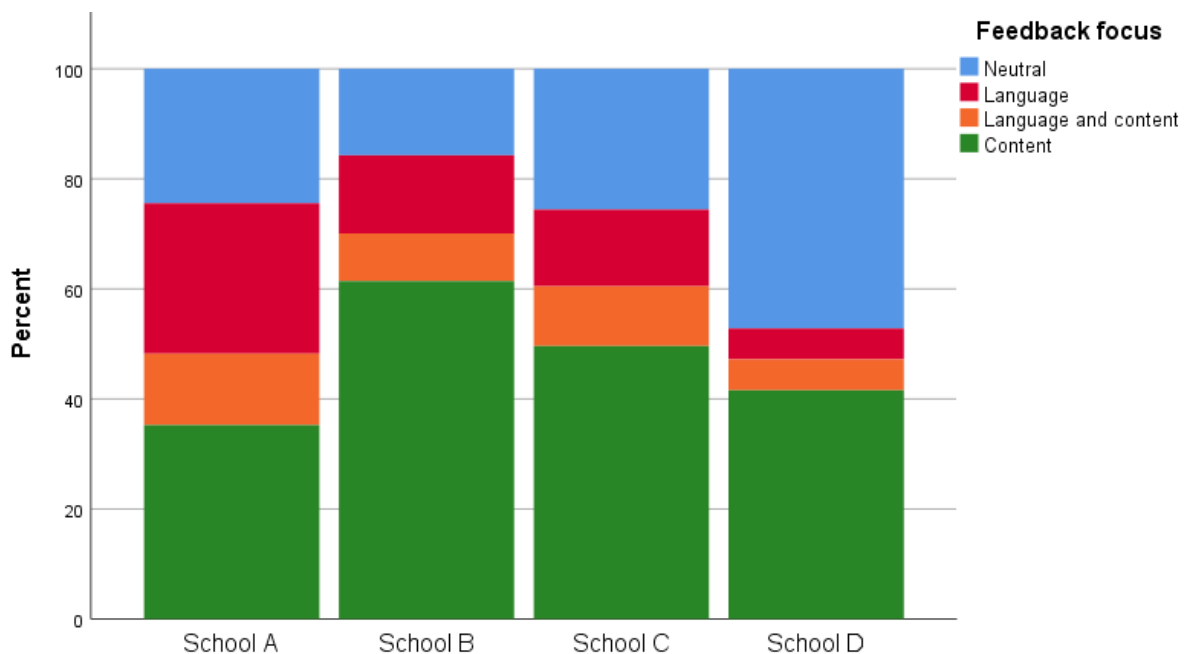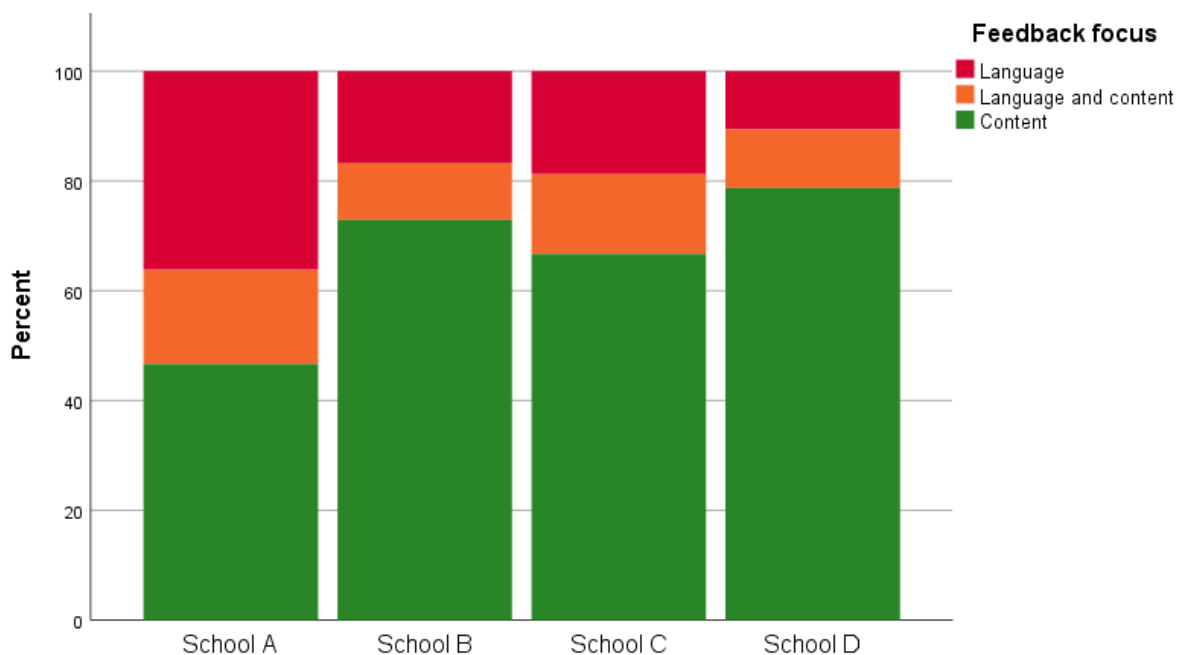

Figure 3. Feedback Focus: Percentages of Total Feedback per School.

Table 17. Chi-Square Comparisons of Feedback Focus between School Types, between Schools within School Types, and between all Schools.

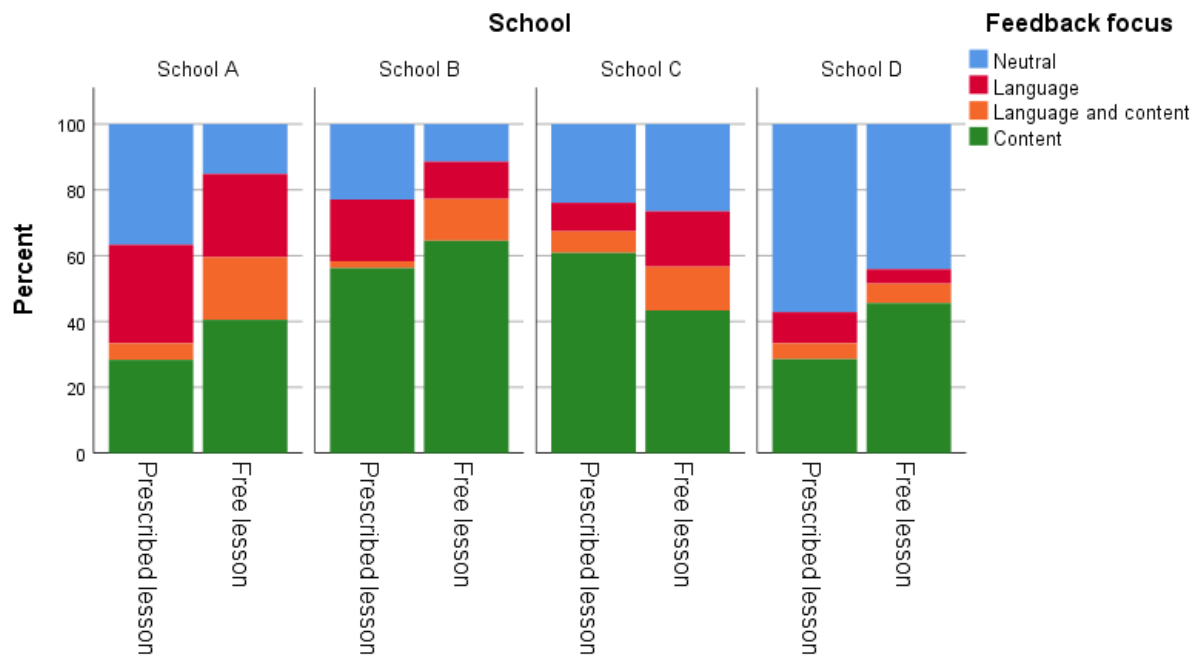| Feedback dataset	Comparison	Chi-Square Test
All feedback	*Feedback focus on language * OTOL[1]/OSOL*	$X^2 (1) = 9.97, p < .001*$
	*Feedback focus on content * OTOL/OSOL*	$X^2 (1) = 0.634, p = .43$
	*Feedback is neutral * OTOL/OSOL[1]*	$X^2 (1) = 12.19, p < .001*$
Excluding	*Feedback focus on language * OTOL[1]/OSOL*	$X^2 (1) = 4.27, p = .039*$
neutral feedback	*Feedback focus on content * OTOL/OSOL[1]*	$X^2 (1) = 5.268, p = .022*$
All feedback	*Feedback focus on language * OTOL A[1]/C*	$X^2 (1) = 9.30, p = .002*$
	*Feedback focus on content * OTOL A/C[1]*	$X^2 (1) = 13.10, p < .001*$
	*Feedback is neutral * OTOL A/C*	$X^2 (1) = 3.11, p = .08$
Excluding	*Feedback focus on language * OTOL A[1]/C*	$X^2 (1) = 15.18, p < .001*$
neutral feedback	*Feedback focus on content * OTOL A/C[1]*	$X^2 (1) = 10.23, p = .001*$
All feedback	*Feedback focus on language * OSOL B[1]/D*	$X^2 (1) = 6.24, p = .013*$
	*Feedback focus on content * OSOL B/D*	$X^2 (1) = 3.75, p = .053$
	*Feedback is neutral * OSOL B/D[1]*	$X^2 (1) = 10.90, p = .001*$
Excluding	*Feedback focus on language * OSOL B/D*	$X^2 (1) = 2.21, p = .14$
neutral feedback	*Feedback focus on content * OSOL B/D*	$X^2 (1) = 2.54, p = .22$
All feedback	*Feedback focus on language * school*	$X^2 (3) = 25.42, p < .001*$
	*Feedback focus on content * school*	$X^2 (3) = 17.38, p < .001*$
	*Feedback is neutral * school*	$X^2 (3) = 27.35, p < .001*$
Excluding	*Feedback focus on language * school*	$X^2 (3) = 22.14, p < .001*$
neutral feedback	*Feedback focus on content * school*	$X^2 (3) = 17.96, p < .001*$

[1]Teacher or group of teachers who used a feedback focus significantly more.

Figure 4. Feedback Focus: Percentages of Total Language- and Content-Focused Feedback per School.

It could be argued that the *neutral feedback* category was not really feedback at all beyond their function in the IRF-sequence, and solely served a communicative function. Therefore, the same analyses were also performed without this category, as reported in Table 17 and visualised in Figure 4. When disregarding neutral feedback, there are still significant differences between the teachers in their use of feedback on the English language and on content. Teacher A gave significantly more feedback on language, and Teachers B and C did significantly less so. Teacher A gave significantly less feedback on content than the other teachers. Due to the significant differences between schools of the same type, further analyses of feedback will focus primarily on within-school and between-school comparisons, rather than between- and within-school-type comparisons.

Figure 5 displays the relative focus of feedback in the prescribed and free lesson at each of the schools. As displayed in Table 18, there was only a significant difference in this context was observed for Teacher A, who gave significantly more neutral feedback and less feedback on content was given during the prescribed lesson than during the free lesson. Figure 6 shows the relative focus of feedback on language and content. When disregarding the neutral feedback, no significant differences emerged between the focus on content or language between the prescribed and the free lesson for any of the schools. Teachers were thus consistent in the relative focus of their feedback on language and content in the two lessons.

Figure 5. Feedback Focus in Each Lesson.

**Table 18**. Chi-Square Comparisons of Feedback Focus between Lessons and between Lesson Parts.

| Feedback dataset	Comparison	Teacher	Chi-Square Test
All feedback	*Feedback focus * prescribed/free lesson*	Teacher A	$X^2 (2) = 12.67, p = .002*$
		Teacher B	$X^2 (2) = 5.05, p = .08$
		Teacher C	$X^2 (2) = 3.69, p = .16$
		Teacher D	$X^2 (2) = 2.10, p = .35$
Excluding neutral feedback	*Feedback focus * prescribed/free lesson*	Teacher A	$X^2 (1) = 1.42, p = .23$
		Teacher B	$X^2 (1) = 0.20, p = .66$
		Teacher C	$X^2 (1) = 3.65, p = .056$
		Teacher D	$X^2 (1) = 0.96, p = .33$
All feedback	*Feedback focus * lesson part*	Teacher A	$X^2 (4) = 12.01, p = .017*$
		Teacher B	$X^2 (2) = 1.95, p = .38$
		Teacher C	$X^2 (2) = 7.57, p = .023*$
		Teacher D	$X^2 (4) = 7.96, p = .093$
Excluding neutral feedback	*Feedback focus * lesson part*	Teacher A	$X^2 (2) = 6.36, p = .042*$
		Teacher B	$X^2 (1) = 1.05, p = .31$
		Teacher C	$X^2 (1) = 2.50, p = .11$
		Teacher D	$X^2 (2) = 3.10, p = .21$
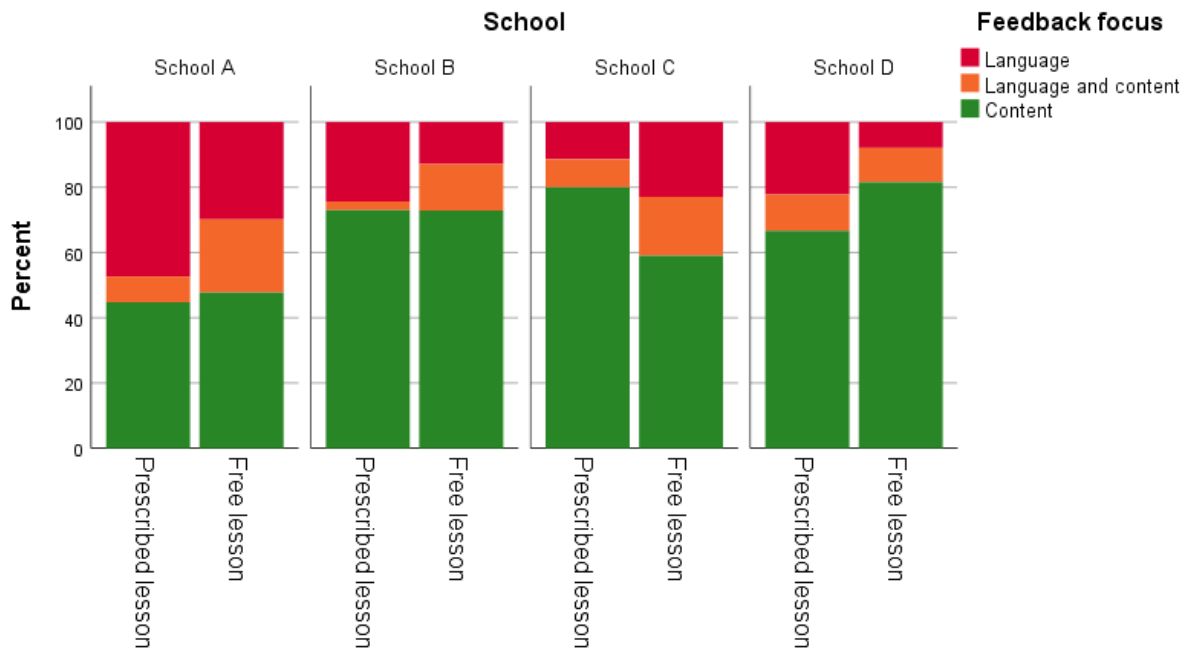
Figure 6. Non-Neutral Feedback Focus in each Lesson.

Next, the teachers' relative feedback focus in the various lesson parts was compared, as displayed in Figures 7 and 8. When including neutral feedback in the analyses, there was a significant difference between the feedback focus by Teacher A and by Teacher C (see Table 18). Both teachers show different patterns of change: In the introductions, Teacher A increased her use of neutral feedback and decreased her focus on content; Teacher C decreased her use of neutral feedback and increased her focus on content. During plenary activities, Teacher A decreased the use of neutral feedback. During discussions, Teacher C increased the use of neutral feedback and decreased her focus on content. For Teachers B and D, there were no significant differences between their relative feedback focus in the different lesson parts. When disregarding neutral feedback and solely looking at content- and language-focused feedback, differences were only significant for Teacher A; she focused significantly more on language and less on content during the introductions, as compared to the other lesson parts. The relative focus of feedback on language or content, as well as the relative use of neutral feedback, across different lesson parts thus seems to be teacher-dependant: Two teachers were consistent in their feedback focus throughout the lessons, while the other two teachers applied significantly different feedback foci in patterns that diverge from each other.
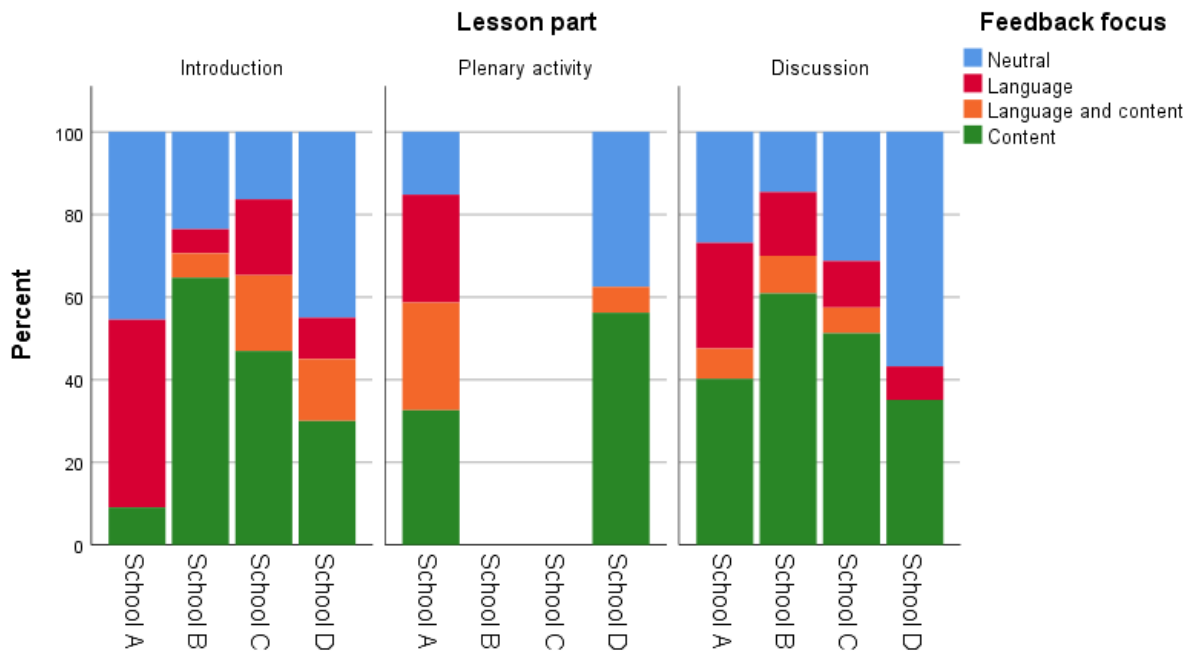
**Figure 7**. Feedback Focus in different Lesson Parts, including Neutral Feedback.
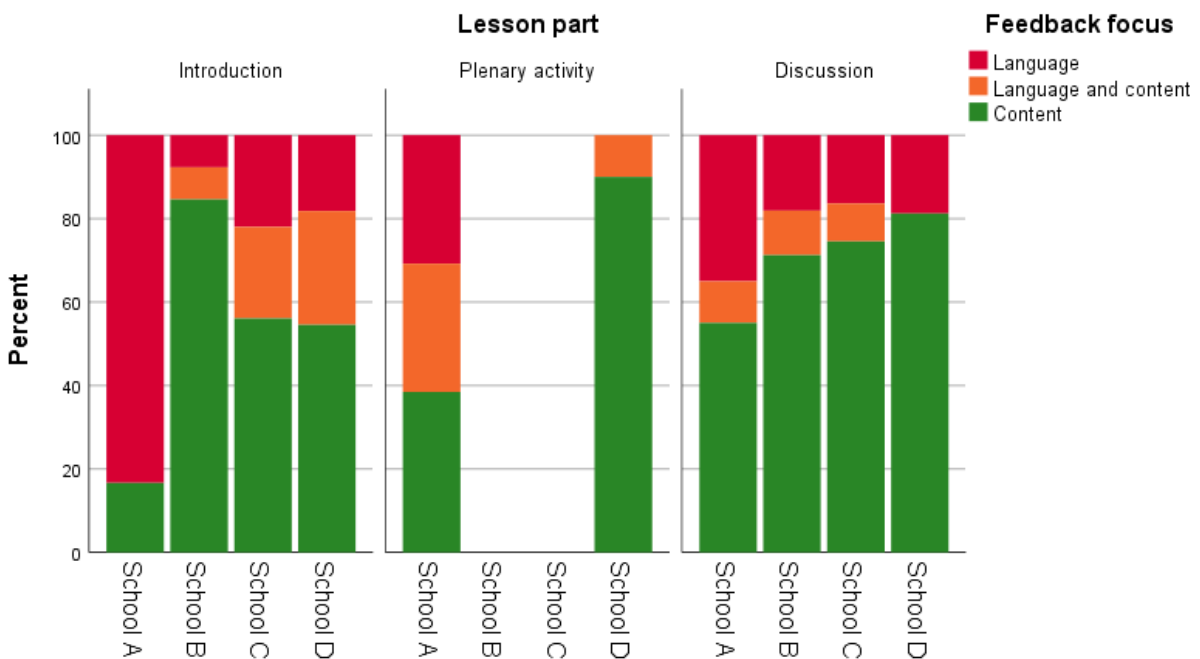
Figure 8. Feedback Focus in different Lesson Parts, excluding Neutral Feedback.

4.3.2 Feedback type

To compare the use of the various feedback types between teachers, the numbers of occurrences of each type of feedback were put into crosstabs to enable statistical analyses. It is important to keep in mind that results to these analyses thus display the proportion of the total number of feedback-type categorisations (i.e. tokens) and not of the total number of feedback utterances. For example, by correcting a child on their sentence structure, a teacher

may simultaneously give positive feedback the content displayed in the student's answer, and corrective feedback on their language use. This one token of feedback then adds up to two tokens of feedback types. It is therefore useful for comparing the frequency of use of the feedback types by teachers, but not for indicating the total amount of feedback that is given, which may seem disproportionally much due to categorisation as multiple feedback types.

Figure 9 shows the percentual use of each feedback category by each teacher. Most teachers gave predominantly general positive feedback. Especially when combining general and specific positive feedback, such as which was done in earlier research, the use of positive feedback category exceeds the other feedback categories tremendously. Other often-occurring feedback categories are corrective and suggestive feedback, followed by didactive feedback. Negative feedback is relatively scarcely given by all teachers.
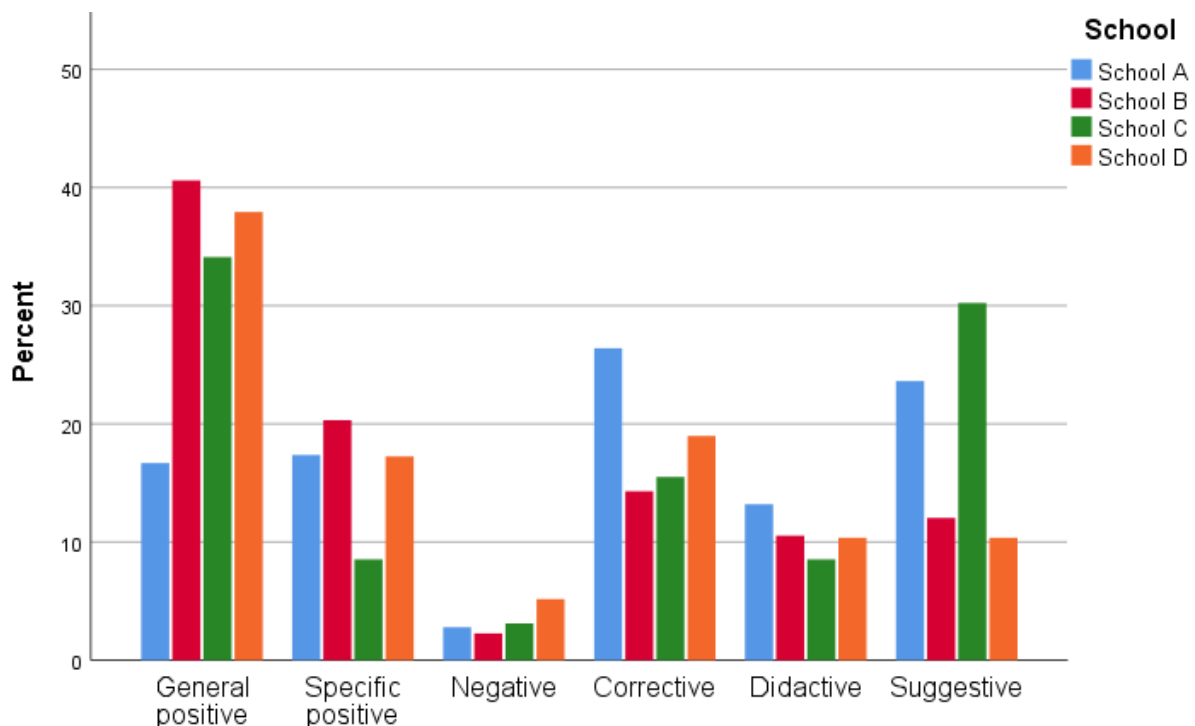

Figure 9. Feedback Types: Percentages of All Language- and Content-Focused Feedback-Type Tokens.

When separating the feedback types focusing on language from those focusing on content, a different picture emerges. Figures 10 and 11 display how often each of the feedback types was respectively used to focus on language and on content by each of the four teachers. At all schools, there was a significant difference in the frequency of use of each of the feedback categories focused on language compared to those focused on content (see Table 19). At all

schools, general and specific positive feedback and negative feedback were used significantly more often as content-focused feedback, and corrective feedback was significantly more often used as language-focused feedback. Furthermore, Teacher B used suggestive feedback significantly more often to focus on language than on content, and Teacher D used didactive feedback significantly more often to focus on content than on language.
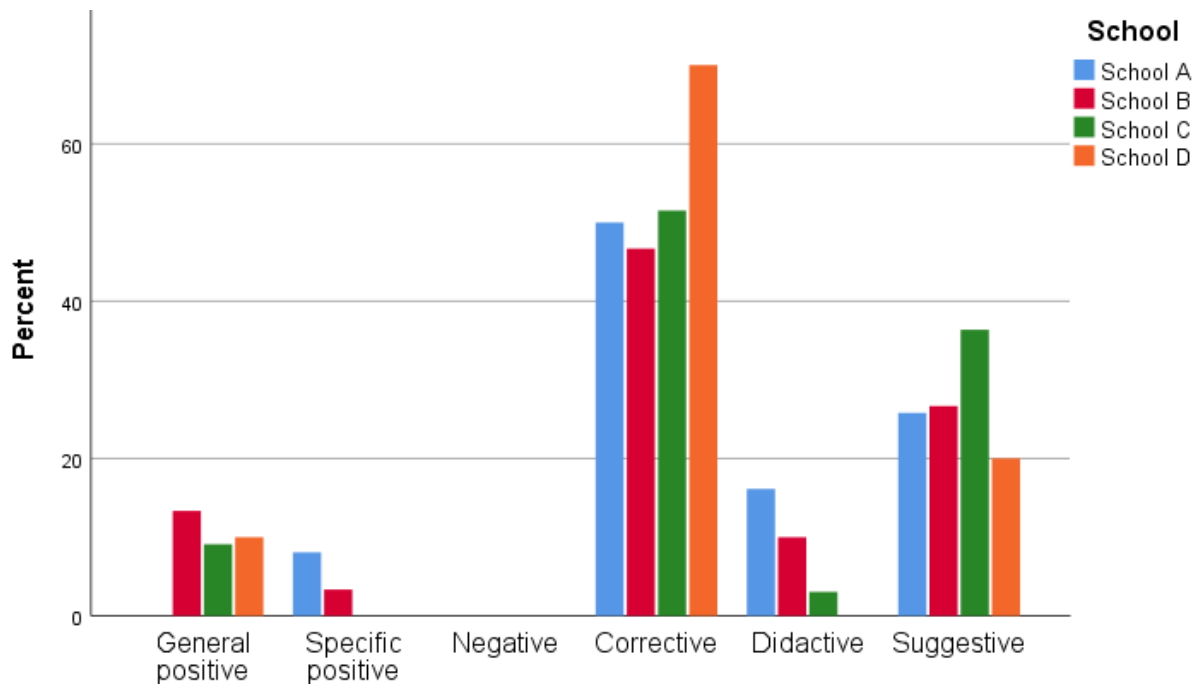
Figure 10. Percentage of Language-Focused Feedback-Type Tokens for each School.
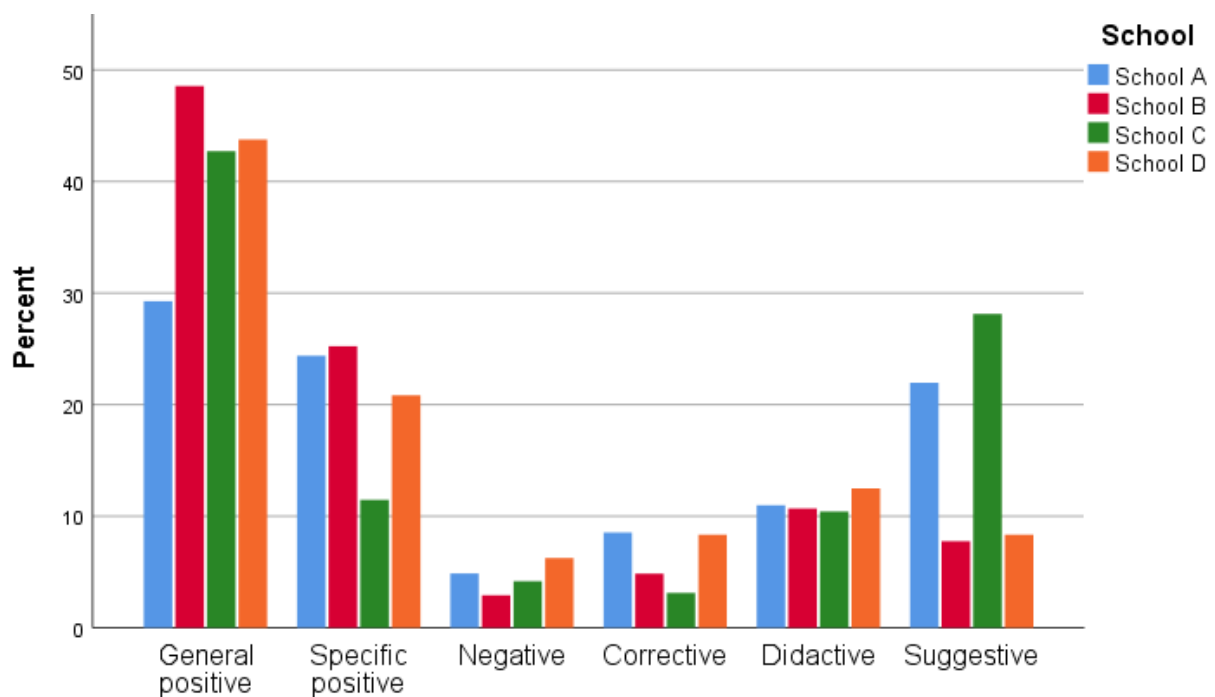
Figure 11. Percentage of Content-Focused Feedback-Type Tokens for each School.

Table 19. Chi-Square Comparisons of Feedback Types.

| Feedback focus	Comparison	Teacher	Chi-Square Test
	*Feedback focus * feedback type*	Teacher A	$X^2 (5) = 50.53, p < .001*$
		Teacher B	$X^2 (5) = 48.80, p < .001*$
		Teacher C	$X^2 (5) = 52.51, p < .001*$
		Teacher D	$X^2 (5) = 24.13, p < .001*$
Language	*Feedback type * OTOL/OSOL*		$X^2 (4) = 8.67, p = .07$
Content			$X^2 (5) = 7.60, p = .18$
Language	*Feedback type * school*		$X^2 (12) = 17.85, p = .12$
Content			$X^2 (15) = 29.09, p = .016*$
Language	*Feedback type * prescribed/free lesson*	Teacher A	$X^2 (3) = 0.50, p = .92$
		Teacher B	$X^2 (4) = 3.54, p = .48$
		Teacher C	$X^2 (3) = 0.74, p = .87$
		Teacher D	$X^2 (2) = 1.84, p = .40$
Content	*Feedback type * prescribed/free lesson*	Teacher A	$X^2 (5) = 5.89, p = .32$
		Teacher B	$X^2 (5) = 16.27, p = .006*$
		Teacher C	$X^2 (5) = 10.74, p = .057$
		Teacher D	$X^2 (5) = 4.17, p = .53$

No significant differences were found between the OTOL schools and the OSOL schools on the relative use of each of the feedback types to focus on language or on content, nor between the four teachers in how often they used each of the feedback types to give feedback on language, as reported in Table 19. All teachers used more corrective feedback than any other type when giving feedback on the English language use of their pupils, followed by suggestive feedback. None of the teachers gave any negative feedback on language. However, there was a significant difference between the teachers in the feedback categories they used to target content (see Table 19). When giving feedback on content, Teacher A gave significantly less general positive feedback and significantly more corrective and suggestive feedback, Teacher B used significantly more general and specific positive feedback and less suggestive feedback, Teacher C gave significantly less specific positive and corrective feedback and more suggestive feedback, and Teacher D used significantly less

suggestive feedback. Overall, most content feedback was of the general positive type, followed by specific positive and suggestive feedback. Negative feedback was used least often. Thus, although there were some individual differences that could not be explained by school type, a general pattern of use of the six types of feedback and their use to focus on language or on content is distinguishable.

Next, chi-square analyses were carried out to compare the feedback types in the two different lessons. There were no significant differences between the types of feedback given in each lesson that were focused on language or for any of the teachers, and for content for three out of four teachers (see Table 19). An exception was formed by Teacher B, who gave significantly more negative and corrective feedback on content during the prescribed lesson than during the free lesson. Thus, teachers were overall mostly consistent in the types of feedback they used in the two lessons.

In the interviews, the teachers also mentioned using forms of corrective feedback when pupils make mistakes in English. All teachers mentioned they usually deal with students' English language errors by repeating what they said in correct English, i.e. by giving *recasts*. Teacher C explains: "*Eigenlijk een beetje op een natuurlijke manier, wat je ook doet met je kinderen als je hoort dat ze een taalfout maken* [Actually in a kind of natural way, like what you also do when children make a language mistake in their first language]." Teacher B also mentions being against negative feedback on language production: "really dislike[s] to fix it, to say 'no, that's not how you say it.' What I do 98 percent of the time is just say it back correctly, just modelling the correct way of saying something. It's just a natural flow, it doesn't interrupt the flow of conversation." This seems to be especially important in the bilingual education setting, as Teacher B adds: "I want to know what they're thinking, because I'm their teacher, and not just their teacher of English, but I'm their teacher of everything. And then language is kind of the second thing that comes along. I very infrequently correct them explicitly, and then I do it very gently and quickly, because it also makes them freeze up." She exemplifies that in one of the observed lessons, "one of the kids was struggling with the word 'about', she said it in the Dutch way: 'It tells what this page is *over*'. In that case I just say it back the right way. Instead of saying, 'No, that is not how you say it', I say 'Oh, you mean…'" Teacher C explains about explicit correction that "*of ik het ook expliciet maak ligt aan de situatie. Als het nodig is dan leg je er nadruk op; als het niet nodig is dan leg je er geen nadruk op.* [whether I also make it explicit depends on the situation. If it's necessary, you will emphasise it, but if it is not necessary, you don't.]" Furthermore, Teacher A notes that "if it's explicit, if it's something I gave direction on and

they mess it up, I'll call them out on it ('wait a minute, what did you just say? Let's try again')," which is a form of suggestive feedback; "sometimes I want to break that habit, so if they constantly keep saying the wrong thing, I want to make them aware of it." This is also recognised by Teacher B, who adds an example of the use of didactic feedback in such a situation: "Now that we are studying verb tenses and speaking in the past, I might do some explicit correction by saying something like 'That's how you say it in the present tense, how would you say it in the past tense?'"
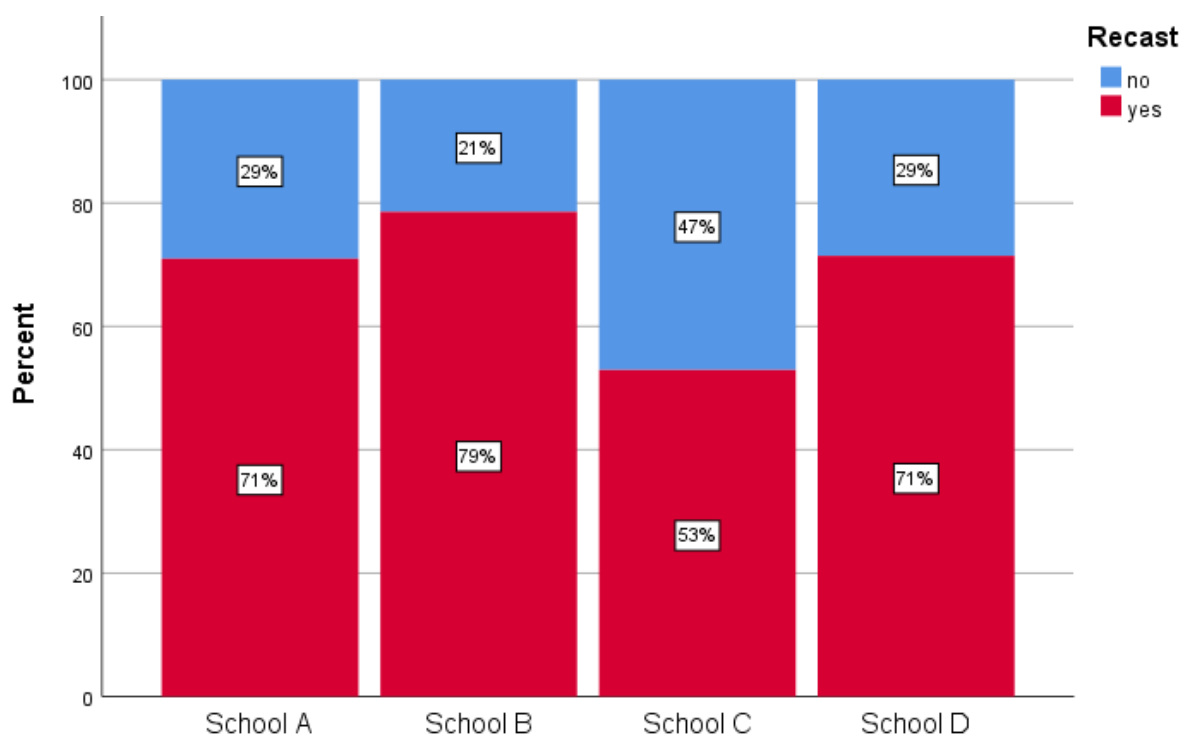
Figure 12. Proportion of Corrective Feedback on Language that is a Recast.

As all teachers named recasts of incorrect English and translations of Dutch utterances as important forms of feedback, the proportions of the teachers' corrective feedback on language that consisted of recasts (i.e. correct repetitions of pupils' utterances by teachers) and of recasts in which the child's Dutch utterance was repeated back to them in English, were calculated and are displayed in Figures 12 and 13, respectively. Indeed, for all teachers, over half of the corrective feedback on language consisted of recasts. Translation recasts made up 9 to 18 percent of all feedback on language by Teachers B, C, and D, and 67 percent by Teacher A. Given the earlier observation that much of the language-focused feedback is corrective feedback, and in line with the explanations given by the teachers in the interviews, recasts thus make up a large proportion of all feedback on language.
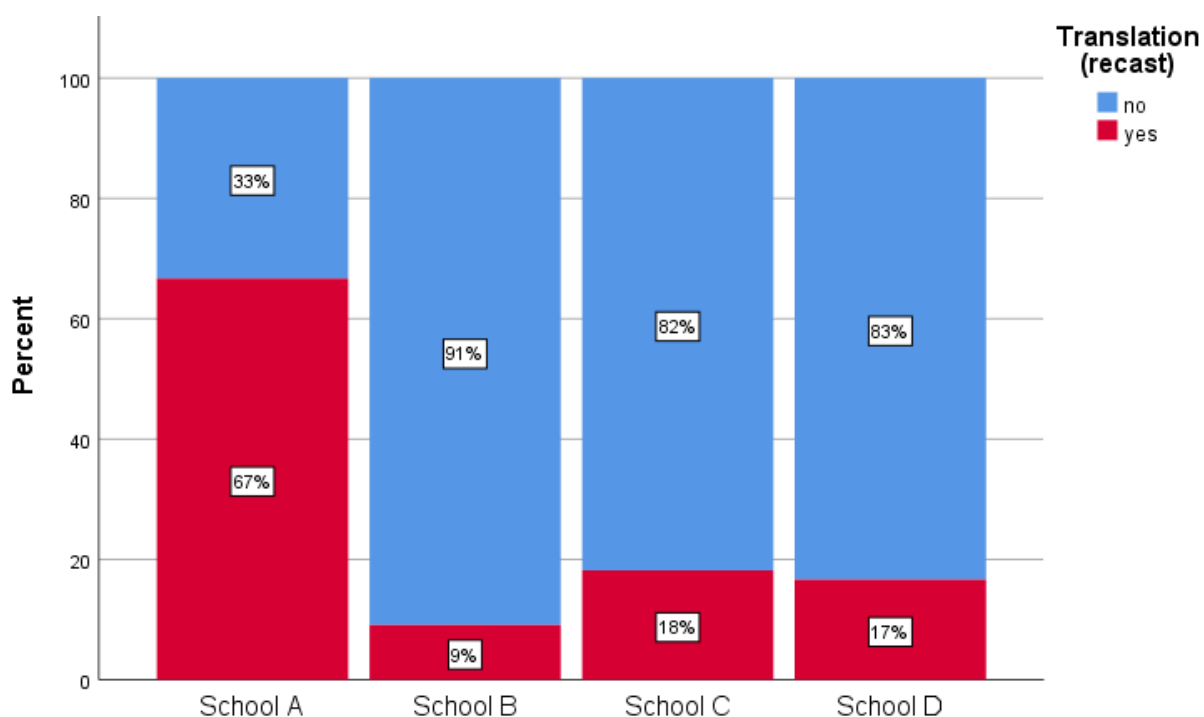
Figure 13. Proportion of Corrective Feedback on Language in which the Teacher Translates the Pupil's Dutch Utterance to English.

4.4 Use of the L1

When comparing the numbers of words spoken in each language, as reported in Table 20 and displayed in Table 21, it becomes apparent that students at all schools used the target language significantly less than their teachers. A comparison between the four teachers' language use reveals that native speaker Teacher A spoke significantly fewer of her total spoken words in the target language compared to her colleagues at the other three schools: X^2 (3) = 450.70, $p < .001$. Pupils at School A similarly spoke significantly less English than their peers at the other schools: X^2 (3) = 374.34, $p < .001$. Thus, individual differences between teachers seem to predict the language used by teacher and students rather than native language of the teacher.

Table 20. Chi-Square Comparison between Languages Spoken by Teachers and Students.

Comparison	Teacher	Chi-square test result
*Teacher/student * number*	Teacher A	$X^2 (1) = 805.62, p < .001*$
of words English/Dutch	Teacher B	$X^2 (1) = 286.99, p < .001*$
	Teacher C	$X^2 (1) = 387.04, p < .001*$
	Teacher D	$X^2 (1) = 216.39, p < .001*$

Table 21. Words Spoken in each Language in Absolute Numbers and Percentages[4].

	English		Dutch		Total	
	Teacher	Student	Teacher	Student	Teacher	Student
School A	4274 (96%)	615 (65%)	202 (5%)	327 (35%)	4487	948
School B	6174 (100%)	950 (95%)	0 (0%)	46 (5%)	6190	1001
School C	4169 (100%)	836 (90%)	0 (0%)	84 (9%)	4190	924
School D	3259 (98%)	330 (84%)	60 (2%)	62 (16%)	3332	395

In the interviews, the teachers expressed their views in important factors for their pupils' language acquisition. Interestingly, Teacher A, who used Dutch significantly more than her colleagues at other schools, expressed the importance she attributes to input: "I give a lot of input. I've switched my philosophy over the years, because I always thought that the kids mainly had to practice speaking a lot, to get good at it and feel comfortable in it. […] My take on it now is that kids need a lot of input, and a lot of repetition." In this way, "I want to make sure that they have enough leading up to be able to actually understand a higher-level input in English about the topic. And then, maybe be able to produce some output at a simple

---

[4] As explained in the method section of this paper, pupils' utterances falling in the *song/drill* category are excluded from this data (cf. Table 4).

level, what they learned" The teacher explains that she will use the language input in many different ways and contexts:

> "It's just when you're learning or acquiring a language, interacting with the kids. Speaking a lot about what I want them to learn about. For example, 'do you want the banana?', 'no, the banana is not yours', 'should I peel the banana for you?', then you're using all this language about a banana, and you never really have to say 'look, this is a banana.'"

Teacher A furthermore adds she values correct use of the language: "They need to hear the correct language models before they can produce it correctly. So, I don't spend a whole lot of time with getting them to talk to each other about topics, because I know they'll be saying things incorrectly and be struggling over their words." Contrarily, Teachers B and C stressed that pupils needed to use the language to reach their ultimate goal of being confident users of the language. Teacher B expresses that she is "not shooting for perfect English, I am shooting for fluency." She elaborates that the pupils "don't have to know all the right ways to say something. What I would love for them is to be comfortable. And to be able to just get it out, one way or the other." This goal is also shared by Teacher D, who aims for pupils to build vocabulary, and to be able to, as well as feel confident about, using it. Teacher C furthermore explicitly names language accuracy a next step, but she prioritises the pupils being confident language users: "And then the next step is to improve their levels, of course, but I think that at first, the most important thing is that they are secure enough to use it, to communicate in that language."

The views of the teachers on their pupils' use of Dutch and English also differed. The pupils at School A spoke most Dutch back to their teacher, which matches her explanation that the L1 has an important role in her classroom, in their way to building their English language skills:

> "Their discussion happens, but it's usually pretty structured. And if I want them to have a free discussion about something, I might say 'talk to each other about this, you can use Dutch, but try to use these words in English'. So that they mix languages a lot, or a little. […] I think it's good to use your mother tongue to stimulate growth and language acquisition."

The amount of Dutch spoken by the pupils may also have been affected by earlier years, as bilingual primary education is still a relatively new, hence still developing, area: "Last year I would speak more Dutch to them, mostly because of laziness and it seemed easier to do it in Dutch, but now I choose to speak English, even if it's more challenging for them at times, and

I haven't noticed that it is." The opposite pattern in language use seems to be the case for non-native Teacher C, who observes "*het voordeel dat ik hen ook voorgaande jaren Engels heb gegeven; dat ze naar mij toe ook meer gewend zijn om Engels te spreken. Ze vonden het ook in het begin van het jaar heel gek om Nederlands te praten tegen mij.* […the advantage of having taught these pupils English in previous years; they are used to speaking English to me. At the start of the year, they found it very odd to speak Dutch to me.]" Like Teacher A, Teacher D notes about the pupils' disuse of the English language:

> "*Ik ga ze niet forceren, zeg maar. Ik zie het een beetje als een stille periode dan. Omdat ze dan wel heel veel opnemen en luisteren en kijken. Ze zijn wel heel hard aan het werk dan, maar dan voelen ze zich nog niet vertrouwd genoeg om het al te spreken* [I am not going to force them, so to say. I view such cases as a kind of *silent period*, because they still absorb much, they are listening and watching. They are working very hard all the same, but they do not feel familiar enough with the language to speak it.]"

In contrast, Teachers B and C stress learning by doing, i.e. the pupils producing output and interacting as an important factor in language acquisition. Teacher C explains that bilingual education gives her students the opportunity to practice using the language in natural situations:

> *Want aan de ene kant heb je natuurlijk die* explicit knowledge*, dat wordt expliciet aangeleerd. Maar daarnaast moeten ze het ook gewoon kunnen toepassen in natuurlijke situaties, en daar zeker genoeg van zijn. Ook er genoeg oefening voor gehad hebben om dat te durven doen. En ik vind dat wel heel belangrijk, dat ze daar ook in de les de mogelijkheid voor krijgen. Dus door ze communicatieve opdrachten te geven en te laten samenwerken.* [Because on the one hand you have *explicit knowledge*, which is learned explicitly. But apart from that they need to be able to apply it in natural situations, and be confident enough about it. They also need to have had enough practice to dare to do so. And I find it very important they get the opportunity to do so in my lessons. So, by giving them communicative exercises and having them work together.]

She also compares it to regular, non-BPE primary education:

> "*Als je het vergelijkt met reguliere basisscholen, waar in groep 7/8 pas Engels wordt gegeven, daar wordt expliciet lesgegeven in hoe je je voor kunt stellen aan iemand, en dergelijke. Maar dat houdt niet automatisch in dat als je naar Engeland gaat, of als je op straat wordt aangesproken door een buitenlander, dat je het ook kunt en durft te*

gebruiken. En dat vind ik wel het hoofddoel van taal: een communicatiemiddel. Dus leer het ook in die natuurlijke situatie. [If you compare it to regular primary education, in which you start learning English around age 10-11; there, the pupils are being taught explicitly how to introduce themselves to others, among other things. But that does not automatically entail that when you go to England, or when a foreigner asks you something in your own country, you are able and confident enough to use the language. And I think that is the primary aim of language: A means of communication. So, you should also learn it in such a natural way."

This is also affirmed by Teacher B, who finds that "the only way to learn a language is to talk; you can't learn a language by not talking." She adds that "if you want to learn a language, you need to speak; if you want to learn to write, you've got to write; and if you want to learn to read, you need to read. So, I try to make most of my teaching not so much about me teaching, but about them doing, and me guiding." Furthermore, in clear contrast to Teachers A and D, she points out that expectations and helping the pupils reach those expectations are important factors in her teaching strategy: "I am scaffolding a lot, such as today is building up to having them be able to produce something themselves. But I do still think it's important to have high expectations, and high expectations of what they can produce – because they *can*, they *will* step up to it. They *can* write books, they *can* write, real writing." Yet, she, too, recognises that students may need their L1 in the process of constructing knowledge: "A lot of times they need to say it in Dutch first, because they can't think and make the English at the same time. So, once they know what they want to say, they can give it a go in English. Some of them can just think in English." Yet, for her, the pupils need to ultimately express that content in English:

"Sometimes, when a child doesn't know how to say something in English, I'll ask them to just say it in Dutch first, and then have the rest of the class help them put it into English, and then have the child that was struggling say it again. That way they're not just passive, but now that they've been handed the way to say it, they need to say it themselves."

Several teachers also mentioned the naturalness of the situation of speaking English with their interlocutor as a factor in their pupils' language choice. For example, native-speaker Teacher B declares that her "Dutch quite frequently isn't good enough to help them translate their thoughts into English," which is an additional reason for her to have the pupils help their struggling classmates translate their thoughts into English: "The good thing about me not speaking Dutch, is that they know that when they talk to me, they really have to try in

English." Teacher A also mentions that "creating genuine situations works the best, so interns come in from other countries that do not speak Dutch (e.g. America, Spain) so the kids *have* to speak English. So as a coordinator, I try to have that at school as much as possible." Yet, she finds that the usefulness of the students' L1 for their language learning process outweighs using this function herself: "at the same time I don't pretend that I don't speak Dutch. […] I think that if you are with them all the time, they have to be able to use their mother language to learn their second language." Teacher B illustrates that she has found a middle way in the pupils' being able to use their L1 to construct knowledge, and still ultimately using English with her and in the group:

> "The minute I'm not around, they switch into Dutch. And of course they do, because they are way more comfortable in Dutch. So when they're meeting in their small groups, I'm not making it a requirement that they speak in English. That's because they're building their understanding, and that's natural they would do it in their L1. Then, when they have to speak to the whole class, they must use English."

Non-native-speaker Teacher C exemplifies how the interlocutor has an effect on pupils' language choice and willingness and confidence to communicate in the target language:

> *[naam] is bijvoorbeeld een kind dat een hele sterke preferentie heeft voor Nederlands. Maar op het moment dat ze geconfronteerd wordt met een kind dat niet goed is in Nederlands, schakelt zij direct over naar het Engels. Dan is het misschien qua niveau wel simpel, maar ze doet het wel. Ze weet wel direct van 'deze situatie vraagt van mij…'.* [[Name] is an example of a child that has a strong preference for Dutch. But when confronted with a child who cannot speak Dutch well, she promptly switches to English. It might be simple in terms of language level, but she still does it. She immediately knows that this is what the situation asks of her.]

This teacher also recognises: "Dan is zo'n les nog best soms een beetje onnatuurlijk voor ze, omdat ik gewoon Nederlands kan. Terwijl, op het moment dat ze weten dat die ander dat niet of minder kan, ze het direct kunnen toepassen. [Such a lesson can then be a bit unnatural for them, because they know I also speak Dutch. Meanwhile, when they know their interlocutor speaks little to no Dutch, they swiftly switch to English.]"

4.4.1 Contexts for using the L1

To find out in what contexts teachers made use of the L1 instead of the target language, the numbers of words in each language and context were calculated and compared between schools, as displayed in Table 23. Due to the large within-group differences in the two types

of schools, the decision was made to focus the statistical analyses on the individual schools, rather than the types of schools.

For each of the speech functions, there was a significant difference between the language choices of the four teachers (see Table 22). What stands out in this data is that Teachers B and C did not use the students' L1 at all, while Teachers A and D used Dutch in some or all contexts. Teacher A used significantly more Dutch than the other teachers in IRF-sequences, Teacher D used significantly more Dutch than the other teachers when giving instructions and when giving explanations, and Teachers A and D both spoke significantly more words in Dutch than the other teachers when correcting students' behaviour. Furthermore, language choices were also examined between speech functions for both teachers who spoke Dutch during the observed lessons, i.e. Teachers A and D (see Table 22). Teacher A spoke significantly more Dutch in IRF-sequences and when correcting students' behaviour than when giving instructions or explanations:. Teacher D spoke significantly more Dutch while correcting pupils' behaviour and when giving explanations compared to the other speech functions, with almost a quarter of all behavioural corrections being done in Dutch.

Table 22. Chi-Square Comparison between Languages Spoken by Teachers and Students.

Teacher	Function of Teacher Talk	Comparison	Chi-square test result
	IRF-sequences	*School * number of words*	$X^2 (3) = 360.78, p < .001*$
	Instructions	*English/Dutch*	$X^2 (3) = 43.45, p < .001*$
	Explanations		$X^2 (3) = 32.82, p < .001*$
	Corrections		$X^2 (3) = 143.09, p < .001*$
Teacher A		*Teacher-talk function * number*	$X^2 (3) = 81.98, p < .001*$
Teacher D		*of words English/Dutch*	$X^2 (3) = 224.10, p < .001*$

Table 23. Numbers of Words Spoken by Teachers in English and Dutch in IRF-sequences, Instructions, Explanations, and Behaviour Corrections.

	School A		School B		School C		School D	
	English	Dutch	English	Dutch	English	Dutch	English	Dutch
IRF-sequence	2611 (94%)	177 (6%)	2571 (100%)	0 (0%)	2324 (100%)	0 (0%)	1600 (99%)	17 (1%)
Instruction	967 (100%)	0 (0%)	2542 (100%)	0 (0%)	1171 (100%)	0 (0%)	1176 (99%)	11 (1%)
Explanation	306 (100%)	1 (0%)	274 (100%)	0 (0%)	588 (100%)	0 (0%)	401 (97%)	13 (3%)
Correction	367 (94%)	24 (6%)	772 (100%)	0 (0%)	83 (100%)	0 (0%)	63 (77%)	19 (23%)

Qualitative analyses of the use of Dutch by Teachers A and D in the data revealed that there were overlapping as well as different situations in which the L1 was used by these two teachers. Firstly, the L1 seemed to have a didactic function. Teacher A regularly repeated Dutch answers back in Dutch; Teacher D did this once. Teacher A actively tried to let the pupils connect what they were learning to their existing knowledge in Dutch. Teacher D also used once translated a target concept in English to Dutch, to help the pupils connect the new knowledge to their prior knowledge. Another use both teachers had for the L1 was to help with classroom management, i.e. when correcting students' behaviour. At School A it was noticeable that as students got more distracted towards the end of the lesson, more use was made of Dutch in behaviour corrections. Finally, for Teacher D, some of the Dutch that was used seemed to be caused by a lacking language proficiency. This teacher seemed to be more impulsive and think faster in Dutch. This was reflected in impulses that seemed to be directed at herself rather than the children (e.g. "*Waarom gaat 'ie nou niet? Ohja.* [Why isn't it working? Oh, right.]," $n = 3$), use of Dutch non-content words of which she did not seem to be aware she was using them (e.g. "*het is* [it is]," $n = 3$), and being unable to find the English word she was looking for (i.e. "*eh… overleg, dus* [eh… so, discuss]," $n = 1$).

In the interview, Teacher D indeed mentions that her own language fluency sometimes stands in her way during lessons: "*Ik struikel soms nog weleens over mijn woorden. En dat is dan mijn eigen onzekerheid. Dus ik vind dat dan soms nog weleens lastig.* [I sometimes stumble over my words. That has to do with my own insecurity. So sometimes I find that difficult.]" She adds that "*ik merk wel dat het mij meer focus kost om een Engelse les te geven, dan dat ik bijvoorbeeld gewoon Nederlands of spelling geef* [I notice it takes more focus to teach English lessons, than for instance when teaching Dutch or spelling]." Yet, this teacher expressed she aimed to teach at least 95% of these lessons in English. A reason for her to use Dutch is when she knows a pupil is really struggles to keep up. She exemplifies:

> "*Ik heb nu ook bijvoorbeeld een dyslectisch meisje in de klas, en die vindt Engels echt heel erg lastig. En die kwam nu ook nog later binnen. Dus dan moet ik soms echt wel even in het Nederlands iets… Maar dat probeer ik wel zo veel mogelijk te voorkomen, om in het Nederlands over te schakelen.* [For instance, I have a dyslexic pupil who really struggles with English, and who also entered later during this lesson. So, in such cases, I sometimes really need to use Dutch. But I try to avoid switching to Dutch as much as possible.]"

Teacher A similarly mentions that although she "speak[s] English most of the time, I use Dutch for extra support." She explains her reasons for using Dutch for behavioural corrections:

> "Sometimes to get them listening, so if I really want them to take me seriously; sometimes I think in English it's easier to tune out, if I'm constantly speaking English. So sometimes to get them back in, I'll say something in Dutch. If it's a behaviour issue, and I feel like they really need to know what I'm talking about and they need to be able to express themselves about what happened, I'll switch to Dutch."

Furthermore, Teacher A uses the pupils' L1 to fulfil didactic functions: "And I use it to make comparisons a lot between the Dutch and English language ('what would this be called in Dutch?'). And to check understanding." She illustrates the importance of this function with an example from her lesson: "So I was teaching about 'precipitation', but half of the kids didn't know the word '*neerslag*'. I want to make sure we're not just skipping over Dutch words they don't know, and only teach them the English. A lot of it is content learning and these may be words that they don't come in contact with on a regular basis." Similarly, this teacher also lets the pupils use the L1 to fulfil a similar function: "And sometimes, if I really want to get an explanation, and know that they understand, I'll tell them to say it in Dutch. Because I don't want them to be struggling with how to say it in English at the cost of being able to show that they have the knowledge." This teacher thus considers the risk of exclusively using English during English-taught lessons too great on the content learning of the pupils, resulting in more Dutch used in her lessons.

As reflected in the data, Teachers B and C both mentioned in the interview that they exclusively use English during English-taught lessons. Teacher B adds as an exception that she "sometimes say[s] '*hoe zeg je dat in het Nederlands?* [How do you say that in Dutch?]*'* That's about the only thing I ever say in Dutch, to do a check for understanding. Other than that it's all in English." Teacher C names as an exception that when she refers to method names in Dutch, she does not translate these to English, and that there are a few words that do not have a good English counterpart and she thus uses the Dutch term, adding: "*Dus dat heb ik ook uitgelegd aan de kinderen: ik ben even bij de Engelse collega's langsgegaan, van goh, hoe zou jij dat zeggen in het Engels? Maar dat bestaat niet in het Engels, dus dat leg je dan uit.* [So, I also explained the children: I went and asked my English colleagues how they would say this in English. But the concept does not exist in English, so you explain that to the pupils]."

5. Discussion and conclusion

In the present study, the interaction between teachers and their pupils in CLIL lessons at bilingual primary schools in the Netherlands was studied. Use was made of the criteria set out in earlier research by Wieringa (2014), who studied teacher-student interaction in a related setting, i.e. Early Foreign Language Education (EFLE) in Dutch primary schools. Identical to Wierenga's (2014) research, verbal dominance of teachers and students, types of questions posed by teachers, and feedback given by teachers were analysed to find out more about the characteristics of this relatively new educational method. Furthermore, the use and function of the pupils' first language by the teachers was investigated. Comparisons between teachers' beliefs with regards to their pupils' language acquisition, as expressed in the interviews, and their teaching practices were made where relevant.

To answer the first research question, i.e. '*what is the verbal dominance of teachers versus students during the plenary parts of lessons in bilingual primary schools?*', the degree of initiative taken by teachers as compared to students, and the number of words spoken by each were compared. Teachers consistently took more initiative than pupils in each lesson and lesson part. Teachers also made more of their initiations in the target language and less in the pupils' L1 as compared to their pupils. This is in line with earlier research in the EFLE setting (Wierenga, 2014). Furthermore, the teachers spoke more words than their pupils. With 69-74% of all words in IRF-sequences and 82-85% of the total number of words spoken by the teachers, and with 88-96% of all initiative being taken by the teachers, the verbal dominance of teachers during the plenary parts of these lessons was significantly higher than that of the students. This was also in line with findings in earlier research in EFLE (Wierenga, 2014).

To answer the second research question, i.e. '*what types of questions do teachers ask during the plenary parts of lessons in bilingual primary schools?*', the frequency of use of three predetermined question categories, and the length of the students' responses to each of these question types were measured and analysed. Analyses of the pupils' responses showed that *open/reference questions* generated the longest responses, followed by *closed/display questions*, and *yes/no questions* generated the shortest responses from pupils. This is supported by earlier findings that open questions lead to more as well as more complex language production by pupils (Graesser, Person & Hu, 2002; as cited by Wierenga, 2014). This may have important implications for the quantity of output produced by pupils, which in turn might affect their language acquisition. However, more research is needed to study these effects, as well as the quality of the output, in this context in more detail. The four

participating teachers differed from each other in the types of questions they asked. All teachers asked fewer *yes/no questions* compared to the other question types in both lessons combined; for two teachers there was an additional least-used question type category, i.e. *closed/display* for one of them, and *open/reference* for the other. Three teachers used predominantly *closed/display questions*, while the other teacher used mostly *open/reference questions*. This is largely in line with findings by Wierenga (2014), in whose study similarly three out of four participating teachers used mostly *open/reference questions* and least *yes/no questions*. As noted by Wierenga (2014), this is contrary to findings by Yang (2010) and Faharian and Rezaee (2012), who both found that teachers asked mostly *closed/display questions*; however, participating teachers in those studies were inexperienced, which may as such have influenced the types of questions they asked.

Three out of four teachers asked significantly different questions in the two observed lessons. Both when controlling for activity (i.e. in the prescribed lesson) and when letting the teacher decide the activities (i.e. in the free lesson), all teachers used fewer or the same number of *yes/no questions* compared to the other question types. One teacher asked mostly *open/reference questions* in both lessons, one teacher asked mostly *closed/display questions* in both lessons, one teacher asked mostly *open/reference questions* in the prescribed lesson and *closed/display questions* in the free lesson, and one teacher asked mostly *closed/display questions* in the prescribed lesson and asked about the same number of each of the question types in the free lesson. There is thus a clear tendency in which question type is least often used, i.e. *yes/no questions*, but not in the type that is most often used by teachers.

The individual differences between the teachers' use of the question types could not be explained by school type and thus native language of the teacher. However, other factors related to the design of the lessons might provide an explanation. Three teachers increased their use of *open/reference questions* during the prescribed lesson compared to the free lesson; two of these teachers decreased their use of *closed/display questions* in this lesson, while the other teacher decreased their use of *yes/no questions*. Although teachers differ in their use of the three question types, their use is not completely arbitrary, as the increase in use of *open/reference questions* at these three schools, and the lack thereof at the other school, can be explained by looking at the lesson content in the prescribed lessons: Pupils were asked to guess the question their classmates had written an answer to. At the school where no increased use of *open/reference questions* was observed in the prescribed lesson, pupils had been asked to describe one of the text features. This meant that their classmates were asked the *closed question*: "which text feature are they describing," instead of the *open question*:

"what was their question." Thus, in this lesson, the content of the lesson had a large influence on the types of questions asked.

Furthermore, some teachers made regular use of 'language drills', in which pupils collectively repeated a word or sentence in English, to practice speaking English in a highly controlled setting. This form of language production was elicited by the teacher in *display questions*, and was linked to the teachers' views on language acquisition: One of the teachers who used this type of question the most in the free lesson viewed accuracy as an important factor in language production, the other teacher viewed vocabulary learning as an important factor. The other two teachers, in contrast, considered language fluency and confidence to speak the language as more important. The only teacher who did not have any data that was categorised as this type of utterance, was the only teacher who used more *open/reference questions* than *closed/display questions* in the two lessons combined. It thus seems that lesson content and teachers' philosophies on their pupils' language acquisition both have an impact on the question types used by the teachers.

Finally, three out of four teachers also changed the question types they asked throughout the lessons. During the introduction, these teachers increased their use of *closed/display questions* and decreased their use of *open/reference questions*. During the discussion, they decreased their use of *closed/display questions*, while they increased their use of one or both of the other question categories. It thus appears that the types of questions asked by teachers does not remain the same throughout the lesson, but it is related to the lesson part for most of the teachers.

To answer the third research question, i.e. '*what types of feedback do teachers provide during the plenary parts of lessons in bilingual primary schools?*', the relative focus of feedback on language and on content, the use of six predetermined feedback categories, and the use of neutral feedback were analysed. At all schools, more feedback was given on the content of the responses than on language use. This is not surprising, given that in bilingual education, the foreign language is used as a tool for learning content, and language is simultaneously learned by using it (De Graaff, 2015), and it is in line with observations and teachers' views in the bilingual secondary school context (Schuitemaker-King, 2013). There were significant differences in the use of neutral feedback between teachers; the teacher who made most use of 'language drills' gave significantly more neutral feedback than the other teachers. This may have been related, as teachers usually did not give feedback after the class had repeated the utterance, but rather often gave neutral feedback (e.g. "okay") before moving on. This teacher also had the lowest English language proficiency, which could also possibly

have affected the amount of corrective feedback on language given by her, as she seemed oblivious of some of the language mistakes her students made and even asked the rest of the class to repeat erroneous utterances. However, studying these observations in more detail fell outside the scope of the present study; more research is needed to support these claims. Significantly less neutral feedback was given by the native speaker who viewed students' output as one of the most important elements in language acquisition (i.e. Teacher B). Although arguably this type of feedback has a purely communicative function and might as such be used to encourage natural use of the language, and be used more rather than less to promote output, this teacher also stressed being her students' "teacher of everything," indicating that she was first and foremost a content teacher, but also corrected and taught language along the way. When feedback contained any kind of truth value on language or content, it was not considered neutral, but it could still serve communicative functions in addition to expressing this truth value and correcting errors. There were also differences between the use of neutral feedback in different lessons and in different lesson parts for some but not all teachers, but no conclusive patterns have been found of the exact use of this type of feedback.

Teachers did not change their feedback focus on language versus content in the two different lessons, and only one teacher changed her focus in the different lesson parts (i.e. Teacher A). This teacher, who expressed valuing language accuracy in her pupils' output, gave more feedback on language and less on content during the introduction compared to the rest of the lesson. When including neutral feedback, more significant differences were found; therefore, as well as due to its qualitatively different nature in lacking a truth value, it is advisable to exclude neutral feedback from the analyses of feedback focus in future research. Although it remains unclear why one teacher changed her feedback focus throughout the lesson, overall teachers thus seemed to remain constant in their feedback focus in different lessons and throughout different parts of lessons.

Although all teachers gave more feedback on content than on language, some differences existed between the teachers. The teachers who gave significantly less feedback on language compared to the other teachers were the same teachers who exclusively used English in their lessons[5], and who mentioned in the interviews that they valued confidence in

---

[5] In the present study, it was not investigated whether the teachers who gave less feedback on language in fact could have given language feedback and thus ignored more, or whether, for instance, the pupils were more proficient due to more oral practice, and feedback on language was thus not necessary.

speaking the language over accuracy; the teacher who gave significantly more feedback on language and less on content was the same teacher who spoke significantly more Dutch, and whose pupils also spoke more Dutch, compared to teachers and pupils at the other schools. This teacher explained in the interview that she attached less importance to having the pupils produce output than on abundant correct English input, but that she did stress accuracy in language production. Furthermore, this was also the teacher who changed her feedback focus significantly throughout the lesson, as she gave more language-focused feedback in the introduction than in the rest of the lesson.

Like in earlier research, the present study found that from the distinguishable feedback types, positive feedback was given most often, while negative feedback was given least often (Wierenga, 2014). General positive feedback was given more frequently than specific positive feedback. When combining these two types of feedback, as was done in earlier research, this category is especially prevalent. The six feedback types were used in different frequencies to focus on language versus on content. Teachers showed largely similar patterns of their use for each of these foci. When giving feedback on content, general positive feedback was the most occurring type, followed by specific positive and suggestive feedback. Feedback on language consisted predominantly of corrective feedback, followed by suggestive feedback. No negative feedback was given on language. This is in line with Wierenga's (2014) finding that teachers gave more corrective feedback during a lesson in L2 English than during a lesson in L1 Dutch. Wierenga (2014) further points out that corrective feedback can be an important tool to promote language acquisition (e.g. Lightbrown & Spada, 1990; Lyster & Ranta, 1997; Schuitemaker-King, 2013). In the interviews, teachers expressed that they often corrected their students' language errors through recasts, and indeed, the major part or corrective feedback on language consisted of recasts. This is in contrast with Schuitemaker-King's (2013) observation that CLIL teachers in bilingual secondary schools in the Netherlands used recasts in only 40% of all lessons, versus 64% in EFL lessons in non-bilingual schools and 43% in EFL lessons in bilingual schools). It was, however, in line with findings from immersion education at a primary school in Canada, where most teachers used recasts for over half of their feedback on language, although importantly, this was found to be the most ineffective feedback category that did not lead to uptake by the learners in over a third of all cases (Lyster & Ranta, 1997). Perhaps, teachers in bilingual or immersion education at the primary level use recasts more often than those who teach at the secondary level. A specific type of recast, in which students' Dutch answers were repeated by the teacher in English, was also named by the participating teachers as a strategy to deal with pupils' use of Dutch. This

made up two thirds of the corrective feedback of the teacher who gave significantly more feedback on language, at the school where pupils spoke the most Dutch during the plenary parts of the CLIL lessons (versus up to 18% at the other schools); this may thus be an explanation for the statistical difference in feedback focus of this teacher as compared to the others.

To answer the fourth research question, i.e. '*what is the frequency and function of the students' L1 during the plenary parts of CLIL lessons?*', the number of words spoken in each language by teachers and pupils was compared, as well as the number of words teachers spoke in each language to fulfil various speech functions; furthermore, teachers' utterances in which they spoke Dutch were qualitatively analysed, and compared to teachers' views as expressed in the interviews. It seemed to be the case that the more teachers spoke Dutch, the more their pupils spoke Dutch during the CLIL lessons, which is in line with reported practices in foreign language classes in secondary education (Haijma, 2013). However, it is not clear whether the teachers' use of the L1 was the result or the cause of the pupils' use of the L1, or both. It is important to note here, though, that pupils' disruptions were only included if they led to interaction with the teacher. This could have influenced the number of Dutch words that were represented in the data, especially at School B, where the teacher had a limited proficiency of Dutch and the pupils spoke the lowest number of Dutch words during the observed lessons. Two of the teachers, of which one was a native speaker of English and the other was not, exclusively spoke English throughout both lessons, which could be linked to their views on language acquisition. The two teachers who spoke Dutch during their lessons, who were similarly a native speaker and a non-native speaker of English, used Dutch to correct students' behaviour, to fulfil a didactic function (e.g. to help pupils build connections to prior knowledge in the L1 or to check students' understanding of the material), and in the case of the non-native speaker, as a result of the teacher's own proficiency in the target language. Of these two teachers, most Dutch was spoken by the native English speaker, who mentioned in the interview that she found it important to use the L1 both for supporting the acquisition of the target language and to ascertain learning the target content knowledge, which also entailed avoiding the possibility that pupils did not know or learn words related to the content in their L1 Dutch. The non-native English speaker furthermore mentioned that she used Dutch when pupils particularly struggled with the English language, for which a dyslexic pupil served as an example, and admitted to struggling with using the language sometimes. In her study in the early foreign language education context, Wierenga (2014) similarly found that teachers with a lower language proficiency did not manage to speak exclusively English

during instructions, based on which she recommended that teachers should have an English language level of at least B2 following the CEFR. Moreover, Jenneskins et al. (2020) report that in the national study of which the present study used a subset of the data, children at schools where the teacher had an English CEFR-level of B2 or higher had a significantly larger receptive and productive vocabulary and better perceptive grammar scores compared to children with less proficient teachers. The same recommendation of a minimum English proficiency of CEFR-level B2 for CLIL teachers in bilingual primary education would be given based on the results in the present study.

In conclusion, the verbal interaction between teachers and pupils in the plenary parts of CLIL lessons in the bilingual primary education (BPE) context that was investigated in the present study, revealed both tendencies of all participating teachers and differences between them. Differences could not be explained by looking at the school type, and thus the native language of the teachers; instead, differences in use of the L1, in the types of questions that were asked, and the feedback focus of the participating teachers seem to be related to their views on factors that positively affect their pupils' language acquisition. This finding has important implications, as educating BPE teachers on language acquisition theories and research findings might therefore have a substantial effect on the quality of their teaching practices and the language acquisition of their pupils, which seems to be supported by one teacher's comment that she changed her view and teaching strategies based on a course on language acquisition.

However, as examining the linguistic quality (e.g. accuracy, complexity) of teachers' and students' utterances fell outside the scope of the present study, more research is needed to learn more about factors that positively influence the language acquisition of pupils in the bilingual primary education setting. For example, future research could investigate the relationship between the teachers' question types and feedback, and the quality of the language produced by pupils (e.g. complexity of sentences, variety of vocabulary, or accuracy of grammar or pronunciation), as well as their speaking confidence. Another limitation to this study is the small sample of teachers and schools that was compared. This gave enough data to find some preliminary results and come up with theories to explain the results; however, more research is needed to support these possible explanations. Furthermore, the present study focused on language used in plenary settings; future research could focus on factors influencing language choice and quality of language production between pupils during group work, as this was indicated by teachers to make up an essential part of their lessons. Also, in the present study, interruptions of pupils were only included in the analyses if they led to

interaction with the teacher. Future research could investigate the language choice and quality of this type of spontaneous reactions from children, as this may be telling of the ease with which they are able to think in and promptly apply the foreign language. It might then also be interesting to take the teacher's proficiency in the pupils' L1 into account, as the interlocutor and their ability to speak each language were factors that were named in the interviews as influential on the pupils' language choice and willingness to communicate in the language. Finally, based on the recommendation done in this study to educate BPE teachers on factors promoting language acquisition in pupils, it would be interesting to study the effects of such interventions on their implementation and as such on the quality of the teachers' lessons.

References

"84% of primary school children study foreign languages" (2017, 25 September). *Eurostat*. Retrieved from https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/EDN-20170925-1

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism 7*(1), 71-87. doi:10.1177/13670069030070010501

Costache, O. (2017). English Language Use in Bilingual Primary Schools: The Relation between Teacher Input and Children's Vocabulary Knowledge (Master's Thesis). Retrieved from Utrecht University Thesis Archive.

Council of Europe. (2001). *The Common European Framework of Reference for Languages*. Cambridge, U.K.: Cambridge University Press. Retrieved from https://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

Cullen, (1998). Teacher talk and the classroom context. *ELT Journal*, *52*(3), 179-187. doi:10.1093/elt/52.3.179

Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Review of Educational Research*, *49*, 222-251. doi:10.2307/1169960

Damhuis, R. (2008). *Gesprekken zijn de spil van het onderwijs [Conversations form the core of education]*. Taalbeleid in uitvoering [Implementation of language policies], lectoral speech. Utrecht, The Netherlands: Marnix Academie. Retrieved from https://www.marnixacademie.nl/Portals/0/Lectorale_rede_Damhuis.pdf

De Graaff, R. (2015). Tweetalig onderwijs en de taal van het vak [Bilingual education and the subject language]. *De Cascade, Tijdschrift voor onderwijs, onderzoek & ontwikkeling*, *12*(30), 26-29. Retrieved from https://dspace.library.uu.nl/bitstream/handle/1874/320699/pages.pdf?sequence=1

De Graaff, R., & Costache, O. (2020). Current developments in bilingual primary education in the Netherlands. In F. Li, K. E. Pollock, & R. Gibb (Eds.), *Child Bilingualism and Second Language Learning: Multidisciplinary perspectives* (pp. 137-165). Amsterdam, The Netherlands: John Benjamin Publishing Company. doi:10.1075/bpa.10.08gra

Dönszelmann, S. (2018). Doeltaal Leertaal: Over bibberen aan de rand van het taalbad en docenten-taalbewustzijn. *Levende Talen Magazine*, *105*(special), 36-37. Retrieved from http://www.lt-tijdschriften.nl/ojs/index.php/ltm/article/view/1814

Farahian, M., & Rezaee, M. (2012). A case study of an EFL teacher's type of questions: an investigation into classroom interaction. *Procedia – Social and Behavioral Sciences*, *47*, 161- 167. doi:10.1016/j.sbspro.2012.06.631

"Foreign language learning statistics" (2018, September). *Eurostat*. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Foreign_language _learning_statistics#Primary_education

Goriot, C. M. M. (2019). *Early-English education works no miracles: Cognitive and linguistic development in mainstream, early-English, and bilingual primary-school pupils in the Netherlands* (Doctoral dissertation). Retrieved from https://repository.ubn.ru.nl/bitstream/handle/2066/201335/201335.pdf?sequence=1

Haijma, A. (2013). Duiken in een taalbad: Onderzoek naar het gebruik van doeltaal als voertaal. *Levende Talen Tijdschrift*, *14*(3), 27-40. Retrieved from http://lt-tijdschriften.nl/ojs/index.php/ltt/article/view/543

Hall, J. K. (2010). Interaction as method and result of language learning. *Language Teaching, 43*, 202-215. doi:10.1017/S0261444809005722

Houben, E. (2017). *CLIL in het basisonderwijs [CLIL in primary education]* (Master's Thesis). Retrieved from Utrecht University Thesis Archive.

Hummel, K. M. (2014). *Introducing Second Language Acquisition: Perspectives and Practises*. Chichester, U.K.: John Wiley & Sons, Inc.

Jenniskens, T., Leest, B., Wolbers, M., Krikhaar, E., Teunissen, C., De Graaff, R., Unsworth, S., & Coppens, K. (2018). *Evaluatie pilot Tweetalig Primair Onderwijs: Vervolgmeting schooljaar 2016/17 [Evaluation pilot Bilingual Primary Education: Follow-up Measurement School Year 2016/17]*. Retrieved from https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2018/04/01 /evaluatie-pilot-tweetalig-primair-onderwijs/Evaluatie+pilot+Tweetalig+Primair +Onderwijs.pdf

Jenniskens, T., Leest, B., Wolbers, M., Bruggink, M., Krikhaar, E., De Graaff, R., Spätgens, T., & Unsworth, S. (2020). *Evaluatie pilot Tweetalig Primair Onderwijs: Vervolgmeting schooljaar 2018/2019 [Evaluation pilot Bilingual Primary Education: Follow-up Measurement School Year 2018-2019]*. Retrieved from https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2020/03/31 /evaluatie-pilot-tweetalig-primair-onderwijs/Rapport+tussenmeting+pilot+tweetalig +primair+onderwijs+groep+5.pdf

Lightbown, P., & Spada, N. (1990). Focus-on-Form and Corrective Feedback in Communicative Language Teaching. *Studies in Second Language Acquisition*, *12*, 429-448. doi:10.1017/S0272263100009517

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake; negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *20*, 37-66. doi:10.1017/S0272263197001034

Menegale, M. (2008). Expanding teacher-student interaction through more effective classroom questions: from traditional teacher-fronted lessons to student-centred lessons in CLIL. In C. M. Coonan (Ed.), *CLIL e l'Apprendimento Delle Lingue: Le Sfide del Nuovo Ambiente di Apprendimento* (pp. 105-128). Venezia: Libreria Editrice Cafoscarina. Retrieved from http://157.138.8.12/jspui/bitstream/11707/736/1 /05Menegale.pdf

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *29*(4), 578-596. doi:10.1093/applin/amm056

Nuffic (2018). Internationalisering in beeld 2018: Feiten en cijfers uit het onderwijs [Internationalisation in the picture 2018: Facts and figures in education]. Retrieved from https://www.nuffic.nl/documents/887/internationalisering-in-beeld-2018.pdf

Nuffic (2020). Dutch primary schools in Nuffic networks. Retrieved February 19, 2021, from https://www.nuffic.nl/en/subjects/facts-and-figures/dutch-primary-schools-nuffic-networks

Paradis, J. (2008). Early bilingual and multilingual acquisition. In P. Auer & L. Wei (Eds.), *Handbooks of applied linguistics: Vol. 5. Handbook of Multilingualism and Multilingual Communication* (pp. 15–44). Berlin: Mouton de Gruyter. Retrieved from https://ebookcentral.proquest.com/lib/uunl/detail.action?docID=364686

Phillipson, R. (2008). Language Policy and Education in the European Union. In S. May (Ed.), *Language policy and political issues in education* (pp. 255-265). New York, N.Y.: Springer. Retrieved from http://www.språkförsvaret.se/sf/fileadmin/PDF /Langpol_education.pdf

Sanz, C. (2000). Bilingual education enhances third language acquisition: Evidence from Catalonia. *Applied Psycholinguistics 21*, 23–44. Retrieved from https://www.researchgate.net/profile/Cristina_Sanz3/publication/231941963_Sanz_C _2000_Bilingual_education_enhances_third_language_acquisition_Evidence_from _Catalonia_Applied_Psycholinguistics_21_23-44/links/004635181a62a1bdd1000000

/Sanz-C-2000-Bilingual-education-enhances-third-language-acquisition-Evidence
-from-Catalonia-Applied-Psycholinguistics-21-23-44.pdf

Saxton, M. (2010). *Child Language: Acquisition and development*. London, U.K.: Sage.

Schmiedtová, B., Von Stutterheim, C., & Carroll, M. (2011). Language-specific Patterns in
Event Construal of Advanced Second Language Speakers. In A. Pavlenko (Ed.),
Thinking and Speaking in Two Languages (pp. 66-107). Bristol, UK: Short Run Press
Ltd.

Schuitemaker-King, J. (2013). Giving corrective feedback in CLIL and EFL classes. *Levende
Talen Tijdschrift*, *2*, 3-10. Retrieved from https://lt-tijdschriften.nl/ojs/index.php/ltt
/article/viewFile/246/240

Spada, N. (2015). SLA research and L2 pedagogy: Misapplications and questions of
relevance. *Language Teaching 48*(1), 69−81. doi:10.1017/S026144481200050X

Swain, M. (2008). The output hypothesis: Its history and its future. *Foreign language
teaching and research*, *40*(1), 45-50. Retrieved from http://www.celea.org.cn/2007
/keynote/ppt/Merrill%20Swain.pdf

Swain, M., & Lapkin, S. (1982). *Evaluating Bilingual Education: A Canadian Case Study*.
Clevedon, U.K.: Multilingual Matters. Retrieved from http://search.ebscohost.com
.proxy.library.uu.nl/login.aspx?direct=true&db=nlebk&AN=18907&site=ehost-live

Tribushinina, E. (2017). Bilingualism [PowerPoint slides]. Retrieved from Blackboard.

Van den Broek, E., De Graaff, R., Unsworth, S., & Van der Zee, V. (2014). *Voorstudie Pilot
Tweetalig Primair Onderwijs [Pre-study Pilot Bilingual Primary Education]*.
Groningen, The Netherlands: GION, Rijksuniversiteit Groningen.

Van Gool, C. (2004). Tweetalig onderwijs: De Europese dimensie in het onderwijs [Bilingual
education: The European dimension in education]. *Levende Talen Magazine*, *91*(6),
26-28. Retrieved from http://www.lt-tijdschriften.nl/ojs/index.php/ltm/article
/download/714/705

Vonk, A. (2007, 23 January). "Tweetalig succesnummer" [Bilingual success case]
[Newspaper article]. *Volkskrant*. Retrieved from https://www.volkskrant.nl/nieuws-
achtergrond/tweetalig-succesnummer~b7b2b026/

Wierenga, J. (2014). *Engels als instructietaal in het basisonderwijs: Een kwalitatief
onderzoek naar verbale interactie tijdens Engelse en Nederlandse instructie [English
as language of instruction in primary education: A qualitative study into verbal
interaction during English and Dutch instruction]* (Master's Thesis). Retrieved from
Utrecht University Thesis Archive.

Wilhelm, F. (2018). Foreign language teaching and learning in the Netherlands 1500–2000: An overview. *The Language Learning Journal*, 46(1), 17-27. doi:10.1080/09571736 .2017.1382053

Yang, C. C. R. (2010). Teacher questions in second language classrooms: An investigation of three case studies. *Asian EFL Journal*, *12*(1), 181-201. Retrieved from https://70.40 .196.162/PDF/March-2010.pdf

Appendix A: Lesson descriptions[6]

School A

Topic: Water cycle

1. Teacher plays sound files, pupils are asked to guess what sound it is, using the phrase "I think I hear…". All sounds are themed around water. After going through all sound files, the teacher shows the right answers, which have to be matched to the right sound in another plenary round, again using the same sentence starter.
2. Children are shown water-themed pictures and are asked to tell what they depict. (Creek, river, ocean, lake, pond, …).
3. **Prescribed lesson** activity
   Children are divided into groups. Each group is given a question. Group members take turns to write answers to the question. Then, each group reads their answers out to the rest of the class, who then guess their question.

Break, children go outside

4. Children get a piece of paper with a word on it. A child is asked what word they have. After answering, they are invited to ask another pupil what word they have (i.e. 'evaporation', …). This goes on for a while, while the teacher tries to set up a video on the water cycle. (Due to some technical issues, this takes longer than anticipated by the teacher, and the class gets a bit noisy at times.)
5. Children watch a short video on the water cycle, including the words they have on their pieces of paper. They are instructed to listen for their words and are told that they will be asked what new things they have learned about the water cycle from the video.
6. The video is discussed plenarily, focussing on the children grasping the concepts and know the words: Evaporation, condensation, precipitation.
7. Children are divided into three groups (evaporation, condensation, and precipitation) and play out the water cycle, while the teacher narrates the story. Children have to say their word out loud and make corresponding movements, in a total physical response (TPR) style activity.

School B

Topic: Text features

1. The words 'text', 'feature', and 'text feature' are discussed.
2. **Prescribed lesson** activity. A difference between this lesson and the fixed lesson at other schools is that here, pupils are only allowed to write one word at a time, and together they need to form sentences. Like at other schools, they are only allowed to talk by asking their group mates for help if they are stuck. They do a test round by plenarily taking turns in adding a word, trying to describe 'title page' without using the words 'title' and 'page'. (While waiting for the last group to finish their answer, pupils sing an English song with their teacher.)

---

[6] Based on Jenniskens et al. (2020).

3. Teacher gives each group a two-page spread from a non-fiction book. The groups are asked to identify as many text features on their pages as they can. The answers are discussed plenarily, in which pupils circle the text feature, and tell the class which text feature they circled and what the function of that text feature is (i.e. how it 'helps us'). This activity aims to prepare the pupils to write their own two-page spread in a future assignment.

School C

Topic: Animal groups

1. Interactive recap of previous knowledge on animal groups from earlier lessons in Dutch.
2. **Prescribed lesson** activity.
3. Teacher introduces the topic animal rights to the class. Pupils write small, personal whiteboards what they think it means; these answers are discussed plenarily.
4. In their groups, pupils discuss which animal they think is most valuable and why. The group needs to convince each other with arguments and reach a joined opinion. The concept valuable is explained as being important; some kids interpret this as important to themselves, or as endangered animals that need to be protected. After the group activity, answers are discussed plenarily.

School D

Topic: Life cycle of the bee

1. Children are given cards with a word related to bees, and are asked to say their word out loud, and find children with the same word. This way, groups are formed for the fixed lesson activity. Groups are asked to read their word out loud to the rest of the class. The rest of the class is asked to repeat the words as well.
2. **Prescribed lesson** activity. *A difference with other schools was that in the plenary discussion, pupils are instructed to discuss in their small groups what they think the question is after listening to the answers to a question, before discussing what the question was plenarily. Also, at this school, only part of the questions was discussed plenarily, while at other schools, all groups had a turn.*
3. The goal of the lesson (i.e. "we are learning about the importance of bees and their life cycle") is introduced.
4. Lists of vocabulary items related to bees are hung on a wall in the hallway. In their groups, pupils need to write as many items as possible within a fixed amount of time. One group member needs to stay in the class, the other needs to run to the hallway and back, and communicate a word and the correct spelling to this child, who writes them down. Pupils are only allowed to do one word at a time. Then, the words are discussed plenarily, and the pupils are asked to repeat each of the words.
5. Pupils are instructed to copy the aim of the lesson from the board into their notebooks, and if they have time left, include the vocabulary items, too.

6. Then, the teacher reads an informative text on the life cycle of the bee to the children. The children are asked to read along or to repeat the sentences. Some vocabulary items are highlighted and discussed while reading the story. The group is also asked to do a physical activity while inserting vocabulary items, such as for 'hive': "Hive, clap clap, hive, click click, hive, hit hit, hive, stamp stamp." Corresponding movements are made (e.g. clicking fingers, stamping feet).

7. The stages in the life cycle of the bee are repeated once more, after which pupils are instructed to draw the life cycle of the bee, and include vocabulary items.

8. The teacher discusses with the class whether they think they have reached the lesson objective.

Appendix B. Prescribed Lesson Questions

School A
- What can you use water for?
- What will sink?
- Where on earth do you find water?
- What can float?
- What can help you stay dry?
- What is waterproof?
- What kind of water falls from the sky?

School B
- What are differences between penguins and monkeys?
- What are differences between spiders and snakes?
- What are differences between fish and birds?
- What are differences between insects and people?
- What are differences between zoo animals and pets?

School C
- What is a caption?
- What is a heading?
- What is a drawing?
- What is a photograph?
- What is a text box?
- What is a label?
- What is a key word?

School D[7]
- What do you need at the beach?
- What kind of product can you find at the bakery?
- What do you need to build a house?

---

[7] The rest of the questions is unknown, as these were the only questions that were discussed in the plenary discussion.