From: Burmeister, Petra; Piske, Thorsten; Rohde, Andreas (Eds.): An Integrated View of Language Development. Papers in Honor of Henning Wode, Wissenschaftlicher Verlag Trier, 2002.
ISBN 3-88476-488-8. www.wvttrier.de

How effective is late partial immersion?
Some findings from a secondary school program in Germany

Petra Burmeister & Angelika Daniel

1. Introduction

This report assesses the effectiveness of late partial immersion at secondary school level with regard to its linguistic outcomes. The data derive from the *Kiel Immersion Project* at Kiel University directed by Henning Wode (see also Kersten et al., this volume; Rohde & Tiefenthal, this volume). Of all things, it was probably Henning Wode's interest in the various types of language acquisition, e.g., first language mono- or bilingualism, naturalistic and instructed second language (L2) acquisition or re-learning of an L2, that finally led his attention to the field of immersion/bilingual education some twenty years ago:

> Depending on the type of program, teaching in all school subjects is wholly or partially carried out in that language; and, in addition to this type of naturalistic exposure, the language may or may not, also be taught. (Wode 1981, p. 41)

In addition to his overriding interest in the linguistic products of immersion (IM) education, Henning Wode has never seized to point out the political situation in Europe that appears to be calling for urgent improvements in the field of foreign language teaching – a demand that IM can obviously satisfy. In 1989, three years before the ratification of the Maastricht Treaty, Henning Wode started promoting immersion education from preschool to secondary school as an effective way to enhance foreign language competence of students in Europe without having to increase the number of language-as-subject periods (e.g., Wode 1990, 1991a,b, 2001). His conviction was grounded on the results of extensive research showing that, through immersion, students are provided with the academic knowledge that is required in the school curriculum, and with a level of foreign language competence that by far exceeds the level reached in traditional language teaching (overviews in, e.g., Genesee 1987; Wode 1995; Wesche, this volume).

In Germany, the IM method, that is, teaching subject matter through an L2 or L3, has been applied in programs that differ with respect to various instructional variables such as the type or status of institution, the languages involved, the clientele that is catered for, the age of entry, the number and choice of subjects taught through the L2, or the goals (for an overview, see Wode 1995). Thus, program types range from 'non-state

schools' predominantely for children from abroad such as the International School or the European Schools, to the typical German model of partial IM education at state secondary schools (*Bilinguale Zweige*), where on average two subjects are taught in a foreign language, mostly in English and French (e.g., Wode 1995; Christ 1996, 1999; Thürmann 2000; Wolff, this volume).

2. The Schleswig-Holstein IM program

In 1991, due to Henning Wode's initiative, the ministry of education in Schleswig-Holstein, northern Germany, implemented the first late partial IM programs at five academic secondary schools (*Gymnasien*). During the past ten years, the number of programs has continually increased to 18 in 2001 (17% of the academic secondary schools in Schleswig-Holstein). The program design is quite similar to the German-English immersion tracks in North-Rhine Westphalia. L2-English is introduced in grade 5, that is, at age ten, and is taught in language-as-subject periods throughout secondary school. Two subjects (geography, history or biology) are taught in English from grade 7 onwards so that the contact time to the L2 amounts to almost 30% of the curriculum – an increase of about 15%. In order to make sure that the students will reach a level of language competence that enables them to cope with IM subject-matter instruction in grade 7, English instruction in the lead-up grades 5 and 6 is boosted by one or two communication-based periods per week.[1] One extra period is granted to the IM subjects in grades 7 to 9.

3. The *Kiel IM Project*

Since the beginning of the *Kiel IM Project*, Henning Wode's research group has undertaken evaluations as to the effectiveness of the Schleswig-Holstein late partial IM programs. The major goal of the Kiel evaluations has been to investigate the linguistic outcomes, that is, to describe the nature and the development of the IM students' L2-English in comparison to non-IM students (e.g., Wode 1994, Wode et al. 1994, Wode et al. 1996, Burmeister 1998a, Wode 1998). In particular, the aim is to determine whether there are differences between IM students and their non-IM controls with respect to the structural categories analysed. For this purpose, oral and written data have been collected at two cross-sections that allow for detailed psycholinguistic analyses of various aspects of the students' L2.

So far, most of the studies based on the Kiel data have focussed on the L2 vocabulary and on aspects of discourse. The results show that the IM students tend to outperform their controls with respect to most measures. This chapter summarizes the most important results from the lexicon and discourse studies with the aim of drawing some conclusions concerning the effectiveness of late partial IM in Schleswig-Holstein.

1 As to the effectiveness of enhanced English instruction in grade 5, see Burmeister 1994, 1998b.

4. General method

4.1 The test: *A Difficult Decision*

The Kiel data has been collected via the 'homemade' communicative test called *A Difficult Decision* (Knust 1992, 1994) designed to elicit both oral and written data. The central component is a text which is based on a simulation game (Klippel 1987). It deals with a group of students and their teacher, who are on a hiking trip in the Scottish Highlands. In the middle of nowhere, one girl breaks her leg and three students start feeling sick. A camp is set up and the children have to figure out how to get help.

The test consists of four tasks. The first two serve as a warm-up and allow students to become familiar with the situation in the test. In the first task, the testees are asked to answer text-comprehension questions. In the second task, they are supposed to describe a rudimentary map of the area which comes with the test-text. In task three, the testees are required to consider different ways out of their desperate situation. In the fourth task, students are asked to write a letter to a friend about what happened to them on their class trip.

4.2 Database

The test *A Difficult Decision* was administered at two cross-sections: from 1992 to 1995 to 12 to 13 year-old students at the end of grade 7 and from 1995 to 1998 to 10th-graders (15 to 16 years of age). The data corpus consists of tests from a total of 1060 students: 543 in grade 7 (144 oral and 399 written tests) and 517 in grade 10 (162 oral and 355 written tests).

Within these cross-sections, two groups were tested at the IM schools: group B with students from the IM classes and group A with non-IM students from the parallel classes. In this context, it is important to keep in mind that IM is optional and the students could not be randomly assigned to the IM or the non-IM group**.** Teachers tend to consider students who enter an IM track as being more motivated and talented and they therefore think that the non-IM classes might not constitute a representative *Gymnasium* group. In order to guard against those biases, a third group (group C) was included which consisted of students from schools without an IM program.

Before testing, each class was divided into above-average, average and below-average students on the basis of the English teachers' judgements of the students' oral communication skills in the L2. The researchers then selected two students per achievement level for the oral testings and formed two groups of three students each, that is, two triads with one student per achievement level. While the two triads were tested in two separate rooms, the rest of the class worked on the written test in their classroom, supervised by their teacher.

The database for most studies undertaken so far are transcripts of the tape-recorded oral versions in which students in groups of three were supposed to discuss what they could do in order to get help. All discussions were attended by a member of the research team, the interviewer, who had been instructed to interfere as little as possible since the main purpose of this task was to elicit spontaneous conversation. The data are then subjected to the various psycholinguistic analyses.

5. Focus on L2-discourse

The test *A Difficult Decision* provides a thematic context that allows students to communicate and negotiate their ideas in a spontaneous way. In order to solve the difficult situation described in the text, they have to make suggestions, describe and explain, express opinions, agree or disagree.

According to the impressionistic judgements by three raters who listened to the recordings of the first two testings, the IM students were 'better communicators'. They outperformed their controls significantly with respect to the rating-scale items *spontaneity*, *involvement in the discussion*, *interaction within the group*, *independence of the interviewer* (Knust 1994).[2]

The goal of the discourse studies undertaken in Kiel is to discover more about how students use their L2 to communicate their ideas in the discussions. Therefore, different discourse-analytical procedures have been developed and applied in order to determine whether they allow for an adequate description and interpretation of the Kiel data and thus help to explain the results of the holistic ratings.

5.1 Cohesion

The first steps in this direction are studies on cohesion (e.g., Krohn 1996, Claussen 1997, Schriever 1997, Mukherjee 1999). Their purpose is to explore whether the analysis of students' utterances with respect to the use of cohesive elements proves to be a useful means of showing qualitative and quantitative differences between IM and non-IM students.

The theoretical and practical framework is based on Halliday & Hasan (1976). According to Halliday & Hasan, a set of sentences or utterances constitute a text when there is texture. Texture in a text is provided for by cohesive relations which are realized by cohesive ties, such as personal pronouns, conjunctions, and nominal or verbal substitutions which refer back to linguistic elements mentioned in the text and thus explicitly mark semantic relationships (Halliday & Hasan 1976, p. 2). Halliday & Hasan's analytical grid has had to be modified in such a way as to allow for an adequate description of the available data. The most important modification concerns

2 The rating-scale, the rating-scale items, and the statistical procedures are described in Knust (1994).

the basic unit of analysis. Whereas Halliday & Hasan consider cohesive relations only between sentences, the basic unit in the Kiel cohesion studies is the clause (e.g., Krohn 1996, Claussen 1997, Schriever 1997, Mukherjee 1999).

5.1.1 Methodology

The analyses are based on oral and written data. The students' texts/utterances are first subdivided into clauses. Then the cohesive elements are classified according to their category. Furthermore, it is indicated which term the respective cohesive element is linked to. The three groups are then compared with regard to the length of their contributions in terms of the number of clauses, the absolute frequency of cohesive elements per category and subcategory, the density of cohesive elements, that is, the number of cohesive elements per clause and the direction of students' links. It should be emphasized that as these studies have been essentially exploratory in nature and only small samples have been analysed, tests of statistical significance have not yet been conducted. Thus, the following findings are still tentative and need further empirical support.

5.1.2 Major findings

The results of the descriptive quantitative and qualitative analyses of the oral and written data of students in grades 7 and 10 show the same consistent trends: The IM students score higher with respect to the number of clauses, the frequency of cohesive devices and the degree of cohesiveness. Even more interesting are the analyses of the direction of interaction in the discussions. As each cohesive item is linked to another one in a given discourse, it can be shown to what speakers have cohesively linked their contributions, e.g., to their own utterances, to those of other students, to the interviewer's remarks, or to the situation. The findings reveal that the IM students talk among themselves the most, whereas their non-IM controls rely more than twice as much on the interviewer's prompts and guidance. As for the performance of the two non-IM groups, comparisons have shown that in most cases, group C (the classes from a school without an IM program) scores higher than group A (the non-IM students from the parallel classes). This confirms the assumption that a non-IM control group from the same school as B may be less representative than an external control group from a school without an IM program (see section 4.2).

5.2 Speaker-shifts: Turn-taking and turn-yielding

Schriever (1999) investigates how speaker-shifts in the oral discussion are characterized. Her database comprises oral discussions from 25 groups of 7th graders, each consisting of three students. The theoretical and practical basis is provided for by Stenström (1994), who, in her turn-taking model, describes interactional strategies that speakers adopt to take, hold, and yield turns. Schriever (1999) modified Stenström's

model in order to account for the given data and investigates how students in the discussions take and yield turns.

How students take a turn and what their respective interactional intentions are, is described by the turn-taking categories *links*, *answers*, *acknowledges*, *continues*, *corrects*, and *echoes* (Schriever 1999). *Links* are classified as turn-takings which are not explicitely elicited and refer to a previous turn either 'grammatically' or via the respective contents. *Links* often begin with a conjunction such as *but* or *because* or with elements such as *I think*. *Answers* are explicitly elicited and realized by *yes* or *no* only or via longer utterances. *Acknowledges* provide a positive or negative follow-up and are often realized by *hmm* or *yeah* or *that's right*. *Continues* describe two instances of turn-taking: Firstly, how speaker A continues his own turn after he has been interrupted or after a pause, and, secondly, if speaker B continues a turn of speaker A. A turn-taking is classified as *correct* if speaker B corrects a mistake of speaker A, which happens most often in the form of an ellipsis. *Echo* is classified as a word-by-word repetition of what has been said in the previous turn.

The turn-yielding categories describe the manner and the interactional intention with which a turn is yielded. Schriever (1999) identifies the following categories in her data: *Questions*, *requests*, *vocabulary questions*, *meta-comments*, and *giving-ups*. *Questions* explicitly elicit answers and refer to the content of the discussion. *Requests* are only used by the interviewer in order to encourage students to carry on with their discussion. *Vocabulary questions* are used to ask the others for an English term. The category *meta-comments* describes those instances of turn-yielding that refer to the organization of the discourse (*speak up, please!*) and not to the discussion content as such. *Giving-ups* are incomplete utterances that signal that the student has nothing more to say and is going to give up his turn.

Schriever (1999) also classifies unsmooth speaker shifts in her data: *Interrupts* are instances where speaker A intentionally interrupts speaker B, whereas unintentional interruptions of the other speaker's utterance are classified as *early starts*.

5.2.1 Methodology

The discussions vary in length and complexity so that a unit of measurement had to be found that allows for comparisons between the groups. As turn-takings only occur at the beginning and turn-yieldings only at the end of utterances, the analyses were therefore independent of the length of the contributions and the unit *turn* was chosen to be the appropriate tool.[3]

3 The criteria that distinguish between a turn and a non-turn in the Kiel data are based on Ruthenberg (1998).

In the transcripts of the discussions, the number of turns is calculated for each participant and the turn-takings and turn-yieldings are classified and listed.[4] It is important to note that only those taking- or yielding-the-turns are included which provide an essential interactive signal. Thus, no taking-the-turn category is listed if, e.g., the sole purpose of the whole turn is to elicit an answer. The results are computed for each individual student and for the groups. In order to allow for intra- and inter-group comparisons, the figures obtained for each category are relativized with respect to the total number of turns per group and with respect to the total number of turns for the respective speaker. In order to evaluate differences between the groups, analyses of variance were carried out.

5.2.2 Major findings

The comparison between the relative number of turns has shown that the IM students talk significantly more than the students in the control groups. The relation between the cases of taking-the-turn and the cases of yielding-the-turn in all three groups shows that most of the speaker-shifts in all three groups are cases of taking-the-turn, that is, most students trigger the interaction via taking-the-turns and very rarely yield turns in the sense of addressing their partners actively.

The analysis of the turn-taking category *links* has shown that the IM students use significantly more links, i.e., twice as many as their peers in the non-IM control groups. This finding corroborates the results of the cohesion studies in that it shows that the IM students are more capable of discussing among themselves and of maintaining a conversation on their own. There is less need for the interviewer to push the IM students via *requests* or *questions*. There is also a significant difference between the IM and the non-IM groups with regard to *answers*. Whereas *answers* constitute over half of the turn-takings in the non-IM groups, only a third of IM students take turns with *answers*. A high number of *answers* could be interpreted as an indication of less initiative on the part of the students. Moreover, it implies a high number of *requests* and *questions* on the part of the interviewer in the respective groups. The analyses with regard to the number of elicitations by the interviewers confirm this assumption (cf. Schriever 1999). The turn-taking types *continues*, *corrects*, and *echoes* are very rarely used by the students and there are no statistically significant differences. *Acknowledges* on the part of the students occur also very rarely but are used significantly more by the IM students, that is, the IM students are more inclined to evaluate the other participants' contributions.

No significant differences between the groups could be shown concerning the cases of unsmooth speaker shifts, i.e., *interrupts* and *early starts*. However, the number of *interrupts* is clearly higher in the IM groups. It seems as if IM students are less willing

4 To maximize the validity of the method and the analyses, at least one additional researcher was involved, who had done her analyses independently of the main researcher. Most steps were repeated to make sure that the results were reliable.

to wait for their turn and have more courage to 'misbehave' in the L2. Students from the non-IM groups tend to wait until they think the previous turn is finished. This can be assumed on the background of the results obtained for the *early starts* which are more frequently used in the non-IM groups.

As regards the turn-yielding categories, there are no significant differences between the groups. In all three groups, *questions* are rarely used. As the testees in the discussion groups depart from the same knowledge level, they apparently do not feel the need to ask each other questions. Thus, most of the student questions are directed towards the interviewer. Even less frequently used are *vocabulary questions*. This seems to indicate that students either had no major difficulties in activating the vocabulary they needed or that they did not dare to ask. The students also adopted 'positive strategies' to overcome vocabulary problems, such as code-switching and paraphrasing. It would be worth comparing the three groups with respect to the use of such strategies. *Meta-comments* and *giving-ups* also occur only very rarely. As in the cohesion studies, the students in the C-groups tend to outperform their peers in the A-groups (see section 5.1.2).

The preliminary results described above suggest that only seven months after the beginning of IM, the 7th-graders in the IM groups are more capable of maintaining the discussion without the help of the interviewer than are their non-IM controls. This is not only due to a large number of turns on the part of the IM students (cf. also *size of the vocabulary* below) but is also reflected in the use of interactive strategies they adopt in the discussions. IM students take their turns more often via *links* and *acknowledges* and less frequently through *answers*, which could be interpreted as an indication of more initiative and greater involvement. In sum, it can be stated that the results of the discourse analyses corroborate the findings obtained in the impressionistic ratings mentioned earlier, that is, the IM students show more spontaneity, involvement in the discussion, interaction within the group, and independence of the interviewer.

6. Focus on L2-vocabulary

L2-vocabulary is another case in point in the evaluation of the late partial IM program in Schleswig-Holstein for two reasons: (1) Of all structural areas, the vocabulary has been the main focus of the linguistic analyses conducted by the Kiel group of researchers. The largest number of students by far has been analysed lexically. This has been done in small-scale studies on the basis of triads (e.g., Kickler 1992, Daniel & Nerlich 1995a,b, Chung 1997, Berg 1998, Brandt 1998, Freese 1999, Daniel 1999), and in various state-of-the-art summaries based on larger sets of students (e.g., Kickler 1995; Daniel & Nerlich 1998; Wode 1999; Rueß 2000; Daniel 2001; Biskup, i. prep.). (2) The overall results indicate that alongside discourse behaviour, the L2-vocabulary seems to benefit most from IM teaching. Therefore, having summarized the main

results in the area of discourse, we now present the findings concerning the productive lexicon.

6.1 Objectives

The overall aim is to assess the effectiveness of late partial IM in Schleswig-Holstein pertaining to the English vocabulary. Note that due to the communicative nature of the test, the emphasis will be on the students' productive vocabulary rather than their passive knowledge. In particular, the following questions are addressed:

- Does IM foster L2-vocabulary?
- Which aspects of the students' vocabulary benefit most from IM teaching?
- Which aspects benefit to a lesser extent or not at all?
- How does L2-vocabulary develop over time?

6.2 Lexical analyses

The lexical analyses as part of the *Kiel IM Project* have been conducted from the outset to keep track of the development of the students' L2-lexicon as thoroughly as possible (e.g., Kickler 1992). When the pilot groups had ultimately reached grade 10, the test was administered at another cross-section which opened up the possibility of cross-sectional as well as longitudinal studies (see above). By the end of the year 2001, numerous analyses had been carried out to investigate the state of the L2-vocabulary in grade 7 (e.g., Kickler 1992, 1995, Schmidt 1994, Daniel & Nerlich 1995a,b, Haupt 1996a,b, Peters 1997, Berg 1998, Brandt, Franz & Stanko 1998, Freese 1998, Daniel 2001), in grade 10 (e.g., Chung 1997; Brandt 1998; Peters 1998; Daniel 1999; Freese 1999; Hammerich 1999; Kölling 1999; Rüsch 1999; Rueß 2000; Stanko 1999; Biskup, i. prep.) and its development from grade 7 to grade 10 (e.g., Daniel 1999, Kröger-Zühlke 2000). Meanwhile, the successive evaluation of the oral data collected in grade 7 has been completed (summary in Daniel 2001). As far as cross-sections in grade 10 and longitudinal analyses are concerned, the database is continuously being enlarged via small-scale analyses and large-scale summaries (state-of-the-art summary in Biskup, i. prep.).

6.3 Database

The results presented below are based on the data collected between 1992 and 1999 in grade 7 by means of the oral test *A Difficult Decision* (see section 4.1). Altogether, 144 IM and non-IM students from five different secondary schools in Schleswig-Holstein were interviewed. They fall into three groups of 48 students each, i.e., (1) a group of IM students (B), (2) non-IM students from the same school as B (A), and (3) non-IM students from a school without an IM program (C) (see section 4.2).

For the purpose of lexical evaluation, the written transcripts of the tape-recorded discussions were modified in such a way as to arrive at an alphabetical word list for each student. Each word list contains in an alphabetical order all the lexemes that were used by a testee in the discussion. As far as lemmatisation is concerned, inflected and irregular forms were subsumed under the base form, e.g., *dogs/dog's* ⇨ *dog*; *going/goes/went/gone* ⇨ *go*; *smaller/smallest* ⇨ *small*, with base forms and irregular forms being counted as one type (see below). Furthermore, the lists contain the frequencies of occurrence for each lexical item, information about its origin and non-target uses and their frequency. These word lists served as a starting point for the analyses.

6.4 Methodology

To be able to assess manifold aspects of the students' productive English vocabulary, a comprehensive lexical grid has been developed (Kickler 1992, 1995, Daniel & Nerlich 1995a, Daniel 2001). The grid comprises both quantitative and qualitative aspects of the lexicon. In particular, the following measures are taken:

- Size of the vocabulary
- Lexical errors
- Lexical sources
- Lexical fields and contextual variants

6.4.1 Size of the vocabulary (number of types and tokens)

As for the size of the vocabulary, two numbers are calculated, namely the tokens, i.e., the total amount of words used in the test including repetitions, and the types, i.e., the amount of different words that are used. Accordingly, an utterance such as *Tim can go to the village and the teacher stays in the camp* contains 13 tokens and, due to the repetition of the definite article, 11 types. Both measures complement each other in that the number of types reflects the apparent size of the vocabulary as activated in the test, whereas the number of tokens reflects the students' ability and willingness to use this word stock in conversation (cf. Wode 1994).

6.4.2 Lexical errors

The aim of the error analysis is to make sure that IM does not promote lexical richness and lexical quantity at the cost to lexical accuracy. At the same time, developmental errors can serve as indicators of the different states of development of the L2-lexicon. For this purpose, non-target forms are identified in the data. To assess the nature and frequency of the errors, the number of different errors (error types) and the total number of errors (error tokens) are calculated for each student and for the groups and related to the total number of tokens. As for more detailed accounts of non-target uses

including discussions of possible error sources and the processes involved, additional studies are being conducted (e.g., Reinhardt 1999; Nerlich 1999; Hustedt, i. prep.).

6.4.3 Lexical sources

A third measure pertains to the origin of the vocabulary. The aim is to trace the words activated in the test back to their likely sources, that is, (1) the text of the test and (2) the vocabulary list in the textbook *English G Band 3* (Schwarz et al. 1987). When it comes to the categorization of the words according to their sources, the following combinations can be identified, each of which is of different importance for the analysis:

- words that are neither mentioned in the test-text nor in the vocabulary list and that are therefore likely to be of different origin, e.g., the media, stays abroad or the IM/foreign language classroom,
- words that are mentioned in the test-text but not included in the vocabulary list,
- words that are not mentioned in the test-text and marked as passive in the vocabulary list,
- words that are mentioned in the test-text and marked as passive in the vocabulary list.

For each group, the average proportions of the total lexicon with respect to the four categories are calculated and compared. The questions that need addressing are: Can IM provide for maximum exposure to the foreign language in terms of the amount and the quality of the input? Does IM promote the students' willingness to expose themselves to the L2 outside school more than English-as-subject does?

6.4.4 Lexical fields and contextual variants

As a measure of lexical richness, lexical fields and synonyms are investigated. The underlying assumption is that high proportions of words which are similar in their meaning suggest a more advanced stage of acquisition in that the students seem to have a comparatively wide array of lexical resources at their disposal, whereas low proportions of synonyms may be taken as an indication of a less differentiated vocabulary. To avoid misunderstandings, the term 'contextual variants' has been introduced for the purpose of the lexical analyses within the research project. We define contextual variants as two or more words that may be used to refer to the same extralinguistic referent in a given context. In the context of the test, pairs and triplets such as *road/path/street/way*, *house/hut*, *go/walk*, *ill/sick*, etc. can be used alternately to refer to a particular item on the map, or to describe an activity to be performed as part of the simulation game, and in this sense they are considered similar in meaning and/or function. The average numbers of contextual variants in relation to the total lexicon are computed for each student and for the groups and the results are compared.

6.5 Major findings

First and foremost, the research results obtained between 1992 and 2001 show that IM teaching affects different aspects of the productive L2-vocabulary differently. For some measures, a clear distinction between IM and non-IM students could be observed, for others, there was little or no difference between the groups (cf. Daniel 2001). The aspects that appear to benefit enormously from IM and for which also statistical support is available are *words from sources other than the test-text or the vocabulary list*, and *contextual variants*. For these two measures, the mean differences between group B and both groups A and C were highly significant.

IM seems to have no impact whatsoever on *lexical errors*, that is to say, it neither increases nor decreases the error rate significantly when compared to traditional English-as-subject. This finding is remarkable since one might suspect that less focus on form in the IM classroom entails an increase in errors. Obviously, IM in Schleswig-Holstein unfolds its potential at no cost to lexical correctness.

Of all measures taken, the IM students received the lowest scores with respect to *words included in the test-text but not in the vocabulary list*. However, it must be recalled that high scores for this aspect suggest a frequent use of the lexical resources provided by the test, whereas low scores indicate that the students must have drawn upon sources not accessible to them during the test, e.g., the vocabulary list of the textbook or words introduced in the IM/foreign language classroom. Given this, the performance of the IM students appears in a different light in that they rely on the test-text to a lesser extent, and in doing so, they display a higher degree of lexical independence. This is corroborated by a significant difference in the mean scores between group B and C.

Furthermore, significant differences between the IM students and either group A or C have been shown with regard to *words which are marked as passive in the vocabulary list and are not contained in the test-text* and *words that are passive only*. In other words, IM demonstrably promotes the use of words from the above-mentioned sources. Seen in isolation, these two findings must not be overvalued. Seen in the context of the other aspects, words from other sources and words from the test-text only, they all point in the same direction: As for lexical sources, the IM students, less than their non-IM counterparts rely on sources available during the test, and more often than their non-IM peers activate words which must have been obtained from unknown sources. It seems highly likely that in this connection the additional input provided in the IM classroom plays an important role, either as a direct source or as an indirect stimulus for the students to expose themselves to the English language outside school, e.g., through stays abroad, pen-friends, pop music, books, or the internet.

As for the number of types and tokens, the differences between the groups are restricted to a descriptive level, that is, there is no clear statistical support for an

influence of IM on the size of the vocabulary. The results for the tokens point to a superiority of the IM students, who on average produced the largest number of words. As for the types, the results can tell little about the effectiveness of IM. Though group C is slightly ahead of the others, we must take into consideration that, due to the specific test-task, the possible number of types to be used in the discussion is rather limited. As a consequence, the group differences with regard to the types may ultimately level out and the measure may lose selectivity.

As for grade 10, several cross-sectional and longitudinal studies have shown that the results seem to point in the same direction (e.g., Daniel 1999, Kröger-Zühlke 2000, Rueß 2000). Meanwhile, the database has been increased to such an extent that tests of statistical significance are presently being carried out. Similar to grade 7, the ultimate aim is to analyse the remaining data in order to arrive at a final report on the productive L2-vocabulary in grade 10.

7. Conclusion

For the past ten years, Henning Wode has called for the introduction of IM programs in German schools on a large scale. The result was, among other things, the launch of a late partial IM program at secondary school level in northern Germany in 1991. In the wake of its success, the first IM tracks have also been opened at primary/elementary level as well as in preschool (see Kersten et al., this volume; Rohde & Tiefenthal, this volume).

The effectiveness of late partial IM has been evaluated as part of the *Kiel IM project* from the very beginning. Above, we have summarized the findings of studies that compare the L2-English language outcomes of IM students with those of non-IM controls with regard to the discourse and the lexicon. The results suggest that already by the end of grade 7, that is, only seven months after the onset of IM, the students have indeed profited from it. The psycholinguistic analyses have shown that the IM students are more talkative, more eager to talk, and need less guidance and encouragement on the part of the interviewer. The IM students' vocabulary is considerably larger and more differentiated as to lexical fields, and more often than for the controls it derives from other sources, such as the IM classroom.

Given that one or two IM school-subjects are a successful means of promoting an L2 at secondary level, why not teach subject matter in the L3, too? In order to allow more time to develop a functionally appropriate level in a third language, Henning Wode advocates what he terms "sequential double IM" (Wode et al. 2001), that is, a link between the introduction of an L2 through early-start IM in preschool, follow-up IM programs in primary school, and IM in the L3 from grade 7 onwards (cf. also Wode et al. 1999). In order to investigate what factors contribute to the success of IM in Schleswig-Holstein – be it in preschool, primary school or secondary school – more research is needed. This should also include investigations within the classroom, e.g.,

the language of the teachers and the students, and outside the classroom, e.g., the influence of the media.

References

Arnau, J. & Artigal, J.M. (eds.) 1998. *Els programes d'immersió: una perspectiva Europea/Immersion programmes: A European perspective*. Barcelona: Publicacions de la Universitat de Barcelona.

Berg, S. 1998. Vergleich des aktiven mündlichen Wortschatzes von bilingual vs. konventionell unterrichteten Schülern der Ricarda-Huch-Schule, Kiel. Unpublished manuscript: Kiel University.

Biskup, Ch. i. prep. Linguistische Untersuchungen zum Stand der Wortschatzentwicklung in der 10. Klassenstufe bilingual unterrichteter Schüler in Schleswig-Holstein.

Brandt, B. 1998. Vergleich des aktiven mündlichen Wortschatzes von deutsch-englisch bilingual unterrichteten Schülern der 10. Klasse ohne verstärkten Vorlauf mit konventionell unterrichteten Schülern der Ricarda-Huch-Schule, Kiel. Unpublished manuscript: Kiel University.

Brandt, B., Franz, S. & Stanko, A. 1998. Vergleich des aktiven mündlichen Wortschatzes von bilingual vs. konventionell unterrichteten Schülern der Ricarda-Huch-Schule und der Hebbelschule in Kiel. Unpublished manuscript: Kiel University.

Burmeister, P. 1994. *Englisch im BILI-Vorlauf: Pilotstudie zur Leistungsfähigkeit des verstärkten Vorlaufs in der 5. Jahrgangsstufe deutsch-englisch bilingualer Zweige in Schleswig-Holstein*. Kiel: l&f Verlag.

Burmeister, P. 1998a. Zur Entwicklung der Fremdsprachenkenntnisse im bilingualen Unterricht: Ergebnisse aus fünf Jahren Forschung. In: Hermann-Brennecke, G. & Geisler, W. (eds.), *Zur Theorie der Praxis & Praxis der Theorie des Fremdsprachenerwerbs*. Münster: Lit Verlag, 101-116.

Burmeister, P. 1998b. The effectiveness of enhanced English instruction in preparation for bilingual instruction in the Schleswig-Holstein late partial immersion program. In: Arnau & Artigal (eds.), 667-674.

Christ, I. 1996. Bilingual teaching and learning in Germany. In: Fruhauf, G., Coyle, D. & Christ, I. (eds.), *Teaching content in a foreign language. Practice and perspectives in European bilingual education*. Alkmaar: Stichting Europrint, 81-100.

Christ, I. 1999. Zur Weiterentwicklung des bilingualen Lehrens und Lernens in deutschen Schulen. Bilinguale Züge - Bilingualer Unterricht - Module "Fremdsprachen als Arbeitssprachen". *Zeitschrift für Interkulturellen Fremdsprachenunterricht* [Online], 4(2), 12 pp. Available: http://www.ualberta.ca/~german/ejournal/christi2.htm.

Chung, T. 1997. Linguistische Untersuchungen zum Englischwortschatz bei Schülern der 10. Klasse im bilingualen Unterricht in Schleswig-Holstein. Unpublished M.A. thesis: Kiel University.

Claussen, I. 1997. Ein Vergleich der Diskursfähigkeiten mono- und bilingual unterrichteter Schüler im Rahmen der Unterrichtserprobung in Schleswig-Holstein. M.A. thesis, unpublished manuscript: Kiel University.

Daniel, A. 1999. *Wortschatzentwicklung im deutsch-englisch bilingualen Unterricht: Von der 7. bis zur 10. Klasse*. Kiel: l&f Verlag.

Daniel, A. 2001. *Lernerwortschatz und Wortschatzlernen im bilingualen Unterricht*. Frankfurt am Main: Peter Lang.

Daniel, A. & Nerlich, B. 1995a. Vergleich des aktiven mündlichen Wortschatzes von bilingual unterrichteten Schülern mit unterschiedlich hohem Stundeneinsatz und von bilingual vs. konventionell unterrichteten Schülern der Ricarda-Huch-Schule, Kiel. Unpublished manuscript: Kiel University.

Daniel, A. & Nerlich, B. 1995b. Vergleich des aktiven mündlichen Wortschatzes von bilingual vs. konventionell unterrichteten Schülern der Friedrich-Paulsen-Schule, Niebüll. Unpublished manuscript: Kiel University.

Daniel, A. & Nerlich, B. 1998. Analysing vocabulary acquisition in the Schleswig-Holstein immersion program. In: Arnau & Artigal (eds.), 647-657.

Freese, M. 1998. Vergleich des aktiven mündlichen Wortschatzes von bilingual vs. konventionell unterrichteten Schülern der Friedrich-Paulsen-Schule und der Hebbelschule in Kiel. Unpublished manuscript: Kiel University.

Freese, M. 1999. Linguistische Untersuchungen zum Wortschatz von bilingual und nicht-bilingual unterrichteten Schülern der 10. Klassenstufe der Ricarda-Huch-Schule und der Hebbel-Schule. Unpublished manuscript: Kiel University.

Genesee, F. 1987. *Learning through two languages: Studies of immersion and bilingual education*. Cambridge, MA: Newbury House.

Halliday, M.A.K. & Hasan, R. 1976. *Cohesion in English*. London: Longman.

Hammerich, C. 1999. Untersuchungen zum Wortschatz englisch-deutsch bilingual unterrichteter Schüler der 10. Jahrgangsstufe (1997) der Friedrich-Paulsen-Schule, Niebüll M.A. thesis, unpublished manuscript: Kiel University.

Haupt, S. 1996a. Vergleich des aktiven mündlichen Wortschatzes von bilingual und konventionell unterrichteten Schülern des Helene-Lange-Gymnasiums, Rendsburg, vs. konventionell unterrichteten Schülern der Hebbelschule, Kiel. Unpublished manuscript: Kiel University.

Haupt, S. 1996b. Vergleich des aktiven mündlichen Wortschatzes von bilingual vs. konventionell unterrichteten Schülern der 7. Klasse der Klaus-Groth-Schule, Neumünster. Unpublished manuscript: Kiel University.

Hustedt, M. i. prep. Vergleichende Untersuchungen zu Wortschatzfehlern im bilingualen Unterricht.

Kickler, K.U. 1992. Lexikalische Aspekte des L2-Wortschatzerwerbs im bilingualen Unterricht. Unpublished manuscript: Kiel University.

Kickler, K.U. 1995. *Wortschatzerwerb im bilingualen Unterricht. Pilotstudie zur Evaluierung der lexikalischen Fähigkeiten bilingual unterrichteter Schüler anhand eines kommunikativen Tests*. Kiel: l&f Verlag.

Klippel, F. 1987. *Intensive discussions and exciting activities sampler. Übungsvorschläge und Arbeitsblätter für einen aktiven Englischunterricht*. Dortmund: Lensing.

Knust, M. 1992. Die Entwicklung eines Tests zur Überprüfung der fremdsprachlichen Leistung von Schülern im deutsch-englisch bilingualen Unterricht der 7. Jahrgangsstufe in Schleswig-Holstein. Unpublished manuscript: Kiel University.

Knust, M. 1994. *"Bili ist echt gut...": Pilotstudie zur Evaluierung von deutsch-englisch bilingualem Unterricht in Schleswig-Holstein*. Kiel: l&f Verlag.

Kölling, D. 1999. Wortschatzanalyse der mündlichen Daten von Schülern einer bilingual unterrichteten 10. Klasse des Helene-Lange-Gymnasiums (Rendsburg) des Jahrgangs 1998. Unpublished manuscript: Kiel University.

Kröger-Zühlke, B. 2000. Longitudinalstudien zur Entwicklung des Wortschatzes von der 7. zur 10. Klassenstufe bilingual unterrichteter Schüler. Unpublished manuscript: Kiel University.

Krohn, G. 1996. Kohäsive Merkmale in mündlichen Texten bilingual englisch-deutsch unterrichteter Schüler im Rahmen der Unterrichtserprobung in Schleswig-Holstein. Unpublished manuscript: Kiel University.

Mukherjee, V. 1999. *Schriftlichkeit im bilingualen Unterricht: Kohäsive Merkmale in schriftlichen L2-Daten bilingual deutsch-englisch unterrichteter Schüler der 7. Jahrgangsstufe*. Kiel: l&f Verlag.

Nerlich, B. 1999. Vergleichende Untersuchungen zu Wortschatzfehlern bei bilingual unterrichteten Schülern der 7. Klassenstufe. M.A. thesis, unpublished manuscript: Kiel University.

Peters, I. 1997. Wortschatzanalyse mündlicher Daten des Tests "A Difficult Decision". Unpublished manuscript: Kiel University.

Peters, I. 1998. Vergleich des aktiven mündlichen Wortschatzes von englisch-deutsch bilingual vs. konventionell unterrichteten Schülern aus 10. Klassen an Gymnasien in Schleswig-Holstein. Unpublished manuscript: Kiel University.

Reinhardt, M. 1999. Fehleranalyse zum Wortschatz bilingual unterrichteter Schüler der 10. Klassenstufe im deutsch-englisch bilingualen Unterricht in Schleswig-Holstein. M.A. thesis, unpublished manuscript: Kiel University.

Rüsch, M. 1999. Vergleich des Wortschatzes von bilingual und nicht-bilingual unterrichteten 10.-Klässlern des Jahrgangs 1998 der Helene-Lange-Schule, Rendsburg, und der Hebbelschule, Kiel. Unpublished manuscript: Kiel University.

Rueß, C. 2000. Der Wortschatz 10. in Schleswig-Holstein bilingual unterrichteter Klassen im Querschnittsvergleich. Unpublished manuscript: Kiel University.

Ruthenberg, T. 1998. Pilotuntersuchung zu gesprächstaktischen Redemitteln bei L1-Sprechern und ihre Anwendungsmöglichkeiten bei L2-Sprechern. M.A. thesis, unpublished manuscript: Kiel University.

Schmidt, H. 1994. Wortschatz – Analytische Auswertung der mündlichen Daten von 1994 von Schülern der Friedrich-Paulsen-Schule, Niebüll (Jahrgang 1994). Unpublished manuscript: Kiel University.

Schriever, K, 1997. Analyse kohäsiver Elemente zum Nachweis kommunikativer Kompetenz von Schülern der 7. Jahrgangsstufe des deutsch-englisch bilingualen Unterrichts in Schleswig-Holstein. Unpublished manuscript: Kiel University

Schriever, K. 1999. Vergleichende Untersuchungen zum Interaktionsverhalten bilingual und nicht-bilingual unterrichteter Schüler in Schleswig-Holstein: Pilotuntersuchungen der Schülerdiskussion des 7. Jahrgangs der Testjahrgänge 1992/93/94. M.A. thesis, unpublished manuscript: Kiel University.

Schwarz, H., Taylor, C. & Vettel, F. 1987. *English G Neue Ausgabe: Band A3 für das 7. Schuljahr an Gymnasien*. Berlin: Cornelsen.

Stanko, A. 1999. Linguistische Untersuchungen zum Wortschatz bilingual und nichtbilingual unterrichteter Schüler der 10. Klassenstufe der Ricarda-Huch-Schule (Kiel) und der Hebbelschule (Kiel) des Jahrgangs 1997. Unpublished manuscript: Kiel University.

Stenström, A.-B. 1994. *An introduction to spoken interaction*. London: Longman.

Thürmann, E. 2000. *Zwischenbilanz zur Entwicklung der bilingualen Bildungsangebote in Deutschland*. http://www.learn-line.nrw.de/angebote/berichtsarchiv/info/ar0001/kap1.html.

Wode, H. 1990. Immersion: Mehrsprachigkeit durch mehrsprachigen Unterricht. *Informationshefte zum Lernen in der Fremdsprache*, 1, Eichstätt/Kiel.

Wode, H. 1991a. Fremdsprachenunterricht für Europa: Psycholinguistische Voraussetzungen für alternative Unterrichtsformen. In: Brusch, W. & Kahl, P. (eds.), *EUROPA-Die sprachliche Herausforderung*. Berlin: Cornelsen, 82-101.

Wode, H. 1991b. Formen und Leistungsfähigkeit mehrsprachigen Unterrichts. In: Raasch, A., Herold D. & Kiupel C. (eds.*), Fremdsprachen lehren und lernen: Perspektiven für ein Europa nach 1992*. Saarbrücken: Ottweiler Druckerei, 35-46.

Wode, H. 1981. *Learning a second language - An integrated view of language acquisition*. Tübingen: Narr Verlag.

Wode, H. 1994. *Bilinguale Unterrichtserprobung in Schleswig-Holstein*. 2 vols. Kiel: l&f Verlag.

Wode, H. 1995. *Lernen in der Fremdsprache: Grundzüge von Immersion und bilingualem Unterricht*. Ismaning: Hueber.

Wode, H. 1998. A European perspective on immersion teaching: The German scenario. In: Arnau & Artigal (eds.), 43-65.

Wode, H. 1999. Incidental vocabulary acquisition in the foreign language classroom. *Studies in Second Language Acquisition*, 21, 243-258.

Wode, H., Kickler, K., Knust, M. & Pries, B. 1994. The Schleswig-Holstein bilingual education project: a preliminary report. In: Laurén, Ch. (ed.), *Evaluating European immersion programs. From Catalonia to Finland.* Vaasa: Vaasan yliopisto, 154-174.

Wode, H., Burmeister, P., Daniel, A., Kickler, K.U. & Knust, M. 1996. Die Erprobung von deutsch-englisch bilingualem Unterricht in Schleswig-Holstein: Ein erster Zwischenbericht. *Zeitschrift für Fremdsprachenforschung*, 7, 15-42.

Wode, H., Burmeister, P., Daniel, A. & Rohde, A. 1999. Verbundmöglichkeiten von Kindergarten, Grundschule und Sekundarstufe I im Hinblick auf den Einsatz von bilingualem Unterricht. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* [Online], 4/2, 17 pp. (http://www.ualberta.ca/~german/ejournal/-wode2.htm).

Wode, H., Hansen, N., Imhoff, Ch. & Lauer, K. 2001. *Trilingual education in Germany: A European extension of Canadian immersion*. Unpublished manuscript: Kiel University.