The Language Learning Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rllj20

Issues in the assessment of bilingually educated students: expressing subject knowledge through L1 and L2

Dana Gablasova[a]

[a] Department of Applied Language Studies and Linguistics, University of Auckland, Auckland, New Zealand

Published online: 14 Mar 2014.

To cite this article: Dana Gablasova (2014) Issues in the assessment of bilingually educated students: expressing subject knowledge through L1 and L2, The Language Learning Journal, 42:2, 151-164, DOI: 10.1080/09571736.2014.891396

To link to this article: http://dx.doi.org/10.1080/09571736.2014.891396

The Language Learning Journal, 2014
Vol. 42, No. 2, 151–164, http://dx.doi.org/10.1080/09571736.2014.891396

Issues in the assessment of bilingually educated students: expressing subject knowledge through L1 and L2

Dana Gablasova*

Department of Applied Language Studies and Linguistics, University of Auckland, Auckland, New Zealand

This article discusses issues related to oral assessment of school knowledge of L2-educated students. In particular, it examines benefits and disadvantages of students being tested in their L1 (their dominant language) and in their L2 (their language of instruction). The study draws on the data from 37 high school students studying in a content and language integrated learning (CLIL) setting in Slovakia. They were tested both in their L1 (Slovak) and their L2 (English) on material which they read in English. Their ability to verbally express content knowledge was assessed in terms of linguistic accuracy, fluency, academic format appropriateness and lexical appropriateness. The results from these 37 students were compared with the performance of 35 students who read the same material in their L1 and were tested in L1. The study shows how the choice of either L1 or L2 can to some extent constrain students' ability to express the knowledge they have.

Introduction

Accurate assessment of students' subject knowledge plays an important role in all types of education. As with students in other programmes, students educated (partly) through the medium of an additional language may have their knowledge tested for different reasons, e.g. to monitor the progress of an individual learner or to assess the effectiveness of a whole programme (Coyle, Hood and Marsh 2010). However, in contrast to regular classes where students are both taught and tested through their L1, assessment of bilingually-educated students has been a notoriously difficult issue, both for educators and researchers. The main problem is related to students' mastery of the language of assessment (whether L1 or L2) and the extent to which their proficiency places constrains on their ability to express the content knowledge they have. Thus, when testing the knowledge of L2-educated students, 'teachers may not be sure whether a student is simply unable to demonstrate knowledge because of a language barrier or whether, indeed, the student does not know the content material being assessed' (Short 1993: 629–630; see also Coyle, Hood and Marsh 2010: 116; Hofmannová, Novotná and Pípalová 2008).

Given the importance of assessment of individuals and evaluation of educational programmes, it is surprising that assessment of bilingually-educated students has not received more principled attention (Byrnes 2008) and remains a 'weak link' (Short 1993: 653) in

*Email: d.gablasova@lancaster.ac.uk. The author is currently at Department of Linguistics and English Language, Lancaster University, Lancaster, UK.

content-based instruction such as content and language integrated learning (CLIL). This study therefore seeks to address some of the questions related to assessing the knowledge of students educated through the medium of L2. In particular, it focuses on the relative merits of choosing either L1 or L2 to elicit students' subject knowledge and the consequences that the choice of language has on students' ability to demonstrate their content knowledge. Moreover, this study focuses on oral elicitation of subject knowledge which is a common means of probing students' knowledge at secondary level both for formal as well as informal assessment (see e.g. Hofmannová, Novotná and Pípalová 2008; Hönig 2009). Yet, compared to written production, little is known about speaking ability of high school students (Menyuk and Brisk 2005). In this study, the assessment of subject knowledge is understood broadly as any formal or informal attempt to elicit the content knowledge that students may have in order to evaluate its quality and/or quantity.

Choosing the language of assessment

So far, there have been several approaches to testing bilingually-educated students. With respect to the language of assessment, four main options can be considered. First, there is testing in the language of instruction, the L2. This approach, often preferred by the students themselves (Hönig 2009; Ullmann 1999), does not require students to transfer their knowledge from one language into another and thus follows the basic pedagogical tenet: test only what you have taught. However, testing in L2 has been found problematic as teachers are sometimes unable to determine whether it is the lack of knowledge or language skills that prevents students from demonstrating their knowledge (Hofmannová, Novotná and Pípalová 2008; Lindholm-Leary and Borsato 2006; Short 1993). Second, testing in L1 has been proposed as an alternative, as L1 is still the dominant language of the majority of the students in bilingual programmes. This, however, has been found equally problematic, as some researchers reported that students had problems with retrieving the knowledge in their non-instructional language or lacked essential subject-specific vocabulary (Abedi, Hofstetter and Lord 2004; Airey 2010). Third, 'translanguaging', using a mixture of both languages to show that the content knowledge is available to students, has been recommended (García 2009) and used in teaching practice in some places (Johnson 1997; Serra 2007) as well as in research (Stohler 2006). Finally, testing in both of the bilingual's languages could be employed (i.e. eliciting the same material in students' L1 as well as L2). This has been so far done only as part of research designs where the language of testing served as an independent variable (cf. Airey 2010; Hincks 2010; Järvinen 2010; Stohler 2006). Some of these studies showed that students (often highly proficient users of L2) benefitted differently from L1 and L2 production. For example, Hincks (2010) reported that students experienced problems with disciplinary vocabulary in their L1 which were not observed in their L2. On the other hand, their speech in L1 was faster and appeared more elaborate (e.g. in terms of metaphor use).

The question of how the choice of a particular language affects the ability of students to express the knowledge they have is especially pressing because of the high-stakes examinations that bilingual students have to undertake in some educational settings. Often, the language of these exams is determined by the authorities without taking into consideration students' preferences or their language proficiency in the language of the assessment. In many cases, CLIL-educated students have to be examined on their subject knowledge in the language of the mainstream curriculum (i.e. L1) (Eurydice 2006). For example, Ullmann (1999) reported that CLIL students studying history and geography through French in a bilingual programme in Britain had to take their exam in English despite

their preference for being tested in their language of instruction (for a similar case in Hungary, see Duff 1997). These practices seem to reflect the belief that knowledge can be transferred between the bilinguals' two languages without any consequences, especially when it is transferred into L1. However, as discussed above, this may not always be the case.

The focus of the research

Undoubtedly, the choice of a particular language (L1 or L2) has an effect on how well students can express the content knowledge they possess. However, little is known about the impact of language factors on students' ability to demonstrate their content knowledge orally. This study therefore examines four linguistic features, different aspects of formal academic language, that directly affect the quantity and comprehensibility of the content knowledge expressed (cf. Airey 2010; Hincks 2010; Morgan 2006; Scarcella 2003).[1] These features are: (a) accuracy; (b) fluency; (c) appropriate academic format; and (d) appropriate vocabulary. The role of the four areas in the communication of knowledge is discussed below.

In addition, this study compares the results from L2-instructed students with the performance of students who both gained and expressed the same knowledge through L1. It is the performance of these students that is taken as a baseline in this study. This baseline not only allows us to better interpret the performance of L2-instructed students, but it also reflects the situations where L2-educated students are assessed according to the same criteria as their L1-taught peers (e.g. in the national standardised tests as reported for example in Eurydice [2006]).

Accuracy

Accuracy in a given language is an essential pre-requisite of formal discourse as it demonstrates adherence to the basic rules and norms of the linguistic code used. From a communicative perspective, if the speaker observes the conventional rules of a language he or she makes it easier for the listener to process the information. On the other hand, not following these rules can affect the informational value of the message which can get distorted. Also, the listener can get distracted by the errors and has to exert more effort to comprehend the message.

Fluency

Fluency is an important feature of oral communication. Effective oral communication of information depends on the availability of and access to the relevant linguistic and content knowledge. Thus, faster speech and a smaller number of reformulations, restarts and fillers can result in more information being communicated and be regarded as indicators of good subject-related discourse competence (Morgan 2006; Snow and Uccelli 2009). Responses with these characteristics can be also considered more communicatively effective as these qualities make it easier for the listener to process the message.

Appropriate academic format

Teachers use different means to probe the accuracy and depth of students' knowledge. Asking for definitions of subject-related terms is a common means of eliciting declarative

subject knowledge both in classroom interaction as well as in written forms of assessment (Temmerman 2009; Watson 1995). Moreover, definition-making skills of students have also been regarded as an indicator of the development of their academic language, i.e. the move from a more conversational towards a more formalised and decontextualised discourse (Iris, Litowitz and Evens 1988).

The definition is an academic language function characterised by a distinctive lexicogrammatical form. The basic, prototypical form of definition is called a formal definition. It should contain a word that establishes the membership of the defined term in a class of similar terms, together with the defining characteristics that allow the term/concept to be distinguished from other words/concepts (Dalton-Puffer 2007; Flowerdew 1992). Research into language development demonstrated that students as young as eleven years of age show a strong preference for giving formal definitions when asked 'What is an X?' (Watson 1995). This tendency has been documented in written as well as oral production (Marinellie 2009). The ability to produce formal definitions has been long regarded as an important aspect of becoming literate. This study used the definitional format to inquire into students' adherence to the norms of academic language.

Appropriate vocabulary

Another aspect of the mastery of a subject matter is the correct use of disciplinary vocabulary that allows communicating the knowledge with an appropriate degree of accuracy and precision (Snow and Uccelli 2009). Technical, subject-specific words carry greater informational load than more general words and can thus affect how much information is 'compressed' into an answer. Moreover, the use of technical words can be regarded as indicative of the depth of the specialist knowledge (Bravo and Cervetti 2009).

Method

Participants

Participants were recruited from two high schools in Slovakia which have implemented CLIL. Thirty-seven students (aged 17–20) from two upper, pre-university years formed the experimental group. Another group of students ($n = 35$) from the same two schools and the same CLIL programmes acted as a control group. The two groups were sampled randomly but were balanced for proficiency (upper-intermediate or advanced users of English), age, gender and years spent in a bilingual school. Students from a CLIL school rather than from a non-CLIL programme were selected as a control group in order to control for factors such as academic performance, socio-economic status and motivation which are usually higher among the students in bilingual than in mainstream programmes.

Tools

Materials used in this study were two medium-length textbook texts (around 850 words each) on geography and history of New Zealand and recorded versions of these texts. The texts contained information previously unknown to the participants as established by a pre-test. In particular, the texts contained 12 subject-specific words, defined or explained in the text, which were new to the participants. The 12 terms were: ampelography, diastrophism, ecocentrism, kumara, moa, moko, pa, perendale, RCD, terroir, transhumance and whanau. Table 1 shows an example of an explanation with which the words appeared in the text.

Table 1. An example of an explanation used in the texts.

	Explanation
Target word	
pa	The different tribes became more aggressive however and inter-tribal warfare became much more frequent over time. This led to the introduction of the pa (a fortified village). An average pa was placed near the top of a hill or cliff and it included ditches and palisades as protection.

The reading materials and recordings were prepared in two versions: in Slovak and in English. It is important to stress that the L2 technical words had a similar form and pronunciation both in English and Slovak thus making the texts fully comparable.

Procedure

In the study, 37 Slovak high school students (i.e. the experimental group, also referred to as the 'bilingual group') read and listened to two expository texts in English which contained 12 subject-related terms previously unknown to them. All participants were first given 10 minutes to read and study the first text and then were played the recording of it while being able to follow the text. Then the same process was repeated with the second text. Listening-while-reading was used in addition to reading, in order to ensure that students read the whole of the text and became familiar with the pronunciation of the target words. After this, participants were given a two-minute non-verbal distractor task which was followed by the test in which they were asked 36 questions about the two texts. Among these, they were asked to provide a definition for each of the 12 terms that appeared in the texts (e.g. 'What is transhumance?'). The questions were computer-administered and participants answered them orally. Half of the questions were asked in English and half in Slovak. The results of the test were compared to the performance of the control group of participants ($n = 35$) who read and listened to the texts in their L1 (Slovak) and were also tested in the same language (the control group only answered the questions in Slovak). Table 2 provides an overview of the two groups and the combination of the languages in which they learned the content and were tested on it.

Analysis

In this study, only participants' answers giving the definitions of the technical words were analysed. The definitional format allowed us to directly compare the production of the two groups as it restricted the content and the format of the answer. The procedure yielded 371 correct definitions from both groups of participants (control group: 144 definitions in L1, bilingual group: 118 definitions in L1 and 109 definitions in L2). For accuracy, fluency

Table 2. Pairings between the language of materials and the language of testing.

Group	Language of materials	Language of testing	Learning-testing language pairing
Control	L1 (Slovak)	L1	L1→L1: Slovak materials – Slovak test
Bilingual	L2 (English)	L1 & L2	L2→L1: English materials – Slovak test
			L2→L2: English materials – English test

and appropriate academic format, a series of independent-samples *t*-tests were used to establish whether the performance of the bilingual group in either of their languages differed from the performance of the control group. Cohen's *d* was used as a measure of effect size for the results that reached statistical significance. For lexical appropriateness, a different approach, discussed below, was adopted.

Results

Accuracy

Both grammatical and lexical accuracy were measured. Grammatical accuracy comprised morphological and syntactic accuracy. False starts or errors that were corrected by the speaker were not included in the analysis. Lexical accuracy was concerned with the accuracy at the word level of the utterance. Lexical errors were operationalised as errors that violated lexical rules of the target language (Llach 2011). This involved either the use of a word from a language other than the target language or a collocational error.

Accuracy in this study was operationalised as the proportion of error-free clauses. The clauses were counted using Analysis of Speech (AS) units (Foster, Tonkyn and Wigglesworth 2000) which were developed to allow for more reliable coding of speech production. To calculate the proportion, all clauses for all definitional attempts provided by one participant were added together. Likewise, all erroneous clauses were added together. The proportion of error-free clauses was then obtained by dividing the number of error-free clauses by the total number of clauses. The linguistic accuracy of participants' responses was compared using an independent-samples *t*-test. As shown in Table 3, there was no difference between the performance of the two groups in L1, where their accuracy was very high (around 97%). However, the accuracy of the bilingual group was on average 15% lower in their second language. This result was of large effect size.

Fluency

Two features of spoken fluency were measured in students' answers: speech rate and the pruned to unpruned speech ratio. Unpruned speech refers to the speech sample as it was produced by a participant, whereas pruned speech refers to the speech from which fillers, repetitions and false starts were removed. The ratio of pruned to unpruned speech was calculated to determine the informational value of the produced speech, i.e. to determine how much of the participants' response contributed towards the meaningful content of the message.

Table 3. Between-group comparison of accuracy.

Testing condition (No. of participants[a])	Descriptive statistics		Comparison with the control group (L1→L1)			
	Mean	*SD*	*t*	*df*	Sig.	*d*
L1→L1 (35)	.9714	.058	–	–	–	–
L2→L1 (36)	.9736	.052	−.166	69	.868	–
L2→L2 (36)	.8253	.131	6.098	48.6	.000**	1.75

[a]As some participants did not provide any correct definitions, their data were treated as missing and were not included in the analyses. This is reflected in the lower number of participants.

The speech rate was defined as the number of syllables produced per second (cf. Hincks 2010). The syllable count was calculated for eight out of the 12 target words. These words were ampelography, moa, RCD and whanau from the words that were defined in L1, and moko, ecocentrism, kumara and transhumance from the words defined in L2. These were selected as they yielded the longest answers. Only those responses that contained more than 10 syllables were included in the calculation. Shorter answers were considered insufficient data for the analysis of speech rate. An average speech rate for each participant was calculated from the four definitions in each language.

Table 4 shows the results of a *t*-test used to compare the oral production of the two groups of participants. As can be seen from the table, when answering in L2, the bilingual group's unpruned speech rate was on average slower than that of the control group. Whereas this may be expected, the performance of the bilingual group in L1 brought perhaps somewhat surprising results. While there was no difference between the two groups when L1 unpruned speech was taken into consideration, there was a difference with respect to the pruned speech (large effect size). This difference was further confirmed and quantified in the pruned to unpruned speech ratio (Table 5).

Table 5 shows that when we look at the ratio of pruned to unpruned speech, while there is no difference between the groups when the bilingual group was speaking in L2, there is a significant difference in their L1 production, which is of medium effect size. This result reveals that while the speed with which the groups were speaking in L1 was the same, the *type* of speech they produced was different. Compared to the control group, the speech of the bilingual group contained significantly more fillers and reformulations. As can be seen from Table 5, when speaking in the language of instruction, both groups produced on average 93% of speech free of distractors; when the bilingual group had to transfer their knowledge into their L1, their answers contained on average 12% of non-informative speech (i.e. 88% of distractor-free speech).

Table 4. Between-group comparison of (A) mean unpruned and (B) pruned speech rate.

Testing condition (No. of participants)	Descriptive statistics		Comparison with the control group (L1→L1)			
	Mean	SD	*t*	*df*	Sig.	*d*
(A) *Unpruned speech*						
L1→L1 (30)	3.71	.84	–	–	–	–
L2→L1 (27)	3.47	.67	1.207	55	.233	–
L2→L2 (23)	2.38	.34	8.557	41.6	.000**	2.65
(B) *Pruned speech*						
L1→L1 (30)	3.52	.76	–	–	–	–
L2→L1 (27)	3.12	.67	2.052	55	.045*	.55
L2→L2 (23)	2.19	.38	8.222	44.8	.000**	2.46

Table 5. Between-group comparison of proportion of pruned to unpruned speech (all $n = 33$).

Testing condition	Descriptive statistics		Comparison with the control group (L1→L1)			
	Mean	SD	*t*	*df*	Sig.	*d*
L1→L1	.930	.053	–	–	–	–
L2→L1	.880	.083	2.450	65	.017*	.60
L2→L2	.934	.063	−.500	65	.619	–

Appropriate academic format

In this section, the quality of participants' definitions is examined. The coding was based on the formal Aristotelian definition which consists of a class word (superordinate) and defining features of the word. Those of participants' definitions that contained an appropriate superordinate (noun) and at least one definitional feature were considered formal. In this way, all definitions were subdivided into formal and non-formal definitions (cf. Flowerdew 1992). The non-formal definitions consisted of, for example, just a synonym or an association of the defined word. The formal definitions were awarded 1 point, the non-formal definitions were scored 0.

Table 6 shows the results of the *t*-test used to compare the number of formal definitions provided by the two groups. The majority of the definitions given by both groups in their L1 were formal. On the other hand, the proportion of formal definitions elicited from the bilingual group through L2 was smaller, with less than a half of the definitions qualifying as formal. This difference was significant and of large effect size.

Lexical appropriateness

Due to the design of the study, only L1 answers of the two groups (i.e. definitions of six words defined in Slovak) were compared.[2] In order to compare the vocabulary used by the control and bilingual groups to convey the definitions of the six words in L1, the following procedure was employed. A frequency list of words used by participants to define each target word was created for each group and the lists were compared. All instances in which participants in one of the groups used a certain word with a relatively higher frequency than the other group were noted and examined. In the instances where the evidence showed that both groups were trying to convey the same informational content albeit through different lexical means, the source of the difference was further analysed.

Using this approach, two instances of a systematic use of an inaccurate term were identified in the responses of the participants from the bilingual group. The first was related to the definition of *pa.* (The explanation of the term as it appeared in the text can be seen in Table 1). The difference in the definitions of the two groups was found with respect to the class word (e.g. village, settlement, etc.) they used. As shown in Table 7, whereas the control group used words that depicted a more historic type of settlement (among the most frequent ones were 'osada' [historic village] and 'osídlenie' [settlement]), the majority of the bilingual group used the word 'dedina' (present-day village) with only one person choosing 'osada'. Whereas both Slovak words 'osada' and 'dedina' are correct translations of the English word 'village', 'dedina' depicts a more contemporary type of settlement, while 'osada' is the more appropriate term when referring to pre-historic dwellings (Pečiar 1959–1968). In this case, the bilingual group opted for the more general, and perhaps

Table 6. Between-group difference in the type of definition produced (all $n = 35$).

Testing condition	Descriptive statistics			Comparison with the control group (L1→L1)			
	Mean	Percent	SD	*t*	*df*	Sig.	*d*
L1→L1	4.29	71.5	1.13	–	–	–	–
L2→L1	3.84	64.0	1.73	1.297	62.4	.194	–
L2→L2	2.78	46.3	1.70	4.438	62.8	.000	1.12

Table 7. Conveying the idea of 'village' in the definition of *pa.*

	Slovak original	English translation
Control group	osada (9); osídlenie (8); dedinka (2); dedina; mesto	(historic) village (9); settlement (8); (present-day) village (dim) (2); (present-day) village; town
Bilingual group	dedina (16); mesto (5); osada; dedinka	(present-day) village (16); town (5); (historic) village; (present-day) village (dim)

Table 8. Conveying the idea of 'rabbit' in the definition of *RCD.*

	Slovak original	English translation
Control group	králik/y (29), zajac (2), králičí (3), zajačí (1)	rabbit/s (29), hare (2), rabbit (adj.) (3), hare (adj.) (1)
Bilingual group	zajac/e (30), králik/y (10)	hare/s (30), rabbit/s (10)

more familiar, translation of the word 'village'. Whereas this could be an acceptable translation of the word in some other contexts, the students failed to select the word appropriate for historical discourse. (It should be noted that the word for 'village' used in the Slovak original was 'osada', thus the L1-instructed participants to some extent repeated the given word in their answers [nine times] but also used a different appropriate word, 'osídlenie' [eight times].)

Second, the two groups systematically differed in the choice of words in the definition of *RCD* which appeared with the following description in the text:

> In 1996 and 1997, a debate emerged in New Zealand about whether a new rabbit virus *RCD,* should be introduced to control the rabbits. The RCD virus seems to be very successful at reducing rabbit population – killing up to 90% of adult wild rabbits. Rabbits die within 30–40 hours of infection, owing to blood clotting in major organs.

In the case of *RCD*, whereas the control group consistently used the word 'rabbit' ('králik') in their definitions, the bilingual group predominately used the word 'hare' ('zajac') instead (see Table 8). The mistake most likely originated from the fact that 'a hare' in Slovak can be colloquially used to refer both to 'a rabbit' and 'a hare' (Pečiar 1959–1968). However, in a specific disciplinary discourse, these two words are used to refer to two different species and cannot be freely interchanged. Thus, as a result of their lexical choice, the participants in the bilingual group produced fewer discipline-appropriate answers, which may present a problem at a pre-university level at which a higher level of accuracy is expected.

Discussion

Accuracy

The results showed that the accuracy of participants in L2 was lower than when they were speaking in their native language. Whereas this finding in itself might not be surprising, it is important to note that the problems with linguistic accuracy in the L2 production of bilingual learners have consequences for students' ability to express their content knowledge

through L2. As a result of our limited processing ability, 'in any cognitive activity, we are able to attend to only some selected information at any given time' (Kellerman and Bialystok 1997: 33). Thus, if a substantial amount of students' attention is still directed to monitoring the accuracy of their speech (as also suggested by the self-corrections observed in students' answers), this can negatively affect their degree of control over the communication of the content knowledge. Compared to students from regular monolingual classes, this may put the L2-taught learners at a disadvantage educationally in situations when their knowledge is assessed through their second language.

Fluency

As for the speech rate, students spoke more quickly in their native than in their additional language and their L1 speech rate was comparable with that of the control group. This finding is in line with the research that measured the speech rate of university students in technical fields (Airey 2010; Hincks 2010). With respect to this finding, Hincks (2010) warned that the slower L2 speech rate will ultimately result in less content being communicated through L2 than L1 in situations where time is limited (e.g. in oral presentations) and thus prevent students from demonstrating the knowledge they have.

In addition, interestingly, when speaking in L1, the speech of students who were transferring their content knowledge from their L2 was more likely to contain 'distractors', i.e. words that did not contribute to the content of the message (e.g. fillers and reformulations). This resulted in greater redundancy and wordiness in the L1 responses of participants from the bilingual group. As a result, when speaking in their first language, students who gained their content knowledge through the medium of English appeared at a disadvantage compared to the students taught through their L1. The hesitations in the speech of the L2-instructed students or the laborious search for the right words could be interpreted incorrectly as signs of a lack of knowledge (Hofmannová, Novotná and Pípalová 2008). For example, Hönig (2009: 75) describes a case when the lack of fluency (signalled among other things by frequent pauses and slow speech) in an oral answer of a student was interpreted as showing 'less confidence in knowledge', despite the fact that the student's subject knowledge was found to be correct and sufficient.

As this study showed, however, some of these disfluencies could be instead attributed to the process of retrieval and production of content knowledge that had been encoded in a different language. Whereas previous research has described the problems of L2-instructed students with respect to L1 lexical gaps in their subject-specific vocabulary (Abedi, Hofstetter and Lord 2004; Airey 2010), the processing difficulties accompanying the transfer of knowledge from one language into another that become apparent in oral communication deserve further attention.

Appropriate academic format

Academic language places various restrictions on the form in which content can be communicated (e.g. the requirements of genre or formal language.). It is these rules that distinguish it from everyday, informal language and which restrict the choice of linguistic means which students can use (Scarcella 2003; Snow and Uccelli 2009). In order to produce acceptable academic language, the speaker has to be both aware of the form required and have resources for satisfying the requirement. Definitional format is one example of such 'restriction'. The L1 data indicated that students in both groups were aware of the expected form and had sufficient linguistic means to adhere to it. However, this was not observed in L2

answers of the bilingual group where the proportion of formal definitions was considerably lower.

The difference between L1 and L2 answers of the bilingual group indicates problems with transferring literacy skills from one language into another. The most likely cause for this is insufficient mastery of the target language, which prevents the students from using the skills known in their L1 in their L2 (Cummins 1991). The additional demands of formal academic language can prevent even very advanced users of language from expressing the content in appropriate form. Naturally, informal language can be used in classroom situation in the pursuit of knowledge development; yet situations should be created in which formal academic language, appropriate cognitively and linguistically, can be also cultivated.

Appropriate vocabulary

Development of subject-specific vocabulary is an integral part of subject learning. Thus, if students show deficiencies in this area it may raise questions about the depth and accuracy of their subject understanding. Results showed that when answering in L1, there were two instances in which the participants from the bilingual group appeared unaware of the correct disciplinary terms and chose more general words for expressing their knowledge. Whether this was caused by the unfamiliarity with the second sense of the word or whether in the limited time span, the more frequent word sense occurred to them more automatically, as a result, the students from the bilingual group appeared less well-grounded in their subject knowledge than the L1-instructed students.

This finding raises questions about the development of L1 subject-specific vocabulary of L2-educated students. It seems that the problem is not limited to merely providing the students with a translation of a term in L1 but concerns a more complex view of lexical development which includes the deepening knowledge of words' meanings and the awareness of the constrains on their use. This is most often learned from exposure (Snow and Uccelli 2009: 128) as could be seen in the case of L1-instructed students. These findings suggest that we should pay special attention to the development of subject-specific vocabulary in students' L1 if there is a possibility that students will be assessed in this language.

Conclusion

Compared with the control group, the bilingual group performed equally well in their L1 with regard to the appropriate format (formal definitions), linguistic accuracy and speech rate. On the other hand, these students performed worse in their L1 than the control group with respect to the proportion of informative speech and their lexical choice. Thus, different aspects of participants' content knowledge have been emphasised through their L1 as opposed to L2. This is in line with previous studies that described different types of constraints that bilingual learners face in each of their languages (e.g. Hincks 2010; Järvinen 2010).

It is important to note that although disfluencies, grammatical inaccuracies, imprecision in lexical choice are common features of everyday spoken language, they are not as frequent in formal monologic academic discourse where the flow of information is expected to be more controlled (Berman and Ravid 2009). However, even formal academic discourse is not governed by hard and fast rules. For this reason the L1 baseline was crucial for interpreting the data from L2-educated students in this study.

The problematic areas identified in students' L1 performance deserve additional attention. Whereas problems related to students' proficiency in L2 are often anticipated, those in L1 might not be recognised so easily. The two types of problems identified in students' L1 (disfluencies and the use of inaccurate terms) might be interpreted more readily as signs of knowledge gaps especially because L1, the students' native language, appears to be fully mastered, e.g. there are no apparent signs such as grammatical mistakes that would suggest linguistic problems. This might put bilingually-educated students at a disadvantage when they are compared to L1-educated students in situations beyond the school context where they might need to demonstrate their content knowledge orally through their L1 (such as job or university entrance interviews).

Further, the study provided evidence that the transfer of content knowledge between bilinguals' two languages is not as unproblematic as the assessment practice in some countries would suggest (Eurydice 2006). Although it is possible that more proceduralised literacy skills such as reading might be transferred more easily given that a certain threshold of proficiency is reached, this does not seem to be the case with declarative subject knowledge which may be more strongly tied to a particular language. This raises questions about appropriate and fair ways of assessing and comparing the content knowledge of students from CLIL and regular classes.

It should be noted that this study highlighted problems with expressing content knowledge through L1 and L2 relatively soon after the learning of this knowledge took place. It is possible that in the long-term some of the problems might be rectified either by a direct pedagogical intervention or as a result of students' growing understanding of the subject (Airey 2010). Nevertheless, the results of the study remind us to be cautious when assessing the content knowledge of students educated through their additional language, especially in situation where the bilinguals' performance might be directly compared with that of students from mainstream education.

This study aimed to contribute to our understanding of the impact of L1 or L2 language skills on knowledge demonstration. It focused in detail on four psycholinguistic and academic-language features which can (negatively) affect the ability to express content knowledge successfully. Content knowledge and its assessment is a very complex phenomenon. This study was able to address only a small part of it. Additional studies are therefore needed to investigate other forms of content knowledge assessment (e.g. through asking for explanations of scientific phenomena or causal relationships) as well as other aspects of academic language involved in knowledge demonstration (e.g. stylistic and genre-related features, grammatical complexity).

Notes

1. The differences in the quantity of content knowledge expressed through L1 and L2 are discussed elsewhere (Gablasova 2012).
2. In contrast to the other aspects of academic spoken communication examined in this study, the lexical analyses required a comparison of the same definitions (i.e. with the same content).

References

Abedi, J., C.H. Hofstetter and C. Lord. 2004. Assessment accommodations for English language learners: implications for policy-based empirical research. *Review of Educational Research* 74, no. 1: 1–28.

Airey, J. 2010. The ability of students to explain science concepts in two languages. *Hermes-Journal of Language and Communication Studies* 45: 35–49.

Berman, R.A. and D. Ravid. 2009. Becoming a literate language user. Oral and written text construction across adolescence. In *The Cambridge Handbook of Literacy*, ed. D.R. Olson and N. Torrance, 92–111. Cambridge: Cambridge University Press.
Bravo, M.A. and A. Cervetti. 2009. Teaching vocabulary through text and experience in content areas. In *Essential Readings on Vocabulary Instructions*, ed. M.F. Graves, 141–52. Newark, DE: International Reading Association.
Byrnes, H. 2008. Assessing content and language. In *Encyclopedia of Language and Education*, vol. 7, 2nd edition, ed. E. Shohamy and N.H. Hornberger, 37–52. New York: Springer.
Coyle, D., P. Hood and D. Marsh. 2010. *CLIL: Content and Language Integrated Learning*. New York: Cambridge University Press.
Cummins, J. 1991. Interdependence of first- and second-language proficiency in bilingual children. In *Language Processing in Bilingual Children*, ed. E. Bialystok, 70–90. Cambridge: Cambridge University Press.
Dalton-Puffer, C. 2007. *Discourse in Content and Language Integrated Learning (CLIL) Classrooms*. Amsterdam: John Benjamins.
Duff, P. 1997. Immersion in Hungary: an EFL experiment. In *Immersion Education: International Perspectives*, ed. R.K. Johnson and M. Swain, 19–43. Cambridge: Cambridge University Press.
Eurydice. 2006. *Content and Language Integrated Learning (CLIL) at School in Europe*. ec.europa.eu/languages/documents/.../clil-at-school-in-europe_en.pdf (accessed 6 June, 2012).
Flowerdew, J. 1992. Definitions in science lectures. *Applied Linguistics* 13, no. 2: 202–21.
Foster, P., A. Tonkyn and G. Wigglesworth. 2000. Measuring spoken language: a unit for all reasons. *Applied Linguistics* 21, no. 3: 354–75.
Gablasova, D. 2012. Learning and expressing technical vocabulary through the medium of L1 and L2 by Slovak-English bilingual high-school students. PhD thesis, The University of Auckland.
García, O. 2009. *Bilingual Education in the 21st Century: A Global Perspective*. Chichester, UK: Wiley-Blackwell.
Hincks, R. 2010. Speaking rate and information content in English lingua franca oral presentations. *English for Specific Purposes* 29, no. 1: 4–18.
Hofmannová, M., J. Novotná and R. Pípalová. 2008. Assessment approaches to teaching mathematics in English as a foreign language (Czech experience). *International CLIL Research Journal* 1, no. 1: 20–35.
Hönig, I. 2009. *Assessment in CLIL.* Unpublished MA thesis, University of Vienna.
Iris, M.A., B.E. Litowitz and M.W. Evens. 1988. Moving towards literacy by making definitions. *International Journal of Lexicography* 1, no. 3: 238–52.
Järvinen, H.M. 2010. Language as a meaning making resource in learning and teaching content. In *Language Use and Language Learning in CLIL Classrooms*, ed. C. Dalton-Puffer, T. Nikula and U. Smit, 145–68. Amsterdam: John Benjamins.
Johnson, R.K. 1997. The Hong Kong education system: late immersion under stress. In *Immersion Education: International Perspectives*, ed. R.K. Johnson and M. Swain, 171–89. Cambridge: Cambridge University Press.
Kellerman, E. and E. Bialystok. 1997. On psychological plausibility in the study of communication strategies. In *Communicaiton Strategies. Psycholinguistic and Sociolinguistic Perspectives*, ed. G. Kasper and E. Kellerman, 31–48. London: Longman.
Lindholm-Leary, K. and G. Borsato. 2006. Academic achievement. In *Educating English Language Learners. A Synthesis of Research Evidence*, ed. F. Genesee, K. Lindholm-Leary, W.M. Saunders and D. Christian, 176–223. Cambridge: Cambridge University Press.
Llach, M.P.A. 2011. *Lexical Errors and Accuracy in Foreign Language Writing*. Bristol, UK: Multilingual Matters.
Marinellie, S.A. 2009. The content of children's definitions: the oral-written distinction. *Child Language Teaching and Therapy* 25, no. 1: 89–102.
Menyuk, P. and M.E. Brisk. 2005. *Language Development and Education: Children with Varying Language Experiences*. New York: Palgrave.
Morgan, C. 2006. Appropriate language assessment in content and language integrated learning. *Language Learning Journal* 33, no. 1: 59–67.
Pečiar, S. (ed). 1959–1968. *Slovník slovenského jazyka* [The dictionary of the Slovak language]. Bratislava, Slovakia: Vydavateľstvo slovenskej akadémie vied.
Scarcella, R. 2003. *Academic English: A Conceptual Framework*. Irvine, CA: University of California Press.

Serra, C. 2007. Assessing CLIL at primary school: a longitudinal study. *International Journal of Bilingual Education and Bilingualism* 10, no. 5: 582–602.
Short, D.J. 1993. Assessing integrated language and content instruction. *TESOL Quarterly* 27, no. 4: 627–56.
Snow, C.E. and P. Uccelli. 2009. The challenge of academic language. In *The Cambridge Handbook of Literacy*, ed. D.R. Olson and N. Torrance, 112–33. Cambridge: Cambridge University Press.
Stohler, U. 2006. The acquisition of knowledge in bilingual learning: an empirical study on the role of language in content learning. *Views* 15, no. 3: 41–6.
Temmerman, M. 2009. Communiative aspects of definitions in classroom interaction: learning to define in class for first and second language learners. *Linguistics and Education* 20: 126–44.
Ullmann, M. 1999. History and geography through French: CLIL in a UK secondary school. In *Learning Through a Foreign Language: Models, Methods and Outcomes*, ed. J. Masih, 96–105. London: Centre for Information on Language Teaching and Research.
Watson, R. 1995. Relevance and definition. *Journal of Child Language* 22, no. 1: 211–22.