International Journal of Bilingual Education and Bilingualism
Vol. 15, No. 2, March 2012, 131–145

Routledge
Taylor & Francis Group

Linguistic and metalinguistic outcomes of intense immersion education: how bilingual?

Nicola Hermanto[a,b], Sylvain Moreno[c] and Ellen Bialystok[b]*

[a]*Department of Psychology, McGill University, Montreal, QC, Canada;* [b]*Department of Psychology, York University, Toronto, ON, Canada;* [c]*Rotman Research Institute, Baycrest Hospital & Department of Psychology, University of Toronto, Toronto, ON, Canada*

(*Received 17 September 2011; final version received 20 December 2011*)

Anglophone children in Grades 2 and 5 who attended an intensive French immersion program were examined for linguistic and metalinguistic ability in English and French. Measures of linguistic proficiency (vocabulary and grammatical knowledge) were consistently higher in English and remained so even after 5 years of immersion education in French. Measures of metalinguistic ability (letter fluency and ignoring semantic anomalies in sentence judgments) in French improved significantly over the two grades studied and closed the gap (letter fluency) or caught up with (sentence judgments) similar performance in English. This dissociation between developmental trajectories for linguistic and metalinguistic development is exactly the pattern expected for fully bilingual children, endorsing immersion education as a route to bilingualism.

Keywords: bilingual education; immersion education; language proficiency; metalinguistic development; advantages of bilingualism; language achievement

French immersion programs were introduced into Canadian public education in 1965 in St. Lambert, Quebec. The programs were designed to teach Anglophone children high levels of French proficiency by presenting the academic curriculum entirely in French, even though these children typically heard no French at home (see Genesee 1981 for different versions of immersion programs). From the beginning, there has been intense research monitoring the educational outcomes of children in these programs, particularly in terms of their development of language and literacy skills (for reviews see Genesee 1984; Safty 1988). The majority of this research has compared the progress of children in French immersion programs to their counterparts in regular English programs for their developing skills in English language and literacy, largely to reassure parents that English was not being sacrificed by education in French. Results tended to show initial delays in English skills (e.g., Barik and Swain 1975, 1976b) that disappeared after several years in the program (e.g., Barik and Swain 1976a, 1978; Kendall et al. 1987; Turnbull, Hart, and Lapkin 2003). There was less research examining the educational outcomes of French proficiency, but results showed moderate progress in French language and literacy (Barik and Swain 1978). One important result was that literacy skills largely (but not completely) transferred across these two languages, so instruction in French literacy led to developments in

*Corresponding author. Email: ellenb@yorku.ca

ISSN 1367-0050 print/ISSN 1747-7522 online

http://dx.doi.org/10.1080/13670050.2011.652591
http://www.tandfonline.com

English literacy (Comeau et al. 1999; Deacon, Wade-Woolley, and Kirby 2007, 2009; Jared et al. 2011). More recently, the concept of immersion education has been applied to different contexts using different languages, with comparable results (Spanish–English immersion: Ballester 2010; Cantonese–Mandarin immersion: Chen et al. 2008; Cantonese–English immersion: Lo and Murphy 2010).

In the majority of this literature, the question was to determine the degree to which language and academic skills developed at grade-appropriate levels in the two languages. An implicit assumption, therefore, was that children in immersion programs potentially resemble monolingual children developing those skills in single-language programs, and the research evaluated the outcomes against that expectation. Since the introduction of these programs much has been learned about these developments in bilingual children, and it turns out they are not identical to those of monolingual children being instructed in their only language. For both formal proficiency in two languages and metalinguistic ability, the development of bilingual children is different from that of their monolingual peers. Therefore, it is not known whether children in immersion programs more closely resemble monolingual children in either the home or school language of immersion students, or bilingual children learning both languages. Put another way, how bilingual are children becoming in immersion programs?

Studies of the lexical and grammatical proficiency of bilingual children generally show that these developments lag behind the levels achieved by monolingual learners of each language. In a recent analysis of 1738 children between the ages of 3 and 10 years, about half of whom were monolingual English speakers, the bilingual children obtained a lower standardized score on the Peabody Picture Vocabulary test (PPVT; Dunn and Dunn 1997) of receptive vocabulary than their monolingual counterparts (Bialystok et al. 2010). However, an analysis of a subset of those children who were 6 years old (Grade 1) that compared knowledge of home-related words (‘squash,’ ‘pitcher’) to school-related words (‘rectangle,’ ‘astronaut’) indicated no difference in school-based vocabulary, confining group differences to words most likely heard at home. All these children were being educated in English and had similar experience with academic vocabulary. Importantly, therefore, bilingual children had no disadvantage relative to their monolingual classmates in the language of school instruction.

Similar results have been found for grammatical development. Most studies investigating developing grammatical knowledge in bilingual children have compared progress in the two languages as a means of determining their degree of interdependence (Paradis and Genesee 1996). Most of these studies have shown that grammatical development proceeds independently in each language and is largely tied to vocabulary acquisition in that language (Conboy and Thal 2006; Marchman, Martínez-Sussmann, and Dale 2004; Simon-Cereijido and Gutiérrez-Clellen 2009), but some report interference across languages for specific structures (Austin 2007). However, studies that have compared the development of grammatical knowledge in bilingual children to that of monolingual children acquiring one of the languages have typically shown delays in the bilingual group (Marinis and Chondrogianni 2010; Müller and Hulk 2001; Paradis 2010). Thus, as is the case with vocabulary development, grammatical proficiency develops more slowly for bilingual children than for comparable monolinguals.

A much larger literature has examined metalinguistic development in bilingual children. Unlike results for developing formal language proficiency, these studies

have shown enhanced metalinguistic awareness in bilingual children for syntactic awareness (Galambos and Goldin-Meadow 1990; Galambos and Hakuta 1988; Ricciardelli 1992), word awareness (Ben-Zeev 1977; Cummins 1978), and, to a lesser extent, phonological awareness (Campbell and Sais 1995; Yelland, Pollard, and Mercuri 1993). There are, however, two constraints on these results that limit their generalizability. The first is that the outcomes depend on the relation between the child's two languages. For example, different patterns of emerging phonological awareness were found for children whose languages were English and Spanish (Bialystok, Majumder, and Martin 2003) than for children whose languages were English and Chinese (McBride-Chang et al. 2004). Moreover, most of the bilingual advantage found for phonological awareness disappeared when alphabetic literacy was introduced (Bruck and Genesee 1995).

Second, the specific task demands used in assessments of metalinguistic awareness are important in determining the outcomes. To this end, there is a distinction between metalinguistic tasks that depend primarily on formal knowledge of language (representation) and those that depend primarily on attentional processing to isolate form and meaning individually (control), a distinction described in detail elsewhere (Bialystok 1993, 2001). For example, a common test of metalinguistic awareness is to ask children to judge whether a sentence is grammatically correct or not, and in some cases, to correct the sentence if there is an error. In some studies, bilingual children outperformed monolingual children in such demonstrations of explicit knowledge of grammar (Galambos and Goldin-Meadow 1990), but in other studies, simple judgments of grammaticality showed no difference between monolingual and bilingual children (Bialystok 1986). A variation of that task is to include sentences that are grammatically correct but contain an irrelevant semantic error (i.e., the sentence is silly), requiring children to ignore the meaning and evaluate the grammar. This judgment does not require much formal knowledge of grammar because the sentence is intact (representation) but requires instead a high level of attentional focus to avoid being distracted by the irrelevant meaning (control). On these problems, bilingual children consistently outperform monolingual children (Bialystok 1986; Cromdal 1999). Using the same task with adults, Moreno et al. (2010) showed that this higher performance is related to an executive function advantage and involved functional brain differences between monolinguals and bilinguals that were shown through electrophysiology using event-related potentials. Thus, problems requiring children to make judgments of sentence grammaticality reveal children's level of ability for both language proficiency (grammatical knowledge when there is a syntactic error) and metalinguistic control (ignoring misleading meaning when the grammar is intact).

Another task used to assess children's linguistic and metalinguistic ability is verbal fluency. In a standardized version, this task is a neuropsychological assessment of the integrity of brain functioning. The task contains two conditions. The first, called category (or semantic) fluency, requires participants to generate as many words as possible in 60 seconds that conform to a category, such as 'animals.' The task is believed to assess vocabulary size (Delis, Kaplan, and Kramer 2001), and bilingual participants generally produce fewer words than comparable monolinguals because of their lower vocabulary in each language (Bialystok, Craik, and Luk 2008). The second, called letter (or phonological) fluency, requires participants to generate as many words as possible in 60 seconds that begin with a given letter, such as 'F.' This task assesses both vocabulary size and executive control (Delis, Kaplan, and

Kramer 2001), and bilingual participants typically perform comparably to monolinguals because their advantages in executive control compensate for their disadvantages in vocabulary size (Bialystok, Craik, and Luk 2008). Therefore, the distinction between the two versions of the task provides a means of distinguishing between performance typically associated with monolingual and bilingual participants. Whereas category fluency provides basic evidence for proficiency level, letter fluency indicates the extent to which bilingualism is involved. Thus, category fluency is part of linguistic proficiency, and letter fluency is part of metalinguistic ability.

In the present study, children's progress in developing English and French linguistic and metalinguistic skills in an intense French immersion environment was examined to determine the degree to which these developing abilities resemble those of bilingual children. We examined children in second- and fifth-grade who were attending a French-language school in an English-speaking community. Therefore, instead of assessing children's development against the usual standards of grade-appropriate levels for each language, children's language and metalinguistic abilities in English and French were used to determine how bilingual they were becoming in this educational context. If children follow the path of monolingual children, then proficiency in the school language (French) will be equivalent to that of the home language (English), and metalinguistic tasks requiring control (judging anomalous sentences and letter fluency) will be more difficult than their counterparts assessing language proficiency (detecting grammatical errors and category fluency). This is the usual situation for children who speak the majority language of the community at home, in this case, English. In contrast, if children follow the path of bilingual children, then proficiency in the home language (English) will be higher than proficiency in the school language (French), and metalinguistic tasks requiring control will be solved at least as easily as those requiring only representational knowledge of the language.

Method

Participants

A total of 83 students from a private school in which all instruction is delivered in French participated in the study. The school environment includes more French language than is typically found in public French immersion schools, because all the staff speaks French and interactions out of class are generally conducted in French. In traditional French immersion programs, French is largely confined to the classroom. Children whose first language or home language was French were excluded from the analyses, although children with other first languages or home languages (six children in Grade 2 and four children in Grade 5) were retained. There were 50 children (28 girls) in Grade 2 with a mean age of 7.7 years and 33 children (25 girls) in Grade 5 with a mean age of 10.6 years.

Measures and tasks

Background measures consisted of information from a detailed questionnaire regarding home language use and socioeconomic status (SES) as well as a test of non-verbal intelligence. Linguistic outcomes were assessed using a measure of receptive vocabulary, performance on the sentence judgment task for correct and

grammatically incorrect sentences, and performance on the verbal fluency task for category fluency. Metalinguistic outcomes were assessed using performance on the sentence judgment task for semantically anomalous sentences and performance on the verbal fluency task for letter fluency.

Language and social background questionnaire (LSBQ)

The LSBQ was a parent report of language use in the home by the child and of social background information, such as age, place of birth, and education levels of the parents. Language use in a number of specified situations (e.g., language spoken at the dinner table, language in which parents watched movies and videos, language spoken by children to parents) was measured on a 5-point Likert scale, ranging from all English (1) to no English (5). Maternal education was used as a proxy for SES and was measured as the proportion out of 5 on a 5-point Likert scale indicating level of education: 1 = not completed high school, 2 = high school diploma, 3 = some post-secondary education, 4 = bachelor's degree, and 5 = graduate or professional degree.

Non-verbal intelligence

The Matrices subtest of the Kaufman Brief Intelligence Test (KBIT-2) was administered to assess fluid reasoning (Kaufman and Kaufman 2004). On each trial, the child was presented with visual stimuli representing either drawings of concrete objects or abstract figures. In the first part, the child saw a target drawing at the center of the page and five additional drawings below it and was asked to identify which of the five stimuli matched the target image. For the other two sections, the child saw an incomplete display of 2×2 or 3×3 visual stimuli with one stimulus missing, and five stimuli below the display. The task was to choose the stimulus to complete the displayed pattern. The testing, scoring, and standardization followed the standard procedure described in the manual.

Receptive vocabulary in English and French

English vocabulary was assessed by the Peabody Picture Vocabulary Test 3rd Edition (PPVT-III; Dunn and Dunn 1997) and French vocabulary by the Echelle de Vocabulaire en Images Peabody (EVIP; Dunn, Theriault-Whalen, and Dunn 1993), a standardized French adaptation of the PPVT. In both tests, the child was shown a page with four pictures while the experimenter said a word, and the task was to point to the picture that best illustrated that word. Raw scores were transformed to standardized scores using an age-corrected norm table. Scores on this task were used as an indication of vocabulary knowledge in each language and contributed to the assessment of formal language proficiency.

Sentence judgment task

Sentences from Atchley and colleagues (2006) were adapted to build 120 English sentence frames that were grammatically correct and meaningful, grammatically incorrect but meaningful, or grammatically correct but semantically anomalous. Thus, a correct sentence such as 'Where does a horse like to run?' could be made syntactically incorrect: 'Where does a horse like to runs?,' or semantically

anomalous, 'Where does a horse like to sail?' The syntactically incorrect version always involved a third-person ending on what should correctly be an infinitive (i.e., *to runs* instead of *to run*). In the semantically anomalous version, the matrix verb appeared in the grammatically appropriate form but introduced an unsuitable pairing of actions with agents (e.g., *animal – sail*). The French version of the task used the same 120 sentence frames as the English version of the task.

Three stimulus lists were created, each containing 40 sentences for each of the three experimental conditions. Items were counterbalanced such that only one version of each sentence was presented on a given list. Thus, each participant heard a total of 120 sentences. The sentences were recorded by a female speaker in a soundproof booth on a Dell Inspiro laptop using Adobe Audition 2. Each trial began with a fixation cross that appeared for 500 ms after which a sentence was presented auditorily. Children were told that they were going to hear sentences and had to decide if it was said the right way or not. They were told that some of the sentences would be silly, but that that was ok, as long as the sentences were said *the right way*. There was a 1450-ms response interval after each sentence before the next trial began. Children indicated their response by pressing one of two buttons, which were counterbalanced for left and right positions across participants. They were given a trial block of 12 sentences and monitored to determine if they had understood the directions, and the trial block was repeated as many times as necessary to ensure successful completion of the task before the experimental blocks began. Accuracy rates and reaction times (RTs) were recorded.

The grammatically correct and meaningful sentences provide a baseline condition that makes little demand on either linguistic or metalinguistic knowledge. The grammatically incorrect but meaningful sentences increase the need for linguistic knowledge of grammatical rules. Children must have sufficient understanding of the grammatical structure of the language to assess whether the sentence is well formed or not. The metalinguistic demands in this case are minimal. The grammatically correct but anomalous sentences, in contrast, make minimal demands on linguistic knowledge because the grammar is correct but instead place high demands on metalinguistic ability. The metalinguistic challenge is to ignore the salient meaning and focus attention on the grammatical form when these values conflict, so the task requires judgments only of form. In previous research, children between 5 and 9 years old were equivalent in judging correct sentences and meaningful sentences with grammatical errors, but bilingual children were better than monolingual children in judging that sentences that were grammatically correct but anomalous were well formed (Bialystok 1986, 1988; Cromdal 1999). Thus, performance on the grammatically incorrect sentences assesses formal language proficiency and performance on the anomalous sentences assesses metalinguistic awareness.

Verbal fluency task

Verbal fluency was assessed using the Delis-Kaplan Executive Function (D-KEF) system (Delis, Kaplan, and Kramer 2001). In both the English and French versions of this task, participants were asked to produce as many words as possible in 60 seconds. For category fluency, they were asked to name members of two categories, clothing items and girls' names, and for letter fluency, to produce words that start with letters F, A, and S. The usual restrictions of the letter task that exclude proper names, numbers, and morphological variations of the same word were

removed because it was believed that the task would be too difficult for the youngest participants. Responses were recorded on a digital recorder. Raw scores were obtained by subtracting incorrect responses (words that did not start with the specified letter or not in the designated categories) and repeated words from the total number of responses. Following previous research with this task, performance on category fluency is interpreted as a reflection of formal linguistic knowledge and performance on letter fluency as a reflection of metalinguistic ability.

Procedure

Parents indicated consent for their children's participation in accordance with the guidelines established by the university ethics board and completed the LSBQ. Children were tested individually at their school in two separate sessions. The first session consisted of the KBIT and English versions of the PPVT, verbal fluency, and sentence judgment tasks, and the second session consisted of French versions of the PPVT, verbal fluency, and sentence judgment tasks. The order of the sessions was counterbalanced across children and the order of the tasks within each session varied randomly.

Results

Table 1 presents the mean scores and standard deviations for the background measures. One-way ANOVAs with grade as a between-subjects factor indicated no differences in SES as measured by mother's years of education, $F < 1$, or non-verbal intelligence as measured by the KBIT, $F < 1$. Mean scores from the LSBQ for home language use, ranging between 1 (Always English) and 5 (Always Other Language), also indicated that homes were predominantly monolingual English. Other home languages included Mandarin (three participants), Cantonese (three participants), Italian (two participants), Turkish (one participant), and Estonian (one participant).[1]

Scores from the English (PPVT) and French vocabulary (EVIP) tests are also presented in Table 1. The data were analyzed with a two-way ANOVA for language and grade. There was a significant main effect of language, $F(1,76) = 186.73$, $p < 0.0001$, $\eta_p^2 = 0.71$, with higher scores on the English test ($M = 104.9$, SD = 17.2) than on the French test ($M = 82.5$, SD = 16.5). There was no main effect of grade, $F < 1$, indicating that there were no significant differences in vocabulary between the two grades (because scores are standardized for age), and no interaction of grade and language, $F(1,76) = 2.69$, n.s.

Table 1. Mean score (standard deviation) for background measures by grade.

Measure	Grade 2 Mean (SD)	Grade 5 Mean (SD)
Language spoken by parents at home	2.0 (1.3)	2.0 (1.2)
Language spoken by child at home	1.7 (1.1)	1.8 (1.0)
SES (maternal education)	0.9 (0.1)	0.9 (0.1)
Non-verbal intelligence (KBIT-2)	104.4 (19.8)	105.8 (12.4)
English receptive vocabulary (PPVT-III)	110.4 (15.5)	112.7 (20.4)
French receptive vocabulary (EVIP)	83.7 (15.2)	80.3 (18.5)

Results from the sentence judgment task are presented in Figure 1. RT data were omitted from the analysis because they were too long to be interpretable, with overall mean RT of approximately 3 seconds.[2] The accuracy data were analyzed with a three-way ANOVA for grade, language, and sentence type. There was a main effect of grade, $F(1,74) = 57.49$, $p < 0.0001$, $\eta_p^2 = 0.44$, indicating that fifth graders achieved higher accuracy scores ($M = 0.87$, SD $= 0.07$) than second graders ($M = 0.67$, SD $= 0.13$). There was a main effect of language, $F(1,74) = 21.96$, $p < 0.0001$, $\eta_p^2 = 0.23$, showing that students performed better overall on the English version of the task ($M = 0.81$, SD $= 0.16$) than the French version ($M = 0.69$, SD $= 0.19$). There was also a main effect of sentence type, F (2, 148) $= 30.48$, $p < 0.0001$, $\eta_p^2 = 0.29$. Grammatically correct sentences were judged more accurately ($M = 0.82$, SD $= 0.14$) than either grammatically incorrect ($M = 0.67$, SD $= 0.18$), $F(1,74) = 97.17$, $p < 0.0001$, $\eta_p^2 = 0.57$, or semantically anomalous sentences ($M = 0.74$, SD $= 0.19$), $F(1,74) = 15.71$, $p < 0.001$, $\eta_p^2 = 0.17$; and semantically anomalous sentences were judged more accurately than grammatically incorrect sentences ($M = 0.67$, SD $= 0.18$), $F(1,74) = 10.63$, $p < 0.002$, $\eta_p^2 = 0.13$.

There were two two-way interactions that limit these main effects. First was a language by grade interaction, $F(1,74) = 10.07$, $p < 0.002$, $\eta_p^2 = 0.12$. To further explore this interaction, tests of simple effects were run to examine language differences within each grade. The analysis showed that second graders were more accurate in responding to English sentences ($M = 0.76$, SD $= 0.17$) than French sentences ($M = 0.59$, SD $= 0.16$), $F(1,47) = 36.62$, $p < 0.0001$, $\eta_p^2 = 0.44$; however, the difference disappeared by fifth grade and the older children judged English ($M = 0.89$, SD $= 0.12$) and French ($M = 0.86$, SD $= 0.09$), F (1,27) $= 1.21$, n.s., sentences equivalently. Secondly, there was a language by sentence type interaction, $F(2,148) = 12.68$, $p < 0.0001$, $\eta_p^2 = 0.15$. To further explore this interaction, tests of

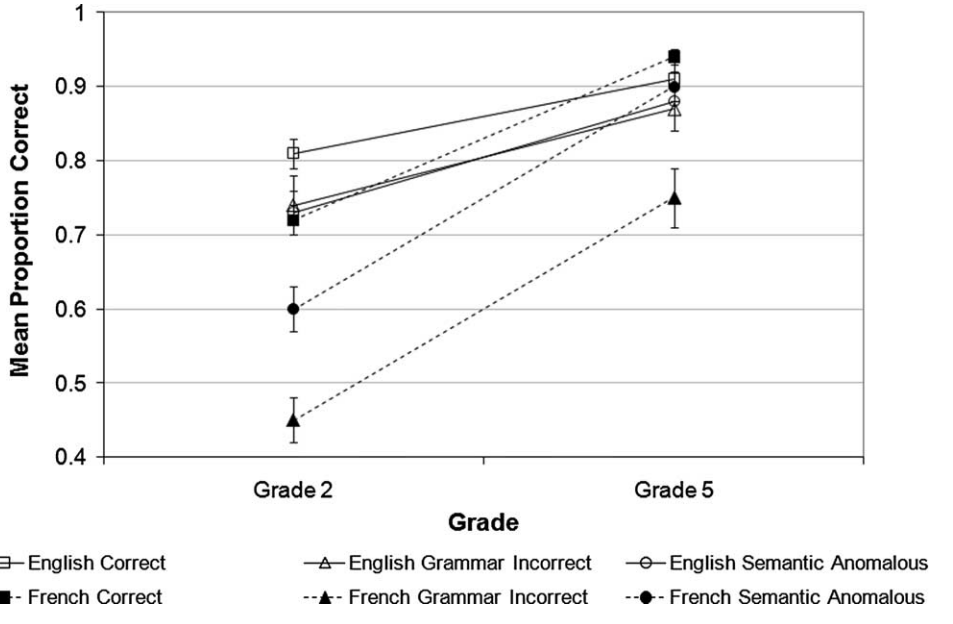

Figure 1. Mean accuracy rates and standard errors for the sentence judgment task in English and French by grade.

simple effects were run to examine language differences at each level of sentence type. There were small but significant differences in which English judgments were more accurate for the grammatically correct, ($M = 0.85$, SD $= 0.18$), $F(1,75) = 5.89$, $p < 0.02$, $\eta_p^2 = 0.07$, and semantically anomalous sentences, ($M = 0.79$, SD $= 0.24$), $F(1,75) = 5.32$, $p < 0.02$, $\eta_p^2 = 0.07$, than the corresponding French judgments ($M = 0.80$, SD $= 0.17$ and $M = 0.71$, SD $= 0.25$, respectively). However, there was a larger difference in accuracy for grammatically incorrect sentences, in which English judgments ($M = 0.79$, SD $= 0.23$) were more accurate than French judgements ($M = 0.56$, SD $= 0.24$), $F(1,75) = 49.10$, $p < 0.0001$, $\eta_p^2 = 0.40$.

Results from the verbal fluency task are presented in Figure 2. The data were analyzed with a three-way ANOVA for grade, language, and fluency type. There was a main effect of grade, $F(1,77) = 114.41$, $p < 0.0001$, $\eta_p^2 = 0.60$, showing that older children produced more responses ($M = 12.1$, SD $= 2.3$) than younger children ($M = 7.3$, SD $= 1.7$). There was a main effect of language, $F(1,77) = 204.23$, $p < 0.0001$, $\eta_p^2 = 0.73$, indicating that participants produced more responses in the English version of the task ($M = 10.6$, SD $= 3.3$) than the French version ($M = 7.7$, SD $= 3.0$). There was also a main effect of fluency type, $F(1,77) = 175.58$, $p < 0.0001$, $\eta_p^2 = 0.70$, showing that participants produced more responses for category fluency ($M = 11.5$, SD $= 4.0$) than for letter fluency ($M = 7.6$, SD $= 2.8$).

There were two two-way interactions that are necessary to interpret these effects. First was an interaction of grade and fluency type, $F(1,77) = 7.37$, $p = 0.01$, $\eta_p^2 = 0.09$. To further explore this interaction, tests of simple effects were run to examine grade differences at each level of fluency type. For letter fluency, second graders produced fewer responses ($M = 6.0$, SD $= 1.7$) than fifth graders ($M = 10.1$, SD $= 2.4$), $F(1,78) = 78.36$, $p < 0.0001$, $\eta_p^2 = 0.33$. Similarly for category fluency, second graders produced fewer responses ($M = 9.3$, SD $= 2.5$) than fifth graders

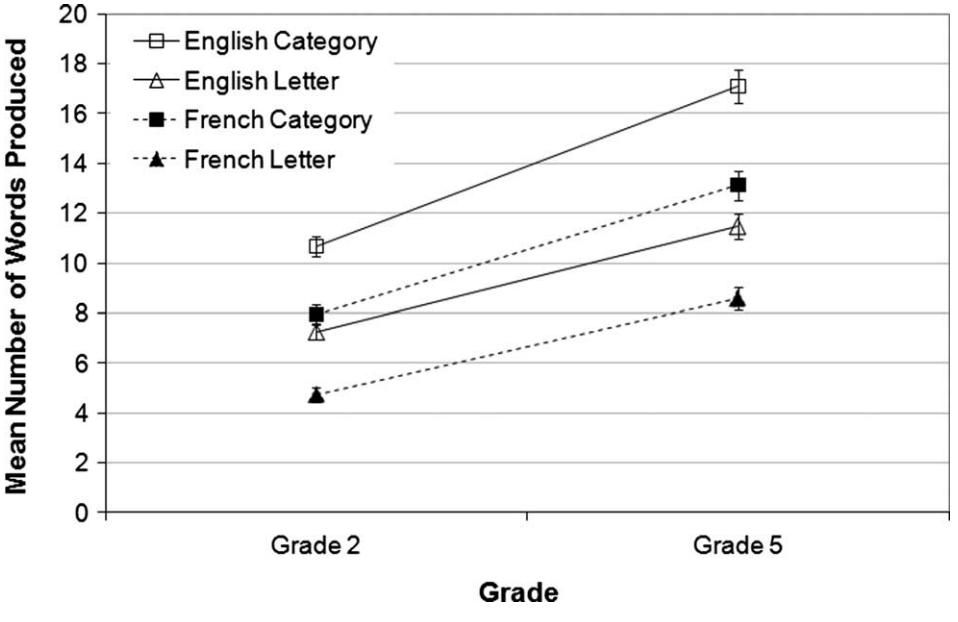

Figure 2. Mean number of words produced and standard errors in the category and letter conditions of the verbal fluency task in English and French by grade.

($M = 15.2$, SD = 3.2), $F(1,78) = 82.06$, $p < 0.0001$, $\eta_p^2 = 0.34$, but the discrepancy was slightly larger in this case. Second was a language by fluency type interaction, $F(1,77) = 4.43$, $p = 0.04$, $\eta_p^2 = 0.05$. To understand this interaction, tests of simple effects examined language differences at each level of fluency type. For letter fluency, children produced more responses in English ($M = 8.9$, SD = 3.2) than in French ($M = 6.2$, SD = 2.8), $F(1,78) = 119.19$, $p < 0.0001$, $\eta_p^2 = 0.60$. Similarly for category fluency, children produced more responses in English ($M = 13.2$, SD = 4.4) than in French ($M = 9.9$, SD = 3.8), $F(1,78) = 126.82$, $p < 0.0001$, $\eta_p^2 = 0.62$, but the difference between languages was larger for this fluency type.

Discussion

The linguistic and metalinguistic skills of second- and fifth-grade Anglophone children attending a French school were examined on English and French measures. The purpose of the study was to investigate the degree to which these children were becoming bilingual in the educational context of an intensive French immersion program. The children lived in middle-class families in an English-speaking environment. Other situations may lead to stronger school-language skills than home-language skills, but that has not been the case with this population.

The French program was similar to that in standard public French immersion education, but the environment included more French in the interactions outside the classroom, providing greater support for children's development of French language proficiency. As expected, fifth-graders outperformed second-graders on all tasks, showing the typical developmental progress of linguistic and metalinguistic abilities over these grades. Nonetheless, the results showed different developmental patterns for linguistic and metalinguistic abilities in the two languages. Although students demonstrated better performance in English than in French across all measures, the relation between standardized scores on the linguistic tasks in the two languages did not change much across the grades, even though children had spent an additional 3 years in a French school environment.

The pattern was most clear for the vocabulary scores where the standardized score for French vocabulary was the same in Grade 5 as it was in Grade 2. Although children's French vocabulary was developing at a normal rate, they were not making gains relative to their English vocabulary growth. The number of words produced on the category fluency test is also an indication of vocabulary knowledge. For the English test, children increased by 6.4 words, producing an average of 10.7 words in Grade 2 and 17.1 words in Grade 5; for the French test, children increased by 5.1 words, producing an average of 8.0 words in Grade 2 and 13.1 words in Grade 5. Not only were more English words produced at both grade levels but also the increase was greater for English than for French.

For grammatical knowledge, children in Grade 5 were 87% accurate in detecting grammatical errors in meaningful sentences in English but only 75% accurate in performing the same judgment in French. This is especially notable because the baseline (grammatically correct meaningful) and metalinguistic (semantically anomalous) conditions were performed equivalently in the two languages by the older children, with the only discrepancy in their ability to detect grammatical violations in French. Thus, children are making greater linguistic progress in the language of home than they are in the language of schooling.

This pattern is similar to that reported in the French immersion literature in which progress in French language and literacy scores is modest when compared with monolingual French norms. It is also consistent with aspects of language development reported for bilingual children who are instructed in a language that is different from the language of the home. In that case, bilingual children educated in English but whose home language was not English obtained lower overall scores on a receptive vocabulary test in English (the language of schooling) than did their monolingual classmates (Bialystok et al. 2010). Unlike the present study, it was not possible to assess the vocabulary of the bilingual children in their home language, but assuming a similar pattern, the present data indicate that the vocabulary scores in the language of schooling would be lower than vocabulary scores in the language of home.

Another way of considering these results is to compare them to those obtained from a group of monolingual English-speaking children in Grades 2 and 5 who were part of a different study but completed the English version of the same tasks (Bialystok, Peets, and Moreno, forthcoming). For grammaticality judgment, Grade 2 children in both studies obtained 74% accuracy in detecting grammatical errors in the English sentences, but by Grade 5, the French immersion students in the present study were more accurate in detecting grammatical errors in English (87%) than were monolingual English-speaking children (78%). For generating English words in category fluency, Grade 2 children in both studies produced about 10 words, but by Grade 5, the children in the present study produced about 17 words, and the monolingual children in the previous study produced 15 words. Thus, in the present study, the most dramatic outcome of immersion education in French is in its accelerated effect on English language proficiency.

The results are different for metalinguistic outcomes. The relevant variables in this case are the ability to agree that semantically anomalous sentences are grammatically correct and to generate words to conform to a phonological cue. Judging anomalous sentences has been shown to be performed more accurately by bilingual children than by monolingual children between the ages of 5 and 9 years (Bialystok 1986, 1988). Consistent with this pattern, in the study by Bialystok, Peets, and Moreno (forthcoming) in which monolingual children in Grades 2 and 5 performed this task in English, children in Grade 2 had equivalent success with both ungrammatical and anomalous sentences (74% and 75% correct, respectively) but in Grade 5, children were more accurate on judging ungrammatical sentences (78% correct) than anomalous sentences (71% correct). In the present study, children in Grade 2 (74%) and Grade 5 (87%) performed both judgments equivalently. Not only were the difficult anomalous judgments (metalinguistic) performed as well as the grammaticality judgments (linguistic) but also the children in Grade 5 at the French school appeared to be more accurate than monolingual English children.

A similar pattern was found for performance on the verbal fluency test. Considering that category fluency indexes language proficiency and letter fluency includes metalinguistic involvement, we can compare performance in English for the children in the present study with the monolingual children tested in the study by Bialystok, Peets, and Moreno (forthcoming). English category fluency scores were approximately equivalent for children in both studies in Grade 2 (10 words) and Grade 5 (17 words). However, for English letter fluency in Grade 2, monolingual children produced 8.1 words and immersion children in the present study produced 7.2 words; in Grade 5, monolingual children produced 10.7 words and immersion

children in the present study produced 11.5 words, reversing the order. Put another way, monolingual children increased their performance by 2.6 words over the 3 years, but immersion children increased their performance by 4.3 words for the same period. This pattern indicates greater progress in a difficult metalinguistic task in English for children in the French immersion environment than for monolingual children who only spoke English. Thus, in both the grammaticality judgment task and verbal fluency task, the condition that signals metalinguistic development improved more in immersion children than in monolingual children. The comparisons with the monolinguals from the study by Bialystok, Peets, and Moreno (in press) are speculative because they were not included in this controlled experiment but provide a benchmark for interpreting the present results.

There are three important conclusions from the present results. The first is evidence for a dissociation between children's developing linguistic and metalinguistic ability in the two languages. English skills were consistently high, possibly higher than those of comparable monolingual children (although that was not directly tested in this study), but French skills were different for the two types of abilities. This distinction has implications for understanding children's development of the linguistic and metalinguistic abilities that form the basis of literacy and for understanding interactions between developing cognitive and linguistic ability in the context of an educational program that is not based on the home language.

The second outcome is that formal proficiency in French as indicated by vocabulary and grammatical knowledge was consistently lower than the comparable ability in English and developed more slowly in spite of all formal education being conducted in French. In contrast, metalinguistic skills were equally good for both languages and developed significantly across the two grades studied. Thus, there is a gap in children's proficiency in the two languages with poorer outcomes on formal measures in the language of schooling, although metalinguistic skills were significantly enhanced in both languages. In this sense, the children in this immersion school are not becoming native speakers of French and are not making significant gains in their level of French proficiency across the 3 years studied. For example, their standardized scores on the French receptive vocabulary test are just over one standard deviation below the population norm in both grades. However, one standard deviation is not a liability and need not impede children's academic progress; the interesting finding is that their English scores are almost one standard deviation *above* the population norm and remain so even after 3 years of instruction only in French. These results indicate the limitation of developing proficiency in a language that is not supported at home.

Finally, the pattern of results for linguistic and metalinguistic performance conforms to that found for bilingual children who are educated in English but speak a non-English language at home (e.g., Barac and Bialystok, forthcoming). Specifically, formal proficiency scores were stronger in the home language than in the language of schooling, but metalinguistic results were better than those found for monolingual children. Moreover, unlike the pattern typically found in monolingual children, metalinguistic tasks requiring control (judging anomalous sentences and letter fluency) were performed better than their counterpart conditions requiring representation. The conclusion, therefore, is that the experience of schooling in French in conjunction with an English home language environment produces patterns of linguistic development typically found for fully bilingual children. Thus, these children are becoming bilingual and bringing with them the positive

outcomes of bilingualism. With continued exposure to and instruction in the formal structure of French, their linguistic skills will surely improve as well.

Acknowledgement

This work was partially supported by grant R01HD052523 from the US National Institutes of Health and by grant A2559 from the Natural Sciences and Engineering Research Council of Canada to EB. We thank the students and staff at the school for taking part in this study. We also wish to thank Kathleen Peets for her contribution to organizing the study, Tashua Case for coordinating the data collection, Camille Blochet for helping with data collection, and Raluca Barac for assistance in the preliminary statistical analyses.

Notes

1. All analyses were repeated, excluding the 10 children who spoke another language at home, and none of the results changed.
2. Reaction time data are intended to provide an index of cognitive processing, but when RTs exceed approximately 1.5 seconds, they no longer indicate cognitive processes.

References

Atchley, R.A., M.L. Rice, S.K. Betz, K.M. Kwasny, J.A. Sereno, and A. Jongman. 2006. A comparison of semantic and syntactic event related potentials generated by children and adults. *Brain and Language, Special Issue: Language Comprehension Across the Life Span* 99: 236–46.

Austin, J. 2007. Grammatical interference and the acquisition of ergative case in bilingual children learning Basque and Spanish. *Bilingualism: Language and Cognition* 10, no. 3: 315–31. DOI: 10.1017/S1366728907003094

Ballester, E.P. 2010. Child L2 development of syntactic and discourse properties of Spanish subjects. *Bilingualism: Language and Cognition* 13, no. 2: 185–216. DOI: 10.1017/S1366728909990447

Barac, R., and E. Bialystok. Forthcoming. Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Development*.

Barik, H.C., and M. Swain. 1975. Three-year evaluation of a large scale early grade French immersion program: The Ottawa study. *Language Learning* 25, no. 1: 1–30. DOI: 10.1111/j.1467-1770.1975.tb00106.x

Barik, H.C., and M. Swain. 1976a. English-French bilingual education in the early grades: The Elgin study through grade four. *Modern Language Journal* 60, no. 1–2: 3–17. DOI: 10.2307/325408

Barik, H.C., and M. Swain. 1976b. Update on French immersion: The Toronto study through grade 3. *Canadian Journal of Education* 1, no. 4: 33–42. DOI: 10.2307/1494622

Barik, H.C., and M. Swain. 1978. Evaluation of a French immersion program: The Ottawa study through grade five. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement* 10, no. 3: 192–201. DOI: 10.1037/h0081548

Ben-Zeev, S. 1977. The influence of bilingualism on cognitive strategy and cognitive development. *Child Development* 48, no. 3: 1009–18. DOI: 10.2307/1128353

Bialystok, E. 1986. Factors in the growth of linguistic awareness. *Child Development* 57, no. 2: 498–510. DOI: 10.2307/1130604

Bialystok, E. 1988. Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology* 24, no. 4: 560–7. DOI: 10.1037/0012-1649.24.4.560

Bialystok, E. 1993. Metalinguistic awareness: The development of children's representations of language. In *Systems of representation in children: Development and use*, ed. C. Pratt and A.F. Garton, 211–33. Oxford: John Wiley and Sons.

Bialystok, E. 2001. *Bilingualism in development: Language, literacy, and cognition*. New York, NY: Cambridge University Press.

Bialystok, E., F.I.M. Craik, and G. Luk. 2008. Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics* 21, no. 6: 522–38. DOI: 10.1016/j.jneuroling.2007.07.001

Bialystok, E., G. Luk, K.F. Peets, and S. Yang. 2010. Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition* 13, no. 4: 525–31. DOI: 10.1017/S1366728909990423

Bialystok, E., S. Majumder, and M.M. Martin. 2003. Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics* 24, no. 1: 27–44. DOI: 10.1017/S014271640300002X

Bialystok, E., K. Peets, and S. Moreno. Forthcoming. Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied Psycholinguistics*.

Bruck, M., and F. Genesee. 1995. Phonological awareness in young second language learners. *Journal of Child Language* 22, no. 2: 307–24. DOI: 10.1017/S0305000900009806

Campbell, R., and E. Sais. 1995. Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology* 13, no. 1: 61–8.

Chen, X., Y. Ku, E. Koyama, R.C. Anderson, and W. Li. 2008. Development of phonological awareness in bilingual Chinese children. *Journal of Psycholinguistic Research* 37, no. 6: 405–18. DOI: 10.1007/s10936-008-9085-z

Comeau, L., P. Cormier, É. Grandmaison, and D. Lacroix. 1999. A longitudinal study of phonological processing skills in children learning to read in a second language. *Journal of Educational Psychology* 91, no. 1: 29–43. DOI: 10.1037/0022-0663.91.1.29

Conboy, B.T., and D.J. Thal. 2006. Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development* 77, no. 3: 712–35. DOI: 10.1111/j.1467-8624.2006.00899.x

Cromdal, J. 1999. Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish–English bilinguals. *Applied Psycholinguistics* 20, no. 1: 1–20. DOI: 10.1017/S0142716499001010

Cummins, J. 1978. Bilingualism and the development of metalinguistic awareness. *Journal of Cross-Cultural Psychology* 9, no. 2: 131–49. DOI: 10.1177/002202217892001

Deacon, S.H., L. Wade-Woolley, and J. Kirby. 2007. Crossover: The role of morphological awareness in French immersion children's reading. *Developmental Psychology* 43, no. 3: 732–46. DOI: 10.1037/0012-1649.43.3.732

Deacon, S.H., L. Wade-Woolley, and J.R. Kirby. 2009. Flexibility in young second-language learners: Examining the language specificity of orthographic processing. *Journal of Research in Reading* 32, no. 2: 215–29. DOI: 10.1111/j.1467-9817.2009.01392.x

Delis, D.C., E. Kaplan, and J.H. Kramer. 2001. *Verbal fluency subtest of the Delis–Kaplan Executive Function System*. San Antonio, TX: The Psychological Corporation.

Dunn, L.M., and L.M. Dunn. 1997. *Peabody picture vocabulary test – third edition*. Bloomington, MN: Pearson Assessments.

Dunn, L., C. Theriault-Whalen, and L. Dunn. 1993. *Echelle de vocabulaire en images Peabody: Adaptation française du Peabody Picture Vocabulary test-revised*. Toronto, ON: PsyCan.

Galambos, S.J., and S. Goldin-Meadow. 1990. The effects of learning two languages on levels of metalinguistic awareness. *Cognition* 34, no. 1: 1–56. DOI: 10.1016/0010-0277(90)90030-N

Galambos, S.J., and K. Hakuta. 1988. Subject-specific and task-specific characteristics of metalinguistic awareness in bilingual children. *Applied Psycholinguistics* 9, no. 2: 141–62. DOI: 10.1017/S0142716400006780

Genesee, F. 1981. A comparison of early and late second language learning. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement* 13, no. 2: 115–28. DOI: 10.1037/h0081168

Genesee, F. 1984. Beyond bilingualism: Social psychological studies of French immersion programs in Canada. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement. Special Issue: Social Psychology Applied to Social Issues in Canada* 16, no. 4: 338–52. DOI: 10.1037/h0080864

Jared, D., P. Cormier, B.A. Levy, and L. Wade-Woolley. 2011. Early predictors of biliteracy development in children in French immersion: A 4-year longitudinal study. *Journal of Educational Psychology* 103, no. 1: 119–39. DOI: 10.1037/a0021284

Kaufman, A.S., and N.L. Kaufman. 2004. *Kaufman brief intelligence test – second edition*. Circle Pines, MN: American Guidance Service, Inc.

Kendall, J.R., G. Lajeunesse, P. Chmilar, and L.R. Shapson. 1987. English reading skills of French immersion students in kindergarten and grades 1 and 2. *Reading Research Quarterly* 22, no. 2: 135–59. DOI: 10.2307/747662

Lo, Y.Y., and V.A. Murphy. 2010. Vocabulary knowledge and growth in immersion and regular language-learning programmes in Hong Kong. *Language and Education* 24, no. 3: 215–38. DOI: 10.1080/09500780903576125

Marchman, V.A., C. Martínez-Sussmann, and P.S. Dale. 2004. The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science* 7, no. 2: 212–24. DOI: 10.1111/j.1467-7687.2004.00340.x

Marinis, T., and V. Chondrogianni. 2010. Production of tense marking in successive bilingual children: When do they converge with their monolingual peers? *International Journal of Speech-Language Pathology* 12, no. 1: 19–28. DOI: 10.3109/17549500903434125

McBride-Chang, C., E. Bialystok, K.K.Y. Chong, and Y. Li. 2004. Levels of phonological awareness in three cultures. *Journal of Experimental Child Psychology*, 89, no. 2: 93–111. DOI: 10.1016/j.jecp.2004.05.001

Moreno, S., E. Bialystok, Z. Wodniecka, and C. Alain. 2010. Conflict resolution in sentence processing by bilinguals. *Journal of Neurolinguistics* 23: 564–79. DOI: 10.1016/j.jneuroling.2010.05.002

Müller, N., and A. Hulk. 2001. Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4, no. 1: 1–21. DOI: 10.1017/S1366728901000116

Paradis, J. 2010. Bilingual children's acquisition of English verb morphology: Effects of language exposure, structure complexity, and task type. *Language Learning* 60, no. 3: 651–80. DOI: 10.1111/j.1467-9922.2010.00567.x

Paradis, J., and F. Genesee. 1996. Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition* 18: 1–25.

Ricciardelli, L.A. 1992. Bilingualism and cognitive development in relation to threshold theory. *Journal of Psycholinguistic Research* 21, no. 4: 301–16. DOI: 10.1007/BF01067515

Safty, A. 1988. French immersion and the making of a bilingual society: A critical review and discussion. *Canadian Journal of Education* 13: 243–62. DOI: 10.2307/1494954

Simon-Cereijido, G., and V.F. Gutiérrez-Clellen. 2009. A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics* 30, no. 2: 315–37. DOI: 10.1017/S0142716409090134

Turnbull, M., D. Hart, and S. Lapkin. 2003. Grade 6 French immersion students' performance on large-scale reading, writing, and mathematics tests: Building explanations. *Alberta Journal of Educational Research* 49, no. 1: 6–23.

Yelland, G.W., J. Pollard, and A. Mercuri. 1993. The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics* 14, no. 4: 423–44. DOI: 10.1017/S0142716400010687