Applied Psycholinguistics **35** (2014), 177–191
doi:10.1017/S0142716412000288

Producing bilinguals through immersion education: Development of metalinguistic awareness

ELLEN BIALYSTOK
York University

KATHLEEN F. PEETS
Ryerson University

SYLVAIN MORENO
Rotman Research Institute of Baycrest

Received: February 13, 2011 Accepted for publication: October 25, 2011

ADDRESS FOR CORRESPONDENCE
Ellen Bialystok, Department of Psychology, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada. E-mail: ellenb@yorku.ca

ABSTRACT
This study examined metalinguistic awareness in children who were becoming bilingual in an immersion education program. The purpose was to determine at what point in emerging bilingualism the previously reported metalinguistic advantages appear and what types of metalinguistic tasks reveal these developmental differences. Participants were 124 children in second and fifth grades who were enrolled in either a French immersion or a regular English program. All children were from monolingual English-speaking homes and attended local public schools in middle socioeconomic neighborhoods. Measures included morphological awareness, syntactic awareness, and verbal fluency, with all testing in English. These tasks differed in their need for executive control, a cognitive ability that is enhanced in bilingual children. Overall, the metalinguistic advantages reported in earlier research emerged gradually, with advantages for tasks requiring more executive control (grammaticality judgment) appearing later and some tasks improving but not exceeding performance of monolinguals (verbal fluency) even by fifth grade. These findings demonstrate the gradual emergence of changes in metalinguistic concepts associated with bilingualism over a period of about 5 years. Performance on English-language proficiency tasks was maintained by French immersion children throughout in spite of schooling being conducted in French.

A large body of research has identified the impact of bilingualism on cognitive and linguistic processing in children (for a review, see Bialystok, 2001). Evidence for the former comes from studies showing superior performance by bilinguals on nonverbal conflict tasks assessing executive control (e.g., Adi-Japha, Berberich-Artzi,

& Libnawi, 2010; Bialystok, 2010; Bialystok & Martin, 2004; Carlson & Meltzoff, 2008) and for the latter from studies showing advantages in metalinguistic awareness (e.g., Ben-Zeev, 1977; Cummins, 1978). However, these studies typically compare children who are monolingual with those who have had lifetime experience with a second language. What is not known is the point in emerging bilingualism at which these advantageous effects emerge. Moreover, these cognitive and linguistic abilities are interrelated in that different metalinguistic tasks require different levels of executive control, with the largest bilingual advantages observed in tasks based on high levels of control (Bialystok, 1988, 1993). The present study examined the development of metalinguistic awareness for children who are becoming bilingual through an educational immersion program using metalinguistic tasks that vary in their demands for executive control. The purpose is to determine how much second-language instructional experience is required for metalinguistic benefits to emerge and the role of task demands on these outcomes in terms of their requirements for executive control. Therefore, children at different levels of language immersion education may demonstrate enhanced performance on different types of tasks.

French immersion (IMM) programs were created in the public school system in Canada in the 1960s as a means for Anglophone children to learn French in a more complete manner than was the case in standard school French-language instruction that consisted of 20–30 min of instruction per day (Genesee, 1984). Typically, children enter the immersion program in kindergarten (age 5 years) and experience the entire school day in French, even though they know no French at all at the time of entry. English is gradually introduced as a subject, usually around Grade 3, but the medium of instruction for curricular content continues to be French (see Genesee, 1996, for a description of the programs). There is a large literature evaluating the outcomes of these programs on children's academic and linguistic achievements (Genesee & Jared, 2008; Lambert, Genesee, Holobow, & Chartrand, 1993; Turnbull, Hart, & Lapkin, 2003). The essential finding is that academic outcomes are strong and English proficiency advances on the usual schedule, although there may be some initial delays especially around English literacy achievement (Genesee & Jared, 2008). Such delays can be explained by the lack of English literacy instruction in the early elementary years.

Because the school day is conducted entirely in French, a language not spoken at home or in the child's community, these children are in the process of becoming bilingual. Moreover, all the children in the program have approximately equivalent experience with bilingualism: they all speak English at home and live in an English-dominant society. French speakers are a minority in this context, and French schooling for native speakers is separate from immersion schooling for English speakers. French is typically not heard in this Anglophone context. Therefore, the immersion program provides a controlled forum for studying emerging bilingualism in order to investigate the process by which the reported advantages for metalinguistic awareness begin to appear.

Evidence for a bilingual advantage in the development of metalinguistic awareness was the first indication that bilingualism might be beneficial for children's development, in contrast to earlier views warning of confusion and retardation (for discussion, see Hakuta, 1986). Studies showed that bilingual children had more advanced metalinguistic concepts than their monolingual peers for word awareness (Ben-Zeev, 1977; Cummins, 1978; Yelland, Pollard, & Mercuri, 1993), grammatical awareness (Galambos & Goldin-Meadow, 1990; Galambos & Hakuta, 1988; Ricciardelli, 1992), and, to a lesser extent, phonological awareness (Bialystok, Majumder, & Martin, 2003; Campbell & Sais, 1995; Rickard Liow & Poon, 1998; White & Genesee, 1996). Because metalinguistic awareness is crucial to academic achievement and literacy acquisition (Adams, 1990), modifications to the development of metalinguistic awareness are important. Although it is clear that fully bilingual children experience a different course of metalinguistic development than monolingual children do, it is necessary to understand at what point in the progression toward bilingualism these effects are expected to appear and what types of bilingual situations lead to them. It is not known, for example, whether the partial bilingualism that is a result of immersion education modifies children's ability in a similar way as found for fully bilingual children. Moreover, metalinguistic awareness is a collection of abilities rather than a single skill (e.g., McBride-Chang, 1995), making it essential to compare metalinguistic tasks that differ systematically on a known dimension.

The early research on metalinguistic development in monolingual and bilingual children revealed an important divide. In several studies (e.g., Bialystok, 1986, 1988; Cromdal, 1999), monolingual and bilingual children were equally capable of detecting grammatical violations in meaningful sentences (e.g., "Apples growed on trees"), a typical measure of metalinguistic functioning in children (de Villiers & de Villiers, 1972). However, determining that sentences are grammatically correct even if they are semantically anomalous (e.g., "Apples grow on noses") requires the child to ignore the misleading meaning and focus on the grammar. The meaning is salient, especially for young children, yet the instructions require them to attend only to the sentence form. Bilingual children between 5 and 9 years old were more accurate than monolingual children in this anomalous condition. Thus, bilingualism did not benefit linguistic knowledge per se (bilingual children were no better than monolinguals in detecting grammatically incorrect sentences) but rather some process associated with access to that knowledge (distinguishing between sentence form and meaning). Therefore, part of the bilingual success on these tasks was an advantage in selective attention to appropriate information (form) and inhibition of misleading information (meaning). These processes of attention and inhibition are part of the executive function (cf., Miyake et al., 2000). Therefore, children's success on these metalinguistic tasks is determined by their level of executive control, an ability enhanced through bilingualism (Bialystok, 2001). Similar results have been found for adults: in a study of monolingual and bilingual adults performing a metalinguistic-judgment task, conditions that involved a greater degree of executive control produced different event-related potentials (ERPs) for the two groups, indicating less conflict for the bilinguals (Moreno, Bialystok, Wodniecka, & Alain, 2010).

At the core of children's linguistic development, however, is not metalinguistic awareness but language proficiency; children gradually master the structure of language, accumulate vocabulary, and develop efficient access to words and concepts. Such measures of proficiency, however, generally favor monolinguals. Considering only receptive vocabulary, higher scores in English vocabulary by monolinguals than comparable bilinguals have been reported,both for adults (Bialystok & Luk, in press; Portocarrero, Burright, & Donovick, 2007) and children (Oller & Eilers, 2002). In a study including over 1,700 children between the ages of 3 and 10, monolingual children obtained significantly higher scores than bilingual children on a test of English receptive vocabulary at every age studied, even though all children obtained scores in the normal range (Bialystok, Luk, Peets, & Yang, 2010). These differences in vocabulary level may interact with metalinguistic performance. Therefore, metalinguistic performance in bilingual children might be enabled through enhanced control but limited through restricted formal linguistic knowledge.

A standard test of linguistic (and metalinguistic) ability is the verbal-fluency task, a neuropsychological instrument used to evaluate not only language proficiency but also the integrity of brain functioning. The test exists in several standardized forms such as the Delis–Kaplan Executive Function Battery (Delis, Kaplan, & Kramer, 2001) and the Controlled Oral Word Association Test (Strauss, Sherman, & Spreen, 2006). Participants are asked to generate as many words as possible in 60 s that conform to a criterion given by either a semantic category (e.g., articles of clothing, animals) or an initial letter (typically f, a, s), but the demands of each condition are different. The category fluency condition conforms to the structure of representations in semantic memory (animals are normally associated with each other), so the main determinant of performance is richness of the lexical representations for that category and rapid access to those lexical items (Delis et al., 2001). The letter fluency criterion does not conform to the structure of linguistic representation: words are not primarily organized in the mental lexicon according to their initial letter. Therefore, the retrieval of appropriate words requires more effortful monitoring, attending, and selecting. In addition, the letter condition generally carries more exclusionary restrictions than the category condition, such as morphological variants, numbers, and proper names, again increasing the need for monitoring and control. Therefore, performance on the letter-fluency task is normally considered to assess both vocabulary knowledge and executive control (Delis et al., 2001), making it a metalinguistic task.

Parallel to the results described above for the sentence-judgment task, the category-fluency task should be solved equivalently by individuals who have similar vocabulary knowledge (cf., ungrammatical sentences) but the letter-fluency task should be performed better by individuals who have higher levels of executive control (cf., anomalous sentences), once vocabulary has been equated. That is, both the ability to judge grammaticality and the ability to generate lexical items are primarily linguistic tasks, with some demands for executive control. In contrast, the ability to use a more arbitrary sorting criterion (i.e., initial letter) and to resolve conflict in sentences that are semantically incorrect yet grammatically correct both involve higher levels of monitoring, as well as linguistic knowledge. However, given the consistency of differences in vocabulary level, bilingual participants should produce fewer words throughout the verbal-fluency task, as vocabulary size is crucial to both the category and letter conditions. This is exactly the pattern found for adults. Bilingual adults typically generate fewer words in the verbal-fluency task than comparable monolinguals (Bialystok, Craik, & Luk, 2008; Gollan, Montoya, & Werner, 2002; Rosselli et al., 2002), but when vocabulary is controlled, bilinguals produce the same number of words as monolinguals in the category condition but more words than monolinguals in the letter condition (Bialystok et al., 2008; Luo, Luk, & Bialystok, 2010). Although this task has not often been used with children, it can potentially distinguish between the two components of metalinguistic ability: language proficiency and executive control. Therefore, children should generate the same number of words in the category condition, but to the extent that the letter condition requires effortful processing, bilingual children should generate more words than monolinguals in this condition, at least once vocabulary has been controlled.

To summarize, bilingualism has been shown to accelerate children's development of metalinguistic awareness, particularly in tasks that require high levels of executive control. This has been demonstrated in the anomalous condition of the sentence-judgment task (for children) and in the letter condition of the verbal-fluency task (for adults), both of which are more effortful than their counterparts (ungrammatical sentences and category fluency, respectively). However, research demonstrating this bilingual advantage has generally compared fully bilingual participants with monolinguals, leaving unanswered how much bilingual experience is necessary to produce these developmental changes.

There is a small number of studies that have included groups of partially bilingual children to determine whether these effects are evident in the absence of full bilingualism. In a study of monolingual and bilingual kindergarteners performing a battery of nonverbal executive control tasks, Carlson and Meltzoff (2008) included a small group of native English speakers who were in the first year of a Spanish immersion education program. They found that the children in the immersion program did not differ from the monolinguals, showing none of the advantages reported for the bilingual group. This lack of effect may be because the children did not have sufficient bilingual experience for an effect to emerge, the cognitive tasks required more bilingualism than is necessary for metalinguistic tasks, or the children were too young for these training effects to make a difference. By contrast, in studies of metalinguistic awareness, Bialystok (1988; Bialystok & Majumder, 1998) found that children in immersion programs performed at a higher level than monolinguals but at a lower level than bilinguals. Therefore, there is some evidence that the immersion experience is sufficient to foster at least some measure of an advantage in metalinguistic development.

These studies showing increments in metalinguistic development for children in immersion education programs or partially bilingual children examined at different grade levels but with only one grade involved in each study. Therefore, it is not known if these educational effects are a function of particular grades or amounts of time spent in an immersion situation. Given the evidence that children may be gradually developing a metalinguistic advantage as they become bilingual, it is important to establish the point at which this advantage emerges. Therefore, we compared the effects of varying levels of emerging bilingualism by assessing children at different stages in immersion education on measures of metalinguistic awareness.

Children in Grade 2 and Grade 5 who were either in a IMM or an English (ENG) program were given a set of metalinguistic tasks. There were three goals. The first was to determine whether a few years in an immersion program led to the metalinguistic awareness ability typically found in bilingual children. The hypothesis was that children in IMM will obtain higher scores on the metalinguistic tasks than comparable children in ENG. This prediction, therefore, is for a main effect of program. The second was to determine the effects of increased bilingual experience on the development of metalinguistic awareness. If more intense or longer experience is necessary for these effects to emerge, then there would be an interaction between program and grade, such that metalinguistic benefits would be found in Grade 5 but not in Grade 2. Therefore, the hypothesis was that there will be a two-way interaction between program and grade. Finally, metalinguistic tasks vary in their linguistic and cognitive complexity, particularly in the level of executive control required for their completion. Following the study of Carlson and Meltzoff (2008), it might be that the executive control advantages of bilingualism require a more intensive environment than is available in an immersion setting. In this case, the metalinguistic tasks that depend heavily on executive control will not be solved better by children in IMM because their bilingualism has not reached a sufficient level. Therefore, the hypothesis is that there will be a three-way interaction between program, grade, and task, with the executive control demands of the task determining the outcome.

Three tests of metalinguistic awareness were used. The first was the Wug Test of morphological awareness developed by Berko (1958). This task assesses children's sensitivity to morphological structure in English and carries little demand for executive control because there is no misleading information or need for effortful processing. The second was grammatical awareness and was assessed through a sentence-judgment task based on the study by Bialystok (1986). Children were instructed to judge sentences only on the basis of grammatical form ("Tell me if the sentence is said the right way") and the relevant conditions are ungrammatical sentences that are meaningful (require "no" response, low control demands) and grammatical sentences that are anomalous (require "yes" response, high control demands). The third task was a verbal-fluency task that contained both category and letter conditions. Category fluency primarily assesses vocabulary knowledge and letter fluency requires both vocabulary knowledge and executive control. For children, the letter-fluency task also requires literacy knowledge because the instruction is given to produce words that begin with a specific letter.

METHOD

Participants

A total of 124 children were recruited from the public school system in Toronto, Canada. Both the IMM and ENG groups lived in the same neighborhoods and attended the same schools. All children were born and raised in Canada and came from monolingual English-speaking families in which parental education minimally included some postsecondary education. There were 62 second-grade students, with 28 in the ENG group (13 girls) and 34 in the IMM group (22 girls), between the ages of 7.0 and 8.4 years, and 62 fifth-grade students, including 28 in the ENG group (12 girls) and 34 in the IMM group (15 girls), between the ages of 9.9 and 11.5 years. Parents provided informed consent prior to children's participation in accordance with the guidelines established by the university ethics board.

Procedure

There was one testing session, lasting approximately 45–60 min. Children were seated beside the experimenter in a comfortable chair in front of a Dell Inspiro laptop computer. The order of the tasks was randomized across participants, and breaks were offered as needed. Children were given gifts at the end of the experiment to thank them for their participation.

Measures

Language and social background questionnaire. This questionnaire 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 was measured on a 5-point Likert scale, ranging from all English to no English usage in the home. Maternal education was used as a proxy for socioeconomic status and was measured by a 5-point Likert scale. The points on the scale indicated levels of education: 1 = *not completed high school*, 2 = *high school diploma*, 3 = *some postsecondary education*, 4 = *bachelor's degree*, and 5 = *graduate or professional degree*.

Peabody Picture Vocabulary Task—III, Form A (PPVT-III). This test (Dunn & Dunn, 1997) is a measure of English receptive vocabulary and was used as an indicator of English-language proficiency. In a study of metalinguistic awareness, it is critical to establish the basic proficiency levels of children as a potential confounding factor. The reported median Cronbach alpha of PPVT-III is 0.95 (Dunn & Dunn, 1997). For each item, a page of four black and white line drawings was shown and a word corresponding to one of them was read by the experimenter. Participants were asked to choose the picture that best described the word. Items are grouped in sets of 12 and arranged in increasing levels of difficulty, and basal and ceiling sets are established by the number of errors made in a set. Raw scores were transformed to standardized scores using an age-corrected norm table. Standardized scores were used in analyses.

Raven's Coloured Progressive Matrices. This test (Raven, Raven, & Court 1998) measures nonverbal visuospatial reasoning. Metalinguistic awareness incorporates both linguistic and nonlinguistic processing, and therefore this nonverbal measure ensures that the performance of the children is not attributable to differences in intelligence. The child is shown two arrays of colored pictures: one that forms a pattern and a second that depicts potential components of the pattern. The child must point to the picture in the second array that best matches the pictures in the first array. Results were calculated as standard scores that were corrected for age.

Wug Test. This is a classic psycholinguistic test (Berko, 1958) to assess morphological development. The test consists of 27 stimuli (6 of the stimuli require two responses) that depict objects and actions, each with an introduction to their name. All the names are nonsense words that follow the rules of English phonology. Following the introduction of the new word, a prompt is given for the child to add the correct morphological form to the nonsense word. For example, the introduction to the first picture is "*Here is a wug. Here is another wug. How many are there?*" followed by the prompt "*There are two* _____." The test not only assesses children's understanding of word endings but also requires children to manipulate their knowledge of linguistic rules to create those endings for the nonsense words. It is this manipulation of implicit grammatical knowledge that makes the task metalinguistic. There are no developmental data for this task, so raw scores out of a maximum of 33 were used for analysis.

Sentence-judgment task. Sentences from Atchley et al. (2006) were adapted to build 120 sentence frames that were grammatically correct, grammatically incorrect but meaningful, and semantically anomalous but grammatical. 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 (e.g., *to munches* instead of *to munch*). 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*).

These materials were used to create three stimulus lists, 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 auditorally. 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 okay, 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.

Verbal Fluency Test from the Delis–Kaplan Executive Function system. Children were asked to produce as many English words as possible in 60 s (Delis et al., 2001). 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 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

Table 1. *Mean score (standard deviation) for background variables by language group*

	Grade 2		Grade 5	
	English	Immersion	English	Immersion
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
SES (maternal education)	4.2 (0.77)	4.3 (0.77)	3.9 (0.76)	4.2 (0.74)
Home language (1 = *all English*)	1.3 (0.49)	1.3 (0.33)	1.2 (0.37)	1.4 (0.58)
Nonverbal IQ (Raven standard[a])	104.0 (15.5)	104.9 (11.9)	96.1 (12.3)	102.1 (10.6)
Vocabulary (PPVT standard[b])	111.5 (11.1)	116.1 (12.2)	108.4 (12.3)	111.6 (9.8)

Note: SES, socioeconomic status.
[a]Raven's Progressive Matrices (Raven, Raven, & Court, 1998).
[b]Peabody Picture Vocabulary Test (Dunn & Dunn, 1997).

specified letter or not in the designated categories) and repeated words from the total number of responses.

RESULTS

All analyses were based on mixed-model analyses of variance (ANOVAs) with grade level (2, 5) and program (ENG, IMM) as between-subjects variables and all other variables as within-subjects variables. Background data are reported in Table 1 and were analyzed by a two-way ANOVA for language group and grade for each measure. There were no significant differences on either factor for socioeconomic status, home language use, or vocabulary scores. For the Raven test, there was a main effect of grade, F (1, 120) = 5.08, $d = 0.40$, $p < .05$, with higher performance by Grade 2 children. Because of these differences in nonverbal IQ, subsequent analyses that included grade were performed as analyses of covariance (ANCOVAs) with Ravens score as a covariate.

Mean scores and standard deviations for the Wug Test are presented in Table 2. A two-way ANCOVA with Ravens as the covariate revealed a main effect of grade, F (1, 119) = 28.41, $d = 0.83$, $p < .0001$, with older children scoring higher than younger children, and main effect of language group, F (1, 119) = 5.56, $d = 0.45$, $p < .02$, with IMM children scoring higher than ENG children, and no interaction, F (1, 119) = 2.22, *ns*.

Results from the sentence-judgment tasks are presented in Table 2. The reaction times (RTs) were long and averaged about 3 s for all sentence types. There was a main effect of grade in which the Grade 5 children (mean RT = 2827 ms, $SD = 277$) responded more quickly than did the Grade 2 children (mean RT = 3194 ms, $SD = 311$), but there were no other significant effects or interactions. RTs this long cannot be interpreted in terms of processing differences, so no further analyses of the RT data were pursued. The accuracy data were first analyzed in a three-way ANCOVA for language group, grade, and sentence type, covarying out effects of IQ. There was a main effect of grade, F (1, 119) = 5.49, $d = 0.43$, $p <$ .03, a main effect of language group, F (1, 119) = 3.99, $d = 0.37$, $p < .05$, and an

Table 2. *Mean (standard deviation) score for the Wug Test, correct responses for sentences in the sentence judgment task, and number of words generated in the verbal fluency task*

	Grade 2		Grade 5	
	English	Immersion	English	Immersion
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Wug Test	20.5 (6.6)	21.4 (6.2)	23.3 (5.9)	27.7 (2.0)
Sentence judgment task				
Grammatical	33.5 (7.4)	32.9 (8.7)	32.9 (12.3)	36.9 (5.4)
Ungrammatical	29.7 (10.2)	31.2 (8.8)	31.3 (10.5)	34.9 (4.9)
Anomalous	30.1 (11.2)	28.9 (11.7)	28.4 (14.5)	36.9 (6.6)
Average	31.1	31.0	30.9	36.3
Verbal fluency task				
Category fluency	10.1 (3.0)	11.4 (2.8)	15.1 (4.6)	15.5 (4.7)
Letter fluency	8.1 (2.4)	6.6 (2.4)	10.7 (3.8)	11.4 (3.5)
Average	9.1	9.0	12.9	13.5

Note: The scores are mean (standard deviation) for the Wug Test (raw scores) out of a maximum of 33, mean (standard deviation) correct responses out of a maximum score of 40 for each of the three types of sentences in the sentence judgment task (grammatical, ungrammatical, and anomalous), and mean (standard deviation) number of words generated in the two conditions of the verbal fluency task.

interaction of grade and group, F (1, 119) = 4.80, $p < .03$. There was also a main effect of sentence type, F (2, 240) = 5.15, $p < .01$, in which correct sentences were more accurate overall than the other two types.

To understand these results in more detail, separate analyses were conducted for each sentence type. In previous research bilingual children outperformed monolinguals only when making judgments about anomalous sentences. Therefore, a clearer picture should emerge when these conditions are examined individually. For grammatically correct sentences, there were no effects of grade, F (1, 119) = 1.50, *ns*, language group F (1, 119) = 0.97, *ns*, or their interaction, F (1, 119) = 1.87, *ns*. For ungrammatical sentences there was an effect of grade, F (1, 119) = 4.52, $d = 0.32$, $p < .04$, with the Grade 5 group outperforming the Grade 2 group, but no effect of language group, F (1, 119) = 3.51, *ns*, or the interaction of grade and language group ($F < 1$). For the anomalous sentences, there was an effect of grade, F (1, 119) = 3.88, $d = 0.32$, $p < .05$, no effect of language group, F (1, 119) = 2.26, *ns*, but a significant interaction of grade and language group, F (1, 119) = 5.0, $p < .03$. Simple effects analyses showed that the two groups performed equivalently on the anomalous items in Grade 2 ($F < 1$), but the IMM children obtained higher scores than the ENG children in Grade 5, F (1, 60) = 9.33, $d = 0.75$, $p < .001$.

To fully analyze the results of the verbal-fluency task, a three-way ANCOVA for language group, grade, and task was conducted, covarying out effects of IQ. There was a main effect of grade, F (1, 119) = 55.0, $d = 1.37$, $p < .0001$, with older children producing more items than younger children. There was also a two-way interaction of task and language group, F (1, 119) = 3.69, $p < .05$, and a three-way interaction of grade, task, and language group, F (1, 119) = 6.10, $p < .02$. Therefore, to understand these results, separate two-way ANOVAs were conducted for each task separately because task was involved in both interactions.

The two-way ANOVA for category fluency indicated a main effect of grade, F (1, 119) = 41.42, d = 1.17, $p < .0001$, but no effect of language group, F (1, 119) = 1.27, *ns*, or interaction ($F < 1$). For letter fluency, there was a main effect of grade, F (1, 119) = 43.24, d = 1.22, $p < .0001$, no main effect of group ($F < 1$), but a significant interaction of language group by grade, F (1, 119) = 3.89, $p < .05$. Examining the simple effects, in Grade 2, there was a main effect of group, F (1, 63) = 6.02, d = 0.63, $p < .01$, in which children in the ENG group generated more words than did children in IMM; for Grade 5, there was no difference between groups in the number of words generated ($F < 1$).

DISCUSSION

This study examined emerging bilingualism in two grades of a IMM program to determine how many years in the program would be sufficient to observe metalinguistic advantages previously reported for full bilinguals. The results showed that all the IMM children outperformed their ENG counterparts in the Wug Test, and the Grade 5 French Immersion children were more accurate in judging the crucial anomalous sentences in the sentence-judgment task. There were no IMM advantages in verbal fluency, and the younger IMM children actually produced *fewer* words than ENG children on the letter task.

The Wug Task revealed a metalinguistic advantage after only 2 years of immersion education. Although it appeared that the gap between children in the two programs was greater in Grade 5 than in Grade 2, the interaction with grade was not significant. The task is metalinguistic in that it requires children to use principles of English morphology to generate new forms, but there are low demands on executive control because there is no salient distracting information (as in the semantically anomalous sentences of the judgment task) or effortful monitoring (as in the fluency task). Therefore, it is similar to tasks used in the original literature that demonstrated bilingual advantages in metalinguistic knowledge, such as the correction of grammatical errors (e.g., Galambos & Goldin-Meadow, 1990). In those tasks, children needed to focus on linguistic information to complete the task, but since there were few distracting cues, there was little demand for high levels of executive control.

The results of the sentence-judgment task also replicated previous research with fully bilingual children. There were no differences among children in the two programs in their ability to judge the correct or grammatically incorrect sentences, supporting the interpretation that they had equivalent linguistic knowledge of grammar. Also consistent with research in the development of metalinguistic awareness, older children were more accurate in determining there was a grammatical error than younger children. For the anomalous sentences, however, the correct response required overriding the misleading but salient anomaly in meaning in order to report that the grammatical structure was intact. As in previous research with bilingual children, the IMM children were more successful than monolingual children, but this effect was not apparent until Grade 5. These children were both older and had spent more time in an immersion program than those in Grade 2, and either (or both) of those factors may be responsible for the results.

Verbal fluency was the most difficult task and produced the most complex results. Children's ability to generate words to the two types of cues interacted with both grade and language experience. Category fluency is normally interpreted as a reflection of the structure of semantic memory and lexical richness. Therefore, individuals with a more organized semantic system and a larger vocabulary generate more words on this task. In previous research with adults, bilinguals produced fewer words than monolinguals if there was a vocabulary difference in the language of testing but the same number of words if vocabulary was matched (Bialystok et al., 2008). The present results are consistent with that pattern; receptive vocabulary scores were equivalent for all children, and more words were produced by older children with no effect of language program. Letter fluency, in contrast, requires monitoring to scan the lexical network with an unusual criterion and involves planning and monitoring to avoid errors and repetitions. Therefore, even without the usual exclusionary criteria, this task carries additional demands for executive control. Thus, once vocabulary is matched, bilingual adults produce more words than monolinguals on the letter-fluency task (Bialysto et al., 2008). In the present study, however, the IMM children in Grade 2 produced *fewer* words than did their monolingual peers. Thus, not only did the young IMM children not show a bilingual advantage but they actually indicated a cost in performance. The letter-fluency task is related to literacy: The cue given to generate the responses is orthographic. In IMM, literacy instruction is offered in French; although the middle-class Anglophone environment undoubtedly provides constant exposure to English text and literacy, it is not formally taught in school until about Grade 3. It is possible, therefore, that the poorer performance of the young IMM children is a reflection of limited literacy experience in English. By Grade 5, all the children produced the same number of words. These IMM children are more experienced in English literacy so the task is more manageable than it is for their younger counterparts, yet there is still no evidence for a bilingual advantage in this task. The executive control required for the letter-fluency task is subtle: search through a lexical space while maintaining a goal in working memory, switch between semantic categories, and monitor errors. Although fully bilingual adults are able to perform this better than monolinguals (Bialystok et al., 2008), it is possible that children are not sufficiently skilled or that 5 years in an immersion program is not adequate experience.

The results of the present study demonstrate emerging advantages to metalinguistic awareness after a brief experience in an immersion education program. After only 2 years in an educational program in which the language of schooling is different from the language of the home, and for which proficiency in that language is still relatively low, children displayed some of the metalinguistic advantages previously associated with fully bilingual children. After 5 years in the program, the metalinguistic advantages were more robust, but these children still did not show the profile of fully bilingual children. It is important that the metalinguistic task that was most closely related to literacy, namely, the letter condition of the verbal-fluency task, was more problematic for the young IMM children than for their monolingual peers. Learning to read in French instead of in English may have delayed the reorganization of the lexicon along orthographic criteria. Overall, the results depended on the demands of the individual metalinguistic tasks, with earlier evidence for an effect of the IMM program on tasks that carried the lowest demands for executive control. As in the study by Carlson and Meltzoff (2008), executive control may require more substantial bilingual experience.

The main objective of the study was to determine whether the advantages previously reported for fully bilingual children could be detected in children who were learning another language in an educational immersion program. The results on that point are positive: one task (Wug) was performed better by the immersion children at both levels, another task (sentence judgment) was performed better by immersion children with much greater advantages for the older and more experienced children, and the third task (verbal fluency) began as a problem for the children in the immersion program but the older children regained ground and performed equivalently to monolingual children. Thus, all three predicted statistical patterns were found, indicating the importance of all three factors: grade, program, and task demands. The results show a continuum in which more experience in using two languages is associated with greater benefit and greater approximation to the pattern reported for bilingualism. It is an evolving system in which experience gradually and continually modifies ability. Language-education programs are not only teaching children language but they are also making them bilingual. The road to bilingualism is incremental, and so are the accrued advantages.

ACKNOWLEDGMENTS

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 E.B.).

REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Adi-Japha, E., Berberich-Artzi, J., & Libnawi, A. (2010). Cognitive flexibility in drawings of bilingual children. *Child Development, 81*, 1356–1366.

Atchley, R. A., Rice, M. L., Betz, S. K., Kwasny, K. M., Sereno, J. A., & Jongman, A. (2006). A comparison of semantic and syntactic event-related potentials generated by children and adults. *Brain and Language, 99*, 236–246.

Ben-Zeev, S. (1977). The influence of bilingualism on cognitive strategy and cognitive development. *Child Development, 48*, 1009–1018.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150–177.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development, 57*, 498–510.

Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology, 24*, 560–567.

Bialystok, E. (1993). Metalinguistic awareness: The development of children's representations of language. In C. Pratt & A. Garton (Eds.), *Systems of representation in children: Development and use* (pp. 211–233). London: Wiley.

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

Bialystok, E. (2010). Global-local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology 46*, 93–105.

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics, 21*, 522–538.

Bialystok, E., & Luk, G. (in press). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition*.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*, 525–531.

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics, 19*, 69–85.

Bialystok, E., Majumder, S., & Martin, M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics, 24*, 27–44.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science, 7*, 325–339.

Campbell, R., & Sais, E. (1995). Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology, 13*, 61–68.

Carlson, S. M., & Meltzoff, A. N., (2008) Bilingual experience and executive functioning in young children. *Developmental Science, 11*, 282–298.

Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish–English bilinguals. *Applied Psycholinguistics, 20*, 1–20.

Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. *Journal of Cross-Cultural Psychology, 9*, 131–149.

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

de Villiers, J. G., & de Villiers, P. A. (1972). Early judgments of semantic and syntactic acceptability by children. *Journal of Psycholinguistic Research, 1*, 299–310.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Galambos, S. J. & Goldin-Meadow, S. (1990) The effects of learning two languages on levels of metalinguistic awareness. *Cognition, 34*, 1–56.

Galambos, S. J., & Hakuta, K. (1988) Subject-specific and task-specific characteristics of metalinguistic awareness in bilingual children. *Applied Psycholinguistics, 9*, 141–162.

Genesee, F. (1996). Second language immersion programs. In H. Goebl, P. H. Nelde, Z. Stary, & W. Wolck (Eds.), *Contact linguistics: An international handbook of contemporary research* (Vol. 2, pp. 493–502). Berlin: Walter de Gruyter.

Genesee, F. (1984). Beyond bilingualism: Social psychological studies of French immersion programs in Canada. *Canadian Journal of Behavioural Science, 16*, 338–352.

Genesee, F., & Jared, D. (2008). Literacy development in early French immersion programs. *Canadian Psychology/Psychologie canadienne, 49*, 140–147.

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish–English bilinguals. *Neuropsychology, 16*, 562–576.

Hakuta, K. (1986). *Mirror of language: The debate on bilingualism.* New York: Basic Books.

Lambert, W. E., Genesee, F., Holobow, N., & Chartrand, L. (1993). Bilingual education for majority English-speaking children. *European Journal of Psychology of Education, 8*, 3–22.

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition, 114*, 29–41.

McBride-Chang, C. (1995). What is phonological awareness? *Journal of Educational Psychology, 87*, 179–192.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Moreno, S., Bialystok, E., Wodniecka, Z., & Alain, C. (2010). Conflict resolution in sentence processing by bilinguals. *Journal of Neurolinguistics, 23*, 564–579.

Oller, D. K., & Eilers, R. (2002). *Language and literacy in bilingual children*. Clevedon: Multilingual Matters.

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology, 22*, 415–422.

Raven, J., Raven, J. C., & Court, J. H. (1998, updated 2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.

Ricciardelli, L. A. (1992). Bilingualism and cognitive development in relation to threshold theory. *Journal of Psycholinguistic Research, 21*, 301–316.

Rickard Liow, S. J., & Poon, K. K. L. (1998). Phonological awareness in multilingual Chinese children. *Applied Psycholinguistics, 19*, 339–362.

Rosselli, M., Ardila, A., Salvatierra, J., Marquez, M., Matos, L., & Weekes, V. A. (2002). A cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience, 112*, 759–776.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary* (3rd ed.). New York: Oxford University Press.

Turnbull, M., Hart, D., & Lapkin, S. (2003). Grade 6 French immersion students' performance on large-scale reading, writing, and mathematics tests: Building explanations. *Alberta Journal of Educational Research, 49*, 6–23.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12*, 233–265.

Yelland, G. W., Pollard, J., & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics, 14*, 423–444.