Review of Educational Research
http://rer.aera.net

A Meta-Analysis on the Effectiveness of Bilingual Programs in Europe

Gabrijela Reljic, Dieter Ferring and Romain Martin

REVIEW OF EDUCATIONAL RESEARCH published online 5 September 2014
DOI: 10.3102/0034654314548514

The online version of this article can be found at:
http://rer.sagepub.com/content/early/2014/09/05/0034654314548514

Published on behalf of

American Educational Research Association

and

http://www.sagepublications.com

Additional services and information for *Review of Educational Research* can be found at:

Email Alerts: http://rer.aera.net/alerts

Subscriptions: http://rer.aera.net/subscriptions

Reprints: http://www.aera.net/reprints

Permissions: http://www.aera.net/permissions

>> OnlineFirst Version of Record - Sep 5, 2014

What is This?

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

Review of Educational Research
Month 201X, Vol. XX, No. X, pp. 1–37
DOI: 10.3102/0034654314548514
© 2014 AERA. http://rer.aera.net

A Meta-Analysis on the Effectiveness of Bilingual Programs in Europe

Gabrijela Reljić
Dieter Ferring
Romain Martin
University of Luxembourg

The effectiveness of bilingual programs for promoting academic achievement of language minority children in the United States has been examined in six meta-analyses. The present meta-analytic study investigates this topic for the first time in the European context. Thorough literature searches uncovered 101 European studies, with only 7 meeting the inclusion criteria. Two studies were excluded from further analyses. Results from the random-effects model of the five remaining studies indicate a small positive effect (g = 0.23; 95% confidence interval [0.10, 0.36]) for bilingual over submersion programs on reading of language minority children. Thus, this meta-analysis supports bilingual education—that is, including the home language of language minority children—in school instruction. However, the generalizability of the results is limited by the small number of studies on this topic. More published studies on bilingual education in Europe are needed as well as closer attention to the size of the effects.

Keywords: meta-analysis, bilingual education, language minority, academic achievement, Europe

In the early 1980s, Cummins (1981) highlighted the fear of parents, educators, and school administrators who perceived bilingualism as a disease that was harming children's learning. Eradication of students' bilingualism seemed to be the best way for them to become "good Americans," so "children were often punished for speaking their L1 in school and were made to feel ashamed of their own language and cultural background" (Cummins, 1981, p. 17). Although there is considerable evidence today regarding the benefits of bilingualism, in a recent article, Bialystok, Craik, and Luk (2012) pointed to the same fear Cummins (1981) was addressing. In their concluding remarks they noted that some educators and clinical practitioners do advise parents to help children avoid linguistic confusion, "but such views are based on fear and anecdote" (p. 248).

Adesope, Lavin, Thompson, and Ungerleider (2010) conducted a systematic review and meta-analysis of the cognitive correlates of bilingualism. Based on the data from 63 studies involving 6,022 participants, the authors concluded that there

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

are bilingual benefits regarding attentional control, working memory, metalinguistic awareness, and abstract and symbolic representational skills. Bilingual speakers constantly control their attention to select the required language, a task in which they outperform their monolingual counterparts (e.g., Bialystok, 1999; Bialystok & Martin, 2004). Bialystok et al. (2012) drew our attention to the fact that bilingualism is associated with a delay in the symptoms of dementia, which can begin up to 4 years later than in their comparable monolingual counterparts (Bialystok, Craik, & Freedman, 2007). Furthermore, bilingual speakers seemed to have greater working memory capacity than their monolingual peers (Engle, 2002; Kane, Bleckley, Conway, & Engle, 2001, as cited in Adesope et al., 2010, p. 209).

There is substantial evidence demonstrating that bilinguals outperform their monolingual peers in tasks of metalinguistic awareness, which is "the ability to treat language as an object of thought" (García, 2009, p. 95). Words are symbols for concepts, and bilingual children seem to be faster in understanding this than monolinguals (e.g., Bialystok, 1986, 2001; Bialystok & Martin, 2004; Lasagabaster, 2000; Ricciardelli, 1992). In sum, the results of Adesope et al.'s (2010) meta-analysis showed that bilinguals seemed to outperform their monolingual peers on measures of metalinguistic awareness and metacognitive awareness (knowledge about one's own thinking; $g = 0.33$) and on measures of abstract and symbolic representation, attentional control, and problem solving ($g = 0.52$). In addition, García (2009) underlined the benefits of divergent thinking, communicative sensitivity, and the ability to learn multiple languages. The notion of linguistic confusion in children is indeed based on fear and anecdote (Bialystok et al., 2012; Cummins, 1981).

Why Bilingual Education?

How best to educate language minority children is becoming an increasingly important topic in today's society. Language minority children are those who come from a "linguistic background which is different (and less represented) to the language spoken and used in the larger community/society" (Murphy, 2010, p. 162). The education of language minority children is not relevant only in the United States, with its long research tradition on this topic. The relaxing of borders in Europe, particularly since the inception of the European Union, has increased its significance for many European countries as well. This importance is also related to the economic competitiveness of the country (The Organisation for Economic Co-operation and Development, 2010; Stanat & Christensen, 2006). It is evident that academic achievement and later chances in the labor market largely depend on the difficulties in understanding the language of school instruction (i.e., school language) in the country in which language minority children live (Stanat & Christensen, 2006). Difficulties with language may also increase the dropout rate, which will then increase the cost of education for language minority children (Grin, 2003).

Aims of the Present Study

The aim of this article is to present a meta-analysis of the effectiveness of bilingual programs in promoting academic achievement of language minority children in Europe. Bilingual education programs

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

> use the language as *a medium of instruction*; that is, bilingual education programs teach content through an additional language other than the children's home language" and provide "*meaningful and equitable education*, as well as an education that builds tolerance towards other linguistic and cultural groups. (García, 2009, p. 6)

The language of instruction in schools is often different from the *home language* for many language minority children. The term *home language* refers to the language of the home, which may be "a mixed language or a set of languages or dialects" (Davies, 2003, p. 18), and it is usually the language that remains once the public language is no longer used by the child (Davies, 2003). In the previous meta-analyses conducted in the United States (e.g., Rolstad, Mahoney, & Glass, 2005; Slavin & Cheung, 2005), the term used to describe home language was *native language*. In the present contribution, we will use synonyms used by different authors (e.g., L1 or first language). We opted for the term *home language* as this term has been increasingly used (García, 2009; Rolstad et al., 2005).

Previous research indicated that language minority children whose home language differs from the school language may be at a disadvantage in reading and mathematics (e.g., Abedi, 2003). International data showed that 15-year-old language minority students who do not speak the school language at home are, on average, 1 year behind their native peers (Christensen & Stanat, 2007). However, the findings from the United States on the academic achievement of former English language learners (ELL), that is, students who, on reaching a benchmark in standardized tests in English, no longer receive language support services (Ardasheva, Tretter, & Kinny, 2012), are very promising. Three studies (Ardasheva et al., 2012; Kim & Herman, 2009; New York City Department of Education, 2009) showed that former ELL students significantly outperformed their native speaker counterparts and current ELL participants in standardized reading and mathematics proficiency tests. For example, in a study by the New York City Department of Education (2009), it was found that over a 5-year period (2003–2008), former ELL students performed significantly higher than their native peers and current ELL students in reading and mathematics (Grades 4 and 8). In addition, it was pointed out that the former ELL group had higher education rates (70.9%) and lower dropout rates (9.7%) than even English-proficient students who were never ELLs (at 63.5% and 13%, respectively; New York City Department of Education, 2009, p. 4).

Psycholinguistic Base of Bilingual Education

Although not the focus of this article, it is important to understand the benefits of bilingual education, which will be facilitated by reviewing some psycholinguistic constructs developed by Cummins (1981, 2000). Cummins (2000) introduced the construct of the interdependence between the two languages of a bilingual speaker. He explained the interdependence hypothesis as follows: "To the extent that instruction in Lx is effective in promoting proficiency in Lx, transfer of this proficiency to Ly will occur provided there is adequate exposure to Ly" (p. 38). If the conditions are right, Cummins (2000) argued, "Transfer across languages is two-way," so that the students who have developed reading in their first language will tend to make stronger progress in acquiring reading in their second language (p. 26).

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

It is important to underline that Cummins (2000) did not posit that the first language needs to be fully developed before the second language is introduced: "Rather, the first language must not be abandoned before it is fully developed, whether the second language is introduced simultaneously or successively, early or late, in that process" (p. 25). Indeed, transfer will not happen automatically, which is why the schools are responsible for guaranteeing the child's adequate exposure to both home and school language (Baetens-Beardsmore, 2009; Cummins, 2000; García, 2009). Empirical evidence for the interdependence hypothesis has been found in several studies (Abu-Rabia, 2001; Chireac, Serrat, & Huguet, 2011; Durgunoğlu, 1998; Hauptman, Mansur, & Tal, 2008; Huguet, 2008; Oller & Vila, 2011).

Equally important are the concepts of basic interpersonal communication skills and cognitive academic language proficiency (CALP; Cummins, 1981). Basic interpersonal communication skill is contextualized language, supported by paralinguistic cues, used in interpersonal situations (e.g., on the playground). CALP is decontextualized language, devoid of pictures and other cues, and it is related to literacy skills. It is conventional wisdom that children learn the school language very fast, but it generally takes a minimum of about 5 to 7 years for immigrant students to catch up to native speakers in CALP (Cummins, 1981, 2000; Hakuta, Butler, & Witt, 2000).

In their report on language minority children and youth, August and Shanahan (2006) reviewed studies that provided evidence supporting a significant relationship between first-language and second-language literacy development. For example, word and pseudoword reading, cognate vocabulary, reading comprehension, reading strategies, spelling, and writing in the second language are significantly related to similar constructs in the first language. August and Shanahan concluded, "Well-developed literacy skills in the first language can facilitate second-language literacy development" (p. 14). Thus, the base for bilingual education programs has been set.

Bilingual Programs

Bilingual programs offering instruction in both the students' home and school language seem to be relatively uncommon in Europe (Stanat & Christensen, 2006). These programs involve teaching content through an additional language other than the children's home language (García, 2009), which is why they should facilitate conceptual language transfer between the home and school language (Cummins, 2000) and promote academic achievement of language minority children (Baetens-Beardsmore, 2009). In addition, bilingual education is beneficial to language minority children because it provides them with the access to school content in a comprehensible way (i.e., their home language), which enhances learning (MacSwan & Rolstad, 2010).

In fact, the countries with the most successful language support programs (e.g., Australia, Canada, and Sweden) have the smallest academic achievement gaps between their language minority and native students, and their second-generation and first-generation language minority students (Stanat & Christensen, 2006). Relatedly, *submersion programs*, in which children are instructed only in school language, seem to be associated with lower levels of second-language proficiency, scholastic underachievement, and psychosocial disorders (Hakuta & Mostafapour, 1996).

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

There is a high dropout rate for language minority students in submersion programs as they need to learn both the school subjects in a weaker language and the language itself at the same time (Bialystok, 2001). Submersion programs impair academic performance and development of the new language, as learners cannot use their home language to make sense of academic material (Cummins, 2000; García, 2009).

On the other hand, bilingual programs that instruct the school material in both the home and school language seem to promote language minority children's academic achievement at no cost to the development of the school language (Kimbrough Oller & Eilers, 2002; Lindholm-Leary, 2001; Lindholm-Leary & Borsato, 2007; Ramirez, 1992; Thomas & Collier, 1997, 2002). This claim is supported by multiple meta-analyses from the United States (Krashen & McField, 2005; Rolstad et al., 2005; Rolstad, Mahoney, & Glass, 2008; Slavin & Cheung, 2005; Willig, 1985; see also below).

However, the research on bilingual education appears to be very controversial because it is tied to the issues of nationalism, immigration, and the politics of multilingualism (Petrovic, 1997). Rossell and Baker (1996, as cited in Rolstad et al., 2005, p. 573) noted that "this field is so ideologically charged that no one is immune from ideological bias or preconceived notion," and "as a result, those attempting to make policy recommendations from the research must carefully read each study and draw their own conclusions;" however, "this does not guarantee that such conclusions will be free from bias, only that they will be free from someone else's bias" (pp. 25–26). Nevertheless, Rolstad et al. (2005) pointed out that description of the evidence is possible and that "a properly conducted meta-analysis will help provide a factual description of program effects across a large range of available studies" (p. 574).

One of the problems that meta-analysts often encounter in the field of bilingual education is the explanation and labeling of the types of bilingual programs. This concern can produce methodological flaws (Takakuwa, 2005) that can lead to ensuing errors in subsequent meta-analyses (Rolstad et al., 2008). Thus, we used the classification system proposed by García (2009) to label bilingual programs (note that some of the programs are specific to Europe only) in a theoretically meaningful way. According to García, monoglossic types start from the monolingual conception, whereas heteroglossic start with the bilingual conception where children are recognized as coming from homes and communities that have some familiarity with bilingualism. Table 1 describes the types of educational programs originating from these two conceptions.

Previous Meta-Analyses in the United States

Meta-analysis is the quantitative procedure used to statistically combine the results from studies on the same topic to reach some general conclusions regarding the effects or outcomes of a given treatment, project, or program (Cooper, 2009; Glass, McGaw, & Smith, 1981). In the United States, several meta-analyses have been conducted to investigate the effects of bilingual programs on academic achievement of language minority children. Table 2 provides a brief summary of the meta-analyses on the effectiveness of bilingual programs in the United States by presenting the number of studies, year of implementation, effect sizes, and the main conclusions.

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 1

Types of bilingual programs

Types of bilingual programs	Description
Monoglossic program types	
Transitional bilingual education types	The child's home language is used in instruction only until the child is fluent in the school language.
Maintenance bilingual education programs	These are mostly used for children who speak their home language at home when they enter school and want to continue to do so while developing proficiency in the school language.
Prestigious bilingual education programs	These teach majority children using two teachers, with one teaching in a school language and the other in the children's home language.
Immersion bilingual education	Usually for a year, language minority children are taught exclusively in the school language.
a. Submersion or sink-or-swim programs	These are otherwise called "English only" programs in the United States. Children are taught only in the school language without taking into account the child's home language.
b. Early immersion bilingual programs	These are mostly present in preschools where a bilingual teacher instructs children in the school language in the first year. In the second year children develop reading in their home language, and in the third year they may spend one half of the time being educated in the school language and the other half in their home language.
c. Late immersion bilingual programs	These usually start after the child has already been instructed in his or her home language.
d. Partial immersion bilingual programs	These immerse children in the school language only for part of the day and not the full day as the total immersion programs do.
Heteroglossic program types	
Immersion revitalization programs	These start at the preschool level and usually relate to indigenous populations within the community.
Developmental bilingual education programs	These are used for nondominant-language groups who develop their minoritized languages and academic proficiency in both the home and school languages.
Polydirectional or two-way (dual language) programs	These teach children in more than one language (e.g., the European schools for civil servants).
CLIL and CLIL-type programs	These focus on content-based language learning, supported by the European Commission (Brinton, Snow, & Wesche, 1989; Snow, Met & Genesee, 1989).
Multiple multilingual education	These involve at least three languages and multilingual groups (e.g., Luxembourg, where students are taught in Luxembourgish, German, and French).

Note. CLIL = content and language integrated learning.
Source. Adapted from García (2009, p. 132), reprinted with permission.

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 2

Meta-analyses on the effectiveness of bilingual programs in the United States

Meta-analysis	No. of studies	Years included	Effect size	Conclusions
Willig (1985)	23	1971–1980	0.33	"Participation in bilingual education programs consistently produced small to moderate differences favoring bilingual education for tests of reading, language skills, mathematics, and total achievement when the tests were in English, and for reading, language, mathematics, writing, social studies, listening comprehension, and attitudes toward school or self when tests were in other languages" (p. 269).
Greene (1998)	11	1972–1991	0.18	Using even more restrictive inclusion criteria based on methodological rigor, Greene (1998) found positive effects for English standardized test scores.
Slavin and Cheung (2003)	17	1971–	0.33	"Although the number of high-quality studies is small, existing evidence favors bilingual approaches, especially paired bilingual strategies that teach reading in the native language and English at different times each day" (p. 247).
Rolstad, Mahoney, and Glass (2005)	17	1985–	0.23	"Bilingual education is consistently superior to all-English approaches, and developmental bilingual education programs are superior to transitional bilingual programs [. . .] bilingual education programs are effective in promoting academic achievement and sound educational policy should permit and even encourage the development and implementation of bilingual education programs" (p. 572). The authors also found a large positive effect ($d = 0.86$) with outcome measures in the home language.
Rolstad, Mahoney, and Glass (2008)	16	1985–	0.23	Due to erroneous information on one of two programs for English language learners included in their meta-analysis in 2005, the authors reanalyzed the studies without the Gersten (1985) studies. The correction resulted in a change in mean effect size from .08 to .19 for all outcome measures, from -.06 to .14 for English reading, from .08 to .17 for English math, and from −.01 to .10 for all Transitional Bilingual Education studies. The authors argue that the most informative result is the effect size for bilingual education ($d = 0.23$), which strengthens the conclusions reached in a previous meta-analysis.
Krashen and McField (2005)			0.26	The authors averaged the effect sizes from five previous meta-analyses (including McField, 2002). Despite slightly different criteria for including studies and different dates of publication, the average effects are "remarkably similar, with students in bilingual education showing a small but consistently positive impact versus those in all-English classrooms" (p. 8). All of these studies were conducted in the United States only and lasted for at least one year. They also note that, in some studies, comparison students were fluent English speakers.

Note. According to Cohen (1988), a "small" effect is $d = 0.20$, a "medium" effect is $d = 0.50$, and a "large" effect is $d = 0.80$.

Krashen and McField (2005) averaged the effects size from six previous meta-analyses (Greene, 1998; McField, 2002 [as cited in Krashen & McField, 2005], Rolstad et al., 2005; Slavin & Cheung, 2003, 2005; Willig, 1985) to obtain the mean effect size of $d = 0.26$. All of the meta-analyses showed a significantly positive effect size (i.e., Cohen's d) in the range of $.18 < d < .28$, with .26 as a mean effect. However, one problem with averaging the effects from these meta-analyses is that many of them share the same studies, which violates the independence assumption (Borenstein, Hedges, Higgins, & Rothstein, 2009). Nevertheless, Krashen and McField (2005, p. 10) concluded that instruction in a child's home language is "a part of the solution, not part of the problem." Moreover, they reported that their results indicate strong support for a bilingual approach to education rather than submersion, thus casting "strong doubt on claims that all-English approaches are superior and should be mandated by law" (p. 10).

Unfortunately, Krashen and McField (2005) did not report the number of the studies for the McField (2002) study nor the confidence intervals (CIs) for the effect size, which is very important information when interpreting the precision of the effect sizes. Moreover, the primary meta-analyses did not report CIs (e.g., Rolstad et al., 2005, 2008). In our opinion, reporting these indicators should be standard practice when this type of research is conducted.

Importance of the Present Meta-Analysis

To the best of our knowledge, no previous meta-analysis has investigated the effectiveness of bilingual programs on academic achievement in Europe, perhaps because they are relatively uncommon (Stanat & Christensen, 2006) or difficult to identify (e.g., written in non-English languages). The present meta-analysis, modeled after the meta-analysis by Rolstad et al. (2005), arises from the need to scrutinize European studies and to contribute to a better understanding of bilingual education in Europe. Based on the previous meta-analyses conducted in the United States, we hypothesized that there will be a significant positive effect for bilingual over submersion programs on the academic achievement of language minority children. Furthermore, the present meta-analysis also increases the awareness of the lack of European studies on the subject and addresses the criticism on the overrepresentation of American studies in major psychological journals (Arnett, 2008).

Finally, more methodological similarities than differences can be found between the current meta-analysis and previous meta-analyses. This issue will be discussed in the article with regard to both the exclusion criteria and the effect sizes. Moreover, we will highlight the practical significance of the effect sizes, which provide extremely important information for practitioners and for the cost–benefit analysis, especially as this element was missing in previous meta-analyses.

Method

Selecting the Studies

In this meta-analysis, there were no time or language limitations placed on publication of the articles. We conducted our search in 2011. Our language scope included English, French, German, Serbian, and Russian. For other languages we

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

needed translators. To avoid biased retrieval by searching only the major journals (Rosenthal, 1995), we used several techniques recommended by Cooper (2009) for retrieving as many studies as possible. Therefore, we employed (a) direct-to-researcher channels (personal contact, mass solicitation, traditional invisible college hubs, electronic invisible college, listservs, bulletin boards, discussion groups), (b) quality control search techniques (professional conference paper presentations and peer-reviewed journals), and (c) secondary searching techniques (research report reference lists, research bibliographies, prospective registers, Internet, reference databases, citation indexes).

For the exhaustive literature search, we used all the above-mentioned techniques and a log for keeping track of a literature search (see Table 3; Cooper, 2009) was created. In this log we recorded the information on the names of the authors contacted through colleagues from the same field, the nature of their responses, and the date. We contacted colleagues from different European universities (e.g., University of Cologne, University of Vaasa, University of Novi Sad, etc.) who provided us with some material (e.g., PhD and master's theses and reports). Regarding quality control and secondary search techniques, we entered the information concerning the organization names, journal titles, years searched, number of documents examined, and number of relevant documents found. The same information was recorded regarding register names, search engines, and database names. We used the following keywords with the top-down approach: bilingual, program, achievement, effectiveness, immigrant, minority children. As there were difficulties in retrieving the studies that were matching the inclusion criteria, in some databases we truncated the search term, for example, using only biling*. In addition, we employed the same words in French, German, Russian, and Serbian, but the results were the same as the searches with the English words. The databases used in the search procedure were Web of Science, PsycINFO, ERIC, Science Direct, and Google Scholar.

Inclusion Criteria

To be included in the meta-analysis, studies needed to meet certain inclusion criteria. Since our aim was to achieve consistency between the inclusion criteria of other meta-analyses on the same topic (Greene, 1998; Rolstad et al., 2005; Slavin & Cheung, 2005; Willig, 1985), we reviewed these studies and noted the inclusion criteria that were followed. Of the seven inclusion criteria that we uncovered, four proved to be applicable in our case:

Criterion 1: Studies needed to be an empirical investigation of the effectiveness of bilingual programs in Europe. The focus was on European studies, and this is the main difference between our and other meta-analyses on the topic.

Criterion 2: In order to compare the types of programs language minority children were assigned to, we needed information on between-program comparison containing statistical data.

Criterion 3: The outcome was a measure of specific academic domain measured by quantitative instruments such as, for example, standardized tests of reading.

Criterion 4: Participants were not involved in special education classes.

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 3

A log for keeping track of a search of the literature (based on Cooper, 2009)

Direct-to-researcher search techniques	Used?[a]	Who was contacted?	Date sent[a]	Date reply received[a]	Nature of reply
Personal contact	Yes	Researcher initials			
	Date 03/31/2010	L. G.	01/05/2009	01/05/2009	Master's thesis and articles
	No	S. G.	02/10/2010	02/10/2010	Sent articles by Hopf and Roth
	Reason	S. U.	02/18/2010	02/18/2010	Sent links for Lecocq studies
Mass solicitation	Yes				
	Date 04/15/2010	Organization name			
	No	University of Novi Sad	01/05/2009	01/05/2009	Master's thesis and articles
	Reason	University of Luxembourg	01/29/2010	01/29/2010	Reports and thesis
		University of Berlin	03/31/2010	03/31/2010	—
		University of Liège	03/31/2010	03/31/2010	PhD thesis
		University of Birmingham	03/31/2010	03/31/2010	—
		University of Hamburg	03/31/2010	03/31/2010	—
		University of Cologne	03/31/2010	03/31/2010	Reports
		University of Barcelona	04/14/2010	04/14/2010	—
		University of Vaasa	04/14/2010	04/14/2010	Books, reports, articles
		Multilingual matters	04/15/2010	—	—
					Retained five theses and 21 book chapters
Traditional invisible college "hubs"	Yes	Researcher initials			
	If no, reason	H. R.	03/31/2010	03/31/2010	Three reports plus contacts
		I. G.	03/31/2010	03/31/2010	Contacts for Dittmar and Sanfuchs
		C. G.	03/31/2010	03/31/2010	Arranging the permit for a report
		J. D.	03/31/2010	—	—
		K L.	03/31/2010	03/31/2010	PhD thesis
		U. N.	03/31/2010	—	—
		S. W.	03/31/2010	—	—
		N. D.	03/31/2010	—	—

(continued)

TABLE 3 (continued)

Direct-to-researcher search techniques	Used?[a]	Who was contacted?	Date sent[a]	Date reply received[a]	Nature of reply
		U. S.	03/31/2010	—	—
		S. J.	03/31/2010	—	—
		H. R.	03/31/2010	04/14/2010	—
		C. L.	04/14/2010	—	Recommended two of his books
		S. B.	04/14/2010	—	—
		J. V.	04/14/2010	04/14/2010	—
		T. S. K.	04/14/2010	—	Recommended her web page
		H. T.	04/19/2010	—	—
		L. V.	09/29/2010	—	—
Electronic invisible college; listservs, bulletin boards, discussion groups	Yes	Organization name			
	Date 04/15/2010	European Centre for Modern Languages	04/15/2010	04/15/2010	Articles, books, reports, websites
	No	Content and Language Integrated Learning	04/15/2010	—	—
	Reason	Centre for Multilingual and Multicultural Research	04/15/2010	—	—
		Piccolingo for Early Foreign Language Learning	04/15/2010	—	Reports
		Verein für frühe Mehrsprachigkeit an Kindertageseinrichtungen und Schulen FMKS	04/15/2010	04/16/2010	Forwarded e-mail to the list (received an invitation to the conference in Magdeburg)

Quality-controlled search techniques	Used?[a]	Organization names or journal titles	Years searched	No. of documents examined	No. of relevant documents found
Professional conference paper presentations	Yes	Organization names			
	Date 04/15/2010	The English Trust for European Education			
	No	International Conference on Minority Languages			

(continued)

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 3 (continued)

Quality-controlled search techniques	Used?[a]	Organization names or journal titles	Years searched	No. of documents examined	No. of relevant documents found
	Reason	Bilingualism Matters			
		European Centre for Modern Languages, Council of Europe	1999–2010	102	10
Peer-reviewed journals	Yes	Journal titles			
	Date 04/22/2010	*Applied Psycholinguistics*	1800–2010	500	0
	Date 11/01/2010	*Bilingualism: Language and Cognition*	1800–2010	320	0
	No	*Bilingual Research Journal*	2001–2005	150	4
	Reason	*Change Transformations in Education*	2000–2010	250	0
		Education Quarterly	2004–2010	38	0
		Education Research & Perspectives	1970–2010	59	0
		English Language & Linguistics	1800–2010	30	0
		English Language Teaching	1800–2010	400	0
		English Today	1980–2010	8	0
		Irish Educational Studies	1980–2010	76	4
		International Journal of Bilingual Education and Bilingualism	1932–2010	13	0
		Journal of Educational Psychology	2002–2010	22	0
		English Teaching: Practice and Critique	1970–2010	98	1
		Journal of Child Language	2000–2010	16	0
		Journal of Language, Identity, and Education	1970–2010	98	1
		Journal of Linguistics	2000–2010	16	0
		Journal of Multilingual and Multicultural Development	1997–2010	132	0

(continued)

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 3 (continued)

Quality-controlled search techniques	Used?[a]	Organization names or journal titles	Years searched	No. of documents examined	No. of relevant documents found
		Language Assessment Quarterly	1996–2010	30	3
		Language, Culture and Curriculum	2004–2010	9	0
		Language in Society	1988–2010	20	1
		Language Learning	1998–2010	4	0
		Language Learning and Development	1948–2010	146	2
		Language Variation and Change	2000–2010	16	0
		Language Teaching	1999–2010	328	0
		Nordic Journal of Linguistics	2009–2010	4	0
		The Bilingual Review	1996–2010	420	4
		Review of Education, Pedagogy, & Cultural Studies	1975–2010	9	1
		Review of Educational Research	1931–2010	124	7
		Scandinavian Journal of Educational Research	1957–2010	4	0
		Studies in Second Language Acquisition	1997–2010	227	1
		Scientific Studies of Reading	1997–2010	5	1
		Zeitschrift für Pädagogik	2001–2010	5	3
					Retained 32

Secondary searching techniques	Used?[a]	Years covered	No. of documents examined	No. of relevant documents found
Research report reference lists	Yes			
	Date 04/22/2010			
	Date 11/01/2010	1980–2010	495	99 of which 33 retained
	No			
	Reason			

(continued)

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 3 (continued)

Secondary searching techniques	Used?[a]		Years covered	No. of documents examined	No. of relevant documents found
Research bibliographies	Yes	Source (name) of bibliography			
	Date 04/22/2010	Two-way bilingual immersion programs by Marcia Vargas	1985–2004	28	8
	Date 11/01/2010	Centre for Applied Linguistics	1987–2006	70	5
	No	Centre for Applied Linguistics	1985–2003	122	6
	Reason	Krashen, Rolstad, MacSwan	1978–2007	47	4
Prospective registers	Yes	Register names			
	Date 04/22/2010	The National Centre for Languages CILT	2002–2010	3	1
	Date 11/01/2010	Centre for Reviews and Dissemination	1992–2010	0	0
	No	Small Grants Research	2006–2007	30	0
	Reason	Research Europe	1996–2010	No access	No access
		Social Science Research Council	1923–2010	9	0
		Bilingual database of the University of Birmingham	1976–2006	8	0
Internet	Yes	Search engines			
	Date 04/22/2010	scholar.google.com	1980–2010	Effectiveness AND bilingual AND programs AND achievement AND Europe Excluded terms: deaf, special needs, literature, history, identity, Hispanic, Asia	4 of 272
	Date 11/01/2010	www.ask.com	Anytime	Effectiveness AND bilingual AND programs AND achievement AND Europe	3 of 200

(continued)

TABLE 3 (continued)

Secondary searching techniques	Used?[a]		Years covered	No. of documents examined	No. of relevant documents found
	No	www.yahoo.com	Anytime	Effectiveness AND bilingual AND programs AND achievement AND Europe Excluded terms: special needs	2 of 183
	Reason				
Reference databases	Yes	Database names			
	Date 04/22/2010	PsycINFO	1800–2010	Bilingual AND program AND achievement AND effective	10 of 27
	Date 11/01/2010	ERIC	1968–2010	Bilingual (OR immersion) AND program AND achievement (OR success) AND effective (OR significant)	5 of 178
	No	FRANCIS	2001–2010	Bilingual (OR immersion) AND program AND achievement (OR success) AND effective (OR significant)	2 of 1,444
	Reason	International Bibliography of the Social Sciences		Bilingual (OR immersion) AND program AND achievement (OR success) AND effective (OR significant)	0 of 34

(continued)

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 3 (continued)

Secondary searching techniques	Used?[a]		Years covered	No. of documents examined	No. of relevant documents found
		Sage journals online		Bilingual (OR immersion) AND program AND achievement (OR success) AND effective (OR significant)	6 of 145
		German Education Portal		Bilingual AND achievement	5 of 1,061
		TEL Multidisciplinary thesis		Biling*	0 of 13
		Wiley online library	Anytime	Bilingual AND achievement	6 of 3,433
Citation indexes	Yes	Index names			
	Date 04/22/2010	Social Sciences Citation Index via ISI Web of Knowledge	1975–2010	Bilingual AND program AND achievement AND effective	3 of 38
	Date 11/01/2010				
	No				
	Reason				

[a]Dates in mm/dd/yyyy format.

Due to the very small number of studies on this topic in the European context, three restrictive criteria had to be released in order to allow us to include valuable empirical data. Specifically, these were the following: (a) participants were enrolled in kindergarten through the end of primary school at the start of the study, (b) the program lasted for at least 1 year, and (c) the expectancy and comparison groups were randomly assigned and matched.

Rolstad et al. (2005) provided us with the rationale for releasing the three remaining criteria: They contended that a meta-analysis needs to include all the relevant studies on the topic and release the focus on empirical effects of particular variables in the analysis. The authors also criticized the best evidence approach used by Slavin and Cheung (2005) based on a systematic literature search, quantification of outcomes as effect sizes, and discussion of primary studies that met the inclusion criteria, because they believed that this approach is based on personal preference about what is included and on the arbitrariness of the selection criteria. The inclusion criteria in Rolstad et al.'s (2005) meta-analysis were the following: (a) participants were involved in K–12 education and were not in special education classes, (b) statistical details were provided, and (c) a description of the treatment and comparison programs was available. After initial searches revealed only a very limited number of European studies on the topic, our inclusion criteria were limited to these categories as well.

We developed a coding guide based on Cooper's (2009) suggestions and included 90 detailed questions. Although we retrieved 101 studies, including peer-reviewed articles (n = 32), unpublished research reports from different European universities (n = 33), unpublished master's and PhD theses (n = 5), conference papers (n = 10), and book chapters (n = 21), only 7 published studies (6 in journals and 1 in a monograph) met the inclusion criteria. We excluded 94 studies because they used only qualitative data and there were no control groups in the program. Figure 1 depicts a flowchart outlining the decision-making process for inclusion or exclusion of studies in the meta-analysis following the recommendations put forth by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Group (Moher, Liberati, Tetzlaff, Altman, & the PRISMA Group, 2009).

The seven studies that met the inclusion criteria are included in Table 4. The countries represented in these studies include Spain (Huguet & Gonzáles, 2002; Lasagabaster, 2000; Serra, 1989), England (Hirst, Hannon, & Nutbrown, 2010), Ireland (O'Muircheartaigh & Hickey, 2008), Norway (Ozerk & Krashen, 2001), and the Netherlands (Verhoeven, 1987). The studies were published in English, with the exception of the studies by Huguet and Gonzáles (2002) and Serra (1989), for which we required the assistance of a Spanish translator.

Coding the Studies

After the studies were selected, we coded their characteristics by assigning them a quantitative description. Broad information reporting the characteristics of the studies based on Rolstad et al.'s (2005) and Cooper's (2009) suggestions included (a) report characteristics (e.g., report's ID, author's name, year of publication), (b) bilingual programs (e.g., type, length, content areas in home language), (c) setting (e.g., ethnic group, country of the study, school type),

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

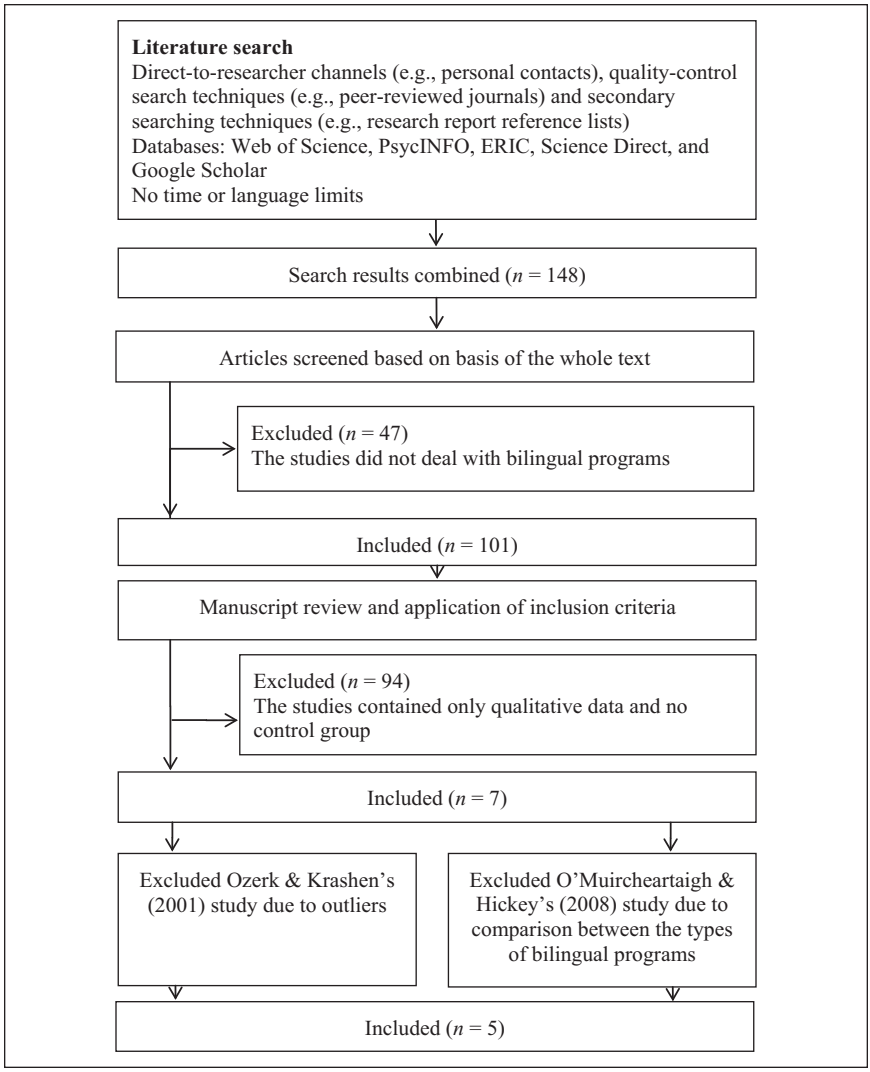

FIGURE 1. *PRISMA flowchart of study selection.*
Note. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
Source. Based on Moher, Liberati, Tetzlaff, Altman, and the PRISMA Group (2009).

(d) participants and sample (e.g., sampling strategy, percentage female, socioeconomic status [SES], grade, age), (e) teachers (credentials in bilingual education, proficiency in students' language, years of experience), (f) research design (e.g., teacher assignment, threats to validity, number of comparisons), (g) outcome

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 4

Background characteristics of the studies in the meta-analysis

Study	Country/minority	Sample	Age	*n*	Outcome	Time
Hirst, Hannon, and Nutbrown (2010)	England/Urdu	16	3–4	1	English literacy	1
Huguet and Gonzáles (2002)	Spain/Asturian	241	12	2	Castilian-Spanish Mathematics	1
Lasagabaster (2000)	Spain/Basque	168	10–13	5	Linguistic creativity	1
O'Muircheataigh and Hickey (2008)	Ireland/Gaelic	97	15	2	Irish	1
Ozerk and Krashen (2001)	Norway/Urdu	67	8–9	4	Civics subjects	1
Serra (1989)	Spain/Catalan	75	4–7	32	Mathematics Oral comprehension Reading Reading speed Reading errors Vocabulary Written expression IQ	1
Verhoeven (1987)	The Netherlands/ Turks	98	6	24	L1/L2 phoneme discrimination L1/L2 productive vocabulary L1/L2 receptive vocabulary L1/L2 sentence imitation	3

Note. *n* = number of comparisons; Time = measurement occasions; L1/L2 = first language/second language.

measure (academic domain, type, language), (h) effect size estimate (pretest and posttest mean and standard deviation, sample size, degrees of freedom, *p* value), and (i) coders (e.g., time spent for coding, notes).

The assessment of risk of bias was performed using our detailed coding guide containing 90 questions. Research design characteristics contained the following questions based on Cooper (2009): (a) type of group assignment (random, convenience, other nonrandom), (b) type of teacher assignment (random, convenience, other nonrandom), (c) the assignment mechanism (self-selected, selected into groups by others on a basis related to outcome, selected into groups by others not known to be related to outcome), (d) use of equating variables (none, prior IQ,

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

prior achievement, native language, SES, other), (e) if equating was done using a statistical process or manually, (f) SES was controlled, (g) threats of validity were reported (confounding, selection bias, history, maturation, repeated testing, instrument change, regression toward the mean, etc.), and (h) number of comparisons in the study. Outcome measures contained the following questions, also based on Cooper (2009): (a) academic domain, (b) type of outcome measure (standardized achievement test, class grades, multiple types of student achievement measures), (c) the score form of the outcome measure (proficiency test, grades, percentile), (d) the language in which outcome was measured, (e) evidence of validity/reliability of outcome measures, and (f) the time when the outcome measure was administered. However, as already explained in the inclusion criteria, there are only very few studies that fulfilled the methodological quality requirements.

We paid special attention when interpreting the information on bilingual programs because "program labels are often oversimplified or misleading" (Rolstad et al., 2005, p. 581). We had one trained coder with a master's degree in psychology. Cohen's kappa of interrater agreement between the raters (the first author and the coder) was very good ($k = 0.93$; Landis & Koch, 1977).

Calculating the Effect Size

Effect size estimates are "the meta-analytic coin of the realm" (Rosenthal, 1995, p. 185) that provide information about "the magnitude and direction of the difference between two groups or the relationship between two variables" (Durlak, 2009, p. 917). As a measure of standardized mean differences and to correct for small sample bias, we used Hedges's *g* (Borenstein et al., 2009) as our effect size indicator. Note that Hedges's *g* has a correction factor (*J*), which is used to convert from Cohen's *d*. This factor uses the degrees of freedom to estimate pooled standard deviation from two independent groups, and it is a less biased estimator than Cohen's *d* (Borenstein et al., 2009). All analyses reported here were performed using the statistical package, Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2005).

As in previous meta-analyses, the first comparison group in the present meta-analysis was the group related to bilingual education pedagogy (i.e., dual language programs, late immersion, and transitional). In this study, we focused only on the comparison between the types of bilingual programs and submersion, which is a "sink-or-swim" program where students are only instructed in the school language, with no inclusion of their home language. Exceptions to this comparison were the studies of O'Muircheartaigh and Hickey (2008), in which early and late immersion programs were compared (both are different types of bilingual programs; our meta-analysis focused on the comparison between a bilingual and a submersion program only), and the Lasagabaster (2000) study, in which dual language was compared to transitional once for the fifth grade and once for the eighth grade, among the usual comparisons versus submersion programs.

Comparison samples were generally not the native speakers, except in one comparison in Ozerk and Krashen's (2001) study, where limited Norwegian speakers were compared to native Norwegian speakers. It is important to distinguish between the groups of language minority children and the native speakers,

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

because analyses may indicate a higher effect size for the native speaker group "unrelated to true achievement differences" (Rolstad et al., 2005, p. 583).

Results

Individual Studies

There is a wide range of variability between the studies in terms of their samples, age of participants, and outcome measures. This is the reason we employed the random-effects model under which "we assume that the true effect size varies from study to study, and the summary effect is our estimate of the mean of the distribution of effect sizes" (Borenstein et al., 2009, p. 6).

When plotting the studies it was possible to identify outliers. In addition, we calculated the weights of each study attributable to the overall effect size. Findings of unusually high standard errors and 95% CIs [1.08, 2.14] pointed to the Ozerk and Krashen (2001) study as an outlier. We conducted the analysis both with and without this study to see if there was a substantial difference in the effect sizes. Moreover, the effect sizes of each study were plotted; however, because of the small number of studies (only five without outliers), a visual display was not informative (Rosenthal, 1995). We decided to present individual effects for each study (Table 5) and a stem and leaf plot of the effect sizes for language outcomes in bilingual versus submersion programs (Figure 2). In this plot, we can also identify outliers (below −1.0 and above 1.0) that were further excluded from the final analysis (i.e., Ozerk & Krashen, 2001).

Synthesis of Studies

In Table 6, we report only the results from the random-effects model. In total, there were 70 comparisons. Although CIs are narrower for fixed-effect models than for random-effects models, the ability to make generalizations for random-effects models is more difficult than for fixed-effect models. Moreover, Cooper (2009) pointed out that many researchers often choose the random-effects model because they "feel the random sampling of studies is more descriptive of their real-world circumstances and also will lead to a more conservative conclusion about the range of impacts the intervention might have" (p. 191). However, nearly identical results were found when we used fixed-effect models.

Table 6 also presents the effect sizes for all the outcomes in which the subgroups are treated as studies. As in all effect size calculations, Hedges's g was used to calculate the effect size. In this particular calculation, Hedges's g was obtained by taking into account all the subgroups with all the outcomes (70 outcomes; see Table 6) in all the studies and treating them as separate studies. However, this is a restricted option because the assumption of statistical independence is not met, indicating that the outcomes are not independent because they belong to the same study; thus, participants are the same (Borenstein et al., 2009). Treating nonindependent outcomes as independent leads to significance test errors (Rosenthal, 1995). Therefore, studies needed to be treated as units.

Before reporting the effect sizes, it is important to mention that three studies contributed the most to the relative weight of the final effect size: Huguet and Gonzáles' study (2002) contributed the most with 36%, followed by Lasagabaster

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 5

Comparisons of the effect sizes (ES) by study

Study	*N* of ES	*M* of ES	*SE* of ES	95% Confidence interval
Hirst, Hannon, and Nutbrown (2010)				
Preschool				
Range of *N*s: 8 vs. 8				
Bilingual vs. submersion				
Early literacy in English	1	1.31	0.53	[0.27, 2.35]
Huguet and Gonzáles (2002)				
Second-year students				
Range of *N*s: *N* = 25 vs. 11, *N* = 130 vs. 93				
Bilingual vs. submersion				
Castilian-Spanish linguistic knowledge	1	0.84	0.37	[0.11, 1.57]
Total bilingual vs. submersion				
Mathematics	1	−0.28	0.14	[−0.55, −0.01]
Lasagabaster (2000)				
Grade 5–8				
Grade 8 *N* = 42 vs. 42, Grade 5 *N* = 42 vs. 42				
Grade 5				
Dual B vs. submersion A				
Average linguistic creativity				
Flexibility	1	0.28	0.22	[−0.15, 0.71]
Fluency	1	−0.16	0.22	[−0.59, 0.27]
Originality	1	−0.05	0.22	[−0.48, 0.38]
Dual B vs. transitional D				
Average linguistic creativity				
Flexibility	1	0.28	0.22	[−0.15, 0.71]
Fluency	1	−0.16	0.22	[−0.59, 0.27]
Originality	1	−0.05	0.22	[−0.48, 0.38]
Transitional D vs. submersion A				
Average linguistic creativity				
Flexibility	1	0.28	0.22	[−0.15, 0.71]
Fluency	1	−0.16	0.22	[−0.59, 0.27]
Originality	1	−0.05	0.22	[−0.48, 0.38]
Grade 8				
Dual B vs. submersion A				
Average linguistic creativity	1	0.51	0.22	[0.08, 0.94]
Flexibility	1	0.63	0.22	[0.20, 1.06]
Fluency	1	0.29	0.22	[−0.14, 0.72]

(continued)

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 5 (continued)

Study	*N* of ES	*M* of ES	*SE* of ES	95% Confidence interval
Originality				
Dual B vs. transitional D				
Average linguistic creativity				
Flexibility				
Fluency	1	−0.30	0.22	[−0.73, 0.13]
Originality				
Transitional D vs. submersion A				
Average linguistic creativity	1	0.57	0.22	[0.14, 1.00]
Flexibility	1	0.53	0.22	[0.10, 0.96]
Fluency	1	0.29	0.22	[−0.14, 0.72]
Originality	1	0.55	0.22	[0.12, 0.98]
O'Muircheartaigh and Hickey (2008)				
Fourth-year secondary school				
Range of *N*s: *N* = 57 vs. 20; *N* = 64 vs. 35				
Early vs. late immersion				
Irish C test 1st level	1	1.04	0.27	[0.51, 1.57]
Irish C test 4th level	1	0.73	0.22	[0.30, 1.16]
Irish junior certificate				
Mathematics				
Ozerk and Krashen (2001)				
Grade 3				
Range of *N*s: 17 vs. 2, 17 vs. 16, 17 vs. 32, 2 vs. 32				
Civic subjects				
Bilingual vs. native bilingual	1	0.69	0.72	[−0.72, 2.10]
Bilingual vs. submersion	1	5.59	0.77	[4.08, 7.10]
Bilingual vs. submersion native	1	1.13	0.32	[0.50, 1.76]
Bilingual vs. submersion	1	0.59	0.72	[−0.82, 2.00]
Serra (1987)				
Grade 2				
Range of *N*s: *N* = 16 vs. 21, 16 vs. 21				
Immersion A vs. submersion C				
IQ	1	−0.64	0.33	[−0.29, 0.01]
Mathematics	1	0.07	0.34	[−0.60, 0.74]
Oral comprehension	1	−0.07	0.33	[−0.72, 0.58]
Reading errors	1	0.72	0.34	[0.05, 1.39]
Reading	1	0.14	0.33	[−0.51, 0.79]
Reading speed	1	−0.66	0.33	[−1.31, −0.01]
Vocabulary	1	−1.15	0.35	[−1.84, −0.46]

(continued)

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 5 (continued)

Study	*N* of ES	*M* of ES	*SE* of ES	95% Confidence interval
Written expression	1	−0.02	0.33	[−0.67, 0.63]
Immersion A vs. submersion D				
IQ	1	−0.35	0.33	[−1.00, 0.30]
Mathematics	1	0.02	0.33	[−0.63, 0.67]
Oral comprehension	1	0.32	0.33	[−0.33, 0.97]
Reading errors	1	0.59	0.33	[−0.06, 1.24]
Reading	1	0.14	0.33	[−0.51, 0.79]
Reading speed	1	−0.50	0.33	[−1.15, 0.15]
Vocabulary	1	−0.84	0.34	[−1.51, −0.17]
Written expression	1	0.53	0.33	[−0.12, 1.18]
Immersion B vs. submersion C				
IQ	1	−0.11	0.33	[−0.76, 0.54]
Mathematics	1	1.25	0.36	[0.54, 1.96]
Oral comprehension	1	0.77	0.34	[0.10, 1.44]
Reading errors	1	0.07	0.33	[−0.58, 0.72]
Reading	1	0.95	0.34	[0.28, 1.62]
Reading speed	1	−0.24	0.33	[−0.89, 0.41]
Vocabulary	1	0.34	0.33	[−0.31, 0.99]
Written expression	1	−0.42	0.33	[−1.07, 0.23]
Immersion B vs. submersion D				
IQ	1	0.13	0.32	[−0.50, 0.76]
Mathematics	1	0.99	0.34	[0.32, 1.66]
Oral comprehension	1	1.04	0.34	[0.37, 1.71]
Reading errors	1	0.06	0.32	[−0.57, 0.69]
Reading	1	0.88	0.34	[0.21, 1.55]
Reading speed	1	−0.05	0.32	[−0.68, 0.58]
Vocabulary	1	0.42	0.33	[−0.23, 1.07]
Written expression	1	0.16	0.32	[−0.47, 0.79]
Verhoeven (1987)				
Grade 1				
Range of *N*s: 24 vs. 74				
Transitional vs. submersion				
L1 phoneme discrimination	3	−0.41	0.24	[−0.88, 0.06]
		−0.02	0.24	[−0.49, 0.45]
		−0.22	0.24	[−0.69, 0.25]
L1 productive vocabulary	3	0.79	0.24	[0.32, 1.26]
		1.06	0.25	[0.57, 1.55]
		0.68	0.24	[0.21, 1.15]

(continued)

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

TABLE 5 (continued)

Study	*N* of ES	*M* of ES	*SE* of ES	95% Confidence interval
L1 receptive vocabulary	3	0.68	0.24	[0.21, 1.15]
		0.75	0.24	[0.28, 1.22]
		1.04	0.25	[0.55, 1.53]
L1 sentence imitation	3	0.53	0.24	[0.06, 1.00]
		0.74	0.24	[0.27, 1.21]
		0.83	0.24	[0.36, 1.30]
L2 phoneme discrimination	3	−0.31	0.24	[−0.78, 0.16]
		−0.12	0.23	[−0.57, 0.33]
		0.25	0.23	[−0.20, 0.70]
L2 productive vocabulary	3	0.02	0.23	[−0.43, 0.47]
		−0.37	0.23	[−0.82, 0.08]
		0.03	0.24	[−0.44, 0.50]
L2 receptive vocabulary	3	−0.27	0.23	[−0.72, 0.18]
		−0.35	0.23	[−0.80, 0.10]
		−0.10	0.24	[−0.57, 0.37]
L2 sentence imitation	3	0.39	0.24	[−0.08, 0.86]
		−0.10	0.23	[−0.55, 0.35]
		0.18	0.23	[−0.27, 0.63]

Note. L1 = first language; L2 = second language.

(2000) with 28%, and Serra (1989) with 30%. The remaining 6% was distributed between the two remaining studies (Hirst et al., 2010; Verhoeven, 1987). Two studies (O'Muircheartaigh & Hickey, 2008; Ozerk & Krashen, 2001) were excluded from the final analysis. Ozerk and Krashen's (2001) study was excluded due to extreme outliers whereas O'Muircheartaigh and Hickey's study (2008) study was excluded due to the comparisons of the different types of bilingual programs. In total, five studies out of the seven that met the inclusion criteria were considered for the final analysis on the outcome of reading.

The effect size for all outcomes in seven studies is $g = 0.38$, indicating a significant, positive small to moderate effect for bilingual programs. However, because there were cases where different outcomes were measured, such as IQ, originality, and flexibility, thus using more specific measures for these outcomes, we decided to group the outcomes by reading and mathematics. There was no effect of bilingual program for mathematics. This result is very limited as there were only two comparisons. On the other hand, all the other outcomes (except IQ, originality, civic subjects, and flexibility) were grouped into reading. These outcomes were: Castillian-Spanish linguistic knowledge (Huguet & Gonzáles, 2002), English early literacy (Hirst et al., 2010), linguistic fluency, linguistic creativity (Lasagabaster, 2000), Irish C test first and fourth levels (O'Muircheartaigh & Hickey, 2008), oral comprehension, reading errors, reading, reading speed,

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

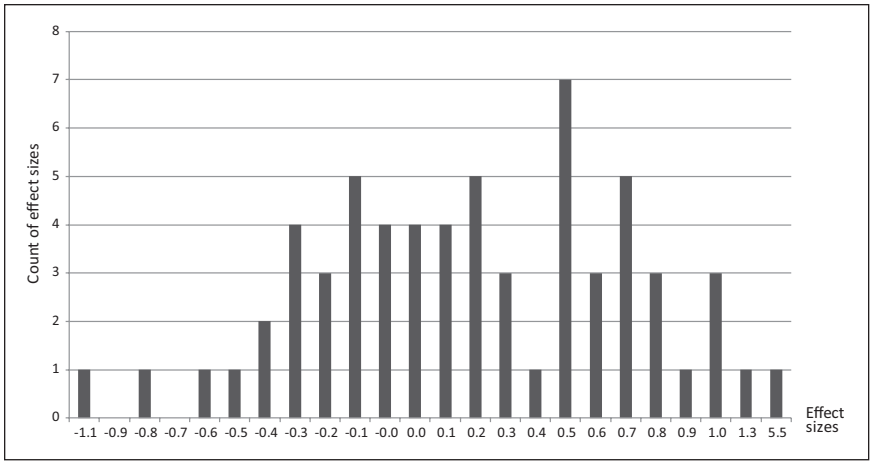

FIGURE 2. *Stem and leaf plot for all the individual effect sizes of the language outcomes in bilingual versus submersion programs including the outliers.*

TABLE 6

Combined effect sizes (ESs) by grouping

Grouping	*N* of ES	*M* of ES	*SE* of ES	95% CI
All outcome measures	70	0.24	0.06	[0.13, 0.40]
All outcomes in studies as units	7	0.38	0.11	[0.17, 0.60]
Without Ozerk and Krashen (2001)	6	0.23	0.08	[0.07, 0.40]
Without Hirst, Hannon, and Nutbrown (2010)	4	0.20	0.05	[0.11, 0.30]
Reading	**5**	**0.23**	**0.07**	[**0.10, 0.36**]
Mathematics	2	0.12	0.41	[−0.68, 0.92]
All outcomes in home language	1	0.52	0.07	[0.38, 0.66]
Language minority vs. language minority students	7	0.36	0.11	[0.15, 0.58]
Bilingual vs. submersion	6	0.32	0.11	[0.11, 0.53]

Note. N = number of outcome measures, ES = effect size, CI = confidence interval. Reading outcome is in boldface. The grouping category "language minority vs. language minority" indicates a comparison between the performances of language minority students in bilingual programs versus language minority students in submersion programs in all outcomes.

vocabulary, written expression (Serra, 1989), productive vocabulary, receptive vocabulary, sentence imitation, and phoneme discrimination (Verhoeven, 1987). These outcomes are considered literacy skills (Whitehurst & Lonigan, 1998). Since literacy involves a wider range of skills than those included in the present

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

outcomes (e.g., conventions of print, knowledge of letters, linguistic (phonological awareness), including phonological sensitivity (e.g., identifying words that rhyme, writing; Whitehurst & Lonigan, 1998), we opted for the term *reading*. Mathematics was measured in the studies conducted by Huguet and Gonzáles (2002), O'Muircheartaigh and Hickey (2008), and Serra (1987) under the label name "mathematics"; thus, the terminology remained the same.

Concerning the outcome of home language, we see that $g = 0.52$ is the largest effect in Table 6. One could further speculate about the possible importance of this effect; however, since it is based only on one comparison, such a speculation would be clearly limited. It is important to mention that out of 13 program comparisons, there were 3 comparisons that compared types of bilingual programs (bilingual vs. native bilingual, dual vs. transitional, and early vs. late immersion). In the present analysis, we excluded these three comparisons for the outcomes of reading and mathematics. O'Muircheartaigh and Hickey (2008) compared students in early versus late immersion. Since our focus was only on the comparisons between bilingual and submersion programs, we excluded this study along with Ozerk and Krashen's (2001).

Since Ozerk and Krashen's (2001) study was identified as an outlier, we conducted the analysis without it, which resulted in a significant, small positive effect size of $g = 0.23$ favoring bilingual education over submersion programs. Moreover, Hirst et al.'s (2010) study was excluded because the program assessed lasted only for a year and it was not a school program. Pakistani 3-year-old children were taught Urdu at home before they went to school. Although this program included home language of children as well as their early English literacy, it was not an official school program. However, we did retain this study for further analysis as it contains valuable information on reading. Even without this study, the effect size for bilingual education is $g = 0.20$, which is small but positive. All the effect sizes reported in Table 6 (except for mathematics) were significant ($p < .001$).

We included only the comparison of language minority in bilingual programs versus language minority students in submersion programs and excluded the comparisons of language minority students to the native speakers. The effect size was $g = 0.36$ for all the outcomes and $g = 0.34$ for reading, again indicating positive small to moderate effects for bilingual programs. Finally, only bilingual versus submersion programs were compared where the effect size was small to moderate and positive ($g = 0.32$) in favor of bilingual programs. The final analysis excluded, as mentioned earlier, Ozerk and Krashen's (2001) study as an outlier and O'Muircheartaigh and Hickey's (2008) study because it compared students in two types of bilingual programs, that is, early versus late immersion.

Publication Bias

Two important analyses included in this meta-analysis are those of homogeneity and publication bias. The most informative analysis included bilingual programs only versus submersion only in reading and without comparisons to native speakers. This analysis—with a small but significant effect size of $g = 0.23$ in favor of bilingual programs—yielded a nonsignificant result for heterogeneity, $Q(4) = 8.71$, $p = 0.07$. Rosenthal (1995) noted that if the test is significant, this alerts the meta-analyst to the likelihood that all the effect sizes are not "cut from the same cloth," the true effects vary, and the meta-analyst is "morally" obliged to

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

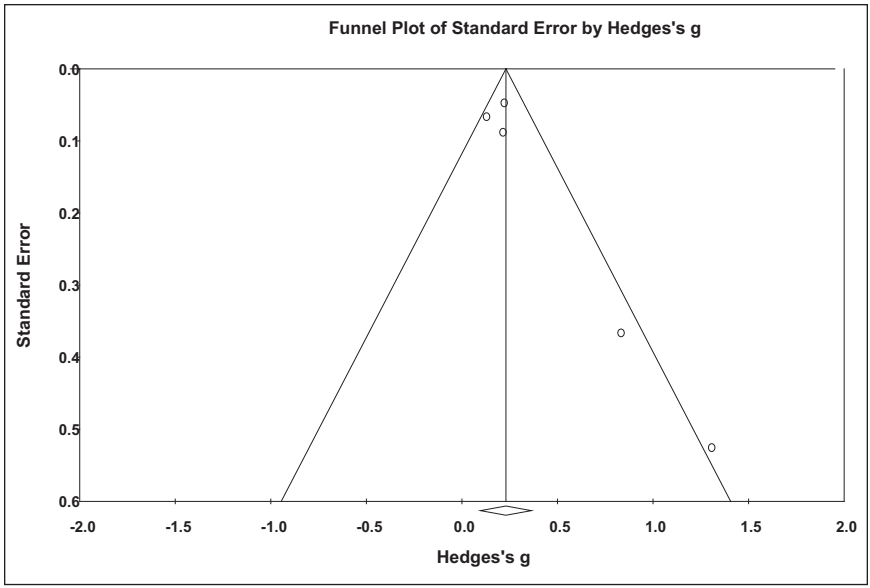

FIGURE 3. *Funnel plot of the five studies included in the meta-analysis.*

search for the moderators (p. 188). However, a nonsignificant *p* value should *not* be taken "as evidence that the effect sizes are consistent, since the lack of significance may be due to low power" (Borenstein et al., 2009, p. 113), which is the case in the present meta-analysis. Indeed, with a small number of studies and/or large study variance (small studies), "even substantial between-studies dispersion might yield a nonsignificant *p*-value" (p. 113).

There are different sources of publication bias (e.g., language bias, availability bias, cost bias, familiarity bias, duplication bias, citation bias; Borenstein et al., 2009). At this point, we refer to the *file drawer problem*: a "well-supported suspicion that the studies retrievable in a meta-analysis are not likely to be a random sample of all studies actually conducted" (Rosenthal, 1995, p. 189). Moreover, the studies that have been published are more likely to have achieved statistical significance than those unpublished studies still in the file drawers as was noted by Sterling (1959) more than 50 years ago. One of the ways to deal with this problem, suggested by Rosenthal (1995), is to "calculate the number of studies averaging the null results that must be in the file drawer before the overall probability of Type 1 error can be brought to any desired level of significance" (p. 189).

The funnel plot of standard errors is, although asymmetric, not informative because of the fact that only 5 studies (Figure 3) are included. Rosenthal's *fail-safe N* (Rosenthal, 1995) indicates that 45 studies are necessary to nullify the effect; however, the *fail-safe N* is based on significance tests that combine *p*-values across studies, and it is widely criticized (Borenstein et al., 2009). Finally, *trim and fill* (Duval & Tweedie, 2000) uses an iterative procedure to remove the

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

most extreme small studies from the positive side of the funnel plot, recomputing the effect size until the plot is symmetric (Borenstein et al., 2009). This method adjusted the effect size from $g = 0.23$ to $g = 0.20$, and this remained significant, meaning that the difference was trivial and the effect size was quite stable. However, "this method strongly depends on the assumptions of the model for why studies are missing, and the algorithm for detection can be influenced by one or two aberrant studies" (Borenstein et al., 2009, p. 286).

Discussion

The present meta-analysis examined European studies on the effectiveness of bilingual programs in promoting academic achievement of language minority children. Based on the previous meta-analyses conducted in the United States, we hypothesized that there will be a significant positive effect for bilingual over submersion programs on the academic achievement of language minority children. Out of 101 studies retrieved, only 7 met the inclusion criteria. Of the 7 studies that met the inclusion criteria, we further excluded 2 studies: 1 due to extreme outliers and 1 due to comparisons done only between the different types of bilingual programs and not between the bilingual versus submersion programs, which was the focus of our study. Results from the random-effects model of the five remaining studies indicated a small positive effect ($g = 0.23$; 95% CI [0.10, 0.36]) for bilingual over submersion programs on the academic achievement of language minority children, particularly in reading. Thus, this meta-analysis supports bilingual education in Europe, which specifically includes the home language of language minority children in school instruction.

Our results are in line with the results of Rolstad et al. (2005, 2008). These authors reported a positive effect for bilingual education of 0.23 *SDs*, whereas Willig (1985) reported 0.20 and Greene (1998) reported 0.21. Within the framework of the present study, we were unable to compare the results in mathematics and assessments in children's home language because of the small number of existing comparisons. The findings of the present meta-analysis are in agreement with those of previous meta-analyses (Greene, 1998; Krashen & McField, 2005; Rolstad et al., 2005, 2008; Slavin & Cheung, 2005; Willig, 1985) in favoring bilingual education that uses the home language of language minority children in the instruction of school subjects to promote their academic achievement.

Furthermore, these results based on only European studies support the superiority of bilingual education programs to the submersion ones outside of the United States. Moreover, this meta-analysis was an opportunity to examine the number of published European studies in this area and leads us to agree with Arnett (2008) that, indeed, American studies are overrepresented in the major psychology journals. Therefore, to help close this gap, we urge European researchers to conduct and publish studies on the effectiveness of bilingual programs on academic achievement. Particularly since the advent of the relaxed borders of the European Union allowing Europeans to move between countries freely, the incidence of children being schooled in countries other than their home country has sharply increased. Therefore, the education and academic achievement of immigrant youth must become a priority of educational systems across Europe. Based on this study, we second Slavin and Cheung's (2005) comments:

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

> The most important conclusion from research comparing the relative effects of bilingual and immersion programs for [. . .] language learners is that there are far too few high-quality studies of this question (p. 273), . . . however, bilingual advocates also argue that without native language instruction [. . .] language learners are likely to lose their native language proficiency, or fail to learn to read in their native language, losing skills that are of economic and social value in the world today. (p. 249)

Practical Significance

Describing effect sizes to general audiences without additional explanation leaves "most people scratching their heads" (Cooper, 2009, p. 212). Cooper cautioned not to apply labels for effects without providing additional contexts. Both Cohen (1988) and Cooper (2009) pointed out that the adjectives "significant," "promising," and "proven" should be treated with caution when it comes to their interpretation in practice. Terms such as *significant, important*," *notable*, and *consequential* are related to the effect size set at $d = 0.25$ (Promising Practices Network, 2007; What Works Clearinghouse, 2007, as cited in Cooper, 2009, p. 208). On the other hand, Lipsey and Wilson (1993) concluded, "We cannot arbitrarily dismiss statistically modest values (even 0.10 or 0.20 *SD*s) as obviously trivial" (p. 1199). In a similar vein, Bloom, Hill, Black, and Lipsey (2008) warned,

> Cohen's widely used "small," "medium," and "large" effect size heuristics and the sweeping claim that an effect size of 0.25 is required for "educational significance" clearly have no general applicability to achievement effect sizes for educational interventions. Their one-size-fits-all character is not sufficiently differentiated to be useful for any specific intervention circumstance and is more likely to result in misleading expectations and interpretations about the respective effect sizes. (p. 30)

As for the present effect size ($g = 0.23$), which is a relatively small effect (equivalent to $r = .10$, explaining about 1% of variance), according to Cooper (2009), it is associated with an increase in success rate from 45% to 55%. For example, if an intervention meant to increase students' reading scores above a proficiency threshold would have this effect size, it would mean that 10 more children in every 100 would meet the minimum requirement. In addition, this would generally mean that children in bilingual programs scored about 20% higher in reading in school language than children in submersion programs. This interpretation would apply to the present meta-analysis as well. Moreover, McGraw and Wong (1992) proposed a common language effect size statistic (CL) where an effect size of 0.2 indicates a value of CL = 0.56. This value signifies the probability that an individual score of the experimental group is superior to a person in the control group, if the two are chosen at random, is 56%. However, can we support these statements?

Hill, Bloom, Black, and Lipsey (2008) noted that the interpretation of the effect size is an issue of particular interest to policy makers and program officials; thus, they explored three of several possible types of benchmarks: (a) normative expectations for change, (b) policy-relevant performance gaps, and (c) effect size results from similar studies. With respect to the normative expectations for change, the authors underlined that the natural growth in test scores declines as students age. Hill et al. concluded that interpreting the magnitude of an

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

intervention effect should be in the context of (a) the intervention being studied, (b) the outcomes being measured, and (c) the samples or subgroups being examined. Therefore, it is difficult to interpret the present effect size of $g = 0.23$ in practical terms, as the present meta-analysis included children from kindergarten to Grade 8, different ethnic groups, and schools from different countries. However, this issue is important, and should be addressed by other meta-analysts.

Limitations of the Study

The limitations of the present meta-analysis are tied to the method itself and to the studies in the field of bilingual education included in this and other meta-analyses. An obvious limitation of the present meta-analysis is the small number of studies. According to Rosenthal (1995), meta-analysis can be applied to "as few as two studies; but when there are very few studies, the meta-analytic results are relatively unstable" (p. 185). Second, since our study sample was small, no attempt for moderator analysis was made. Multivariate approaches can be used in meta-analysis for providing more accuracy (e.g., robust variance estimation in meta-regression; see Hedges, Tipton, & Johnson, 2010). However, because there was a lack of information on correlation between dependent variables and between treatments, no multivariate analysis was conducted. It would be both interesting and valuable to investigate duration of the program as a moderator variable and include quasi-experimental studies as well.

Other limitations to this and other meta-analyses in this field are related to the methodological flaws of the primary studies in bilingual education. Takakuwa (2005) pointed out that among the common mistakes in studies of bilingualism are overgeneralization, invalid use of tests of statistical significance, and inappropriate use of analysis of covariance. For example, the schools in which studies were conducted were usually not randomly selected, which means that the subjects selected nonrandomly are not samples of a population but rather of the population itself, and the results should not be generalized beyond participants. In addition, most studies used comparisons between bilingual and monolingual children and/or focused on balanced bilinguals who have native-like control of two languages, and did not consider nonbalanced bilinguals "who have disparate abilities in the two languages" (Diaz, 1985, p. 1377, as cited in Takakuwa, 2005, p. 2229).

Although meta-analysis imposes discipline in the process of summing up the research findings and protects against overinterpretation, it seems difficult to capture qualitative distinctions between studies in this way. The research on bilingual education has been predominately quantitative (Hadi-Tabassum, 2006). Critics of bilingual education state that there has been relatively little in-depth examination of the contextual factors: (a) the quality of education in the minority language, (b) the effects of bilingual education on intergroup relations between students, (c) how bilingual education programs define the relationship between language and power, and (d) how that relationship may affect both language majority and language minority students in bilingual education classrooms (Valdes, 1999, as cited in Hadi-Tabassum, 2006, p. 1). Both quantitative and qualitative research can offer a clearer understanding of the complexity of bilingual education, which is not "simply a politically neutral instructional phenomenon but rather is implicated in national and international competition between groups for material and symbolic resources" (Cummins, 2010, p. xiv). Particularly, this means that

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

interventions and longitudinal studies are desperately needed. The sooner they start, the better things will be for those who struggle with the language they are being instructed in but do not fully understand.

Future Steps

Finally, is bilingual education a luxury we cannot afford? García (2009) called attention to the study performed under the California legislature in the United States that investigated bilingual programs and found no budgetary advantage for English-only programs:

> The incremental cost was about the same each year ($175 to $214) for bilingual and English immersion programs, as compared with $1,198 for English as a second language (ESL) "pullout" programs. The reason was simple—the pullout approach requires supplemental teachers, whereas in-class approaches do not (Parrish, 1994; Crawford, 1998, 4). (p. 392)

Furthermore, Patrinos and Velez (2009), for example, concluded that students in bilingual schools in Guatemala have "higher attendance and promotion rates, and lower repetition and dropout rates," which result in "considerable cost savings," estimated at "$5 million, equal to the cost of primary education for 100,000 students" (p. 594). Similarly, Samway and McKeon (1999) argued that "*not* providing bilingual education can be costly in human terms" (p. 13). "The economic benefits of producing truly bilingual citizens may far outweigh any programmatic costs incurred" (p. 14).

To conclude, we want to highlight the effect size ($g = 0.23$) in the present meta-analysis that examined five European studies on the effectiveness of bilingual programs in promoting academic achievement of language minority children. This effect size is very similar to those found in the five previous meta-analyses in the United States that all provide support in favor of bilingual education. Today, we understand that the maintenance of minority languages is crucial for benefits to a child's confidence and academic performance, maintenance of linguistic diversity, social benefits, tolerance, social cohesion, both local and national economic benefits, national security, and conflict avoidance, as well as for preserving local identity and linguistic heritage (Sallabank, 2006). The findings of the present meta-analysis also underline this importance.

Note

The authors would like to thank Adrijana Gospić, Brad Bushman, Martin Brunner, Thomas Boll, Christine Merrell, Joan Barbara Travers Simon, David Labaree, Larry Hedges, Tove Skutnabb-Kangas, Bojan Reljić, Lisa Trierweiler, and Lucija Andre for their valuable comments. This research was supported by the Fond National de la Recherche, Luxembourg, as a doctoral dissertation of the first author.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Abedi, J. (2003). *Impact of student language background on content-based performance: Analysis of extant data*. Los Angeles, CA: CRESST.

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

Abu-Rabia, S. (2001). Testing the interdependence hypothesis among native adult bilingual Russian-English students. *Journal of Psycholinguistic Research*, *30*, 437–455. doi:10.1023/A:1010425825251

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*, 207–245. doi:10.3102/0034654310368803

Ardasheva, Y., Tretter, T. R., & Kinny, M. (2012). English language learners and academic achievement: Revisiting the threshold hypothesis. *Language Learning*, *62*, 769–812. doi:10.1111/j.1467-9922.2011.00652.x

Arnett, J. J. (2008). The neglected 95: Why American psychology needs to become less American. *American Psychologist*, *63*, 602–614. doi:10.1037/0003-066X.63.7.602

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth*. Mahwah, NJ: Lawrence Erlbaum.

Baetens-Beardsmore, H. (2009). Bilingual education: Factors and variables. In O. García (Ed.*), Bilingual education in the 21st century: A global perspective* (pp. 137–159). Malden, MA: Wiley-Blackwell.

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

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, *70*, 636–644. doi:10.1111/1467-8624.00046

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

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45*, 459–464. doi:10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*, 240–250. doi:10.1016/j.tics.2012.03.001

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. doi:10.1111/j.1467-7687.2004.00351.x

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Manpower Demonstration Research Corporation working papers on research methodology*. New York: Manpower Demonstration Research Corporation. Retrieved from http://files.eric.ed.gov/fulltext/ED503202.pdf

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis* (Version 2) [Computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley.

Brinton, D., Snow, M. A., & Wesche, M. (1989). *Content-based second language instruction*. New York, NY: Newbury House.

Chireac, S. M., Serrat, E., & Huguet, A. (2011). Transferencia en la adquisición de segundas lenguas. Un estudio con alumnado rumano en un contexto bilingüe [Transfer in second language acquisition. A study with Romanian schoolchildren in a bilingual context]. *Revista de Psicodidactica*, *16*, 267–289. doi:10.1387/RevPsicodidact.1230

Christensen, G., & Stanat, P. (2007). *Language policies and practices for helping immigrants and second-generation students succeed*. Washington, DC: Migration Policy Institute.

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). New York, NY: Academic Press.
Cooper, H. (2009). *Research synthesis and meta-analysis: A step-by-step approach*. Thousand Oaks, CA: Sage.
Cummins, J. (1981). Empirical and theoretical underpinnings of bilingual education. *Journal of Education*, *163*, 16–30.
Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.
Cummins, J. (2010). Introduction to Volume 5: Bilingual education. In J. Cummins & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. Bilingual education* (pp. i–xi). New York, NY: Springer.
Davies, A. (2003). *The native speaker: Myth and reality*. Clevedon, England: Multilingual Matters.
Durgunoğlu, A. Y. (1998). Acquiring literacy in English and Spanish in the United States. In A. Y. Durgunoğlu & L. Verhoeven (Eds.), *Literacy development in a multilingual context: Cross-cultural perspectives* (pp. 135–145). Mahwah, NJ: Lawrence Erlbaum.
Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*, 917–928.
Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. doi:10.1111/j.0006-341X.2000.00455
García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Oxford, England: Wiley-Blackwell.
Gersten, R. (1985). Structured immersion for language minority students: Results of a longitudinal evaluation. *Educational Evaluation and Policy Analysis, 7*, 187–196.
Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
Greene, J. J. (1998). *A meta-analysis of the effectiveness of bilingual education*. Claremont, CA: Thomas Rivera Policy Institute.
Grin, F. (2003). Language planning and economics. *Current Issues in Language Planning*, *4*, 1–66. doi:10.1080/14664200308668048
Hadi-Tabassum, S. (2006). *Language, space and power. A critical look at bilingual education*. Clevedon, England: Multilingual Matters.
Hakuta, K., & Mostafapour, E. F. (1996). Perspectives from the history and politics of bilingualism and bilingual education in the United States. In I. Parasnis (Ed.), *Cultural and language diversity and the deaf experience* (pp. 38–50). New York, NY: Cambridge University Press.
Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* Santa Barbara: University of California Linguistic Minority Research Institute.
Hauptman, S., Mansur, F., & Tal, R. (2008). A trilingual teaching model for developing academic literacy skill in classical Arabic (L1), Hebrew (L2) and English (FL) in Southern Israel. *Journal of Multilingual and Multicultural Development*, *28*, 181–197. doi:10.2167/jmmd530.0
Hedges, L., Tipton, E., & Johnson, C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65. doi:10.1002/jrsm.5

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, W. M. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*, 172–177. doi:10.1111/j.1750-8606.2008.00061.x

*Hirst, K., Hannon, P., & Nutbrown, C. (2010). Effects of preschool bilingual family literacy programme. *Journal of Early Childhood Literacy*, *10*, 183–208. doi:10.1177/1468798410363838

Huguet, A. (2008). Interdependencia lingüística y práctica educativa. Un estudio con escolares de origen inmigrante [Linguistic interdependence and educational practice: A study with schoolchildren]. *Infancia y Apprednizaje*, *31*, 283–301. doi:10.1174/021037008785702901

*Huguet, A., & Gonzáles, A. (2002). Conocimiento linguistico y rendimiento escolar. Implicaciones educativas en un contexto bilingue [Linguistic knowledge and school performance. Educational implications in a bilingual context]. *Infancia y Apredinzaje*, *25*, 53–68. doi:0.1174/021037002753508520

Kim, J., & Herman, J. L. (2009). A three-state study of English learner progress. *Educational Assessment*, *14*, 212–231. doi:10.1080/10627190903422831

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

Krashen, S., & McField, G. (2005). What works? Reviewing the latest evidence on bilingual education. *Language Learner*, *1*(2), 7–10.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310

*Lasagabaster, D. (2000). The effects of three bilingual education models on linguistic creativity. *International Review of Applied Linguistics in Language Teaching*, *38*, 1–15. doi:10.1515/iral.2000.38.3-4.213

Lindholm-Leary, K. L. (2001). *Dual language education*. Clevedon, England: Multilingual Matters.

Lindholm-Leary, K. L., & Borsato, G. (2007). Academic achievement. In F. Genesee, K. Lindholm-Leary, W. M. Saunders, & D. Christian (Eds.), *Educating English language learners: A synthesis of research evidence* (pp. 176–222). New York, NY: Cambridge University Press.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*, 1181–1209. doi:10.1037/0003-066x.48.12.1181

MacSwan, J., & Rolstad, K. (2010). The role of language in theories of academic failure for linguistic minorities. In J. Petrovic (Ed.), *International perspectives on bilingual education: Policy, practice, and controversy* (pp. 173–193). Charlotte, NC: Information Age.

McField, G. (2002). Does program quality matter? *A meta-analysis of selected bilingual education studies* (Unpublished doctoral dissertation). University of Southern California, Los Angeles, United States.

McGraw, O. K., & Wong, S. P. (1992). A common language effect size statistics. *Psychological Bulletin*, *111*, 361–365. doi:10.1037/0033-2909.111.2.361

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*(7), e1000097. doi:10.1371/journal.pmed.1000097

Murphy, V. A. (2010). The relationship between age of learning and type of linguistic exposure in children learning a second language. In E. Macaro (Ed.), *The continuum companion to second language acquisition* (pp. 158–179). London, England: Continuum.

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

New York City Department of Education. (2009). *Diverse learners on the road to success: The performance of New York City's English language learners*. New York, NY: Office of English Language Learners.

Oller, J., & Vila, I. (2011). Interdependencia entre conocimiento de catalán y castellano y efectos de la lengua familiar en la adquisición de las lenguas escolares [Interdependence between knowledge of Catalan and Spanish and effects of the home language on school language acquisition]. *Cultura y Educacion*, *23*, 3–22. doi:10.1174/113564011794728588

*O'Muircheartaigh, J., & Hickey, T. (2008). Academic outcome, anxiety and attitudes in early and late immersion in Ireland. *International Journal of Bilingual Educational and Bilingualism*, *11*, 558–576. doi:10.1080/13670050802149184

The Organisation for Economic Co-operation and Development. (2010). *Pisa 2009 results: What students know and can do—Student performance in reading, mathematics and science* (Vol. 1). Retrieved from http://www.oecd.org/pisa/pisaproducts/48852548.pdf

*Ozerk, K., & Krashen, S. (2001). Subject matter teaching in bilingual education: Impact on bilingual and monolingual students. *ITL: Review of Applied Linguistics*, *131-132*, 1–10.

Patrinos, A. P., & Velez, E. (2009). Costs and benefits of bilingual education in Guatemala: A partial analysis. *International Journal of Educational Development*, *29*, 594–598. doi:10.1016/j.ijedudev.2009.02.001

Petrovic, J. (1997). Balkanization, bilingualism, and comparisons of language situations at home and abroad. *Bilingual Research Journal*, *2*, 233–254. doi:10.1080/15235882.1997.10668662

Ramirez, D. (1992). Executive summary. *Bilingual Research Journal*, *16*, 1–62. doi:10.1080/15235882.1992.10162628

Ricciardelli, L. A. (1992). Creativity and bilingualism. *Journal of Creative Behavior*, *26*, 242–254. doi:10.1002/j.2162-6057.1992.tb01183.x

Rolstad, K., Mahoney, K., & Glass, G. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy*, *19*, 572–594. doi:10.1177/0895904805278067

Rolstad, K., Mahoney, K., & Glass, G. (2008). The big picture in bilingual education: A meta-analysis corrected for Gersten's coding error. *Journal of Educational Research & Policy Studies*, *8*(2), 1–15.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118*, 183–192. doi:10.1037/0033-2909.118.2.183

Sallabank, J. (2006). Prospects for linguistic diversity in Europe and beyond: Views from a small island. In C. Leung & J. Jenkins (Eds.), *Reconfiguring Europe: The contribution of applied linguistics* (pp. 41–63). London, England: Equinox.

Samway, K. D., & McKeon, D. (1999). *Myths and realities: Best practices for language minority students*. Portsmouth, NH: Heinemann.

*Serra, J. M. (1989). Resultados academicos y desarrollo cognitivo en un programa de immersion dirigido a escolares de nivel socio-cultural bajo [Academic results and cognitive development in an immersion program for school children of a low socio-cultural status]. *Infancia y Apredizaje*, *47*, 55–66.

Slavin, R. E., & Cheung, A. (2003). *Effective reading programs for English language learners. A best evidence synthesis*. Baltimore, MD: Johns Hopkins University.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, *75*, 247–284. doi:10.3102/00346543075002247

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014

Snow, M. A., Met, M., & Genesee, F. (1989). A conceptual framework for the integration of language and content in second/foreign language instruction. *TESOL Quarterly*, *23*, 201–217. doi:10.2307/3587333

Stanat, P., & Christensen, G. (2006). *Where immigrant students succeed: A comparative review of performance and engagement of PISA 2003*. Paris, France: OECD.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, *54*, 30–34. doi:10.1080/01621459.1959.10501497

Takakuwa, M. (2005). Lessons from a paradoxical hypothesis: A methodological critique of the threshold hypothesis. In J. Cohen, K. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 2222–2232). Somerville, MA: Cascadilla Press.

Thomas, W., & Collier, V. (1997). *School effectiveness for language minority students*. Washington, DC: National Clearinghouse for Bilingual Education.

Thomas, W., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long term academic achievement: Final report*. Retrieved from http://www.usc.edu/dept/education/CMMR/CollierThomasComplete.pdf

*Verhoeven, L. (1987). *Ethnic minority children acquiring literacy*. Dordrecht, Netherlands: Foris.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*, 848–872. doi:10.1111/j.1467-8624.1998.tb06247.x

Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, *55*, 269–318. doi:10.3102/00346543055003269

Authors

GABRIJELA RELJIĆ is a research assistant at the University of Luxembourg, Research Unit INSIDE, University of Luxembourg, Route de Diekirch, L-7220 Walferdange, Luxembourg; e-mail: *gabrijela.reljic@uni.lu*. She studied pedagogy at the University of Novi Sad (Serbia) and leadership and developmental youth work at the University of Jönkoping (Sweden). She obtained both her master's and PhD degrees at the University of Luxembourg. Her main research areas are early education, literacy, psycholinguistics, bilingualism, and youth work.

DIETER FERRING is a professor of developmental psychology and geropsychology, and head of the Integrative Research Unit on Social and Individual Development (INSIDE) at the University of Luxembourg, Campus Walferdange, Route de Diekirch, 7220 Walferdange, Luxembourg; e-mail: *dieter.ferring@uni.lu*. His main research areas and publications cover life span development and aging.

ROMAIN MARTIN is a professor of psychology and educational science at the University of Luxembourg and professor in charge of the Education, Culture, Cognition, and Society (ECCS) research unit, Campus Walferdange, Route de Diekirch, 7220 Walferdange, Luxembourg; e-mail: *romain.martin@uni.lu*. His main research areas are psychological and educational measurement, technology-based assessment, new technologies of information, and communication in an educational context.

Downloaded from http://rer.aera.net at UNIVERSITAT AUTÒNOMA DE BARCELONA on October 29, 2014