British Journal of Educational Psychology (2011), 81, 629–653

The British Psychological Society

www.wileyonlinelibrary.com

Pedagogical strategies for teaching literacy to ESL immigrant students: A meta-analysis

Olusola O. Adesope[1]*, Tracy Lavin[2], Terri Thompson[2] and Charles Ungerleider[3]

[1]Washington State University-Pullman, USA
[2]Directions Evidence and Policy Research Group, Canada
[3]University of British Columbia, Canada

Background. Many countries rely on immigrants for population growth and to maintain a skilled workforce. However, many such immigrants face literacy-related barriers to success in education and in the labour force.

Aims. This meta-analysis reviews experimental and quasi-experimental studies to examine strategies for teaching English literacy to immigrant students.

Method. Following an exhaustive and systematic search for studies meeting pre-determined inclusion criteria, two researchers independently extracted data from 26 English as a Second Language (ESL) studies involving 3,150 participants. These participants consisted of ESL immigrant students in kindergarten through grade 6 who were exposed to English literacy instructional interventions. Measured outcomes were reading and writing.

Results and Conclusions. Mean effect sizes vary from small to large, depending on instructional interventions and outcome constructs. Across several different grade levels, settings, and methodological features, pedagogical strategies used in teaching ESL to immigrant students are associated with increased competence in reading and writing. Collaborative reading interventions, in which peers engage in oral interaction and cooperatively negotiate meaning and a shared understanding of texts, produced larger effects than systematic phonics instruction and multimedia-assisted reading interventions. The results show that the pedagogical strategies examined in this meta-analysis produced statistically significant benefits for students in all grade levels. The findings also show that students from low socio-economic status (SES) background benefit from ESL literacy interventions. However, significant heterogeneity was found in each subset. Educators and policy makers are encouraged to consider specific school contexts when making decisions about optimal pedagogical strategies. It is possible that contextual factors as well as ESL learner characteristics may influence the effectiveness of these strategies. To ensure literacy acquisition for immigrant students whose primary

*Correspondence should be addressed to Olusola O. Adesope, Department of Educational Leadership and Counseling Psychology, Washington State University, Pullman, WA, 99164-2114, USA (e-mail: olusola.adesope@wsu.edu).

DOI:10.1111/j.2044-8279.2010.02015.x

language is not English, it is important to continue to research successful literacy practices in ways that better inform educators and policy makers.

In the face of shrinking birth rates, many Western countries have come to rely on immigration for population growth and to maintain a skilled workforce. The pace of immigration to Western countries has increased in recent years (Slavin & Cheung, 2005). For example, the Census Bureau's 2006 American Community Survey[1] reported that about 37.5 million immigrants (approximately 12.5% of the entire US population) reside in the United States and about 20% of the US population speaks a language other than English at home.

Similar demographic patterns have been observed in Canada. The 2006 census report by Statistics Canada showed that immigration was responsible for two thirds of Canadian population growth between 2001 and 2006, that immigrants make up nearly 20% of Canada's population and labour force, and that the number of immigrants arriving each year will continue to rise for the foreseeable future (Chui, Tran, & Maheux, 2007). For the majority of these immigrants, English is a second language. In the United Kingdom, the pace of immigration has increased considerably over the last two decades, and similar increases in immigration have occurred in other English-speaking countries, such as Australia and New Zealand.

Do ESL immigrant students face obstacles in school?

As a result of immigration patterns, students learning English as a second language (ESL) make up a substantial portion of the student body in many English-speaking countries, particularly in the large urban centres that attract large numbers of immigrants (Carrasquillo & Rodríguez, 1996). For example, in Canada, 61% of students in the Vancouver school district and over 40% in the Toronto school district speak a language other than English at home (TDSB, 2009; VSB, 2009).

Learning in a second language presents a unique set of challenges, but it is unclear whether ESL immigrant students do or do not face particular obstacles in school. On one hand, results of international tests (e.g., PISA[2]) indicate that the majority of immigrant students do not perform as well as non-immigrant children (Worswick, 2001). The effects of immigration and ESL status play out differently across different academic domains. Even for immigrant students who do show evidence of academic difficulties these are often restricted to reading and writing, with mathematical skills among these students falling within the same range as their non-immigrant colleagues (e.g., Worswick, 2001). On the other hand, some studies suggest that immigrant students achieve at or above the levels of their non-immigrant colleagues (Thiessen, 2007; Zhou, 1997), are less likely to drop out of high school (White & Glick, 2000), and are more likely to pursue post-secondary studies (Dinovitzer, Hagan, & Parker, 2003; Vernez & Abrahamse, 1996).

In the United States, the National Center for Education Statistics (2004) reported mixed findings in the study of language minorities. 'In general, language minority youth and young adults lagged behind their counterparts who spoke only English at home on most education and economic indicators. However, among those who finished high

[1] *Full reports of the US Census Bureau's American Community Survey can be found at http://www.census.gov/acs/www/*

[2] *PISA, the Programme for International Student Assessment, is designed to provide policy-oriented international indicators of the skills and knowledge of 15-year-old students in three literacy domains: reading, mathematics, and science.*

school, no differences were found by English-speaking ability in the percentage that enrolled in postsecondary education' (p. iv). Overall, immigrant students are at least as successful as non-immigrants and sometimes more successful (Thiessen, 2007), but some groups of ESL immigrant students face steep challenges and achieve at dramatically lower levels than other students (Garnett, 2008). Moreover, many immigrant students have trouble leveraging their academic success in the labour market (Gilmore, 2008).

Given the extent to which some English-speaking countries rely on immigrants to remain competitive in the global knowledge economy, it is critical that educators and policy makers are able to develop, fund, and implement effective ESL literacy programmes to ensure that immigrant children are well positioned to join a skilled workforce when they complete their education. With many immigrant children arriving in their adopted countries with limited functionality in English, an important issue to address in providing a solid educational foundation is how best to tackle the literacy needs of immigrant students learning English as a second language.

Over the past few decades, there have been significant interest and intense debate among educational researchers, policy makers, and educators about effective approaches to literacy instruction for English language learners (Baker & de Kanter, 1981; Rossell & Baker, 1996; Slavin & Cheung, 2005; Willig, 1985, 1987). Much of the debate has hinged on the question of whether bilingual approaches (which begin with literacy instruction in students' native language) or English-only methods are most effective. While Rossell and Baker's (1996) review showed that bilingual education through the use of native language instruction was not beneficial, Greene's (1997) meta-analysis of some of the studies reviewed by Rossell concluded that teaching literacy to ESL learners through a mediated use of their native language has moderate beneficial effects. Other researchers have also reviewed the evidence and concluded that bilingual (or native language) approaches produce a small but beneficial effect on second language proficiency outcomes (Shanahan & Beck, 2006; Slavin & Cheung, 2005).

The present review looks beyond this debate to focus on specific pedagogical strategies for teaching English literacy to ESL immigrant students. The evidence on a wide variety of strategies for teaching literacy to ESL immigrant students was examined. These pedagogical strategies include: cooperative reading, systematic phonics instruction, guided reading, multimedia-assisted reading, free and structured writing, and various forms of intensive reading interventions. Specifically, the following research questions were addressed:

(1) What are the effects of different literacy strategies on ESL immigrant students' literacy skills, including reading and writing?
(2) Do the effects of literacy strategies vary across different countries and types of communities in which they are implemented?
(3) Do the effects of literacy strategies differ across ESL students with different first languages; in different grades; and in different socio-economic circumstances?
(4) Does the manner in which literacy strategies are implemented have an effect on ESL literacy outcomes?

In the following section, findings from four well-documented pedagogical practices for teaching literacy are summarized (August & Shanahan, 2006). These findings informed the choice of practices considered in the current meta-analysis.

Pedagogical practices for teaching literacy

Cooperative reading

There is evidence in the literature that proficient literacy in a second language can be achieved when peers engage in interactions and cooperatively negotiate meaning and shared understanding (Calderón, Hertz-Lazarowitz, & Slavin, 1998; Losey, 1995; Slavin, 1988). Researchers have used different cooperative learning strategies as literacy interventions. Two such strategies are dyad reading (Almaguer, 2005) and Cooperative Integrated Reading and Composition (Stevens, Madden, Slavin, & Farnish, 1987). Dyad reading strategies include cooperative read aloud exercises involving a lead reader, who is able to read well, and an assisted reader, who does not read as well. This cooperative peer-assisted reading strategy 'allows the sharing of reading experiences among students' (Almaguer, 2005, p. 511). Eldredge (1995) found that dyad reading can improve the reading fluency and comprehension of assisted readers.

Cooperative Integrated Reading and Composition (CIRC) is a popular form of cooperative learning used for comprehensive reading and writing instruction for students in grades 2 through 8 (Bramlett, 1994; Calderon *et al.*, 1998; Skeans, 1991; Stevens *et al.*, 1987). CIRC includes story-related activities, direct instruction in reading comprehension, as well as other reading activities. Although there are several variants of CIRC, a prominent method involves assigning students to four-member heterogeneous learning teams that work on different cooperative activities, including: peer reading; identification of major elements in a story; summarizing of stories; activities geared towards practice of reading comprehension strategies (e.g., spelling, decoding, and vocabulary); and creative writing (Calderon *et al.*, 1998). Generally, the process of deploying CIRC includes teacher instruction, team practice, and peer assessment.

Systematic phonics and guided reading

In systematic phonics instructions, learners are taught how to read and write using the correspondences between graphemes (letters that represent sounds in print) and phonemes (sounds of oral language). The overarching goal of phonics instruction is to help beginning readers understand how letters (graphemes) are linked to sounds (phonemes) to form letter-sound correspondences; that is, the systematic correspondences between how letters are written and sounded orally. Aside from applying this to improve reading and decode new words, this knowledge also helps beginning readers develop competence in spelling.

There is plentiful evidence that systematic phonics instruction can improve reading literacy for beginning readers (August & Shanahan, 2006; Ehri, Nunes, Stahl, & Willows, 2001; Ehri, Nunes, Willows, Schuster, *et al.*, 2001; National Reading Panel, 2000). For example, in a meta-analysis of 66 treatment-control comparisons, phonics instruction produced an overall moderate effect size, $d = .41$, showing a significant advantage of phonics instruction over other forms of instruction, including non-systematic phonics, and no phonics (e.g., whole language). Although the effects of phonics instruction have been widely documented in the literature, there is limited understanding of the effectiveness of this approach in teaching literacy to immigrant students whose first language is not English.

Multimedia-assisted reading

A number of multimedia-assisted reading interventions have been reported in the literature (Koskinen *et al.*, 2000; Neuman & Koskinen, 1992). Some of the

multimedia-assisted reading interventions are: incidental word learning through the use of captioned television; audio taped or read-along instructional materials; and computer-based interventions for reading and writing (e.g., Fast ForWord™, Word Processing). Literacy skills among beginning learners can be improved by viewing television programmes with captions. Such captioned programmes can create a rich multi-sensory learning environment by allowing students to hear the words, see the words as captions, and experience the contextual meaning of the words by watching the images on the television. Similarly, read-along multimedia materials allow students to listen to stories on audio tapes while concurrently reading the stories in the accompanied book. This approach can help students learn to associate the spoken words with the printed words, thus enriching the students' vocabulary and word recognition (Wray, 2004). Another multimedia programme used to enhance the reading ability of immigrant learners is Fast ForWord™ (Troia, 2004). This computer programme is primarily designed for students whose reading ability falls below their grade levels. The programme includes exercises in perceptual and spoken language comprehension designed to facilitate the development of language skills, especially communicative competence.

Writing

While a number of pedagogical strategies focused on the reading, speaking, and listening skills of ESL students have been developed, there has been limited investigation of ESL students' writing skills (Edelsky & Smith, 1989). Research in writing for ESL students is less well established and does not have the same strong empirical foundation as research in reading (Gomez, Parker, Lara-Alecio, & Gomez, 1996). However, because of the recent movement in education towards greater accountability, there is a growing emphasis on standardized testing, which often includes intensive writing (Walshaw, 2007). Different writing strategies have been used to improve the writing skills of learners, including diaries to promote metalinguistic awareness (Simard, 2004), as well as free and structured writing (Gomez *et al.*, 1996). Interventions that focus on free writing typically emphasize the process of writing rather than the product. Free writing allows students to receive feedback from teachers and their colleagues without being graded. Conversely, in the structured writing learning environment, topics to write about are usually assigned to students and writing is graded based on syntactic and lexical correctness (Shanahan & Beck, 2006).

Method

The current meta-analysis synthesized empirical studies that investigated literacy instruction among ESL immigrant students. Established procedures for conducting meta-analyses (Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985; Lipsey & Wilson, 2001) were followed, as outlined below.

Inclusion/exclusion criteria

Following a preliminary examination of empirical studies and reviews, criteria to capture studies with experimental or quasi-experimental designs investigating best practices for teaching English to immigrant students were developed. To be included in the meta-analysis, each study had to describe original data and meet the following criteria:

(1) The population had to be immigrant ESL students in an English-speaking country. Studies with students learning second languages other than English were excluded.
(2) The population had to be non-clinical. Studies of students with learning disabilities were excluded.
(3) The sample had to include an experimental group taught with a clearly reported strategy or intervention and a control group taught with traditional methods.
(4) The study sample had to include students in kindergarten through grade 6.
(5) Measured outcomes had to be clearly reported for both experimental and control groups. These included reading and/or writing, bi-literacy and vocabulary acquisition related to reading and writing. Studies in which speaking was the only outcome measured were excluded.
(6) Studies had to report sufficient data to allow for effect size calculations. Studies with insufficient data for effect size calculations were excluded[3].
(7) Studies had to be publicly available either online or in library archives[4].

Data sources and search strategies

Studies considered for this meta-analysis were located through a comprehensive and systematic search involving several databases and search strategies. Electronic searches were conducted in the following databases: Academic Search Premier, Education Full Text, ERIC (including British and Australian ERIC), Linguistic and Language Behavior Abstracts, PsycINFO and Web of Science. Multiple search strategies that incorporated the controlled vocabulary of each database for the following general terms were used: ESL students; immigrant, reading strategies; cognitive benefits. The reference lists of several earlier reviews of literature in second language instruction (e.g., Grabe, 1991; Norris & Ortega, 2000) were also 'hand searched'.

Outcomes of the search strategy

A total of 5,185 articles were identified. Duplicates were removed and the remaining articles were subjected to two different phases of screening. In the first phase of the screening process, the titles, abstracts, and keywords of these articles were reviewed by one of two researchers. Those that did not meet the selection criteria were excluded, resulting in a total of 256 included articles after the first phase. In order to determine the reliability of this process, 45 of these articles were randomly selected and reviewed independently by both researchers. Cohen's Kappa inter-rater agreement for this double-screening process was .88.

In the second phase of the screening process, two researchers independently read the full texts of each of the 256 articles retained after the first screening phase. Articles that, on closer inspection, did not meet all seven of the inclusion criteria were excluded during this phase. Another inter-rater reliability analysis was conducted to determine agreement between researchers on articles excluded. Cohen's Kappa was .92. The researchers resolved all discrepancies through discussion.

A total of 20 articles met all the inclusion criteria and were retained for coding and data extraction. Because some articles reported multiple studies, a total of 26 effect

[3] *Studies that did not report basic descriptive statistics but did include inferential statistics were included.*

[4] *For multiple reports of the same study (e.g., dissertation, conference proceedings, and journal article), the version published as a journal article was used for coding.*

sizes were extracted from the 20 articles. The extracted data were entered into EPPI-Reviewer, an extensive online coding application developed by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre; Thomas & Brunton, 2006) at the Institute of Education, University of London.

Effect size extraction and data analysis

Before data analyses were conducted, effect sizes were extracted from the 26 studies that met all inclusion criteria. Effect size is a standardized metric obtained by calculating the difference between the means of the intervention (ESL immigrant students taught with a specific strategy) and control (ESL immigrant students taught with traditional method) groups, divided by the pooled standard deviation of the two groups. Hedges (1981) observed that estimates may yield inflated effect sizes when samples are small. To correct for such bias in effect size estimation, especially with small sample sizes (Lipsey & Wilson, 2001), the obtained Cohen's d values were converted to Hedges' g, an unbiased estimate (Hedges & Olkin, 1985) of the standardized mean difference effect size. When other statistics like F or t were provided, these were also used to derive effect sizes or, in some cases, to verify the obtained d (Cooper & Hedges, 1994).

Standard meta-analytic guidelines and equations were followed in all data analyses (Cooper & Hedges, 1994; Hedges & Olkin, 1985; Lipsey & Wilson, 2001; Nesbit & Adesope, 2006). All data analyses were computed using *Comprehensive Meta-Analysis* 2.2.048 software (Borenstein, Hedges, Higgins, & Rothstein, 2008).

An aggregate effect size was computed from the weighted effect sizes to derive an overall weighted mean estimate of the effect of the intervention. This allowed more weight to be assigned to studies with larger sample sizes. The standard error of the Hedges' unbiased estimate of the mean effect size was then computed. A 95% confidence interval was computed around each mean weighted effect size to determine statistical significance. Confidence intervals spanning a range above zero were interpreted as indicating a statistically detectable result favouring the use of a pedagogical strategy over the use of traditional reading instruction for ESL immigrant students. In this meta-analysis, the upper and lower confidence intervals were also used to detect between-levels differences among different categories of analyses. Specifically, when the effects for different categories are similar, the confidence intervals overlap; if they are dissimilar, the confidence intervals do not overlap.

An important aspect of a meta-analysis involves determining whether the various effect sizes that are averaged into a mean value all estimate the same population effect size. This assumption of homogeneity of variance was tested by the Q statistic. When all findings share the same population value, Q has an approximate χ^2 distribution with $k - 1$ degrees of freedom, where k is the number of effect sizes. When Q exceeds the critical value of the χ^2 distribution, (i.e., $p < .05$), the mean effect size is assumed to be significantly heterogeneous, meaning that individual effect sizes do not estimate a common population mean (Lipsey & Wilson, 2001).

Results

In total, 20 articles encompassing an overall sample of 3,150 participants were found to meet all specified selection criteria, and data from 26 individual studies were extracted from these 20 articles. To generate a distribution of statistically independent effect sizes,

a single effect size was calculated for each set of dependent effect sizes by calculating a weighted average.

Table 1 shows important characteristics of each study, including the grade level of participants, the total number of participants involved in each study, first language spoken by participants, the type of strategies or intervention used in teaching ESL, length or duration of intervention, the unbiased effect size Hedges' *g,* and the 95% lower and upper confidence intervals around each unbiased effect size. 'Mixed first language' refers to studies in which first language varied across participants. For example, participants in Koskinen *et al.*'s (2000) study were drawn from 17 countries and spoke 16 different languages. According to the breakdown of independent effect sizes in Table 1, about 27% of the studies included in this meta-analysis produced moderate effect sizes[5] while 23% produced large effect sizes showing superior advantage of using the pedagogical strategies we examined over traditional methods of reading. Approximately, 27% of the overall studies produced small effect sizes in favour of the pedagogical strategies while 15% yielded small effect sizes in favour of traditional methods of reading. Only 8% of the studies produced null results.

Comprehensive Meta-Analysis software was used to evaluate the effect of removing potentially outlying effect sizes from the distribution and examining whether this will significantly influence the homogeneity results (Hedges & Olkin, 1985). Two potential outlying studies were removed one at a time. In each case, the re-calculated results did not increase the fit of the remaining effect sizes to a simple model of homogeneity. Hence, a decision was made to include all 26 studies in the meta-analysis.

Overall analysis

Collapsing across categories of pedagogical strategies, a moderate effect size was observed, $g = .41$ ($k = 26$, $N = 3{,}150$). The standard error of the mean is 0.04. This effect size represents an overall advantage of about .4 of a standard deviation for the use of different pedagogical strategies over traditional methods of teaching reading and writing. The 95% lower and upper confidence intervals around the mean effect size are .33 and .48, respectively, thus showing a statistically detectable overall mean effect size. However, homogeneity was rejected for this overall effect, suggesting large variability among the individual effect sizes that constitute the overall result ($Q_{\text{total}} = 123.93$, $p < .001$). The Q statistic does not provide information about the extent of heterogeneity, only about its statistical significance. Therefore, the I^2 statistic has been recommended as a complement to the Q statistic to properly interpret the overall results of homogeneity tests (Higgins & Thompson, 2002; Huedo-Medina, Sánchez-Meca, Marin-Martinez, & Botella, 2006). When the value of Q is greater than its associated degrees of freedom, I^2 is calculated by dividing the difference between Q and its degrees of freedom by Q itself and multiplying by 100 (when $Q \leq k - 1$, $I^2 = 0$). Thus, I^2 represents the percentage of the total variability in a set of effect sizes that results from between studies variability. Higgins and Thompson (2002) recommend that values of I^2 around 25%, 50%, and 75% should be interpreted as low, medium, and high heterogeneity, respectively.

In the present meta-analysis with 25 degrees of freedom, the I^2 estimate is 79.83%, thus confirming that the distribution of effect sizes is highly heterogeneous and that the

[5] *Throughout the result section, we adopted Cohen's (1988) interpretation of the magnitude of effect sizes. An effect size that is less than or about .2 is regarded as 'small', an effect size that is within the range .20–.50 is regarded as 'moderate', and an effect size that is greater than or equal to .8 is regarded as 'large'.*

Table 1. Characteristics of coded studies and associated effect sizes

Study	Grade	*N*	First Language[a]	Intervention	Duration of Intervention (hrs)	Effect Size (*g*)	95% *CI*	
							Lower	Upper
Almaguer (2005)	3	80	Spanish	Dyad reading with CRAB[b]	22.5	.45*	.01	.89
Calderón *et al.* (1998)	2–3	93	Spanish	BCIRC[c]/Writing	160.0	.45*	.04	.86
Calderón *et al.* (1998)	2–3	178	Spanish	BCIRC/Reading	160.0	.48*	.20	.80
Carlo *et al.* (2004)	5	254	Spanish	Word learning strategy	40.0	.56*	.28	.81
Denton *et al.* (2004)	2–5	33	Spanish	Read well	20.0	.23	−.45	.90
Denton *et al.* (2004)	2–5	60	Spanish	Read naturally	20.0	−.01	−.55	.45
Ehri, Dreyer, Flugman, and Gross (2007)	1	186	Spanish	Reading rescue	65.0	.68*	.37	.99
Gerber *et al.* (2004)	K	43	Spanish	Phonics Instruction – Core Intervention Model	3.0	.21	−.40	.83
Gomez *et al.* (1996)	6	72	Spanish	Free writing and structured writing	3.6	.34	−.15	.81
Koskinen *et al.* (2000)	1	162	Mixed	Multimedia (audio model) reading	42.0	−.10	−.45	.25
Koskinen *et al.* (2000)	1	159	Mixed	Multimedia (audio model) reading	42.0	−.26	−.62	.10
Lightbown and Spada (1990)	5–6	98	French	Form-focussed instruction and feedback	500.0	.14	−.35	.63
Lightbown, Halter, and White (2002)	6	114	French	Comprehension-based training	Unknown	−.25	−.41	.35

Table 1. (*Continued*)

Study	Grade	*N*	First Language[a]	Intervention	Duration of Intervention (hrs)	Effect Size (*g*)	95% *CI*	
							Lower	Upper
Linan-Thompson, Vaughn, Prater, and Cirino (2006)	2	106	Spanish	Group Guided Reading – RTI[d]	116.7	1.63*	1.20	2.08
Simard (2004)	6	81	French	Diary writing	13.3	.70*	.22	1.17
Slavin and Madden (1999)	4–5	255	Mixed	Success for all and Exito Para Todos	Not Reported	.87*	.61	1.12
Stuart (1999)a	K	112	Not stated	Jolly Phonics and Big Books -Reading	60	.27	−.10	.64
Stuart (1999)b	K	112	Not stated	Jolly Phonics and Big Books -Writing	60	.80*	.37	1.14
Stuart (2004)a	2	101	Not stated	Jolly & Late Phonics and Big Books – Reading	60	.51*	.09	.93
Stuart (2004)b	2	101	Not stated	Jolly & Late Phonics and Big Books – Writing	60	1.51*	1.04	1.97
Troia (2004)	1–6	168	Spanish	Fast ForWord Language Multimedia Program	33.3	.25	−.05	.56
Vaughn *et al.*-1 (2006a)	1	64	Spanish	Systematic group instruction with teacher	116.7	.50*	.01	1.00
Vaughn *et al.*-2 (2006a)	1	62	Spanish	Systematic group instruction with teacher	116.7	−.18	−.67	.32
Vaughn *et al.* (2006b)	1	40	Spanish	Small-group Reading Comprehension	95.8	.51	−.11	1.13
Vaughn *et al.* (2006c)	1	91	Spanish	Proactive Reading Comprehension	116.7	.28	−.13	.69
Waxman *et al.* (1994)	1–5	325	Spanish	ESLCA[e] and EUOT[f]	45.0	.00	−.23	.24

[a]First language spoken by participants in each study.
[b]CRAB – Comprehensive Reading Assessment Battery.
[c]BCIRC – Bilingual Cooperative Integrated Reading and Composition.
[d]RTI – Response To Intervention.
[e]ESL in the Content Areas.
[f]Effective Use of Time.
*$p < .05$; CI = confidence interval.

mean effect size may not be representative of the population. Significant heterogeneity indicates the necessity to further examine the study features that may be responsible for the variability in effect sizes. Hence, moderator analyses were conducted on the following factors: type of intervention and outcome; country and type of community in which studies were conducted; sample characteristics; implementation features; and methodological features. We used a fixed-effects model to compute within- and between-levels statistics.

Effects of literacy strategies on reading and writing among ESL immigrant students

Table 2[6] shows mean effect sizes broken down by: type of literacy intervention; intensity and duration of interventions; outcome constructs; and subcategories of outcome constructs. All the interventions produced statistically detectable mean effect sizes except multimedia-assisted reading strategy that, with only two studies, produced a small and non-significant effect size. Collaborative reading produced the largest effect of all the reading interventions, with a weighted mean effect size of .48 across five homogeneous studies. Diary writing and structured writing interventions produced a moderate effect size ($g = .53$) on writing.

Both reading and writing outcomes produced statistically detectable mean effect sizes. In the reading subcategories, comprehension as well as mixed comprehension and decoding produced statistically detectable mean effect sizes, but the four studies on word decoding interventions did not yield a statistically significant effect size – possibly as a result of the small number of studies in this category. Because the total between studies variance was statistically detectable (Q_B [2] = 15.84, $p < .05$), *post-hoc* analyses were conducted among subcategories of reading outcomes. These analyses showed that studies with mixed comprehension and decoding produced significantly larger effects than studies involving only comprehension- or decoding-based outcomes. Table 2 also shows the results of studies with interventions of different intensities and of different durations. Significant effects were statistically detected, regardless of the intensity of the intervention. There were, however, differences across different durations of interventions: studies conducted over a period of 3 months or less produced larger effects than those conducted over longer periods.

Effects of literacy strategies in different countries and types of communities

Table 3 shows that the effects of different literacy strategies were statistically detectable in the United States and in the United Kingdom, but not in Canada – though there were only three studies from Canada. Two Canadian studies by Lightbown and colleagues (1990; 2002) examined the effectiveness of audio-lingual instruction compared with reading/listening instruction and found no significant differences between these types of instructions. The third study conducted in Canada (Simard, 2004) showed a significant effect of using diaries to promote metalinguistic reflection among elementary school students ($g = .70$). There was significant heterogeneity across countries, indicating that the variability among effect sizes was greater than that expected from sampling error

[6] *Table 2 and subsequent tables include the number of participants (N) in each category, the number of independent effect sizes included in the computed mean effect size (k), the weighted Hedges' unbiased mean effect size (g) and its standard error (SE), the 95% lower and upper confidence intervals around the mean, and the results of a test of homogeneity (Q) with its associated probability value (p) and the magnitude of heterogeneity (I^2).*

Table 2. Weighted mean effect sizes for ESL studies by teaching methods (interventions), outcomes, and duration

	Adjusted effect size (*g*)				95% confidence interval		Test of homogeneity		
	N	*k*	*G*	*SE*	Lower	Upper	*Q*	*p*	I^2(%)
Intervention									
Collaborative reading	648	5	0.48*	0.08	.32	.64	1.00	.91	0.00
Systematic Phonics Instruction and Guided Reading	1,647	14	0.40*	0.05	.30	.50	73.75	.00	82.37
Multimedia-Assisted Reading (Audio, Tape, Computer, etc.)	330	2	0.10	0.12	−.13	.33	2.18	.14	54.11
Writing (Structured & Diary)	525	5	0.53*	0.10	.34	.72	37.86	.00	89.44
Total within (Q_W)							114.78	.00	
Total between (Q_B)							9.14	.03	
Total (Q_T)							123.93	.00	79.83
Outcome constructs									
Reading	2,532	20	0.38*	0.04	.30	.46	84.14	.00	77.42
Writing	618	6	0.51*	0.09	.34	.68	37.98	.00	86.83
Total within (Q_W)							122.11	.00	
Total between (Q_B)							1.81	.18	
Total (QT)							123.93	.00	79.83
Reading outcome subcategory									
Comprehension	1,829	13	0.32*	0.05	.22	.42	63.78	.00	81.18
Decoding	248	4	0.18	0.13	−.07	.43	1.05	.79	0.00
Mixed comprehension and decoding	1,073	9	0.60*	0.06	.48	.73	43.26	.00	81.51

Table 2. (*Continued*)

	Adjusted effect size (*g*)				95% confidence interval		Test of homogeneity		
	N	*k*	*G*	*SE*	Lower	Upper	*Q*	*p*	I^2(%)
Total within (Q_W)							108.09	.00	
Total between (Q_B)							15.84	.00	
Total (Q_T)							123.93	.00	79.83
Intensity of intervention (hrs)									
≤20 hrs	289	5	0.31*	0.12	.07	.54	4.62	.33	13.48
>20 but ≤100 hrs	1,800	12	0.35*	0.05	.25	.45	62.07	.00	82.28
>100 hrs	692	7	0.50*	0.08	.34	.66	36.43	.00	83.53
Not reported	369	2	0.58*	0.11	.37	.80	14.76	.00	93.23
Total within (Q_W)							117.89	.00	
Total between (Q_B)							6.04	.11	
Total (Q_T)							123.93	.00	79.83
Durations of intervention (months)									
≤3 months	1,175	11	0.57*	0.06	.45	.69	36.90	.00	72.90
>3 months	1,818	13	0.33*	0.05	.23	.43	72.20	.00	83.38
Not reported	157	2	0.03	0.17	−0.29	0.36	0.45	.50	0.00
Total within (Q_W)							109.56	.00	
Total between (Q_B)							14.37	.00	
Total (Q_T)							123.93	.00	79.83

*$p < .05$.

Table 3. Weighted mean effect sizes for ESL studies by geographical location and region

			Adjusted effect size (*g*)		95% confidence interval		Test of homogeneity		
	N	*k*	*G*	*SE*	Lower	Upper	*Q*	*p*	I^2(%)
Country									
United States of America	2,431	19	0.38*	0.04	.29	.46	90.07	.00	80.02
Canada	293	3	0.22	0.13	−.03	.47	5.65	.06	64.61
United Kingdom	426	4	0.70*	0.10	.50	.90	17.71	.00	83.06
Total within (Q_W)							113.44	.00	
Total between (Q_B)							10.49	.01	
Total (Q_T)							123.93	.00	79.83
Region									
Urban/Inner city	1,546	13	0.43*	0.05	.33	.54	47.83	.00	74.91
Suburban	321	2	−0.18	0.13	−.43	.08	0.38	.54	0.00
Rural	248	2	0.32*	0.13	.07	.57	0.53	.46	0.00
Not stated	1,035	9	0.54*	0.07	.41	.67	49.57	.00	83.86
Total within (Q_W)							98.31	.00	
Total between (Q_B)							25.62	.00	
Total (Q_T)							123.93	.00	79.83

*$p < .05$.

alone (Q_B [2] = 10.49, $p < .05$); therefore, *post-hoc* analyses were conducted among pairs of countries. Results showed that studies conducted in the United States and Canada were not significantly different from one another but were significantly different from studies conducted in the United Kingdom. The UK studies produced significantly higher effect sizes than studies conducted in the United States and Canada.

Table 3 also shows the results of analyses categorized according to the region where studies were conducted. The studies conducted in urban and rural regions produced statistically significant mean effect size, while those conducted in suburban regions did not produce statistically detectable effect sizes. Since the total between studies variance was significant, *post-hoc* tests were conducted. Results showed that studies conducted with participants in suburban region were different from those of other categories: all other mean effect sizes were significantly higher than the mean effect for suburban studies.

Effects of literacy strategies among samples with different characteristics

Table 4 shows the mean weighted effect sizes and associated confidence intervals and homogeneity statistics broken down by different sample characteristics: grade level, first language, and socio-economic status (SES) of participants. For most of these categories, the effect size distributions were significantly heterogeneous, indicating that the variability among effect sizes was greater than expected from sampling error alone.

As much as possible, effect sizes were calculated for each grade when multiple grades were included within a study (e.g., Lightbown, Halter, & White, 2002). However, many of the studies with participants in multiple grades did not report sufficient statistical information to calculate effect size per grade (e.g., Calderón *et al.*, 1998; Denton,

Table 4. Weighted mean effect sizes for ESL studies by sample characteristics

			Adjusted effect size (*g*)		95% confidence interval		Test of homogeneity		
	N	*k*	*G*	*SE*	Lower	Upper	*Q*	*p*	I^2(%)
Grade Level									
Kindergarten and grade1	1,031	10	0.27*	0.07	.15	.38	29.96	.00	69.96
Grades 2 and 3	659	6	0.77*	0.08	.60	.93	33.86	.00	85.24
Grades 4–6	874	6	0.52*	0.07	.38	.66	18.61	.00	73.13
Mixed grades	586	4	0.09	0.09	−.08	.25	2.07	.56	0.00
Total within (Q_W)							84.50	.00	
Total between (Q_B)							39.42	.00	
Total (Q_T)							123.93	.00	79.83
First language of participants									
Spanish	1,855	16	0.39*	0.05	.29	.48	57.25	.00	73.80
French	293	3	0.22	0.13	−.03	.47	5.65	.06	64.61
Mixed languages and others	1,002	7	0.50*	0.07	0.36	0.63	57.12	.00	89.50
Total within (Q_W)							120.02	.00	
Total between (Q_B)							3.90	.14	
Total (Q_T)							123.93	.00	79.83
Socio-Economic Status (SES)									
Low	2,294	18	0.37*	0.04	.29	.46	83.60	.00	79.67
High	254	1	0.55*	0.13	.28	.81	0.00	1.00	0.00
Not stated	602	7	0.49*	0.09	.31	.66	37.81	.00	84.13
Total within (Q_W)							121.41	.00	121.41
Total between (Q_B)							2.52	.28	2.52
Total (Q_T)							123.93	.00	79.83

*$p < .05$.

Anthony, Parker, & Hasbrouck, 2004; Slavin & Madden, 1999; Troia, 2004, Waxman, DeFelix, & Martinez, 1994). Literacy strategies produced heterogeneous but statistically detectable mean effect sizes for all educational grades, excluding studies conducted with participants in multiple grades. *Post-hoc* analyses revealed that studies conducted with students in grades 2 and 3 were significantly different from studies conducted with students in other grades. Studies with students in grades 2 and 3 produced the largest effect size ($g = .77$).

Spanish was the first language of participants in over half of the studies included in the meta-analysis: among these studies, the mean effect size was moderate and statistically detectable. Many of these studies were conducted in the Southern part of the United States, particularly Texas. Conversely, participants in the three studies with French-speaking students did not benefit significantly from the literacy interventions they were exposed to. In studies where participants spoke mixed languages, there was a statistically detectable mean effect size.

Many of the studies in Table 4 were conducted with participants whose SES was low. The total between studies variance was not statistically detectable (Q_B [2] = 2.52, $p = .28$), indicating that there was no significant difference across levels of socioeconomic

Table 5. Weighted mean effect sizes for ESL studies by study characteristics

	Adjusted effect size (*g*)				95% confidence interval		Test of homogeneity		
	N	*k*	*g*	*SE*	Lower	Upper	*Q*	*p*	I^2(%)
Setting									
Classroom	2,531	22	0.48*	0.04	0.40	.57	102.36	.00	79.48
Pull-out room	126	2	0.16	0.18	−.18	0.51	3.69	.05	72.92
Mixed classroom and pull-out room	493	2	0.10	0.09	−.09	.28	1.62	.20	38.46
Total within (Q_W)							107.67	.00	
Total between (Q_B)							16.25	.00	
Total (Q_T)							123.93	.00	79.83
Assignment of participants									
Random	2,006	16	0.31*	0.05	.22	.41	77.43	.00	80.63
Non-random	932	8	0.68*	0.07	.54	.81	20.76	.00	66.28
Others	212	2	0.03	0.15	−.27	.33	0.30	.58	0.00
Total within (Q_W)							98.49	.00	
Total between (Q_B)							25.43	.00	
Total (Q_T)							123.93	.00	79.83
Intervenor									
Researcher	612	7	0.63*	0.09	.46	.79	50.56	.00	88.13
Teacher	1,877	16	0.33*	0.05	.23	.43	38.99	.00	61.53
Mixed researcher and teacher	661	3	0.43*	0.08	.26	.59	25.28	.00	92.09
Total within (Q_W)							114.83	.00	
Total between (Q_B)							9.10	.01	
Total (Q_T)							123.93	.00	79.83

*$p < .05$.

status. These results suggest that low SES students benefited significantly from the ESL literacy interventions and to a similar degree as their higher SES counterparts.

Effects of ESL literacy strategies across studies with different characteristics

In most of the studies summarized in Table 5, interventions and outcome assessments were conducted in classroom settings, while a few were conducted in pull-out rooms away from students' regular classrooms. The classroom studies produced a moderate and significant mean effect size, while the mean effect sizes associated with studies conducted in pull-out rooms or in a mix of classrooms and pull-out rooms were not statistically significant.

Table 5 also shows the mean effect sizes associated with different approaches to assigning participants to experimental groups. Mean effect sizes were statistically detectable in all categories except in two studies that did not have a clearly defined procedure for assigning participants to groups. Non-random assignment of participants to treatment interventions was associated with a much higher mean effect size than random assignment. Results from Table 5 show that the different literacy strategies examined in this analysis produced statistically detectable effect sizes irrespective of whether the intervention was implemented by researchers or teachers. However, the advantages

Table 6. Weighted mean effect sizes for ESL studies by different methodological features

		Adjusted effect size (*g*)			95% confidence interval		Test of homogeneity		
	N	*k*	*g*	*SE*	Lower	Upper	*Q*	*p*	I^2(%)
Confidence in effect size derivation									
Medium	985	8	0.52*	0.07	.39	.65	36.93	.00	81.04
High	2,165	18	0.35*	0.05	.26	.44	82.74	.00	79.45
Total within (Q_W)							119.67	.00	
Total between (Q_B)							4.26	.04	
Total (Q_T)							123.93	.00	79.83
Treatment fidelity									
Low	493	2	0.10	0.09	−.09	.28	1.62	.20	38.46
High	2,657	24	0.47*	0.04	.39	.55	109.11	.00	78.92
Total within (Q_W)							110.74	.00	
Total between (Q_B)							13.19	.00	
Total (Q_T)							123.93	.00	79.83
Reliability									
Reported	1,496	13	0.39*	0.06	.29	.50	64.06	.00	81.27
Not reported	1,654	13	0.42*	0.05	.32	.52	59.76	.00	79.92
Total within (Q_W)							123.83	.00	
Total between (Q_B)							0.10	.75	
Total (Q_T)							123.93	.00	79.83
Validity									
Reported	993	10	0.36*	0.07	.23	.49	60.15	.00	85.04
Not reported	2,157	16	0.43*	0.05	.34	.52	63.07	.00	76.22
Total within (Q_W)							123.23	.00	
Total between (Q_B)							0.70	.40	
Total (Q_T)							123.93	.00	79.83

*$p < .05$.

are more apparent when literacy interventions are implemented by researchers than teachers.

Effects of ESL literacy strategies across different methodological features

Table 6 shows how effect sizes varied with the methodological features of different studies. The studies were categorized according to the researchers' confidence in the computed effect sizes; the treatment fidelity; and the reliability and validity of instruments used in the primary studies. Confidence in deriving effect sizes was associated with moderate effect sizes whether researchers' confidence was medium or high.

Treatment fidelity is the attention to detail with which the interventions were implemented and monitored. This included consideration of whether teachers were properly trained to provide the interventions; whether participants were properly instructed on how to participate in the intervention activities; and the extent to which participants' engagement in learning activities was monitored. Results in Table 6 show that the majority of the studies reported high treatment fidelity producing a moderate

statistically detectable effect size. The two studies with low treatment fidelity did not produce a statistically detectable effect size. Table 6 also shows that mean effect sizes were statistically detectable irrespective of whether the reliability and validity of measures were reported or not.

Despite our attempt to explain the variability in the overall results by conducting moderator analyses, results in Tables 2–6 are not indicative of a more homogeneous model. Hence, the results of this meta-analysis should be interpreted with caution.

Publication bias

All the studies that met our inclusion criteria were published in peer-reviewed journals. Hence, we could not investigate the potential moderating effect of the source of publication (published versus unpublished). The non-inclusion of unpublished studies in this meta-analysis heightens the potential for publication bias. There is an inherent bias against publishing non-statistically detectable results in peer-reviewed publications. While non-significant results are often included in conference proceedings and presentations, they are less likely to be published in refereed journals (Orwin, 1983; Rosenthal, 1979). This is referred to as the 'file-drawer problem' and becomes apparent in meta-analyses, where there is a tendency to exclude unpublished and grey literature, thereby potentially skewing meta-analytic findings towards a positive mean effect size. This poses a threat to the validity of results obtained from meta-analysis.

Meta-analysts have developed different methods to examine the validity of results obtained from meta-analyses against the potential for file-drawer bias (Rothstein, Sutton, & Borenstein, 2005). We computed two statistical tests in Comprehensive Meta-Analysis to examine the potential for publication bias, that is, the number of *file drawer* studies with null results needed to invalidate the effect found in this meta-analysis. First, we computed Orwin's Fail-Safe *N* using equation 1:

$$k_0 = k\left[\frac{\overline{ES}_k}{\overline{ES}_c} - 1\right], \tag{1}$$

where $k_0|$ is the number of effect sizes with a value of zero needed to reduce the mean effect size to $\overline{ES}_c|$; k is the number of studies involved in generating the mean effect size; $\overline{ES}_k|$ is the weighted mean effect size; and $\overline{ES}_c|$ is the criterion effect size level. A criterion level of 0.05 was used to estimate the number of studies with null results needed to nullify the effects found in this work. The Orwin's fail safe *N* was found to be 186 studies, meaning that 186 missing null studies would be required to bring the mean effect size found in this meta-analysis to a trivial level of 0.05.

A 'classic fail-safe *N*' test was also computed to determine the number of null effect studies needed to raise the *p* value associated with the average effect above an arbitrary alpha level (set at $\alpha = .05$). This test revealed that 720 additional studies will be required to invalidate the overall effect found in this meta-analysis. With 26 overall studies analyzed, both statistical tests show that the number of null or additional studies needed to nullify the overall effect size found in this meta-analysis is larger than the $5k + 10$ limit suggested by Rosenthal (1995). This suggests that the results of this meta-analysis are robust and the validity is not threatened by publication bias.

Discussion

The meta-analysis presented here examined a sample of studies that met established inclusion criteria, resulting in extraction of data from 26 studies. The analysis encompassed a broad view of immigrant ESL students in kindergarten to grade 6 who were exposed to different pedagogical strategies designed to teach English literacy. Measured outcomes were reading and writing.

It is imperative to reiterate that the overall results of combining all the studies did not produce a homogeneous effect size. Further, attempts to explain the variability in the overall results by conducting moderator analyses on the type of intervention and outcome, country and type of community in which studies were conducted, different sample characteristics, implementation, and methodological features did not produce homogeneous effect sizes. The variability in the moderator analyses confirms that the mean effect sizes may not be representative of the population and thus considerably limits the certainty of the conclusions that can be drawn from the results of this meta-analysis. Hence, the results should be interpreted with caution because all within-levels analyses in Tables 2–6 are heterogeneous.

The overall findings provide evidence for collaborative reading ($g = .48$), systematic phonics instruction ($g = .40$), and diary writing ($g = .53$), suggesting that these strategies are effective for teaching English literacy to ESL immigrant students. The moderator analyses highlight a number of additional details relevant to the overall findings.

In particular, the results reveal that the mean effect size was not significant for the four studies in which word decoding was the outcome measure. Because the statistical power for this category is low ($k = 4$) and the confidence intervals are wide ($-.07$ to $.43$), it cannot be concluded that the pedagogical strategies we examined did not benefit ESL immigrant students on decoding outcome. Indeed, our result indicates that, on decoding measure, ESL students benefited with the traditional reading programme as much as with the pedagogical strategies. We speculate that the limited number of studies in this category may have precluded finding a statistically detectable effect.

The moderator analyses also revealed that ESL students from low SES ($g = .37$) backgrounds benefit from literacy interventions in much the same way as their higher SES counterparts. SES is often a strong predictor of academic achievement (Hobbs, 1990; Sirin, 2005) with low SES students generally falling behind higher SES students. Low SES students have been found to hold more negative attitudes towards learning, rate themselves as less confident readers, and derive less enjoyment from reading than their high SES peers (Clark & Akerman, 2006). Low SES students have also reported having fewer books at home, more limited access to educational materials (books, computers, magazines), and as a result of low parental education, they have reported receiving less encouragement from parents to read (Clark & Akerman, 2006). It is encouraging, therefore, to find that the effectiveness of ESL literacy strategies we examined in this meta-analysis extends to low SES students.

The results also show that the pedagogical strategies examined in this review produced statistically significant benefits for students in all grade levels between kindergarten and sixth grade. This is equally encouraging, as ESL immigrant students enter English-speaking schools at all grade levels – not just the early primary grades in which literacy development is a primary focus. Although our study is not a comparative analysis of the differential effects of pedagogical strategies for teaching ESL students and

native English language speakers, based on the report of the National Reading Panel (NRP, 2000) as well as our introductory review of literature in this area, the pedagogical strategies we investigated appear to be at least as effective for native English speakers. Taken together, these findings should persuade teachers and policy makers that the well-planned use of these pedagogical strategies may benefit many students across grade levels.

The moderator analyses also indicate that the mean effect sizes associated with interventions implemented in pull-out rooms or a mix of classroom and pull-out room were not statistically significant. As there are only two studies in each of these categories, it is possible that this is simply too few from which to draw any conclusions. Further research on the benefits of providing ESL literacy interventions outside of the regular classroom is warranted, but the current analysis indicates that available resources would be more wisely directed towards classroom-based interventions. The results also indicate that the mean effect size of the three studies conducted in Canada was not statistically significant. The absence of a significant effect among the Canadian studies should be interpreted with caution as there are only three studies in this category. This finding does not necessarily indicate that the ESL literacy strategies used in Canadian schools are ineffective. Rather, the non-significant effect size for the Canadian studies likely results from the limited number and diversity of Canadian studies. Given the importance of immigration to Canada's demographic and work force renewal, further research on immigrant ESL literacy in the Canadian context is warranted.

Educators and policy makers have a strong interest in developing successful strategies for teaching literacy to immigrant students. In countries, such as Canada, the United States, and the United Kingdom, that rely on large numbers of incoming immigrants to sustain population growth and a skilled labour force, the education of immigrant children is an important issue that presents ongoing challenges. For many of these children, literacy acquisition proceeds in a language that is not their mother tongue. To ensure successful literacy acquisition for immigrant students whose primary language is not English, it is important to identify and review successful literacy practices so that educators and policy makers may be properly informed.

The evidence presented in this meta-analysis suggests that cooperative reading, systematic phonics instruction, and diary and structured writing interventions have the potential to enhance the teaching of English literacy to ESL immigrant students. Overall analyses reveal that, across several educational levels, settings and methodological features, pedagogical strategies used in teaching ESL to immigrant students are associated with increased competence in reading and writing. However, educators and policy makers are encouraged to consider specific school contexts when making decisions about optimal pedagogical strategies. It is possible that contextual factors as well as ESL learner characteristics may influence the effectiveness of these strategies. For example, the current analysis suggests that fidelity of treatment is important to achieving positive outcomes as studies with high treatment fidelity produced a moderate effect size ($g = .47$). Therefore, teachers need adequate training before implementing new literacy strategies. Similarly, although there is plentiful evidence that collaborative learning is beneficial (Johnson & Johnson, 1994; O'Donnell, Hmelo-Silver, & Erkens, 2006); successful outcomes require teachers' skillful development and mediation of structured activities that allow students to work together in small groups while fostering individual accountability.

Limitations and future research directions

The current work is constrained by a few limitations, which can be addressed in future research. First, few of the pedagogical strategies have been investigated across a sufficient number of contexts to make broad statements about their generalizability, and drawing detailed conclusions is hampered by the wide heterogeneity in study findings as well as the relatively small number of studies currently available. For example, all but seven of the independent studies reported in this meta-analysis were conducted in the United States mostly with Spanish-speaking ESL students. In our search for studies exploring pegadogical strategies for teaching English literacy to immigrant students, only three studies conducted with students in grades 7–12 were found, hence our decision to constrain the literature included in this meta-analysis to those conducted with students in kindergarten up to grade 6. In sum, more high-quality studies are needed to increase the generalizability of the results and to fully describe the benefits of using these strategies to enhance literacy acquisition across different population of students with varying linguistic and cultural backgrounds.

Second, there is evidence that the relationship between educational achievement and immigration depends to some extent on immigrants' countries of origin (Cheng *et al.*, 1993; Yau *et al.*, 1993). For example, in some provinces in Canada (especially Ontario and Quebec), immigrant students from Asian countries are more successful in school than are immigrants from African countries (Yau *et al.*, 1993) who, in turn, are more successful than immigrant students from the Caribbean (Cheng *et al.*, 1993). Many of the studies included in this meta-analysis did not report the immigrants' countries of origin. Additional research is needed to explore the relationship between literacy acquisition and immigrant students' countries of origin and to examine the factors underlying this relationship.

Finally, the studies we reviewed in this meta-analysis did not address the relationship between literacy proficiency in the first language and the effects of different literacy strategies. At least two major hypotheses have been offered to explain the relationship between acquisition of literacy in a second language and literacy proficiency in a first language (Clark, 1980; Cummins, 1979). The *linguistic interdependence hypothesis* predicts that reading ability in the first language transfers to the second language reading, whereas the *linguistic threshold hypothesis* posits that ability to read in the first language only transfers to second language reading when learners attain a certain level of morphological, syntactic, and lexical proficiency in the second language. As participants' first language literacy proficiency was not explicitly reported in any of the studies included in this meta-analysis, the current work cannot test these theories. In future research, data on participants' first language literacy proficiency would provide valuable information to help contextualize the effects of literacy strategies among ESL immigrant students.

References

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

*Almaguer, I. (2005). Effects of dyad reading instruction on the reading achievement of Hispanic third-grade English language learners. *Bilingual Research Journal*, *29*, 509–526. doi:10.1080/15235882.2005.10162849

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.

Baker, K. A., & de Kanter, A. A. (1981). *Effectiveness of bilingual education: A review of the literature*. Washington, DC: Office of Planning, Budget and Evaluation, U.S. Department of Education.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2008). *Comprehensive meta-analysis* (Version 2.2.048). Englewood, NJ: Biostat.

Bramlett, R. K. (1994). Implementing cooperative learning: A field study evaluating issues for school-based consultants. *Journal of School Psychology*, *32*, 67–84. doi:10.1016/0022-4405(94)90029-9

*Calderón, M., Hertz-Lazarowitz, R., & Slavin, R. (1998). Effects of bilingual cooperative integrated reading and composition on students making the transition from Spanish to English reading. *Elementary School Journal*, *99*, 153–165. doi:10.1086/461920

*Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., *et al*. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*, 188–215. doi:10.1598/RRQ.39.2.3

Carrasquillo, A. L., & Rodríguez, V. (1996). *Language minority students in the mainstream classroom*. Avon, England: Multilingual Matters.

Cheng, M., & Others. (1993). *The 1991 every secondary student survey, part II: Detailed profiles of Toronto's secondary school students. no. 204. Canada; Ontario.*

Chui, T., Tran, K., & Maheux, H. (2007). *Immigration in Canada: A portrait of the foreign-born population, 2006 Census*. Statistics Canada, Catalogue no. 97-557-XIE. Retrieved from http://www12.statcan.ca/english/census06/analysis/immcit/pdf/97-557-XIE2006001.pdf

Clark, M. (1980). The short circuit hypothesis of ESL reading—or when language competence interferes with reading performance. *The Modern Language Journal*, *64*, 203–209. doi:10.2307/325304

Clark, C., & Akerman, R. (2006). *Social inclusion and reading: An exploration*. UK: National Literacy Trust.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper, H. M., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49* 222–251. doi:10.2307/1169960

*Denton, C. A., Anthony, J. L., Parker, R., & Hasbrouck, J. E. (2004). Effects of two tutoring programs on the English reading development of Spanish-English bilingual students. *Elementary School Journal*, *104*, 289–305. doi:10.1086/499754

Dinovitzer, R., Hagan, J., & Parker, P. (2003). Choice and circumstance: Social capital and planful competence in the attainments of immigrant youth. *Canadian Journal of Sociology*, *28*, 463–488. doi:10.2307/3341838

Edelsky, C., & Smith, K. (1989). Is that writing – or are those marks just a figment of your curriculum. In G. Manning & M. Manning (Eds.), *Whole language: Beliefs and practices, K-8* (pp. 183–193). Washington, DC: National Education Association.

*Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal*, *44*, 414–448. doi:10.3102/0002831207302175

Ehri, L. C., Nunes, S. R., Stahl, S A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, *71*, 393–447. doi:10.3102/00346543071003393

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, *36*, 250–287. doi:10.1598/RRQ.36.3.2

Eldredge, J. L. (1995). *Teaching decoding in holistic classrooms*. Englewood Cliffs, NJ: Merrill/Prentice Hall.

Garnett, B. (2008). *A critical examination of the academic trajectories of ESL youth*. Unpublished doctoral dissertation. University of British Columbia, Vancouver, Canada.

*Gerber, M., Jimenez, T., Leafstedt, J., Villaruz, J., Richards, C., & English, J. (2004). English reading effects of small group intensive intervention in Spanish for K-1 English learners. *Learning Disabilities Research & Practice*, *19*, 239–251. doi:10.1111/j.15405826.2004.00109.x

Gilmore, J. (2008). *The Canadian immigrant labour market in 2006: Analysis by region or country of birth*. (Catalogue No. 71-606-X2008002). Ottawa: Statistics Canada.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage.

*Gomez, R., Jr., Parker, R. I., Lara-Alecio, R., & Gomez, L. (1996). Process versus product writing with limited English proficient students. *Bilingual Research Journal*, *20*, 209–233.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, *25*, 375–406. doi:10.2307/3586977

Greene, J. P. (1997). A meta-analysis of the Rossell and Baker Review of bilingual education research. *Bilingual Research Journal*, *21*, 1–24.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107–128. doi:10.2307/1164588

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558. doi:10.1002/sim.1186

Hobbs, D. (1990). School based community development: Making connections for improved learning. In S. Raferty & D. Mulkey (Eds.), *The role of rural schools in community development* (pp. 57–64). Mississippi State, MS. Southern Rural Development Center.

Huedo-Medina, T. B., Sánchez-Meca, F., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: I^2 or Q statistic? *Psychological Methods*, *11*, 193–206. doi:10.1037/1082-989X.11.2.193

Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperation, competition, and individualization* (4th ed). Englewood Cliffs, NJ: Prentice Hall.

*Koskinen, P. S., Blum, I. H., Bisson, S. A., Phillips, S. M., Creamer, T. S., & Baker, T. K. (2000). Book access, shared reading, and audio models: The effects of supporting the literacy learning of linguistically diverse students in school and at home. *Journal of Educational Psychology*, *92*, 23–36. doi:10.1037/0022-0663.92.1.23

*Lightbown, P. M., Halter, R. H., & White, J. L. (2002). Comprehension-based learning: The limits of 'do it yourself'. *Canadian Modern Language Review*, *58*, 427–464. doi:10.3138/cmlr.58.3.427

*Lightbown, P. M., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition*, *12*, 429–448. doi:10.1017/S0272263100009517

*Linan-Thompson, S., Vaughn, S., Prater, K., & Cirino, P. T. (2006). The response to intervention of English language learners at risk for reading problems. *Journal of Learning Disabilities*, *39*, 390–398. doi:10.1177/00222194060390050201

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Losey, K. M. (1995). Mexican American students and classroom interaction: An overview and critique. *Review of Educational Research*, *65*, 283–318. doi:10.3102/00346543065003283

National Center for Education Statistics (2004). Language minorities and their educational labor market indicators – recent trends. Retrieved from http://nces.ed.gov/pubs2004/2004009.pdf

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of scientific research literature on reading and its implications for reading instruction*. Bethesda, MD: National Institutes of Health.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*, 413–448.

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, *27*, 94–106. doi:10.2307/747835

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528. doi:10.1111/0023-8333.00136

O'Donnell, A. M., Hmelo-Silver, C., & Erkens, G. (2006). *Collaborative learning, reasoning, and technology*. Mahwah, New Jersey: Lawrence Erlbaum.

Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157–159. doi:10.2307/1164923

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638–641. doi:10.1037/0033-2909.86.3.638

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

Rossell, C. H., & Baker, K. (1996). The educational effectiveness of bilingual education. *Research in the Teaching of English*, *30*, 7–69.

Rothstein, H., Sutton, A., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Hoboken, New Jersey: Wiley. doi:10.1002/0470870168

Shanahan, T., & Beck, I. L. (2006). Effective literacy teaching for English-language learners. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth* (pp. 415–488). Mahwah, NJ: Erlbaum.

*Simard, D. (2004). Using diaries to promote metalinguistic reflection among elementary school students. *Language Awareness*, *13*, 34–48. doi:10.1080/09658410408667084

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A Meta-analytic review of research 1990–2000. *Review of Educational Research*, *75*, 417–453. doi:10.3102/00346543075003417

Skeans, S. E. S. (1991). The effects of cooperative integrated reading and composition: Fidelity of implementation, and teacher concerns on student achievement. *Dissertation Abstracts International 53*(02), 0455A. (UMI No. 9217026).

Slavin, R. E. (1988). Cooperative learning and student achievement. *Educational Leadership*, *46*, 31–33.

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

*Slavin, R. E., & Madden, N. (1999). Effects of bilingual and English as a second language adaptations of success for all on the reading achievement of students acquiring English. *Journal of Education for Students Placed at Risk*, *4*, 393–416. doi:10.1207/s15327671espr0404_3

Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative integrated reading and composition: Two field experiments. *Reading Research Quarterly*, *22*, 433–454. doi:10.2307/747701

*Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology*, *69*, 587–605. doi:10.1348/000709999157914

*Stuart, M. (2004). Getting ready for reading: A follow-up study of inner city second language learners at the end of key stage 1. *British Journal of Educational Psychology*, *74*, 15–36. doi:10.1348/000709904322848806

TDSB (2009). Toronto District School Board. *Facts and figures*. Retrieved from http://www.tdsb.on.ca/communications/tdsbfacts.html

Thiessen, V. (2007). Academic performance, human capital skill formation, and post-secondary education: A comparison of First Nation, immigrant, visible minority, and other native-born Canadian youth. *Canadian Labour and Skills Researcher Network. Working paper Number1*. Retrieved from http://www.econ.ubc.ca/clsrn/publications/workingpapers/Thiessen_workshop_cover%20page.pdf

Thomas, J., & Brunton, J. (2006). EPPI-Reviewer 3.0: *Analysis and management of data for research synthesis*. EPPI-Centre Software. London: Social Science Research Unit, Institute of Education.

*Troia, G. A. (2004). Migrant students with limited English proficiency: Can fast ForWord language-super™ make a difference in their language skills and academic achievement? *Remedial and Special Education*, *25*, 353–366. doi:10.1177/07419325040250060301

*Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., *et al*. (2006b). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal*, *43*, 449–487. doi:10.3102/00028312043003449

*Vaughn, S., Linan-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., *et al*. (2006a). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities*, *39*, 56–73. doi:10.1177/00222194060390010601

*Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., *et al*. (2006c). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *Elementary School Journal*, *107*, 153–180. doi:10.1086/510653

Vernez, G., & Abrahamse, A. (1996). *How immigrants fare in U.S. Education*. Santa Monica, CA: RAND. Retrieved from http://www.rand.org/content/dam/rand/pubs/monograph_reports/2007/MR718.pdf

VSB (2009). Vancouver School Board – supporting ESL students. Retrieved from http://www.vsb.bc.ca/supporting-esl-students

Walshaw, M. (2007). Responding to calls for greater accountability. *Mathematics Education Research Journal*, *19*, 1–2.

*Waxman, H. C., DeFelix, J. W., & Martinez, A. (1994). Effects of implementing classroom instructional models on English language learners' cognitive and affective outcomes. *Bilingual Research Journal*, *18*, 1–22. doi:10.1080/15235882.1994.10162665

White, M. J., & Glick, J. E. (2000). Generation status, social capital, and the routes out of high school. *Sociological Forum*, *15*, 671–691. doi:10.1023/A:1007515100190

Willig, A. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, *55*, 269–317. doi:10.2307/1170389

Willig, A. (1987). Examining bilingual education research through meta-analysis and narrative review: A response to Baker. *Review of Educational Research*, *57*, 363–376. doi:10.2307/1170465

Worswick, C. (2001). School performance of the children of immigrants in Canada, 1994–98. Ottawa, ON: Statistics Canada. Retrieved from http://www.statcan.gc.ca/pub/11f0019m/11f0019m2001178-eng.pdf

Wray, D. (2004). *Teaching literacy: Using texts to enhance learning*. London: David Fulton.

Yau, M., & Others. (1993). *The 1991 every secondary student survey, part III: Program level and student achievement no. 205*. Toronto: Toronto Board of Education.

Zhou, M. (1997). Growing up American: The challenge confronting immigrant children and children of immigrants. *Annual Review of Sociology*, *23*, 63–95. doi:10.1146/annurev.soc.23.1.63

Received 13 July 2009; revised version received 12 November 2010